Beta-decomposition for the Volume and Area of the Union of Three-dimensional Balls and Their Offsets

Deok-Soo Kim,*[a] Joonghyun Ryu,[b] Hayong Shin,[c] and Youngsong Cho[d]

Given a set of spherical balls, called atoms, in three-dimensional space, its mass properties such as the volume and the boundary area of the union of the atoms are important for many disciplines, particularly for computational chemistry/biology and structural molecular biology. Despite many previous studies, this seemingly easy problem of computing mass properties has not been well-solved. If the mass properties of the union of the offset of the atoms are to be computed as well, the problem gets even harder. In this article, we propose algorithms that compute the mass properties of both the union of atoms and their offsets both correctly and efficiently. The proposed algorithms employ an approach, called the *Beta-decomposition*, based on the recent theory of the beta-complex. Given the beta-complex of an atom set, these algorithms decompose the target mass property into a set of primitives using the simplexes of the beta-complex. Then, the molecular mass property is computed by appropriately summing up the mass property corresponding to each simplex. The time complexity of the proposed algorithm is $O(m)$ in the worst case where m is the number of simplexes in the beta-complex that can be efficiently computed from the Voronoi diagram of the atoms. It is known in $\mathbb{R}^3$ that $m = O(n)$ on average for biomolecules and $m = O(n^2)$ in the worst case for general spheres where n is the number of atoms. The theory is first introduced in $\mathbb{R}^2$ and extended to $\mathbb{R}^3$. The proposed algorithms were implemented into the software **BetaMass** and thoroughly tested using molecular structures available in the Protein Data Bank. BetaMass is freely available at the Voronoi Diagram Research Center web site.

DOI: 10.1002/jcc.22956

Introduction

Suppose that we are given a set $A = \{a_1, a_2, \ldots, a_n\}$ of spherical balls $a_i = (p_i, r_i)$ in $\mathbb{R}^3$ with the center p_i and radius r_i. We call a_i a *vdW-atom* and r_i a *vdW-radius* where "vdW" denotes the "van der Waals." Set A is called a *vdW-molecule* (or simply molecule) and its boundary ∂A is called the *vdW-boundary*. The area of the vdW-boundary is called the *vdW-area*. The region of space contained within the vdW-boundary is called the *vdW-region* and the volume of the vdW-region is called the *vdW-volume*. To be physicochemically realistic, we assume that a pair of vdW-atoms may intersect but an atom is not allowed to be completely contained within another atom.

Suppose that $A^O = \{a_1^O, a_2^O, \ldots, a_n^O\}$ is a set of enlarged atoms $a_i^O = (p_i, r_i^O)$ where p_i is the center and $r_i^O = r_i + \delta$, $\delta > 0$, is the radius. In geometric modeling and computational geometry community, A^O and ∂A^O are called the *offset model* and the *offset surface* of A by the offset amount δ, respectively. In computational chemistry and the related disciplines, the boundary ∂A^O is frequently referred to by the name of *Lee-Richrads (accessible) surface* since Lee and Richards introduced it as the *(solvent) accessible surface* in 1974.[1] In this context, δ corresponds to the radius of a spherical probe approximating a solvent molecule. The most common solvent molecule is water and its probe is usually assumed as a spherical ball with 1.4 Å radius. In this article, we call the volume of the region of space contained within ∂A^O the *offset-volume* and the area of ∂A^O the *offset-area*. Another common type of surface is the *Connolly surface*, also frequently called the *solvent-excluded surface*. For the complete definition of these surfaces and their naming in literature, we recommend readers to refer to Ref. [2].

In this article, we present algorithms for computing two types of molecular mass properties in a unified framework: the vdW-properties and the offset properties. The computation of these mass properties is important for various applications, particularly for computational chemistry/biology and structural molecular biology because these values are important parameters for understanding various biological phenomena and molecular functions. Hence, fast computation of their correct values is desirable. While there have been many studies due to its importance, mathematically correct and computationally efficient method with implementation for computing these mass properties is still hard to find.

Computing mass properties correctly and efficiently is not as easy as it may seem because some vdW-atoms may intersect.

[a] *D.-S. Kim*
Department of Industrial Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, South Korea
E-mail: dskim@hanyang.ac.kr
[b] *J. Ryu*
Voronoi Diagram Research Center, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, South Korea
[c] *H. Shin*
Department of Industrial and Systems Engineering, KAIST, 335 Gwahak-ro, Yuseong-gu, Daejeon 305-701, South Korea
[d] *Y. Cho*
Voronoi Diagram Research Center, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, South Korea

Contract/grant sponsor: National Research Lab grant; Contract/grant number: 2011-0020410; Contract/grant sponsor: National Research Foundation, Korea

Dodd and Theordorou (1991) reported that the accessible surface (i.e., the offset surface) of a relatively small molecule defined by a 1.4 Å radius probe had over 80,000 instances of distinct eight-fold intersections (i.e., eight atoms have a common intersection) and even had a 17-fold intersection.[3] Correct and efficient computation of molecular mass properties requires two issues to be handled:

- The combinatorics issue: the issue of the combinatorial structure of the intersections among atoms, and
- The decomposition issue: the issue of decomposing the mass properties into a set of primitives.

In this article, we propose two algorithms that employ an approach called the *Beta-decomposition* that correctly and efficiently compute molecular mass properties. The **Beta-decomposition-vdW** algorithm computes the vdW-volume and the vdW-area and the **Beta-decomposition-offset** algorithm computes the offset-volume and the offset-area, all in $O(n)$ time on average for a set of n vdW-atoms in $\mathbb{R}^3$. To be more specific, the proposed algorithms compute both the vdW and the offset mass properties in $O(m)$ time in the worst case, where m represents the number of simplexes in the zero beta-complex in $\mathbb{R}^3$. It is known that $m = O(n^2)$ in the worst case for a set of n general spheres and $m = O(n)$ on average for molecules. We emphasize the following: Provided that the vdW mass property is computed, the marginal computation necessary for the offset mass property is tiny.

The idea of the beta-decomposition is very simple as follows. Assuming that the zero beta-complex of a molecule is available, the algorithms use the topology of the beta-complex for handling the "combinatorics issue" and use both the geometry and topology of the beta-complex for handling the "decomposition issue." Then, the correct mass property can be obtained by the appropriate summation of the mass property of the decomposed primitives.

We first present the **Beta-decomposition-vdW** algorithm and then extend it to the **Beta-decomposition-offset** algorithm. The correctness and efficiency of the proposed algorithms and their implementations are verified through an experiment using molecular structure data available in the Protein Data Bank (PDB).[4] To verify the correctness of the solutions computed by the proposed algorithms, we implemented the Monte Carlo simulation in two different approaches: a straightforward approach to the Monte Carlo simulation (MCS-I) and an efficient approach to the Monte Carlo simulation (MCS-II). Both MCS-I and II were tested with sufficiently many random points to verify the quality of solutions produced by the Beta-decomposition algorithm. The program implementing the proposed algorithms is called **BetaMass** and is also a part of **BetaMol**, the Windows-based molecular modeling and BioCAD software based on the beta-complex theory. The Linux version of **BetaMass** was also implemented. Both are freely available from the web site of Voronoi Diagram Research Center (VDRC).[5] The idea of the beta-decomposition algorithm can also be easily used for the weighted alpha-complex as well.

We call the Voronoi diagrams or the power diagram the *primal structure* and the Delaunay triangulation, the regular triangluation, or the quasi-triangulations the *dual structure*. We note that the beta-decomposition algorithm differs from other previous studies falling into the cell-decomposition approach in that the beta-decomposition algorithm decomposes the vdW-molecule using the simplexes in the (subset of the) dual structure, not the cells in the primal structure.

This article is organized as follows: The "Related Works" section reviews previous, related studies. The "Voronoi Diagram, Quasi-triangulation, and the Beta-Complex" section presents some background materials: the Voronoi diagram for an atom set, the quasi-triangulation, and the beta-complex. The "Area of the Union of Disks in $\mathbb{R}^2$" section presents the beta-decomposition algorithm for computing the two-dimensional volume (i.e., the area) of the union of a set of disks in the plane and provides the idea of the Beta-decomposition algorithm for three-dimensional mass properties. The "Volume of the Union of Balls in $\mathbb{R}^3$" section presents the algorithm for computing the volume of the union of the three-dimensional atoms. The "Boundary of the Union of Balls in $\mathbb{R}^2$ and $\mathbb{R}^3$" section presents the algorithm for computing the boundary area of the union of both two- and three-dimensional atoms. The "Offset Volume" section presents the algorithm for computing the volume of offset model of the three-dimensional molecule. The "Experiments" section presents the experimental results which verify the correctness and efficiency of the Beta-decomposition algorithm. Then, the article concludes in the "Conclusion" section.

Related Works

There were many studies for computing the volume and the boundary area of a set of spherical atoms in $\mathbb{R}^3$, particularly in computational chemistry and computational molecular biology. The first generation of studies was nonanalytic: the Monte Carlo simulation or enumeration of grid points. Shrake and Rupley (1973) reported an algorithm for computing the offset-volume by counting the number of sample points of each atom contained within the offset surface of any other atom.[6] They used 92 sample points for each offset sphere. Other studies used different types of grids.[7–12] Connolly (1985) also used a variation of counting grid points when he first computed the correction term for cusps in the Connolly volume.[13] The algorithm by Eisenhaber et al. (1995) was a typical example of using Monte Carlo simulation.[14] Even today, Monte Carlo simulation is used for computing the vdW-volume and vdW-area.[15, 16]

However, the computational cost for getting a high quality solution using the Monte Carlo simulation was prohibitive and motivated researchers to develop an analytic approach. Approximation was the first effort in this line of study. For example, Lee and Richards (1971) cut the accessible surface (i.e., the offset surface) of a molecule with a number of parallel planes with a predefined spacing so that the set of arcs was defined on each plane from the boundary of the model. Then, they derived an approximation formula for the boundary area within a slab between the two consecutive planes and added them up to get the estimation of the boundary area of the molecule. Obviously, this idea could be used for computing the volume as well.

We want to note here that there is no significant difference between the computation of the mass properties for van der Waals molecule and one for its offset model from the algorithmic point of view in that both take a set of spherical balls as input.

Combinatorial structure of the molecule

A general approach for analytically computing the correct volume of the union of balls requires the two technical issues (stated in "Introduction" section) to be resolved: the combinatorics issue and the decomposition issue. The combinatorics issue has been studied using various types of Voronoi diagrams of different distance measures and/or their derivative structures. The ordinary Voronoi diagram of points where each point represents an atom center was first used by Bernal and Finney (1967)[17] and Richards (1974)[1] for studying molecules. To reflect the size difference among atoms, Gellatly and Finney (1982) used radical planes between each pair of neighbor atoms (where the radical plane was identical to the power bisector between two different-sized atoms in the power diagram[18]). Observing that there might be difficulties such as engulfing cells in the power diagram of molecule, Gerstein et al. (1995) proposed to use a spherical segment as the bisector between two atoms.[19] Goede et al. (1997) proposed to use the Voronoi cell of the Voronoi diagram of atoms, frequently called the additively weighted Voronoi diagram in the computational geometry community, in the estimation of atomic volume,[20] and Will (1998) reported an algorithm for computing each Voronoi cell in the Voronoi diagram of atoms.[21] Eventually, Kim et al. (2005, 2006) successfully designed and implemented the edge-tracing algorithm for computing the complete Voronoi diagram of atoms.[22,23]

Looking at the other side of combinatorics, Edelsbrunner devised the theory of the alpha-shape and alpha-complex which nicely represented the proximity among points.[24] To apply the concept to molecules consisting of atoms with different radii, Edelsbrunner devised the concept of the weighted alpha-shape (and weighted alpha-complex) which was based on the power diagram and the regular triangulation.[25] The weighted alpha-shape (and the weighted alpha-complex) correctly represents the intersection information among atoms, but not the proximity information among non-intersecting atoms in the Euclidean distance sense. Hence, to use the weighted alpha-shape correctly for application problems based on the Euclidean distance metric in molecular biology, it is necessary to transform each problem into an intersection problem among appropriately inflated or shrunken atoms and its power diagram should be completely re-computed. The recently developed theory of the beta-complex overcomes this drawback of the weighted alpha-complex so that the Voronoi diagram is computed only once for the entire lifetime of a given atom set and stored in the quasi-triangulation format for any type of application problems for biomolecules.[26,27] The beta-complex not only efficiently solves geometry/topology problems in computational chemistry/biology but also brings otherwise computationally infeasible problems into a feasible space.

Decomposition of mass properties

In resolving the decomposition issue, there have been two main approaches: the cell decomposition approach and the approach based on the inclusion-exclusion principle.

Cell-decomposition. The cell-decomposition approach decomposes the molecular volume into a set of mutually exclusive regions where the volume of each region can be easily computed. Bernal and Finney (1967) first used the ordinary Voronoi diagram of points for computing the volume of liquid by summing up the volume of the Voronoi cell of each atom.[17] However, this approach had a problem with the boundary atoms because they might have either an unbounded or a too large Voronoi cell. Connolly's algorithm (1985) for the Connolly volume was also primarily based on the decomposition of the Connolly volume into the collection of four types of volume primitives.[13] In fact, Connolly computed the boundary area of the Connolly surface also by decomposing the boundary into convex patches, concave pathes, and saddle patches and applying the Gauss-Bonnet theorem.[28] Perrot et al. modified this algorithm for computing the area of accessible surface.[29] Others studies in[1,20,21] also fall into this category of using different types of Voronoi diagrams.

More significant studies of cell-decomposition mostly compute the intersection of each atom with its power cell because each atom is associated with a convex polyhedral cell in the power diagram. Then, the volume of the union of atoms is obtained by summing up the atomic volume within all power cells. Hence, the core problem is how to compute the volume of an atom intersecting with more than one half-space. Avis et al. (1988) reported a mathematical observation (without reporting implementation) for this approach.[30] Fraczkiewicz and Braun (1998) provided a formula for computing the molecular area and its derivative using the Gauss-Bonnet theorem after dividing a molecule using power bisectors.[31]

The algorithm reported by Dodd and Theodorou (1991) was significant.[3] After computing the intersection between each atom with the corresponding power cell, they further decomposed the atomic volume within a power cell into a set of finite cones, relatively few cones with possibly different shapes. They implemented and tested their algorithm using several molecules. Irisa's algorithm (1996)[32] was also similar to Dodd and Theodorou in both its concept and significance. McConkey et al. (2002) reported a similar algorithm but with the consideration of cases where an atom's center did not belong to the corresponding power cell was also provided.[33] Cazals' algorithm (2009) implemented using the CGAL library[34,35] also belongs to this category.[36,37] To our knowledge, Cazals' program was the most reliable and efficient implementation for the union of balls before the beta-decomposition scheme that is proposed in this article.

It is interesting and important to observe that all the previous algorithms of cell-decomposition schemes decomposed mass properties using the primary structure (i.e., the Voronoi or power cells). The proposed beta-decomposition algorithm falls into the cell-decomposition approach but differs from the previous studies in that the proposed beta-decomposition algorithm

decomposes the molecular mass property using the simplexes of the beta-complex that is an immediate consequence of the dual structure of the Voronoi diagram.

Inclusion-exclusion principle. It is known that the union of a set of elements can be found using the inclusion-exclusion principle. However, it is known to be difficult to get a correct solution in practice when the set is large because of the combinatorial explosion of terms in the expanded formula. For a set of n elements, the total number of terms in the inclusion-exclusion formula is $2^n - 1$. Dodd and Theodorou pointed out that the combinatorial explosion problem is a significant challenge in biomolecules.[3] To properly use this approach, therefore, it is critical to reduce the number of terms in the expanded formula.

Using the inclusion–exclusion principle for molecular mass properties started from Kratky (1978) who proved that the intersection of n disks of an identical size in $\mathbb{R}^2$ can always be reduced to the contribution from the intersections among up to three disks.[38] Based on this observation, Gibson and Scheraga (1987) reported an algorithm for computing the molecular volume and area for unequal atomic radii by removing the intersection terms lying entirely within some atom.[39] Due to the lack of systematic methodology, however, they had to actually compute all possible combinations of intersection to test whether each intersection occurs among a group of atoms. This leads to difficulties both for deriving the correct formula of various intersection cases and the expensive computational cost. Their algorithm incorporated cancelation of the intersection terms up to among five spheres. They actually implemented the algorithm and tested it with two tiny molecules: neo-pantane ($C_5H_{12} = CH_3C(CH_3)_2CH_3$ with four methyl groups) and benzene (C_6H_6). By approximating each methyl group (CH_3) with a spherical ball with a radius of 2.125 Å and the central C atom with a ball with a radius of 2.06 Å, they reduced the neo-pantane problem into a five-spheres problem. In a similar manner, the benzene problem was transformed into a six-spheres problem (where not all six spheres necessarily intersected at the same time). In fact, Dodd and Theodorou reported that the Gibson and Scheraga algorithm failed whenever six- or higher-fold intersections became significant contributors to the inclusion-exclusion formula.[3] We also had a similar experience with our implementation of the Gibson and Scheraga algorithm. Pavani and Ranghino (1982) reported an algorithm for computing molecular volume considering the intersections among up to three atoms and the cancelation of some intersection terms in the inclusion-exclusion formula.[40] Chkhartishvili (2001) also reported a formula for computing the volume of the intersection of three spheres with different radii.[41]

Naiman and Wynn (1992) generalized Kratky's observation to arbitrary dimensions: they found that the union of d-dimensional balls could be computed by the inclusion-exclusion formula containing only at most the intersection terms among $d+1$ balls if a simplicial complex which conveyed the intersection information among all balls was available.[42] They explicitly stated that the Delaunay triangulation was such a simplicial complex in the particular case of identically sized balls. They also showed how to select such a subset of the entire terms using the simplicial complex. Petitjean (1994) reported an analytic algorithm to compute 3- and 4-fold intersections among atoms and used them in the inclusion-exclusion formula.[43] He ran the implementation on 63 small compounds for computing the molecular volume and area. Edelsbrunner (1995) showed a way to reduce the number of terms more effectively using the alpha-shape of spherical balls[44,45] and reported its implementation in Ref. [46]. However, we were not able to obtain the software for the benchmark test.

We emphasize that the inclusion-exclusion principle requires formulae for all possible cases of intersections among $d+1$ atoms in $\mathbb{R}^d$. Even in $\mathbb{R}^3$, it still remains a challenge to verify that the reported formulae such as those reported in[39,46] correctly cover all possible cases of the intersections among the four atoms. Hence, a simpler and verifiable approach is desirable.

Voronoi Diagram, Quasi-Triangulation, and the Beta-Complex

A brief review of the geometric/mathematical constructs related to the Beta-decomposition is in order. $A = \{a_1, a_2, \ldots, a_n\}$ is the set of vdW-atoms $a_i = (p_i, r_i)$ in $\mathbb{R}^3$. In this article, A denotes a vdW-molecule. The *Voronoi diagram* $\mathcal{VD}$ of A is defined as $\mathcal{VD} = \{\mathrm{VC}(a_1), \mathrm{VC}(a_2), \ldots, \mathrm{VC}(a_n)\}$, where $\mathrm{VC}(a_i)$ denotes the *Voronoi cell* for a_i defined as $\mathrm{VC}(a_i) = \{x \in \mathbb{R}^3 | d(x,p_i) - r_i \leq d(x,p_j) - r_j \text{ for } i \neq j\}$ and $d(x,y)$ denotes the Euclidean distance between x and y. $\mathcal{VD}$ is represented by the quadruplet $(V^{\mathcal{V}}, E^{\mathcal{V}}, F^{\mathcal{V}}, C^{\mathcal{V}})$: $V^{\mathcal{V}} = \{v_1^{\mathcal{V}}, v_2^{\mathcal{V}}, \ldots\}$, $E^{\mathcal{V}} = \{e_1^{\mathcal{V}}, e_2^{\mathcal{V}}, \ldots\}$, $F^{\mathcal{V}} = \{f_1^{\mathcal{V}}, f_2^{\mathcal{V}}, \ldots\}$, and $C^{\mathcal{V}} = \{c_1^{\mathcal{V}}, c_2^{\mathcal{V}}, \ldots c_n^{\mathcal{V}}\}$ are the sets of the Voronoi vertices (V-vertices), Voronoi edges (V-edges), Voronoi faces (V-faces), and Voronoi cells (V-cells) in $\mathcal{VD}$, respectively. The computation of $\mathcal{VD}$ takes $O(n^3)$ time in the worst case for n general spheres but it takes $O(n)$ time on average for molecules consisting of n atoms, both in $\mathbb{R}^3$. For the details of the algorithm, see Refs. [22] and [23]. One issue to note is when five or more atoms define a common V-vertex. This case is possible from theoretical point of view and occurs when the boundaries of the atoms are equi-distant from the V-vertex. While this situation may occur in reality and cause a serious computational problem, such a degeneracy has been recently well-solved by the exact computation technique proposed by Sugihara's group.[47] There is a rich set of literature on the Voronoi diagram of the three-dimensional sphere set regarding on its definition, algorithms, and applications.[22,48–57]

Given the Voronoi diagram, the quasi-triangulation $\mathcal{QT}$ of A is the dual structure of $\mathcal{VD}$. Each V-vertex maps to a tetrahedral cell simplex (q-cell); each V-edge maps to a triangular face simplex (q-face); each V-face maps to an edge simplex (q-edge); each V-cell maps to a vertex simplex (q-vertex). The conversion from $\mathcal{VD}$ to $\mathcal{QT}$ (or vice versa) takes $O(m)$ time in the worst case where m is the number of the q-simplexes in the quasi-triangulation (or equivalently, the number of the topological entities in $\mathcal{VD}$). The conversion does not include any floating-point arithmetic. A quasi-triangulation is precisely represented by a quadruplet $(V^{\mathcal{Q}}, E^{\mathcal{Q}}, F^{\mathcal{Q}}, C^{\mathcal{Q}})$: $V^{\mathcal{Q}} = \{v_1^{\mathcal{Q}}, v_2^{\mathcal{Q}}, \ldots v_n^{\mathcal{Q}}\}$, $E^{\mathcal{Q}} = \{e_1^{\mathcal{Q}}, e_2^{\mathcal{Q}}, \ldots\}$, $F^{\mathcal{Q}} = \{f_1^{\mathcal{Q}}, f_2^{\mathcal{Q}}, \ldots\}$, and $C^{\mathcal{Q}} = \{c_1^{\mathcal{Q}}, c_2^{\mathcal{Q}}, \ldots\}$

are the sets of the q-vertices, q-edges, q-faces, and q-cells in $\mathcal{QT}$, respectively. Each simplex in $\mathcal{QT}$ is called a *q*-simplex. The topology of quasi-triangulation is stored in the (extended) Inter-world data structure that requires only $O(n)$ memory where n is the number of atoms in A.[27,58] Unlike the Delaunay triangulation and the regular triangulation, the quasi-triangulation is not necessarily a simplicial complex due to anomalies and small worlds. For details, see Refs. [58], and [59].

Given a quasi-triangulation, the idea of the beta-complex and beta-shape can be explained as follows (but not computed in this way). Think of a three-dimensional space filled with soft matter with some spherical rocks of varying radii scattered within the matter. Carving out the matter with an omnipresent spherical cutter, called a *probe*, whose radius is β, will result in a shape which is called a *beta-hull*. Suppose that we have a beta-hull for a molecule A. We straighten the surface of the beta-hull so that each spherical triangle on the boundary of the beta-hull becomes a planar triangle where its vertices are the centers of the atoms touching the spherical triangle. Then, the circular arcs of the beta-hull are transformed to edges of line segment. The straightened object bounded by planar facets is the *beta-shape* $\mathcal{S}_\beta$ of A corresponding to the probe radius β. The boundary of the beta-shape, denoted by $\partial\mathcal{S}_\beta$, is represented by a triplet $(V^{\mathcal{S}}, E^{\mathcal{S}}, F^{\mathcal{S}})$, where $V^{\mathcal{S}}$, $E^{\mathcal{S}}$, and $F^{\mathcal{S}}$ are the sets of the β-vertices, β-edges, and β-faces, respectively. A beta-complex $\mathcal{C}_\beta$ is a subset of the quasi-triangulation $\mathcal{QT}$ where each simplex in $\mathcal{C}_\beta$ lies within or on the boundary of the underlying space of the corresponding beta-shape. This intuitive description about the beta-complex and beta-shape is similar to that of the alpha-shape for points in Ref. [24]. Given a molecule A and its quasi-triangulation, a unique beta-complex (and therefore a unique beta-shape as well) is defined for a given probe radius β. The computation of the beta-complex takes $O(\log m + k)$ time in the worst case if binary search is used for querying simplexes in the quasi-triangulation, where m and k represent the numbers of simplexes in the quasi-triangulation and the beta-complex, respectively.[27] A more generalized approach to the simplex query was recently proposed.[60] It is known that $m = O(n^2)$ in the worst case for a set of n general spheres but $m = O(n)$ on average for molecules consisting of n atoms in $\mathbb{R}^3$. In this article, the beta-complex and beta-shape are those defined for $\beta = 0$ unless otherwise stated and they are called the *zero beta-complex* and the *zero beta-shape*, respectively. See Refs. [26, 27, 60] for the formal definitions and algorithms of the beta-shape and beta-complex.

A beta-complex $\mathcal{C}_\beta$ is represented by a quadruplet $(V^{\mathcal{C}}, E^{\mathcal{C}}, F^{\mathcal{C}}, C^{\mathcal{C}})$ where $V^{\mathcal{C}}$, $E^{\mathcal{C}}$, $F^{\mathcal{C}}$, and $C^{\mathcal{C}}$ are the sets of the β-vertices, β-edges, (triangular) β-faces, and (tetrahedral) β-cells, respectively. We call each simplex in a beta-complex a *β-simplex*. For notational convenience, a q-edge (or β-edge) may be represented by $e = (a_i, a_j)$, $i \neq j$, where the centers of a_i and a_j define e. Similarly, a q-face (or β-face) and a q-cell (or β-cell) can be represented by $f = (a_i, a_j, a_k)$, $i \neq j \neq k$, and $c = (a_i, a_j, a_k, a_l)$, $i \neq j \neq k \neq l$, respectively. The orientations of e, f, and c are consistently maintained with respect to the orientations of the corresponding V-entities in $\mathcal{VD}$. Each β-simplex $\sigma \in \mathcal{QT}$ takes one of the four bounding states for a given β value: *exterior, singular, regular*, or *interior*. σ is singular at β if it does not bound any higher-dimensional β-simplex in $\mathcal{C}_\beta$. Hence, it is exposed to air in its entirety and belongs to $\partial\mathcal{S}_\beta$. σ is regular at β if it bounds a higher-dimensional β-simplex in $\mathcal{C}_\beta$. Hence, it is also exposed to air and also belongs to $\partial\mathcal{S}_\beta$. σ is interior at β if it is the intersection between higher-dimensional β-simplexes. Otherwise, σ is exterior. Hence, it is not exposed to air and does not belong to $\partial\mathcal{S}_\beta$. Therefore, the set of singular β-simplexes and regular β-simplexes altogether define the boundary of the beta-shape. Each q-simplex in $\mathcal{QT}$ may or may not become a member of the beta-complex depending on the given β value. Each q-simplex is associated with an interval, called the *β-interval*, of real value. The collection of the β-intervals corresponding to the bounding states for each q-simplex is called the *β-span*. Therefore, the computation of the beta-complex (and/or the beta-shape) is in fact the search of appropriate q-simplexes whose β-interval contains the given value of β. This search can be efficiently done via binary search if the β-spans of all q-simplexes are sorted. For the details about their definitions and algorithms, please refer to Refs. [26] and [27].

One important note: The beta-complex is very useful in solving geometry- and/or shape-related problems in computational chemistry and computational biology because it has both of the following dual properties:

- *Precise Proximity:* The beta-complex has the precise proximity information among all atoms both within and on the boundary of a molecule where the boundary is defined by a probe.
- *Concise Abstraction:* The beta-complex has only the topology information of the nearest neighbors for each atom in the form of the connectivity among the β-vertices, β-edges, β-faces, and β-cells.

Additionally, the beta-complex has the following property:

- *Multiresolution:* The beta-complex can be defined with respect to the probe of desired radius.

Hence, the beta-complex can be used to efficiently solve any type of shape-related problem requiring arbitrary level of precision if the radius of each atom is additionally available. The computation of the vdW-volume and vdW-area discussed in this article is such an example. In addition, the beta-complex can also be used to efficiently solve any shape-related problem requiring only the approximation of the shape is sufficient. An example falling into this category is to find an optimal superposition of two protein structures. The multiresolution, also called the level-of-detail, capability facilitates the beta-complex used for problems requiring different granularity of approximation. Figure 1 of the Supporting Information shows an example of an atom set in the plane and its Voronoi diagram, quasi-triangulation, beta-complex, and beta-shape.

Area of the Union of Disks in $\mathbb{R}^2$

Suppose that $A^2 = \{a_1, a_2, \ldots, a_n\}$ is a set of planar, circular atoms $a_i = (p_i, r_i)$ in $\mathbb{R}^2$. We want to compute the area of the two-dimensional molecular region (i.e. the union of the atoms) which is a polygon bounded by arcs in $\mathbb{R}^2$. We decompose the entire molecular region into a number of subregions, called

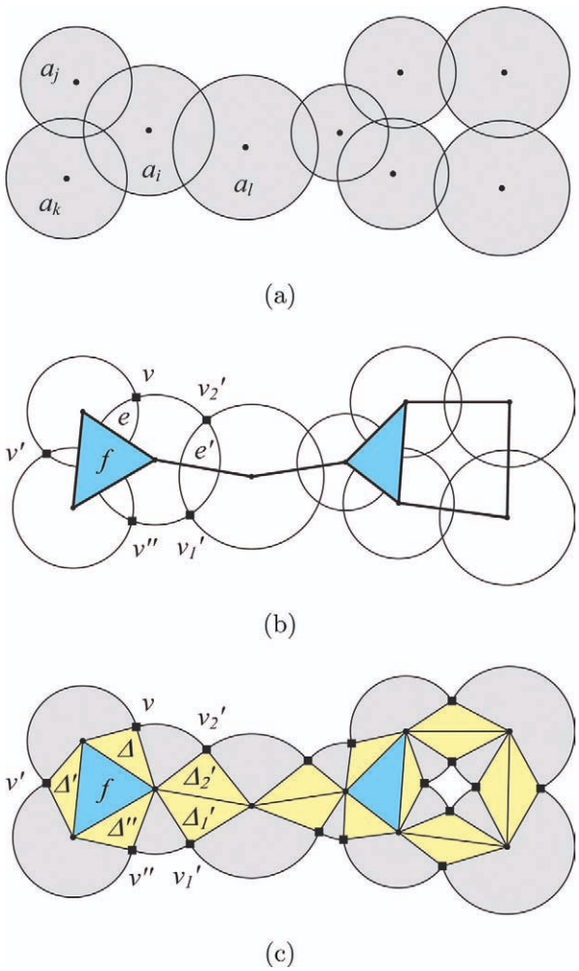

Figure 1. The beta-decomposition idea for a two-dimensional molecule. a) The molecule, b) the zero beta-complex and the m-vertices, and c) the decomposition of the molecular region via the three types of area primitives.

primitives, whose areas can easily be computed. Then, the area of the entire molecular region is computed by the summation of the areas of the primitives. We explore this idea $\mathbb{R}^2$ here and extend it to $\mathbb{R}^3$ in later sections.

A well-established relationship exists between the simplexes on the boundary of the beta-shape and the vertices and edges on the boundary ∂A^2 of the molecular region. Suppose that $a_i \cap a_j \neq \emptyset$ where a_i and a_j are distinct atoms contributing to ∂A^2. Then, $\partial a_i \cap \partial a_j$ defines one or two vertices on ∂A^2. We call such a vertex on ∂A^2 an *m-vertex* where "m" denotes the "model." The arcs on ∂A^2 are called *m-edges*. A face on the the molecular boundary can be similarly defined in $\mathbb{R}^3$ and called an *m-face*. The following lemma is from Lemma 4 of Ref. [61].

Lemma 1. *In $\mathbb{R}^2$, a β-vertex maps to one or more m-edge(s), a regular β-edge maps to an m-vertex, and a singular β-edge maps to two m-vertices.*

Figure 1a shows a molecule consisting of eight atoms in the plane. Figure 1b additionally shows its zero beta-complex and some m-vertices (e.g., v, v_1', and v_2') denoted by the black rectangles. Figure 1c shows the molecular region decomposed into a number of primitives. We define three *area primitives* as follows:

- Area of the *interior triangle* (*Tri*3): The area of the interior (triangular) β-face (in blue) of the beta-complex and denoted by *Tri*3 because its three vertices are atom centers.
- Area of the *apex-exposed triangle* (*Tri*2): The area of the triangle (in yellow) corresponding to a regular β-edge on the boundary of the beta-shape and denoted by *Tri*2 because its two vertices are atom centers.
- Area of the *exposed atomic occupation* (*Fan*1): The area of the circular atomic region exposed to air (in gray) corresponding to a β-vertex on the boundary of the beta-shape. We denote it by *Fan*1 because it corresponds to a single atom and has a shape similar to an oriental fan.

Note that each area primitive corresponds to a simplex in the beta-complex. Then, the area of the entire molecular region can be correctly computed if the area of each area primitive is correctly computed and summed up. The combinatorial point of view is particularly important in this computation.

Area of the interior triangle (*Tri*3): β-face

Let $f = (a_i, a_j, a_k)$, $i \neq j \neq k$, be an interior (triangular) β-face. The computation of the area of f is trivial if f is an ordinary triangle. However, one should be careful with interior β-faces for certain cases.

Figure 2a shows a molecule consisting of five atoms and its Voronoi diagram and Figure 2b shows the corresponding quasi-triangulation. The five atoms in the figure do not intersect each

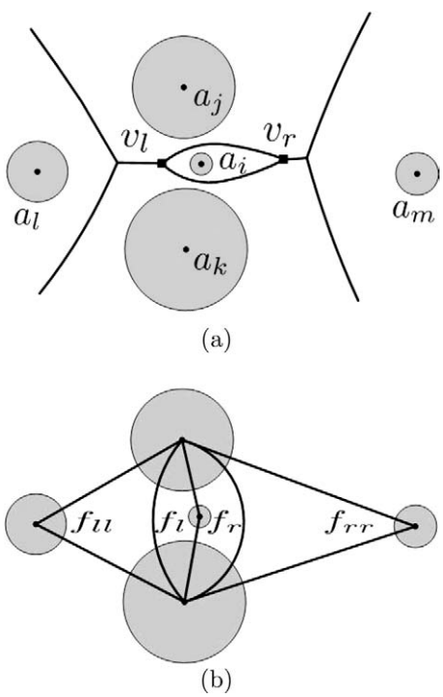

Figure 2. An interior triangle with a negative area. a) the Voronoi diagram of a molecule and b) the corresponding quasi-triangulation containing a triangular face simplex with a negative area.

other. However, consider this example such that the radii of the atoms are reduced by a constant from the real atoms where one intersects another. Note that the two Voronoi vertices v_l and v_r are defined by the three atoms a_i, a_j, and a_k. Hence, there are two q-faces, f_l and f_r, in the quasi-triangulation defined by the centers of these atoms: $f_l = (a_i, a_j, a_k)$ and $f_r = (a_i, a_k, a_j)$. Note that the two q-faces f_l and f_r share two common q-edges because the two V-vertices v_l and v_r are connected by two distinct V-edges. These two triangles look identical in the Euclidean space because they have identical sets of three q-vertices and therefore their areas are identical. On the other hand, the orientations of these triangles viewed in the Euclidean space are opposite to each other. To be specific, f_l is counterclockwise oriented and f_r is clockwise oriented in the Euclidean space. Hence, one q-face (f_l in this example) has a positive area, while the other (f_r in this example) has a negative area. We need to clarify such a case because it may arise in real situations. We borrow Lemma 2 from Ref. [27].

Lemma 2. *Suppose that three atoms define two (adjacent) q-faces f' and f'' in $\mathbb{R}^2$. Then, f' has a positive area if, and only if, f'' has a negative area, and vice versa.*

We call such a pair of q-faces (i.e., f' and f'' in Lemma 2) the *twin q-faces* and the pair of V-vertices v_l and v_r above the *twin V-vertices*. If twin q-faces are members of the beta-complex, they are also called the twin β-faces. When there are such twin β-faces in the zero beta-complex, it turns out that mechanically adding all signed areas automatically cancels out the positive and negative triangle pairs. Twin q-faces are adjacent to each other because they share two q-edges. In addition, it is also possible that a beta-complex has a single β-face with a negative area. In such a case, mechanically adding the negative area also produces a correct solution.

Area of apex-exposed triangle (*Tri*2): β-edge

Consider $e = (a_i, a_j)$, $i \neq j$, is a regular β-edge. The β-edge e in Figure 1b is an example because e bounds a β-face (a blue triangle). Suppose that the m-vertex v is the intersection between ∂a_i and ∂a_j. Then, a triangle $\triangle$ in Figure 1c is defined by e and v. Note that e is shared by two incident triangles: the yellow one ($\triangle$) and the blue one, as shown in Figure 1c. The yellow triangle is called an *apex-exposed triangle* because one of its vertices (i.e., v) is exposed to air.

Consider a singular β-edge $e' = (a_i, a_l)$ in Figure 1b. In this case, e' maps to two m-vertices v'_1 and v'_2 (by Lemma 1) which are the intersections between ∂a_i and ∂a_l. Then, two triangles $\triangle'_1$ and $\triangle'_2$ are defined by v'_1 and v'_2 as shown in Figure 1c, respectively. Hence, both $\triangle'_1$ and $\triangle'_2$ are apex-exposed triangles (yellow triangles). This observation therefore proves the following lemma.

Lemma 3. *In $\mathbb{R}^2$, a regular β-edge defines an apex-exposed triangle and a singular β-edge defines two apex-exposed triangles.*

It is obvious to show that the two apex-exposed triangles corresponding to a singular β-edge are congruent to each other and share the β-edge.

Area of exposed atomic occupation (*Fan*1): β-vertex

Suppose that v is a β-vertex and a is the corresponding atom. If v is singular, a does not intersect any other atom and therefore all points on the boundary of a are exposed to air. Hence, the entire atom a contributes to the molecular region. If v is interior, the entire a is already accounted for in the molecular region by its incident (interior) β-faces. If v is regular, a is partially accounted for in the molecular region by some interior β-faces and some apex-exposed triangles.

Consider an atom a_j in Figure 1a where $a_i \cap a_j \cap a_k \neq \emptyset$. The β-vertex corresponding to a_j is regular. Then, the portion of the atomic area which corresponds to the exposed atom boundary of a_j can be computed by subtracting the already-accounted-for areas within the atom a_j where the already-accounted-for areas correspond to the incident (blue) interior triangles (*Tri*3's) and the incident (yellow) apex-exposed triangles (*Tri*2's). For example, see the blue triangle (f) and the two yellow triangles ($\triangle$ and $\triangle'$) in Figure 1c which intersect a_j. Hence, the exposed region of a_j can be computed by subtracting these triangles from a_j. The remaining atomic region is thus named as an *exposed atomic occupation* (*Fan*1) and may consist of one or more circular sectors attached to an atom center. In particular, it consists of more than one sector if and only if the beta-shape is non-manifold at the β-vertex corresponding to the atom. For example, a_j has one sector and each of a_i and a_l has two distinct sectors.

Let t be a triangle. Suppose that the center of an atom α defines a vertex of t. Then, $t \cap \alpha$ is called the *α-wedge* of t. The triangle t may be either an interior β-face or an apex-exposed triangle. Note that the area of an interior β-face f already includes the area of the three wedges corresponding to the vertices of f. Let $Fan1_v$ be the exposed atomic occupation of an atom a which corresponds to a β-vertex v. Let Area(X) be the area of a shape X. Then,

$$\text{Area}(Fan1_v) = \text{Area}(a) - \sum_{f \in F_v^{\mathcal{C}}} \text{Area}(a \cap f) - \sum_{e \in E_v^{\mathcal{S}}} \text{Area}(a \cap \triangle) \tag{1}$$

where $f \in F_v^{\mathcal{C}} \subseteq F^{\mathcal{C}}$ is a β-face incident to v, $e \in E_v^{\mathcal{S}} \subseteq E^{\mathcal{C}}$ is a β-edge incident to v, and $\triangle$ is an apex-exposed triangle defined by e and the corresponding m-vertex. Recall that $E^{\mathcal{C}}$ and $F^{\mathcal{C}}$ are the sets of the β-edges and the β-faces in the beta-complex.

Area of the entire molecular region

Lemma 4. *Tri3's, Tri2's, and Fan1's altogether tessellate the entire molecular region of A^2 (See the Supporting Information for the proof).*

Let $Tri3_f$ denote that f is an interior β-face and $Tri2_e$ is the apex-exposed triangle(s) corresponding to a β-edge e. Then, Lemma 4 proves the following theorem.

Theorem 5. *The area of the molecular region of A^2 in $\mathbb{R}^2$ is computed by*

$$\text{Area}(A^2) = \sum_{f \in F^{\mathcal{C}}} \text{Area}(Tri3_f) + \sum_{e \in E^{\mathcal{S}}} \text{Area}(Tri2_e) + \sum_{v \in V^{\mathcal{S}}} \text{Area}(Fan1_v) \quad (2)$$

where $F^{\mathcal{C}}$ is the β-face set of the zero beta-complex, and $V^{\mathcal{S}}$ and $E^{\mathcal{S}}$ are the sets of β-vertices and the β-edges of the corresponding zero beta-shape, respectively.

Lemma 6. *Given the zero beta-complex of a molecule A^2 in $\mathbb{R}^2$, $\text{Area}(A^2)$ can be computed in $O(m)$ time in the worst case where m is the number of simplexes in the zero beta-complex of A^2 (See the Supporting Information for the proof).*

Because $|F^{\mathcal{C}}| = |E^{\mathcal{S}}| = |V^{\mathcal{Q}}| = O(n)$ in the worst case in $\mathbb{R}^2$ by the equations in the proof of Lemma 4, the following holds.

Corollary 7. *Given the zero beta-complex of a molecule A^2 in $\mathbb{R}^2$, $\text{Area}(A^2)$ can be computed in $O(n)$ time in the worst case where n is the number of atoms of A^2.*

Note: One important note on the numerical stability for the beta-decomposition: There can be two approaches to cancel out the areas of twin β-faces as follows.

- Include twins: Add up the areas of all β-faces including all twin β-faces.
- Less twins: Skip all twin β-faces when adding up the areas of β-faces.

It may at first seem that the "include twins" approach might be better because the β-faces with the opposite signs will eventually cancel out anyway. However, we do not recommend this approach from the numerical stability point of view because the area computation uses floating-point arithmetic. Let Δ^+ and Δ^- be the positive and negative areas of twin β-faces, respectively. It turns out that the absolute value of the area of twin β-faces tends to be small compared to that of the other triangles. Sometimes, it can be tiny. In other words, $|\Delta^+|$ and $|\Delta^-|$ are usually very small when the distribution of disk sizes is from the distribution of atom sizes in molecules. Let $\sum$ be the summation of the areas of all β-faces (including Δ^+) except Δ^-. Hence, $\sum$ is usually a very big number compared to $|\Delta^-|$. Then, $\sum -|\Delta^-|$ can cause a loss of many significant bits in the computation result from the numerical analysis point of view. Thus, we recommend to employ the "less twins" approach in writing codes. The difference of the computational requirement between the two approaches is insignificant.

Volume of the Union of Balls in $\mathbb{R}^3$

We now extend the idea to $\mathbb{R}^3$. Suppose that a molecule A is given in $\mathbb{R}^3$. A close relationship exists between the β-simplexes on the boundary of the zero beta-shape and the m-vertices, m-edges, and m-faces on the vdW-boundary ∂A. The following lemma is from Lemma 9 of Ref. [61].

Lemma 8. *In $\mathbb{R}^3$, a β-vertex maps to an m-face(s), a β-edge maps to an m-edge(s), and a β-face maps to an m-vertex(es). A regular β-face maps to an m-vertex and a singular β-face maps to two m-vertices.*

Lemma 8 states the relationship between the boundary of the zero beta-shape and the vdW-boundary. Hence, the interior tetrahedral β-cell is additionally required in order to account for the interior of the vdW-molecule in the computation of the vdW-volume. Similar to $\mathbb{R}^2$, the vdW-region in $\mathbb{R}^3$ can be decomposed into a number of primitives where the volume of each primitive can be easily computed. We define four types of volume primitives in $\mathbb{R}^3$ as follows:

- Volume of the interior tetrahedron (*Tetra*4): The volume of an interior β-cell in the beta-complex and denoted by *Tetra*4 because all four of its vertices are atom centers.
- Volume of the apex-exposed tetrahedron (*Tetra*3): The volume of a tetrahedron corresponding to a β-face on the boundary of the beta-shape and denoted by *Tetra*3 because its three vertices are atom centers.
- Volume of the edge-exposed go-stone (*Gostone*2): The volume of the intersection between two atoms, called a go-stone, corresponding to a β-edge on the boundary of the beta-shape and denoted by *Gostone*2 because it is related with two atoms.
- Volume of the exposed atomic occupation (*Fan*1): The volume of the atomic region exposed to air corresponding to a β-vertex on the boundary of the beta-shape and denoted by *Fan*1 because it is related with a single atom. We still call it "Fan" even though it is a three-dimensional object with its shape possibly quite different from an oriental fan.

Note that each volume primitive corresponds to a β-simplex in the zero beta-complex and the vdW-volume can be correctly computed if the volume of each volume primitive is correctly computed and summed up.

Figure 3 shows an example illustrating the primitives in $\mathbb{R}^3$. For presentation convenience, the example in Figure 3a has only three atoms and its zero beta-complex consists of three regular β-vertices, three regular β-edges, and one singular β-face but no β-cell. There are two m-vertices (at the intersection of the atom boundaries) corresponding to the singular β-face. Figure 3b shows the two apex-exposed tetrahedra (Tetra3), corresponding to the two m-vertices. Figure 3c shows the two apex-exposed tetrahedra merged together. Figure 3d shows the go-stone (Gostone2) corresponding to one of the three regular β-edges. Figure 3e shows both the go-stone and the two apex-exposed tetrahedra. Figure 3f shows the go-stone subtracted by the two apex-exposed tetrahedra. Figure 3g shows the three go-stones together after they are all subtracted by the apex-exposed tetrahedra. Figure 3h shows one of the atoms and the apex-exposed tetrahedra. Figure 3i shows the atom subtracted by the apex-exposed tetrahedra. Figure 3j shows all of these primitives together.

Interior tetrahedron (*Tetra*4): β-cell

Let $\tau = (a_i, a_j, a_k, a_l)$ be an interior (tetrahedral) β-cell of the zero beta-complex where the centers of four atoms a_i, a_j, a_k, and a_l

define τ. Let $I = (x_i, y_i, z_i, 1)^T$, $J = (x_j, y_j, z_j, 1)^T$, $K = (x_k, y_k, z_k, 1)^T$, and $L = (x_l, y_l, z_l, 1)^T$. (x_i, y_i, z_i) is the center of a_i and $(x_i, y_i, z_i, 1)^T$ denotes the transpose of the vector $(x_i, y_i, z_i, 1)$. Then, the signed volume of τ is given by the determinant

$$\mathrm{Vol}(\mathit{Tetra4}_\tau) = \frac{1}{6} \mid I \; J \; K \; L \mid . \tag{3}$$

The order of the vertices in the tetrahedron is significant because exchanging two rows or columns in the determinant changes the sign of $\mathrm{Vol}(\mathit{Tetra4}_\tau)$. This implies that the same set of four vertices may define two distinct tetrahedra of the same volume but with opposite signs.

According to the beta-complex theory, a (tetrahedral) β-cell in the beta-complex in $\mathbb{R}^3$ may have either a positive signed volume or a negative signed volume.[27] If there is a β-cell τ with a negative volume, there is also another β-cell τ' which occupies exactly the same Euclidean region as τ but with a positive volume. Hence, an observation similar to Lemma 2 for $\mathbb{R}^2$ can be made for $\mathbb{R}^3$. The following lemma is from Ref. [27].

Lemma 9. *Suppose that four atoms define two (adjacent) q-cells τ' and τ'' in $\mathbb{R}^3$. Then, τ' has a positive volume if, and only if, τ'' has a negative volume, and vice versa.*

In Lemma 9, the two q-cells τ' and τ'' are called the *twin q-cells* and they may share two, three, or sometimes all four q-faces. When there are twin β-faces in the zero beta-complex, we can simply add the signed volumes of all the β-cells to compute the correct vdW-volume. A quasi-triangulation may have a hierarchy of worlds where there can be one or more small world(s) underneath the root world.[27,58,59] Being the subset of a quasi-triangulation, the zero beta-complex has a root world which may or may not contain one or more small world(s). By definition, the vdW-volume corresponding to the atoms in small worlds in the quasi-triangulation is a subset of the vdW-volume corresponding to the atoms of the root world. Hence, we must ignore the β-simplexes constituting small worlds in the computation of vdW-volume regardless it has a positive or negative volume. The following lemma immediately stands.

Lemma 10. *Let τ be a β-cell in the zero beta-complex of a molecule A. If τ does not belong to the root world in the beta-complex, τ does not contribute to the vdW-volume of A.*

Therefore, in this article, we assume that all β-cells are in the root world. In fact, all protein models that we tested have the root world only in their zero beta-complexes. It turns out that not only most protein models but also randomly generated spherical ball sets usually have extremely few twin β-cells.

Apex-exposed tetrahedron (*Tetra*3): β-face

Consider $f = (a_i, a_j, a_k)$ is a regular β-face. Then, f bounds a β-cell, say τ. Suppose that the m-vertex v corresponding to f is the intersection among ∂a_i, ∂a_j, and ∂a_k. Then, a tetrahedron $\triangle$ is defined by f and v. Note that f is shared by two incident tetrahedra: $\triangle$ and τ. We call the tetrahedron $\triangle$ an *apex-exposed tetrahedron* because one of its vertices is exposed to air.

Consider $f' = (a'_i, a'_j, a'_k)$ to be a singular β-face. Then, f' bounds no interior β-cell. In this case, f' maps to two m-vertices v'_1 and v'_2 at the intersection among $\partial a'_i$, $\partial a'_j$, and $\partial a'_k$. Then, a tetrahedron $\triangle'_1$ is defined by f' and v'_1 and another tetrahedron $\triangle'_2$ is defined by f' and v'_2. Both are apex-exposed tetrahedra. This observation is consistent with Lemma 8 and stated as the following lemma.

Lemma 11. *A regular β-face f produces an apex-exposed tetrahedron and a singular β-face produces two apex-exposed tetrahedra.*

Note that the two apex-exposed tetrahedra corresponding to a singular β-face are congruent to each other and share the β-face. Hence, the volume of the apex-exposed tetrahedra corresponding to a singular β-face can be computed by twice the volume of either apex-exposed tetrahedron. An interior β-face (which is not on the boundary of the beta-shape) does not make any apex-exposed tetrahedron.

Edge-exposed go-stone (*Gostone*2): β-edge

Consider a β-edge $e = (a_i, a_j)$ where $a_i \cap a_j \neq \emptyset$, $i \neq j$, where e is on the boundary of the beta-shape. Suppose that e is singular. If the exposed atomic occupation of both a_i and a_j independently contribute to the vdW-volume (without considering the fact $a_i \cap a_j \neq \emptyset$), the volume corresponding to $a_i \cap a_j$ is twice accounted for. Hence, if $a_i \cap a_j \neq \emptyset$, the volume of $a_i \cap a_j$ should be subtracted.

We call $a_i \cap a_j$ the *go*-stone $\mathit{Gostone}2(a_i, a_j)$ because the shape of $a_i \cap a_j$ is similar to the shape of a go-stone with a sharp circular edge. See Figure 3. The shape of $\mathit{Gostone}2(a_i, a_j)$ can be asymmetric around the plane passing through the circular edge. Let π be the plane, called the *radical plane*, passing through the circular edge of a go-stone. Then π cuts both a_i and a_j to define two spherical caps and the volume of the $\mathit{Gostone}2(a_i, a_j)$ can be computed as the sum of these two caps:

$$\mathrm{Vol}(\mathit{Gostone}2(a_i, a_j)) = \mathrm{Vol}(\kappa_i) + \mathrm{Vol}(\kappa_j) \tag{4}$$

where κ_i and κ_j are the spherical caps of a_i and a_j cut by π, respectively. The volume of a spherical cap κ is easily computed by

$$\mathrm{Vol}(\kappa) = \frac{1}{3}\pi h^2 (3r - h) \tag{5}$$

where h is the height of the spherical cap which is defined on the sphere with a radius r[62] (See Fig. 4). Note that eq. (5) also holds for the spherical cap larger than a hemisphere (i.e., when $h > r$).

Suppose that e' is a *regular* β-edge. Then, not the entire $\mathit{Gostone}2(a_i, a_j)$ but its (angular) subset should be subtracted. This is because one or more angular segment around e' are not exposed to air. The go-stones in Figure 3g belong to this case. See Figure 5. Consider that the black circular dot at the center of the circle denotes a regular β-edge $e = (a_i, a_j)$, $i \neq j$, that

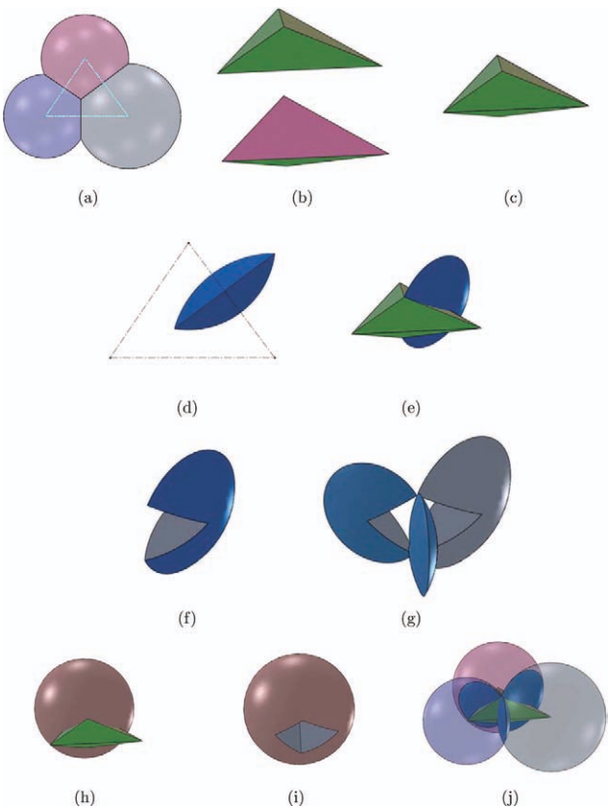

Figure 3. An example of the primitives in $\mathbb{R}^3$. a) Three mutually intersecting atoms, b) two apex-exposed tetrahedra shown separately, c) the two apex-exposed tetrahedra together, d) the go-stone corresponding to one of the regular β-edges, e) both the go-stone and the two apex-exposed tetrahedra, f) the go-stone subtracted by the apex-exposed tetrahedra, g) the three go-stones together after they are all subtracted by the apex-exposed tetrahedra, h) one of the atoms and the apex-exposed tetrahedra, i) the atom subtracted by the apex-exposed tetrahedra, and j) all of these primitives together.

is perpendicular to the plane π, the two broken line segments f_l and f_r denote the β-faces that are incident to e, the blue triangle denotes an interior tetrahedron, and the two yellow triangles denote apex-exposed tetrahedra. The solid circular arc with angle θ denotes the m-edge between the two m-vertices v_l and v_r and is defined as the intersection between the atoms

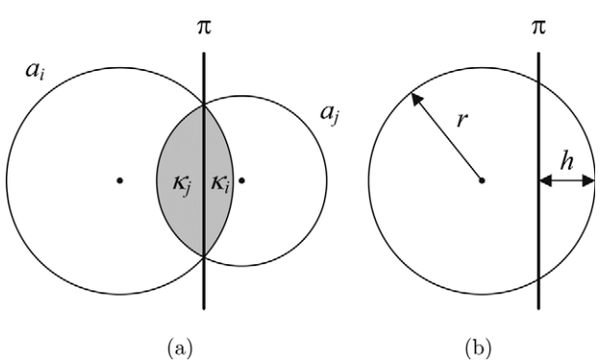

Figure 4. The volume of a go-stone. a) A go-stone decomposed into two spherical caps and b) the height of a spherical cap.

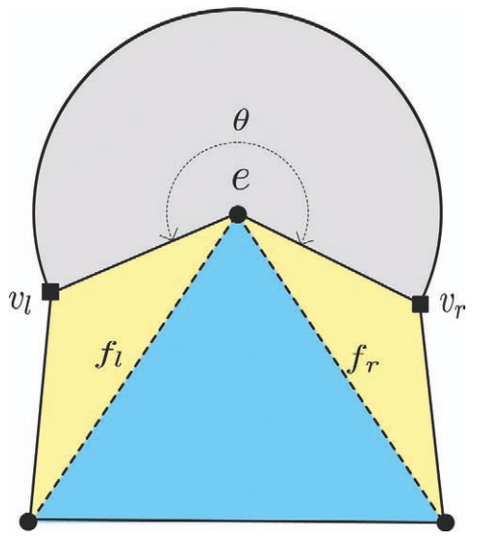

Figure 5. The cross-section view of the go-stone volume contributed by a regular β-edge.

a_i and a_j. Hence, this is the circular edge lying on the boundary of the *Gostone*2(a_i, a_j) and exposed to air.

Note that *Gostone*2(a_i, a_j) is divided into two parts: the *not-exposed-to-air* part and the *exposed-to-air* part. While the not-exposed-to-air part is already correctly accounted for by *Tetra*4's and *Tetra*3's incident to the β-edge e, the exposed-to-air part is twice accounted for in the vdW-volume. We call the exposed-to-air part of the go-stone the *edge-exposed go-stone*. Hence, the edge-exposed go-stone which corresponds to the solid circular arc in Figure 5 should be subtracted.

Knowing the volume of the entire *Gostone*2, the volume of an edge-exposed circular cones segment requires computing the *exposure angle* θ of an m-edge. Let t_l be the triangle defined by the m-vertex v_l and the β-edge e. Similarly, another triangle t_r is defined by the m-vertex v_r and e. Then, θ is the angle between the two triangular faces t_l and t_r.

A regular β-edge can be associated with more than one m-edge (Lemma 8) if the corresponding beta-shape is nonmanifold at the β-edge. Hence, in general, an exposure angle for each m-edge corresponding to a β-edge should be computed and summed up to compute the correct amount of exposed go-stone to be subtracted. Therefore, the volume of the edge-exposed go-stone of a β-edge e is given as

$$\text{Vol}(\textit{Gostone}2_e) = \text{Vol}(\textit{Gostone}2(a_i, a_j)) \times \frac{\Theta}{2\pi}. \qquad (6)$$

where $\Theta = \sum_k \theta_k$ where θ_k is the exposure angle of an m-edge e_k for a β-edge e. An interior β-edge does not make any edge-exposed go-stone.

Exposed atomic occupation (*Fan*1): β-vertex

Suppose that v is a β-vertex and a is the corresponding atom. If v is singular, the entire a contributes to the vdW-region. If v is interior, the entire a is already accounted for in the vdW-region

by its incident (interior) β-cells. If v is regular, a is partially accounted for in the molecular region. Recall its counterpart in $\mathbb{R}^2$ where we subtracted the appropriate atom wedges from the area of the entire atomic region. The computation of *exposed atomic occupation* (*Fan*1) in $\mathbb{R}^3$ is similar.

Suppose that an atom a defines one vertex of the β-cell τ. Then, $\tau \cap a$ is called the a-wedge of τ. Then, the volume of an a-wedge can be computed as

$$\frac{1}{3}r^3((\alpha+\beta+\gamma)-\pi) \tag{7}$$

where α, β, and γ are the interior angles of the spherical triangle $\partial a \cap \tau$ and r is the radius of a Ref. [62]. Note that the volume of an interior tetrahedron includes the volumes of four wedges of the corresponding β-cell. Then, the volume of *Fan*1 corresponding to v is defined as

$$\mathrm{Vol}(\mathit{Fan}1_v) = \mathrm{Vol}(a) - \sum_{\tau \in C_V^{\mathcal{C}}} \mathrm{Vol}(a \cap \tau) - \sum_{f \in F_V^{\mathcal{S}}} \mathrm{Vol}(a \cap \triangle) \tag{8}$$

where $\tau \in C_v^{\mathcal{C}} \subseteq C^{\mathcal{C}}$ is a β-cell incident to a β-vertex v, $f \in F_v^{\mathcal{S}} \subseteq F^{\mathcal{S}}$ is a β-face incident to v, and $\triangle$ denotes an apex-exposed tetrahedron defined by f and the corresponding m-vertex. In eq. (8), $a \cap \triangle$ also denotes a wedge.

Volume of the entire molecular region

Theorem 12. *The vdW-volume* $\mathrm{Vol}(A)$ *is computed by*

$$\mathrm{Vol}(A) = \sum_{\tau \in C^{\mathcal{C}}} \mathrm{Vol}(\mathit{Tetra}4_\tau) + \sum_{f \in F^{\mathcal{S}}} \mathrm{Vol}(\mathit{Tetra}3_f) - \sum_{e \in E^{\mathcal{S}}} \mathrm{Vol}(\mathit{Gostone}2_e) + \sum_{v \in V^{\mathcal{S}}} \mathrm{Vol}(\mathit{Fan}1_v) \tag{9}$$

where $C^{\mathcal{C}}$ is the β-cell set in the root world of the zero beta-complex and $F^{\mathcal{S}}$, $E^{\mathcal{S}}$, and $V^{\mathcal{S}}$ are the sets of the β-faces, β-edges, and β-vertices on the boundary of the zero beta-shape, respectively.

Theorem 12 holds because $\sum_{e \in E^{\mathcal{S}}} \mathrm{Vol}(\mathit{Gostone}2_e)$ is twice accounted for by $\sum_{\tau \in C^{\mathcal{C}}} \mathrm{Vol}(\mathit{Tetra}4_\tau) + \sum_{f \in F^{\mathcal{S}}} \mathrm{Vol}(\mathit{Tetra}3_f) + \sum_{v \in V^{\mathcal{S}}} \mathrm{Vol}(\mathit{Fan}1_v)$. Equation (9) correctly computes the vdW-volume regardless of whether the zero beta-shape is manifold or nonmanifold. The following corollary holds by the similar proof of Lemma 6.

Corollary 13. *Given the zero beta-complex of a molecule A in* $\mathbb{R}^3$*, the vdW-volume* $\mathrm{Vol}(A)$ *can be computed in* $O(m)$ *time in the worst case where m is the number of simplexes in the root world of the zero beta-complex of A.*

Corollary 14. *Given the zero beta-complex of a molecule A in* $\mathbb{R}^3$*, the vdW-volume* $\mathrm{Vol}(A)$ *can be computed in* $O(n)$ *time on average and* $O(n^2)$ *time in the worst case where n is the number of atoms in A.*

Note: Let ξ be the circle such that $\xi = \partial a_i \cap \partial a_j$. Let π be the plane containing ξ and δ be the disk defined as $\delta = \pi \cap (a_i \cap a_j)$. Let $\mathit{Cone}_i = \mathit{Cone}(a_i, \delta)$ be a circular cone with its apex at the center of a_i and the base δ. The formula for the volume of Cone_i is known. Cone_j is similarly defined. Let $\mathit{Cones}2 = \mathit{Cone}_i \cup \mathit{Cone}_j$. Then, eq. (9) can be rewritten as

$$\mathrm{Vol}(A) = \sum_{\tau \in C^{\mathcal{C}}} \mathrm{Vol}(\mathit{Tetra}4_\tau) + \sum_{f \in F^{\mathcal{S}}} \mathrm{Vol}(\mathit{Tetra}3_f) + \sum_{e \in E^{\mathcal{S}}} \mathrm{Vol}(\mathit{Cones}2_e) + \sum_{v \in V^{\mathcal{S}}} \mathrm{Vol}(\mathit{Fan}1'_v). \tag{10}$$

where $\mathit{Fan}1'_v = \mathit{Fan}1_v - \mathit{Cones}2_e$. Equation (10) provides an interpretation of the Beta-decomposition in the additive form with respect to the β-simplexes. **Beta-decomposition-vdW** algorithm in the Supporting Information summarizes the computation of the vdW-volume using eq. (9).

Boundary of the Union of Balls in $\mathbb{R}^2$ and $\mathbb{R}^3$

We also want to compute the area of the vdW-boundary. Similar to the vdW-volume, the main idea is to decompose the vdW-boundary into boundary segments using the zero beta-shape. Note that the β-simplexes in this section are those of the boundary of the zero beta-shape, not the beta-complex. We first present an algorithm in $\mathbb{R}^2$ and extend the idea to $\mathbb{R}^3$.

Length of the vdW-boundary in $\mathbb{R}^2$

Reconsider the planar molecule A^2 in Figure 1a where its boundary ∂A^2 consists of a set of circular arcs. We want to compute the total length of ∂A^2. The boundary primitive is the circular arc(s) on ∂A^2 exposed to air where its generating circle ∂a corresponds to a β-vertex v on the boundary of the beta-shape. This primitive is called the *exposed atomic (occupation) boundary (EAB)*.

If v is singular, the entire ∂a is exposed to air. Suppose that v is regular and another atom a' exists where $a \cap a' \neq \emptyset$. Then, the exposed atomic boundary *EAB* of a is given by

$$\partial a - (\partial a \subset a'). \tag{11}$$

Consider a_j in Figure 1a where $a_i \cap a_j \cap a_k \neq \emptyset$. The exposed atomic boundary of a_j is obtained by $\partial a_j - (\partial a_j \subset (a_i \cup a_k))$. Note that $\partial a_j \subset (a_i \cup a_k) = \partial a_j \cap (\tau \cup \triangle_1 \cup \triangle_2)$ where τ is the interior triangle (in blue) and $\triangle_i$ and $\triangle_k$ are the apex-exposed triangles (in yellow) incident to the center of a_j. The exposed atomic boundary may consist of more than one arc if the beta-shape is nonmanifold at v. For example, see, in Figure 1b that the exposed atomic boundary of a_i consists of two distinct arcs. Let $\mathrm{Length}(X)$ be the length of the curve segment X.

Lemma 15. *Let a be an atom in* A^2 *corresponding to the* β*-vertex v on the boundary of the beta-shape. Then, the length of the exposed atomic occupation boundary EAB of a is given by*

$$\mathrm{Length}(\mathit{EAB}_a) = \mathrm{Length}(\partial a) - \Sigma_{f \in F_V^{\mathcal{C}}} \mathrm{Length}(\partial a \cap f) - \Sigma_{e \in E_V^{\mathcal{S}}} \mathrm{Length}(\partial a \cap f') \tag{12}$$

where $F_v^{\mathcal{C}}$, $E_v^{\mathcal{S}}$, f, and f' are those defined in eq. (1).

Corollary 16. *The total length of ∂A^2 can be computed by adding* Length(*EAB*) *for the atoms corresponding to all β-vertices on the boundary of the zero beta-shape in $O(m)$ time in the worst case where m represents the number of β-simplexes in the zero beta-complex of A^2.*

Area of the vdW-boundary in $\mathbb{R}^3$

We now extend the idea in $\mathbb{R}^2$ to $\mathbb{R}^3$. Recall that a β-vertex maps to a m-face(s) on ∂A in $\mathbb{R}^3$. Let area(X) be the area of X if X is a surface segment and the area of ∂X if X is a volumetric object. We have two types of *area primitives* corresponding to the β-vertices and β-edges on the boundary of the beta-shape as follows:

- Area of the exposed atomic (occupation) boundary *(EAB)*: The area of $\partial a \subset \partial A$ corresponding to a β-vertex.
- Area of the (edge-)exposed go-stone boundary *(EGB)*: The area of $\partial a \subset \partial A$ corresponding to a β-edge.

Suppose that a is an atom corresponding to a β-vertex v. If v is singular, a does not intersect any other atom and therefore the entire atom boundary is exposed to air. If v is regular, there exists another atom intersecting a. The area of *EAB* of a is obtained by subtracting the boundary intersection of a with the other atom from ∂a. Hence, formulae similar to eq. (11) and (12) hold.

Lemma 17. *Suppose that v is a regular β-vertex on the boundary of the zero beta-shape and corresponds to an atom a. Then, the area of EAB corresponding to the β-vertex v is defined as*

$$\begin{aligned}\text{Area}(EAB_a) = \text{Area}(\partial a) - \Sigma_{\tau \in C_V^{\mathcal{C}}} \text{Area}(\partial a \cap \tau) \\ - \Sigma_{f \in F_V^{\mathcal{S}}} \text{Area}(\partial a \cap \triangle)\end{aligned} \quad (13)$$

where $C_v^{\mathcal{C}}$, $F_v^{\mathcal{S}}$, and $\triangle$ are those defined in eq. (8).

Note that $\partial a \cap \tau$ is a spherical triangle and Area($\partial a \cap \tau$) is given by

$$r^2((\alpha + \beta + \gamma) - \pi) \quad (14)$$

where α, β, and γ are the interior angles of the spherical triangle and r is the radius of an atom.[62] The same formula applies to $\partial a \cap \triangle$.

Consider a β-edge $e = (a_i, a_j)$, $i \neq j$, where $a_i \cap a_j \neq \emptyset$. Then, *Gostone*2($a_i, a_j$) should be appropriately accounted for. Suppose that e is singular. Then, Area(*Gostone*2(a_i, a_j)) should be subtracted once because it is counted twice via *EAB* for both a_i and a_j. Note that

$$\text{Area}(Gostone2(a_i, a_j)) = 2\pi(r_i h_i + r_j h_j) \quad (15)$$

where r_i is the radius of a_i and h_i is the height of the cap on a_i.[62] If e is regular, the amount of subtraction is not the entire Area(*Gostone*2(a_i, a_j)) but a subset of Area(*Gostone*2(a_i, a_j)) denoted by the *(edge-)exposed go-stone boundary* (*EGB*) for the same reason discussed in "Edge-Exposed Go-Stone (*Gostone2*): β-Edge" section. Hence, the area is given by

$$\text{Area}(EGB_e) = \text{Area}(Gostone2(a_i, a_j)) \times \frac{\Theta}{2\pi} \quad (16)$$

where Θ is given in eq. (6). Hence, the following theorem stands.

Theorem 18. *The vdW-area* Area($\partial \mathcal{A}$) *can be computed by*

$$\text{Area}(\partial \mathcal{A}) = \sum_{v \in V^{\mathcal{S}}} \text{Area}(EAB_v) - \sum_{e \in E^{\mathcal{S}}} \text{Area}(EGB_e) \quad (17)$$

where $V^{\mathcal{S}}$ and $E^{\mathcal{S}}$ are the sets of the β-vertices and β-edges on the boundary of the zero beta-shape.

Note that *EAB* and *EGB* have an one-to-one correspondence to *EAO* and *EEG* in "Volume of the Union of Balls in $\mathbb{R}^3$" section, respectively.

Corollary 19. *Given the beta-complex of a molecule A in $\mathbb{R}^3$, the vdW-area* Area(A) *can be computed in $O(m)$ time in the worst case where m represents the number of β-simplexes in the zero beta-complex of A.*

Offset Volume

Recall that $A^O = \{a_1^O, a_2^O, \ldots, a_n^O\}$ denotes a set of offset atoms $a_i^O = (p_i, r_i^O)$ from $a_i = (p_i, r_i) \in A$, where $r_i^O = r_i + \delta$, $\delta > 0$. In this section, we present algorithms for computing *Vol*(A^O) and *Area*(A^O) based on the discussions above. Let $\mathcal{VD}^O$ and $\mathcal{QT}^O$ be the Voronoi diagrams and the quasi-triangulation of A^O, respectively, and $\mathcal{C}_\beta^O$ and $\mathcal{S}_\beta^O$ be the zero beta-complex and the zero beta-shape of A^O corresponding to $\beta = 0$, respectively. In this section, we will discuss only the computation of the offset-volume because the offset-area can be similarly computed. There can be three approaches for computing offset-volume.

Algorithm 1: Naive approach

The first, a naive approach would be to blindly use **Beta-decomposition-vdW** algorithm for computing *Vol*(A^O) after explicitly replacing the atom set A by the offset atom set A^O. In this approach, we first create an explicit offset model A^O for A and δ. Then, we compute the Voronoi diagram $\mathcal{VD}^O$ and the quasi-triangulation $\mathcal{QT}^O$ for A^O. Then, we compute the zero beta-complex $\mathcal{C}_\beta^O$ and the zero beta-shape $\mathcal{S}_\beta^O$ for A^O. Lastly, we compute *Vol*(A^O) and *Area*(A^O) using the **Beta-decomposition-vdW** algorithm with $\mathcal{C}_\beta^O$ and $\mathcal{S}_\beta^O$. While this approach computes the correct volume of A^O, it requires the computation of $\mathcal{VD}^O$, $\mathcal{QT}^O$, and $\mathcal{C}_\beta^O$. The computational requirement of this approach is as much as the one of **Beta-decomposition-vdW**. We call this naive algorithm the **Beta-decomposition-offset-1** algorithm and denote its computational requirement by T_1.

Algorithm 2: Translation of β-intervals

Suppose that x is a V-entity (such as the V-vertex, V-edge, V-face, or V-cell) in the Voronoi diagram $\mathcal{VD}$ for a set A. This implies that the distances of x from the boundaries of the atoms related to the definition of x are identical. Then, the addition of a constant offset amount δ to the radii of these atoms still keep the same property. Note that δ can be either positive or negative. This proves the following lemma.

Lemma 20. *$\mathcal{VD} \equiv \mathcal{VD}^O$ and $\mathcal{QT} \equiv \mathcal{QT}^O$.*

Lemma 20 can be used to improve the naive algorithm. Note that, even if $\mathcal{QT}$ and $\mathcal{QT}^O$ are identical, the corresponding β-intervals of a q-simplex $\sigma \in \mathcal{QT}$ and its corresponding q-simplex $\sigma' \in \mathcal{QT}^O$ are not identical. Figure 6a shows a set of four two-dimensional atoms $A = \{a_1, a_2, a_3, a_4\}$ and its quasi-triangulation $\mathcal{QT}$. Figure 6b shows A^O for some $\delta > 0$ and the corresponding quasi-triangulation $\mathcal{QT}^O$. See Figure 6a. Consider the β-edge e and its β-intervals. Suppose that β_1 is the radius of the minimum empty ball tangent to a_1 and a_3, β_2 is the radius of the minimum empty ball tangent to a_1, a_3, and a_4, and β_3 is the radius of the minimum empty ball tangent to a_1, a_2, and a_3. Suppose that $\beta_1 < \beta_2 < \beta_3$. Then, the β-interval for e to be singular is $\beta_1 \le \beta < \beta_2$ because e is defined by itself without bounding any triangular face simplex yet when β is contained in this interval; the β-interval for e to be regular is $\beta_2 \le \beta < \beta_3$ because the β-face $f_l = (a_1, a_3, a_4)$ is now defined and e is an edge of f_l when β is contained in this interval; the β-interval for e to be interior is $\beta_3 \le \beta < \infty$ because another β-face $f_r = (a_1, a_2, a_3)$ is also now defined and e is shared by both f_l and f_r when β is contained in this interval.

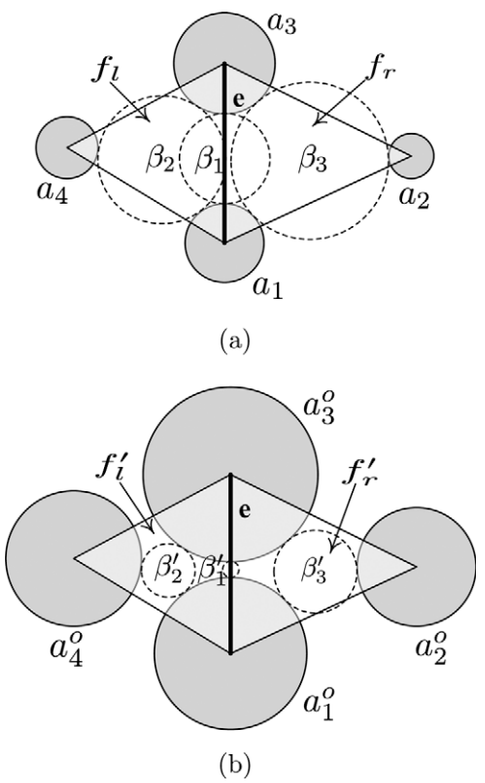

Figure 6. The quasi-triangulations for an atom set A and its offset A^O ($\beta_1' < \beta_2' < \beta_3' < \beta_1 < \beta_2 < \beta_3$). a) $A = \{a_1, a_2, a_3, a_4\}$ and b) $A^O = \{a_1^O, a_2^O, a_3^O, a_4^O\}$.

See Figure 6b which shows A^O and its quasi-triangulation $\mathcal{QT}^O$ (which is identical to $\mathcal{QT}$). We want to investigate the β-intervals of the same β-edge e in $\mathcal{QT}^O$. Suppose that β_1' is the radius of the minimum empty ball tangent to a_1^O and a_3^O, β_2' is the radius of the minimum empty ball tangent to a_1^O, a_3^O, and a_4^O, and β_3' is the radius of the minimum empty ball tangent to a_1^O, a_2^O, and a_3^O. By the similar investigation as earlier, the β-intervals of e are given as follows: The β-interval for the singular state is $\beta_1' \le \beta < \beta_2'$; the β-interval for the regular state is $\beta_2' \le \beta < \beta_3'$; the β-interval for the interior state is $\beta_3' \le \beta < \infty$. Note that $\beta_1' < \beta_2' < \beta_3'$.

We make the following observations:

- The number of the β-intervals constituting the β-spans of each simplex σ in $\mathcal{QT}^O$ is preserved as σ in $\mathcal{QT}$.
- The β-intervals for the simplex σ in $\mathcal{QT}$ and σ in $\mathcal{QT}^O$ are different.

Theorem 21. *Let $\mathcal{I} = [\beta', \beta'')$ be the β-interval of a state of a simplex $\sigma \in \mathcal{QT}$. Then, for an offset amount δ, the β-interval for the same state of the same simplex $\sigma \in \mathcal{QT}^O$ is given by*

$$\mathcal{I}^O = [\beta' - \delta, \beta'' - \delta). \tag{18}$$

See the Supporting Information for the proof. We call this operation in eq. (18) the *translation of β-intervals*. Because a single visit to each simplex in $\mathcal{QT}$ is sufficient to modify the β-intervals for A^O, the following corollary holds.

Corollary 22. *The β-spans of all q-simplexes in $\mathcal{QT}^O$ can be computed from $\mathcal{QT}$ in $O(m)$ time in the worst case where m is the number of simplexes in $\mathcal{QT}$.*

Note that if the simplexes in $\mathcal{QT}$ is ordered in some criterion of the β-interval values, the same order is identically preserved for the simplexes in $\mathcal{QT}^O$. Once $\mathcal{QT}^O$ is available with translated β-intervals, the zero beta-complex for A^O can be computed by simply searching the simplexes in $\mathcal{QT}^O$ whose singular, regular, and interior β-intervals contain $\beta = 0$ value. The algorithm **Beta-decomposition-offset-2** in the Supporting Information computes the offset-volume using the zero beta-complex $\mathcal{C}_\beta^O$ of the offset model A^O.

Lemma 23. *Beta-decomposition-offset-2 algorithm takes $T_2 = O(m + \log m + k)$ time in the worst case, where m and k are the numbers of simplexes in the quasi-triangulation and the zero beta-complex, respectively.*

This lemma holds as follows. The translation of the β-interval for all simplexes takes $O(m)$ time in the worst case. Then, $O(\log m + k)$ time is necessary in the worst case if the binary search is used for searching the simplexes of the beta-complex from the quasi-triangulation, as reported in Ref. [27]. Therefore, the **Beta-decomposition-offset-2** algorithm can compute

the offset-volume more efficiently than **Beta-decomposition-offset-1** algorithm if $\mathcal{QT}$ is available. Note that, in Lemma 23, $k \ll m$.

Algorithm 3: Direct search of β-simplexes

Given $\mathcal{QT}$, it is possible to compute the zero beta-complex $\mathcal{C}_\beta^O$ for A^O directly from $\mathcal{QT}$ without translating all β-intervals of all simplexes in $\mathcal{QT}$. Thus, the computation time can be further reduced. Recall that $\mathcal{C}_\beta$ and $\mathcal{C}_\beta^O$ are the zero beta-complexes for A and A^O, respectively. Careful interpretation of Theorem 21 reveals the relationship between the β-values and the β-intervals for $\mathcal{C}_\beta$ and $\mathcal{C}_\beta^O$ stated as follows:

Theorem 24. *Let $\delta > 0$ be the offset amount. Suppose that $\mathcal{C}_{\beta^*}$ and $\mathcal{C}_{\beta^*}^O$ are the beta-complexes for A and A^O both corresponding to the β-value of β^*, respectively. Then, $\mathcal{C}_{\beta^*}^O \equiv \mathcal{C}_{\beta^*+\delta}$.*

This theorem holds because eq. (18) states that a simplex $\sigma \in \mathcal{QT}^O$ belongs to the beta-complex $\mathcal{C}_\beta^O$ if $\beta' - \delta \leq \beta^* < \beta'' - \delta$. Hence, a simplex $\sigma \in \mathcal{QT}$ belongs to $\mathcal{C}_\beta^O$ if and only if σ has the following interval $\beta' \leq \beta^* + \delta < \beta''$. Theorem 24 states that $\mathcal{C}_\beta^O$ corresponding to β^* can be directly computed from $\mathcal{QT}$ by computing $\mathcal{C}_\beta$ corresponding to $\beta^* + \delta$. The algorithm **Beta-decomposition-offset-3** in the Supporting Information computes $\mathcal{C}_\beta^O$.

Lemma 25. Beta-decomposition-offset-3 *algorithm takes $T_3 = O(\log m + k^O)$ time in the worst case, where m and k^O are the numbers of simplexes in the quasi-triangulation and in the zero beta-complex for A^O, respectively.*

Note that $k < k^O \ll m$ for most biomolecular applications, in particular for $\delta = 1.4$ Å which corresponds to water molecule. In theory, $k \leq k^O \leq m$. It is obvious that the time complexity of the three algorithms have the following relation:

$$T_3 < T_2 < T_1. \tag{19}$$

The difference between T_3 and T_2 is the linear term and may not be very significant. This fact is verified from the experiment which will be shown in the next section. The difference between T_2 and T_1 is somewhat significant.

Experiments

The proposed algorithms were implemented using C++ based on the beta-complex library developed by the Voronoi Diagram Research Center (VDRC)[5] and thoroughly tested using 100 protein models chosen from the Protein Data Bank (PDB)[4, 63, 64] (Table A1). The experiment was performed by the Linux version of *BetaMass* on a cluster computer consisting of 118 nodes at VDRC: Each node has an AMD Opteron dual core 2.2 GHz with 2 GB RAM and dual CPU's for each node. The test data set was carefully chosen so that the model sizes (i.e., the number of atoms) were well distributed. The hydrogen atoms corresponding to the HOH atom field in the PDB file were ignored.

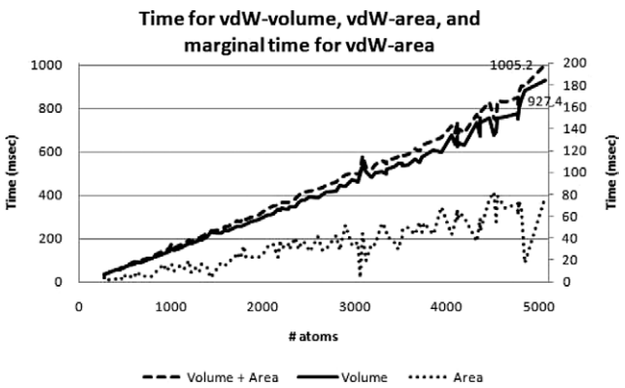

Figure 7. Computation time of the beta-decomposition algorithm for 100 models in PDB: The broken (upper-most) curve for both the vdW-volume and the vdW-area; the solid (middle) curve for the vdW-volume only; the dotted (bottom) curve for the marginal computation to get the vdW-area after the vdW-volume is computed (the right vertical axis).

Figure 7 shows the computation time (in the unit of msed) for the vdW-volume and vdW-area of the 100 test models. The horizontal axis denotes the model size and the vertical axis denotes the computation time in the unit msec. The upper-most, broken curve shows the time for computing both the vdW-volume and the vdW-area, and the middle, solid curve shows the time required to compute the vdW-volume only. For these two curves, the left vertical axis applies. The bottom, dotted curve shows the marginal time required to compute the vdW-area after the vdW-volume is computed and the right vertical axis applies. In other words, the dotted curve is the difference between the upper-most and the middle curves. We observe the following:

1. The curves are strongly linear as theoretically proved.
2. The largest tested model (1rf8 with 5071 atoms) takes about 1 sec to compute both the vdW-volume and the vdW-area. The marginal time for the vdW-area of this model is about 0.08 sec.
3. The marginal computation required to compute the vdW-area in addition to the vdW-volume is tiny. This is because an additional tiny computation is sufficient for obtaining the vdW-area once the topology traversal among the simplexes of the beta-complex is performed for the vdW-volume.

Note that the quasi-triangulation is assumed to be available via preprocessing. In fact, we store the quasi-triangulation of all protein models available in PDB in the quasi-triangulation data base (QTDB) publicly available at VDRC.[5] Figure 7 does not include the computation time for the Voronoi diagram and the quasi-triangulation.

The vdW-volumes and vdW-areas computed by the Beta-decomposition algorithm are given in Table A1. See Column A for the vdW-volume and Column B for the vdW-area. To verify the correctness of the proposed algorithm (and its implementation) for computing the vdW mass properties, we also computed the vdW-volumes of the 100 test models using Monte Carlo simulation (See Column G), called the MCS-I. The purpose of MCS-I is to verify the quality of the solution obtained by the

Beta-decomposition algorithm, not to make an assessment for the efficiency of the algorithm. Hence, we implemented MCS-I in the most obvious way.

In the MCS-I, we first computed the bounding box B of each model and its volume Vol(B). Then, we created a random point q within the box and tested whether q is contained in the union of atoms or not. The containment test is done by checking whether q is contained by any atom of the model. Suppose that m_S points, out of m tested points, are contained in the union of atoms. Then, the vdW-volume $Vol_{\mathrm{MCS}-I}$ estimated by the Monte Carlo simulation is defined as $Vol_{\mathrm{MCS}-I} = \frac{m_S}{m} Vol_B$.

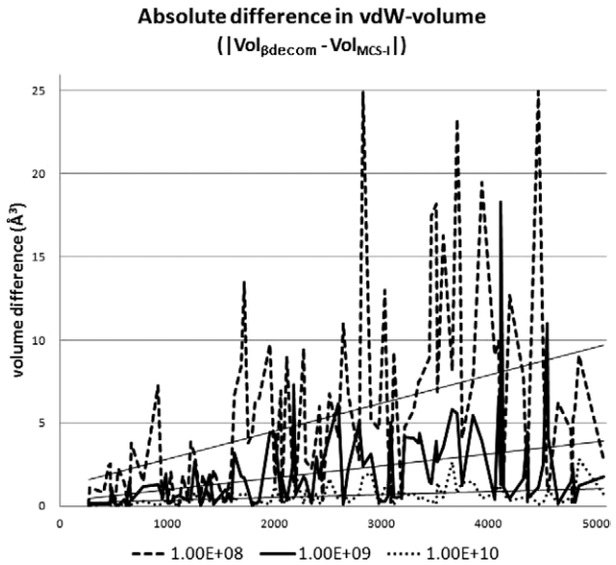

Figure 8. The absolute differences of the vdW-volumes between the beta-decomposition algorithm and the Monte Carlo simulation (MCS-I) with 10^8, 10^9, and 10^{10} sample points.

Let $Vol_{\beta\mathrm{decom}}$ be the volume of each model computed by the beta-decomposition algorithm. Figure 8 shows $|Vol_{\beta\mathrm{decom}} - Vol_{\mathrm{MCS}-I}|$ in three curves. The horizontal axis denotes different test models (sequentially ordered in their sizes) and the vertical axis denotes the volume difference. Each data point in Figure 8 corresponds to $|Vol_{\beta\mathrm{decom}} - Vol_{\mathrm{MCS}-I}|$ of each model. Hence, Figure 8 shows the distribution of absolute differences of the Beta-decomposition from the MCS-I. The top-most curve corresponds to 10^8 random points for each model in the MCS-I. The middle and the lowest curves correspond to 10^9 and 10^{10} random points, respectively. Note that, in a given curve, the absolute difference increases as the molecule size increases and decreases as the number of sample points increases. Column G of Table A1 contains the volumes computed by the MCS-I with 10^{10} random points for each model. Column H shows the signed absolute difference $Vol_{\beta\mathrm{decom}} - Vol_{\mathrm{MCS}-I}$ and Column I shows the signed relative difference $\frac{Vol_{\beta\mathrm{decom}} - Vol_{\mathrm{MCS}-I}}{Vol_{MCS-I}} \times 100$ in percentile.

Figure 9 shows the relative difference $\frac{Vol_{\beta\mathrm{decom}} - Vol_{\mathrm{MCS}-I}}{Vol_{\mathrm{MCS}-I}} \times 100$ for the simulation corresponding to 10^{10} sample points. Note that the relative difference was mostly less than 0.01% and very stable. These graphs strongly indicates that the proposed

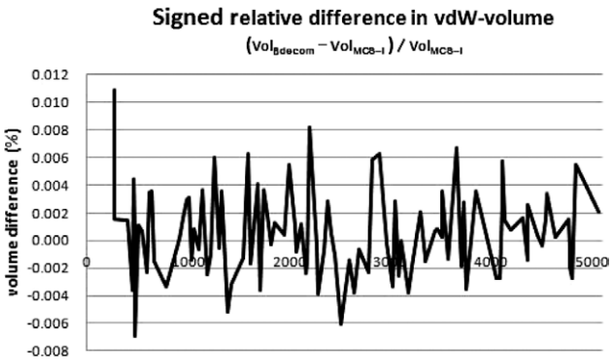

Figure 9. Signed relative difference of the vdW-volumes between the beta-decomposition algorithm and the Monte Carlo simulation (MCS-I) with 10^{10} sample points.

algorithm and its implementation computes the correct volume of the union of atoms in $\mathbb{R}^3$.

The validation of the code implementing MCS-I is also of importance. The "Experiments" section of the Supporting Information describes the convergence of MCS-I code for three molecules and a test model consisting of 100 nonintersecting random balls (Figs. 2 and 3 in the Supporting Information).

To understand the characteristics of the proposed algorithm, we studied the number of volume primitives and the area primitives involved in the computation for the 100 test data models. The area primitives and some volume primitives have identical topological traversals: Both Area(*Fan*1) and Vol(*Fan*1) are defined for each β-vertex and it is necessary to access its incident β-cells and its incident apex-exposed tetrahedra; Area(*Gostone*2) and Vol(*Gostone*2) are defined for each β-edge and it is necessary to compute the angle of the m-edge around the β-edge by accessing the appropriate β-faces incident to the β-edge. The only difference between Area() and Vol() is the formulae to evaluate the functions. Therefore, given the vdW-volume, the vdW-area can be obtained with an additional tiny computation of applying the area formulae.

Figure 10 shows the number of the volume primitives used for the 100 test models. From the bottom, each curve denotes the number of primitives Tetra4, Fan1, Tetra3, and Gostone2, respectively. The top-most curve (of triangles) denotes the number

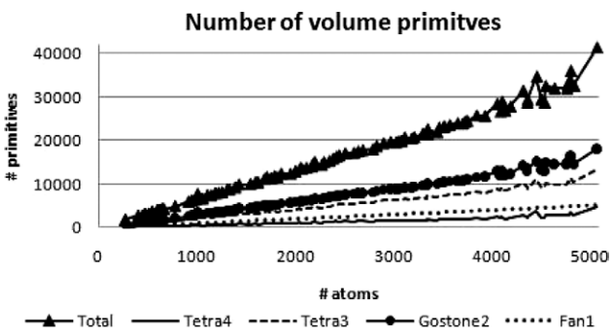

Figure 10. The number of the primitives processed for computing vdW-volume and vdW-area for the 100 test models.

of the total volume primitives: Tetra4 + Tetra3 + Gostone2 + Fan1. From the figure, it is clear that the number of primitives involved in the computation is strongly linear to the molecule size. We also note the following: On average, 6.94 volume primitives per atom are used in the vdW-volume computation and 4.13 area primitives per atom are used in the computation of the vdW-area.

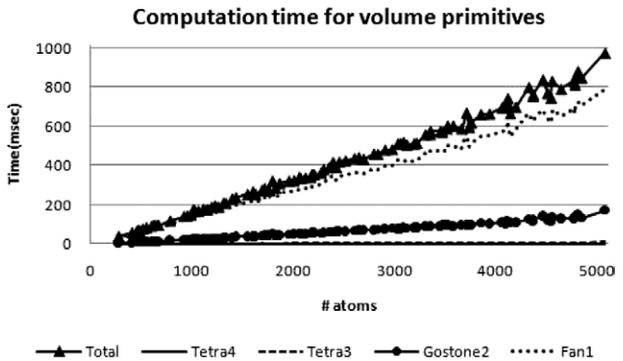

Figure 11. Computation times used for the primitives in computing the vdW-volumes for the 100 test models.

Figure 11 shows the computation time for the volume primitives. We do not include the computation time for the area primitives in this graph because we want to relate Figure 11 to 10. From the bottom, each curve denotes Tetra4, Tetra3, Gostone2, and Fan1. The top-most curve denotes the total computation time for the entire vdW-volume of a model: Tetra4 + Tetra3 + Gostone2 + Fan1. We measured the computation time instead of counting the number of pointer accesses in the topology structure to avoid unnecessary technical complications. As expected, all curves are strongly linear. The curves of Tetra3 and Tetra4 are very low and similar because each of these primitives has only one pointer access and a simple formula is applied. The phenomenon where the Fan1 curve is located high in the graph deserves an explanation. Each primitive for Fan1 and Gostone2 requires further topology traversal in the neighborhood after the initial traversal is performed. To be specific, a Fan1 primitive first accesses all β-cells incident to the β-vertex corresponding to the atom, and then all β-faces should also be accessed because they define Tetra3's. On the other hand, the Gostone2 primitive first accesses all β-faces incident to the β-edge and then computes the sweeping angle. Hence, Gostone2 is simpler than Fan1. This is why the Fan1 curve is placed high in the graph.

The experiment for the offset is in order. Figure 12 shows the time taken for computing the offset-volume of the largest atom (1rf8, 5071 atoms) for 25 different values of the offset amount δ = 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 20.0, 30.0, 40.0, and 50.0 Å. We will not discuss about the offset-area. The horizontal axis denote different values of δ and the vertical axis denotes the computation time in msec. The two more efficient methods i.e., the translation of the β-spans (TB) and the direct search of the β-simplexes (DS) were implemented and tested. In the

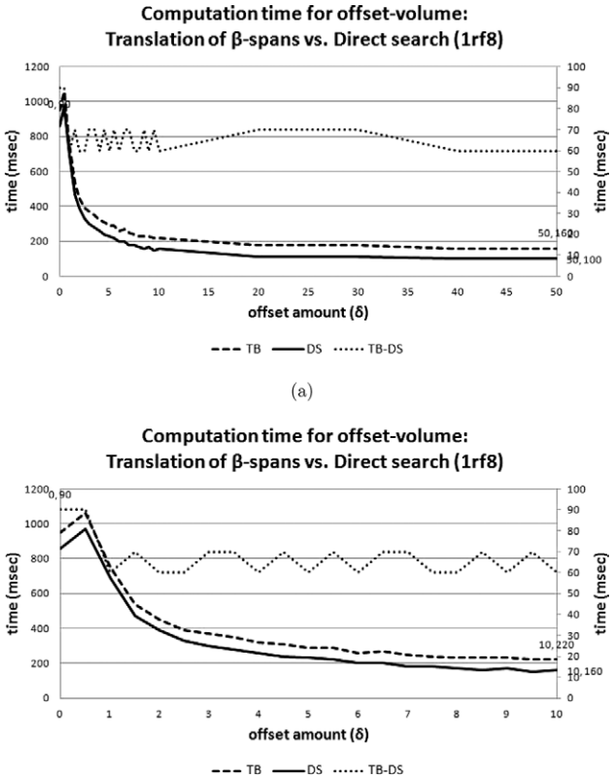

Figure 12. Computation time of the direct search (DS) and the translation of β-span (TB) methods for the offset-volumes (1rf8, 5071 atoms): TB in the broken curve and DS in the solid curve. The difference, TB - DS, in the dotted curve. a) Computation time measured at the 25 δ's (from 0 to 50 Å) and b) the first 21 δ's (from 0 to 10 Å).

figure, the left vertical axis denotes the time taken by these two algorithms and applies to the lower two curves: the broken curve denotes TB and the solid curve denotes DS. Note that these curves shows the marginal computation time after the vdW-volume is computed (which means that the beta-complex is already available). As theoretically shown previously, the DS-method is slightly faster than the TB-method. The upper, dotted curve denotes the difference between the two curves (i.e., TB - DS) and the right vertical axis applies. The horizontal window of δ in Figure 12a is between 0 and 50 Å and one for Figure 12b is between 0 and 10 Å. Note that both curves have peaks around $\delta = 0.5$ Å and then monotonically decrease as δ increases.

Figure 13a shows the time in sec taken by DS-method for computing both the offset-volume and the offset-area for the 20 selected models. Their id.'s are $5, 10, 15, \ldots, 100$ in the 100 test models. There are five curves: $\delta = 0.5, 1.4, 3.0, 5.0$, and 10.0 Å from top to bottom. Note that this time is for the marginal computation taken after the vdw-volume and the vdW-area are computed (i.e., after the zero beta-complex is available). Note two observations: (i) The computation time is tiny compared to the vdW-volume computation; (ii) all the curves show the linear increase. Figure 13b shows the marginal computation time for the offset area after the offset volume is computed.

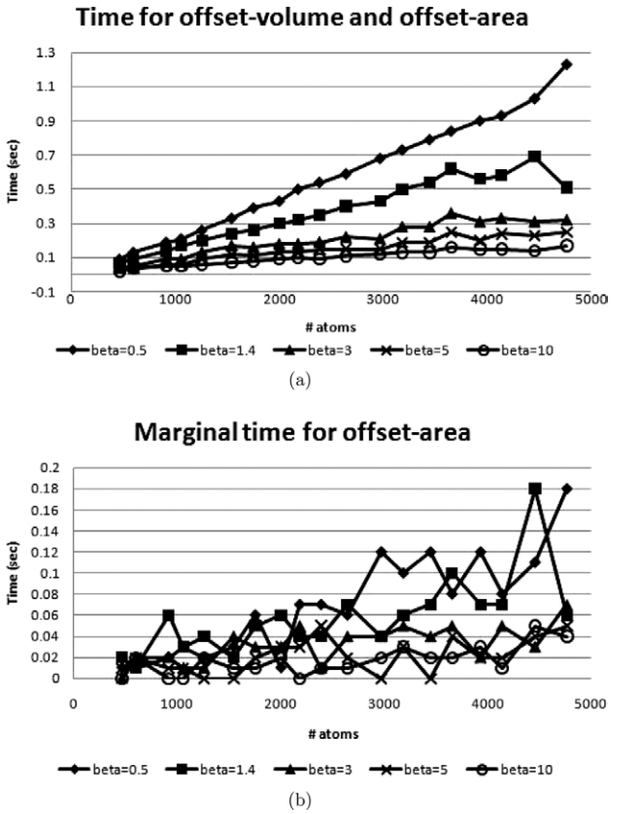

Figure 13. Computation time of the DS-method for the offset-volume and the offset-area (for the 20 selected models). a) The time for both the offset-volume and the offset-area and b) the marginal time needed for the offset-area after the offset-volume is computed.

Table B1 summarizes the statistics of the offset-volumes and the offset-areas computed by the DS-method for the five offset amounts $\delta = 0.5, 1.4, 3.0, 5.0$, and 10.0 Å. In the table, we report test statistics for only five selected models, called the offset test model set, out of the 100 test models due to space limit. Their id.'s are 20, 40, 60, 80, and 100 in the 100 test set models. For each of the offset test models, we computed the offset-volumes and offset-areas by the DS-method and compared them with the computation results using another type of Monte Carlo simulation denoted by MCS-II. Each row in Table B1 denotes as follows. "β-decom" denotes the mass property computed by the DS-method; "MCS-II" denotes the mass property computed by MCS-II; "Abs diff" denotes the absolute difference $|\beta-\text{decom} - \text{MCS}-II|$; "Rel diff" denotes the relative difference $\frac{|\beta-\text{decom}-\text{MCS}-II|}{\text{MCS}-II} \times 100$; and "$\beta$c+V+A" denotes the computation of the beta-complex, the volume, and the area.

The algorithm for MCS-II was as follows: For the offset-volume, we generated grid points in the bounding box of the offset-model and stored them in a grid structure partitioning the bounding box. With the grid structure, each sample point was tested against those offset balls intersecting the grid cell containing itself. The acceleration technique made MCS-II run much faster than MCS-I did. On the other hand, the offset area in MCS-II was measured as follows. We first produced a set of uniformly distributed points on the boundary of each offset atom using the spherical sampling technique proposed by Shao and Badler[65] that was based on the Archimedes' theorem. Given a point q on the boundary of an offset atom o, if q was not contained by some offset atom other than o, q contributed to the offset area. Provided the neighborhood information among offset atoms, counting such sample points could be efficiently done. For each run of MCS-II, 10^9 random points were used. Figure 5 in the Supporting Information shows the computation time for MCS-II.

Figure 14a shows the absolute difference $|V^O_{\text{DS}} - V^O_{\text{MCS}-II}|$ of the offset volume for the five selected models in Table B1 where V^O_{DS} and $V^O_{\text{MCS}-II}$ denote the offset volumes computed by DS-method and by MCS-II method, respectively. Each curve in the figure corresponds to a particular value of the offset amount δ. Figure 14b similarly shows the absolute difference $|A^O_{\text{DS}} - A^O_{\text{MCS}-II}|$ of the offset area where A^O_{DS} and $A^O_{\text{MCS}-II}$ denote the offset areas computed by DS-method and by MCS-II method, respectively. Figure 14c shows the relative difference $\frac{|V^O_{\text{DS}} - V^O_{\text{MCS}-II}|}{V^O_{\text{MCS}-II}} \times 100$ of the offset volume. Figure 14d similarly shows the relative difference for offset area. Note that the relative differences are very tightly bounded and this observation verifies the correctness of both the theory and the implementation. In the MCS-II result, the numerical integration based on regular grid used in volume computation converges faster than the Monte Carlo integration for area computation, as explained in many Monte-Carlo literatures such as Ref. [66].

Conclusions

Computing the mass properties such as the volume and area of vdW-molecule and its offset, usually referred to by the Lee-Richards (solvent) accessible surface, has long been an important research issue in the computational chemistry, computational molecular biology, and structural biology. There were many studies from the grid-based approach and Monte Carlo simulation to analytic approach. However, this seemingly well-defined problem has not been well-solved yet.

In this article, we propose an approach, called the *beta-decomposition*, that is based on the recent theory of the beta-complex which can be quickly computed from the quasi-triangulation, the dual of the Voronoi diagram of the atoms in molecules. As the name suggests, the Beta-decomposition approach decomposes the desired mass property of the entire molecule into a set of primitives. Then, the correct mass property can be obtained by an appropriate summation of these primitives with the proper consideration of redundancies among the primitives.

Two algorithms were presented: **Beta-decomposition-vdW** algorithm computes the vdW-volume and the vdW-area; **Beta-decomposition-offset** algorithm computes the offset-volume and the offset-area. These algorithms compute the mass properties in $O(m)$ time in the worst case, where m represents the number of simplexes in the beta-complex. We emphasize that the computation time for the mass properties vdW-volume,

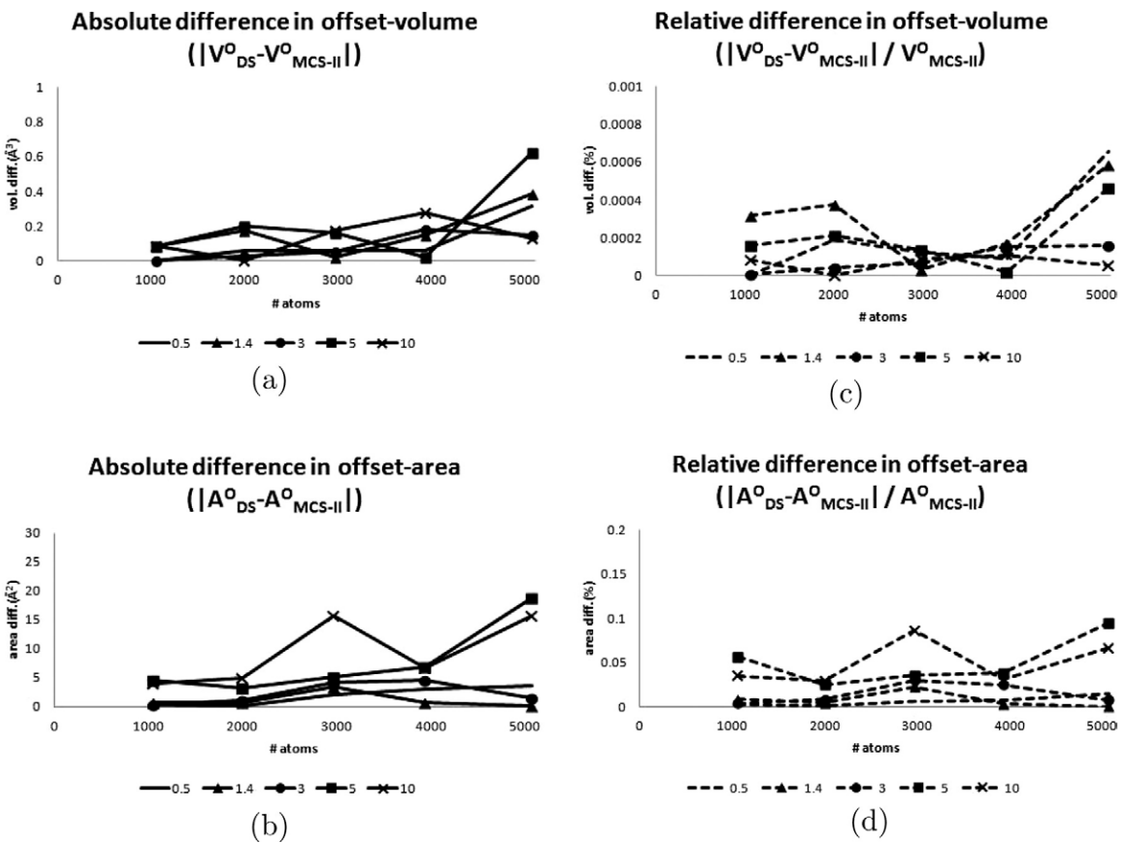

Figure 14. Differences for the offset-volumes and offset-areas of the five test models computed using the direct search (DS) method for the offset amount $\delta = 0.5, 1.4, 3.0, 5.0,$ and 10.0. 10^9 random points were used. a) the absolute difference of the offset-volume, b) the relative difference of the offset-volume, c) the absolute difference of the offset-area, and d) the relative difference of the offset-area.

vdW-area, offset-volume, and offset-area altogether takes very small if the quasi-triangulation is available. The correctness and efficiency of the proposed algorithms and their implementation are also verified through an experiment using model data publicly available from the Protein Data Bank (PDB).[4] The software BetaMass implementing the proposed Beta-decomposition algorithm is freely available from the VDRC web site at //http:voronoi.hanyang.ac.kr.

Appendix A: Test Data Set

Table A1. The vdW-volume and vdW-area of the 100 test PDB models computed by the Beta-decomposition algorithm using the VDRC cluster computer.

		Beta-decomposition		Time (sec)				MCS-I		
No.	ID (No. of atoms)	vdW-vol (A)	vdW-area (B)	VD (C)	QT (D)	Beta-cmplx +A+B (E)	Total (C+D+E) (F)	MCS-vol (G)	A-G (H)	$\frac{H}{G} \times 100$ (I)
1	1c26 (268)	2893.820	3503.820	1.31	0.1	0.04	1.45	2893.505	0.315	0.011
2	2nls (270)	2931.660	3504.170	1.2	0.1	0.04	1.34	2931.614	0.046	0.002
3	2erw (401)	4308.170	5074.170	1.93	0.17	0.06	2.16	4308.106	0.064	0.001
4	1zx6 (447)	4823.530	5810.410	2.32	0.19	0.06	2.57	4823.704	−0.174	−0.004
5	1mhn (464)	5029.590	6043.980	2.27	0.21	0.06	2.54	5029.368	0.222	0.004
6	2igd (467)	5037.880	6102.680	2.33	0.21	0.07	2.61	5037.865	0.015	0.000
7	1zlm (476)	5114.750	6119.590	2.5	0.21	0.07	2.78	5115.102	−0.352	−0.007
8	1y0m (507)	5450.180	6560.810	2.68	0.22	0.08	2.98	5450.119	0.061	0.001
9	2g7o (543)	5893.890	7087.480	2.74	0.23	0.08	3.05	5893.852	0.038	0.001
10	3b7h (597)	6493.190	7780.360	3	0.27	0.09	3.36	6493.340	−0.150	−0.002
11	1t6f (616)	6695.750	8118.480	3.09	0.27	0.09	3.45	6695.515	0.235	0.004
12	2o37 (643)	6944.050	8305.020	3.39	0.28	0.1	3.77	6943.804	0.246	0.004
13	1zpw (663)	7203.330	8693.330	3.45	0.3	0.1	3.85	7203.438	−0.108	−0.001

(*Continued*)

Table A1. (Continued)

		Beta-decomposition		Time (sec)				MCS-I		
No.	ID (No. of atoms)	vdW-vol (A)	vdW-area (B)	VD (C)	QT (D)	Beta-cmplx +A+B (E)	Total (C+D+E) (F)	MCS-vol (G)	A-G (H)	$\frac{H}{G} \times 100$ (I)
14	1j27 (778)	8490.750	10156.300	4.19	0.35	0.12	4.66	8491.035	−0.285	−0.003
15	1d4t (913)	9896.960	11898.600	4.84	0.42	0.14	5.4	9896.946	0.014	0.000
16	1qkd (944)	10122.000	11987.000	4.68	0.44	0.15	5.27	10121.838	0.162	0.002
17	1tp6 (983)	10590.600	12681.400	5.12	0.45	0.15	5.72	10590.288	0.312	0.003
18	2ggr (1008)	6682.870	6917.760	3.57	0.51	0.19	4.27	6682.660	0.210	0.003
19	1lz1 (1028)	11105.800	13220.400	5.28	0.48	0.17	5.93	11105.956	−0.156	−0.001
20	1dc9 (1057)	11461.600	13894.500	5.58	0.48	0.16	6.22	11461.498	0.102	0.001
21	1ezg (1106)	11885.600	14137.100	4.83	0.52	0.17	5.52	11885.674	−0.074	−0.001
22	2op6 (1144)	12414.400	14984.700	5.94	0.54	0.19	6.67	12413.950	0.450	0.004
23	2esk (1186)	12837.300	15319.000	6.39	0.56	0.19	7.14	12837.617	−0.317	−0.002
24	1t4q (1221)	13229.700	15818.400	6.51	0.57	0.19	7.27	13229.830	−0.130	−0.001
25	1jyh (1256)	13564.400	16197.100	6.83	0.6	0.2	7.63	13563.589	0.811	0.006
26	1yck (1305)	14027.800	16662.500	6.93	0.64	0.2	7.77	14027.878	−0.078	−0.001
27	2obi (1329)	14422.800	17211.300	7.17	0.65	0.21	8.03	14422.288	0.512	0.004
28	1wu3 (1389)	15093.000	18024.700	7.57	0.68	0.22	8.47	15093.780	−0.780	−0.005
29	2a8f (1426)	15510.400	18513.900	7.44	0.71	0.23	8.38	15510.890	−0.490	−0.003
30	2h3l (1542)	16637.000	20054.200	8.34	0.75	0.24	9.33	16637.208	−0.208	−0.001
31	2o7h (1596)	17445.700	21088.100	8.55	0.8	0.26	9.61	17444.607	1.093	0.006
32	2ggv (1621)	17523.900	21050.200	8.63	0.8	0.25	9.68	17524.190	−0.290	−0.002
33	2ge7 (1686)	18312.200	21879.300	8.97	0.84	0.27	10.08	18311.451	0.749	0.004
34	1k1b (1712)	18541.200	22235.400	9.02	0.84	0.28	10.14	18541.871	−0.671	−0.004
35	2yz1 (1752)	18920.600	22688.500	9.5	0.85	0.28	10.63	18919.913	0.687	0.004
36	4eug (1788)	19233.700	22849.300	9.96	0.9	0.3	11.16	19233.343	0.357	0.002
37	1i8k (1819)	19636.100	23531.200	9.75	0.9	0.29	10.94	19636.161	−0.061	0.000
38	2gpo (1857)	20357.300	24463.400	10.18	0.94	0.31	11.43	20357.031	0.269	0.001
39	1rav (1952)	21026.600	25092.300	10.66	0.96	0.33	11.95	21026.505	0.095	0.000
40	1fa8 (2000)	21598.200	25794.800	10.9	1.06	0.37	12.33	21597.012	1.188	0.006
41	1xqo (2054)	22232.500	26523.400	11.53	1.03	0.34	12.9	22232.268	0.232	0.001
42	1xba (2068)	22551.800	26968.500	11.52	1.03	0.33	12.88	22551.991	−0.191	−0.001
43	3bxy (2113)	22950.500	27613.000	11.65	1.05	0.34	13.04	22950.220	0.280	0.001
44	2h2r (2167)	23190.200	27579.900	11.76	1.11	0.35	13.22	23190.749	−0.549	−0.002
45	2cwc (2181)	23513.500	28016.300	12.11	1.09	0.36	13.56	23513.846	−0.346	−0.001
46	1syq (2199)	23962.700	28590.500	12.07	1.2	0.42	13.69	23960.735	1.965	0.008
47	1y2t (2268)	24401.600	29348.000	12.55	1.14	0.37	14.06	24401.707	−0.107	0.000
48	1ym5 (2292)	24865.400	29688.100	12.83	1.16	0.38	14.37	24866.365	−0.965	−0.004
49	1lf1 (2348)	25139.800	29972.000	12.84	1.18	0.39	14.41	25139.977	−0.177	−0.001
50	1y9u (2387)	25826.700	30999.900	13.27	1.24	0.4	14.91	25825.967	0.733	0.003
51	2guv (2415)	25635.200	30345.500	13.56	1.26	0.41	15.23	25635.063	0.137	0.001
52	1qxh (2448)	26473.900	31479.700	13.57	1.25	0.41	15.23	26474.183	−0.283	−0.001
53	2f6l (2507)	27098.300	32380.400	13.95	1.27	0.41	15.63	27099.927	−1.627	−0.006
54	1fhl (2597)	27846.500	33002.100	14.42	1.31	0.46	16.19	27846.883	−0.383	−0.001
55	1t45 (2642)	28688.400	34261.400	14.84	1.36	0.45	16.65	28689.473	−1.073	−0.004
56	1xix (2687)	28994.800	34586.400	15.06	1.38	0.44	16.88	28994.989	−0.189	−0.001
57	1x7f (2789)	30295.200	36423.200	15.55	1.46	0.45	17.46	30295.897	−0.697	−0.002
58	2g85 (2829)	30535.200	36371.900	15.72	1.47	0.47	17.66	30533.416	1.784	0.006
59	2ab0 (2900)	31525.800	37682.700	16.19	1.49	0.48	18.16	31523.816	1.984	0.006
60	1rh9 (2970)	32051.800	38281.800	16.73	1.56	0.5	18.79	32051.865	−0.065	0.000
61	2car (3026)	32748.500	39136.900	16.95	1.59	0.5	19.04	32749.597	−1.097	−0.003
62	2goi (3054)	32815.500	39201.200	16.71	1.58	0.52	18.81	32814.564	0.936	0.003
63	1d2k (3082)	33190.100	39575.700	17.41	1.62	0.51	19.54	33190.933	−0.833	−0.003
64	2i49 (3114)	33565.500	39875.100	17.61	1.62	0.52	19.75	33565.511	−0.011	0.000
65	1xh3 (3183)	34039.800	40712.700	17.89	1.66	0.52	20.07	34041.087	−1.287	−0.004
66	1eqp (3211)	34317.600	40827.600	18.28	1.67	0.53	20.48	34318.246	−0.646	−0.002
67	2fts (3298)	35876.300	43095.100	18.38	1.73	0.56	20.67	35875.540	0.760	0.002
68	2p19 (3344)	36392.400	43950.600	18.68	1.72	0.55	20.95	36392.840	−0.440	−0.001
69	2i3f (3348)	36436.800	43402.700	19.29	1.77	0.57	21.63	36437.360	−0.560	−0.002
70	1swh (3446)	36729.800	43816.500	19.55	1.86	0.58	21.99	36729.559	0.241	0.001
71	1ugq (3464)	37306.200	44284.500	19.64	1.83	0.6	22.07	37305.891	0.309	0.001
72	2f82 (3510)	38143.200	45606.500	19.67	1.88	0.6	22.15	38143.110	0.090	0.000
73	1xwg (3519)	38451.400	46209.900	20.11	1.84	0.58	22.53	38450.015	1.385	0.004
74	1xg2 (3573)	38623.900	46313.400	20.1	1.9	0.6	22.6	38624.408	−0.508	−0.001
75	2ol7 (3653)	39600.300	47856.100	20.34	1.92	0.6	22.86	39597.650	2.650	0.007
76	1zvt (3705)	40352.200	48585.900	20.58	1.91	0.61	23.1	40352.954	−0.754	−0.002
77	1r2t (3729)	40495.600	48561.100	21.08	1.97	0.62	23.67	40494.459	1.141	0.003

(*Continued*)

Table A1. (Continued)

		Beta-decomposition		Time (sec)				MCS-I		
No.	ID (No. of atoms)	vdW-vol (A)	vdW-area (B)	VD (C)	QT (D)	Beta-cmplx +A+B (E)	Total (C+D+E) (F)	MCS-vol (G)	A-G (H)	$\frac{H}{G} \times 100$ (I)
78	1qb5 (3745)	40551.100	48491.000	20.16	1.98	0.62	22.76	40552.515	−1.415	−0.003
79	1orj (3847)	41817.800	49689.800	22.36	2.24	0.76	25.36	41816.313	1.487	0.004
80	1lbw (3930)	42447.900	50180.100	22.48	2.09	0.71	25.28	42447.459	0.441	0.001
81	1p7w (4056)	26588.600	29349.200	17.17	2.32	0.72	20.21	26589.333	−0.733	−0.003
82	1f60 (4090)	44394.300	53172.200	23.26	2.16	0.68	26.1	44395.523	−1.223	−0.003
83	2zwu (4106)	36561.800	42708.600	20.82	2.43	0.82	24.07	36560.773	1.027	0.003
84	1m0z (4113)	44675.200	53331.400	23.77	2.13	0.68	26.58	44672.652	2.548	0.006
85	1edq (4136)	44534.200	53075.500	24.09	2.2	0.69	26.98	44533.534	0.666	0.001
86	1mn6 (4191)	45076.000	53512.700	24.33	2.26	0.71	27.3	45075.667	0.333	0.001
87	2pjh (4319)	29105.900	31462.000	17.78	2.47	0.81	21.06	29105.420	0.480	0.002
88	1dqz (4360)	46805.900	55461.400	25.15	2.36	0.74	28.25	46806.548	−0.648	−0.001
89	1qtw (4363)	28490.600	31799.700	18.53	2.57	0.78	21.88	28489.848	0.752	0.003
90	1pfn (4456)	29242.200	30618.000	18.01	2.57	0.88	21.46	29242.089	0.111	0.000
91	1zrs (4503)	48649.000	58407.900	25.94	2.42	0.76	29.12	48649.186	−0.186	0.000
92	1eai (4536)	48959.000	58687.800	25.23	2.43	0.75	28.41	48958.256	0.744	0.002
93	1ls1 (4549)	29358.000	32342.300	19.31	2.67	0.84	22.82	29357.007	0.993	0.003
94	1o4x (4640)	31517.500	34324.100	19.26	2.64	0.83	22.73	31517.434	0.066	0.000
95	1war (4765)	31706.700	35945.500	21.56	2.8	0.95	25.31	31706.211	0.489	0.002
96	1ils (4772)	44054.400	51331.500	24.37	2.54	0.82	27.73	44054.622	−0.222	−0.001
97	3f86 (4781)	30793.300	33833.400	21.75	2.86	0.88	25.49	30793.895	−0.595	−0.002
98	2aiy (4794)	32458.500	33694.800	19.73	2.75	0.98	23.46	32459.388	−0.888	−0.003
99	2gas (4838)	52300.900	62291.000	28.58	2.57	0.82	31.97	52298.055	2.845	0.005
100	1rf8 (5071)	33277.400	33337.500	21.94	2.97	1.04	25.95	33276.701	0.699	0.002

The MCS-vol(ume) is computed by the Monte Carlo Simulation (MCS-I) with 10^{10} random points.

Appendix B: Offset-Volumes of Five Selected Models

Table B1. The offset-volumes of the five selected models computed by the direct search (DS) method for the five offset amounts: $\delta = 0.5, 1.4, 3.0, 5.0,$ and 10.0 Å.

					Offset amount (δ)				
#′	#	ID	Offset	Analysis	0.5	1.4	3.0	5.0	10.0
1	20	1dc9	Vol	β-decomp	18334.109	26814.342	38372.590	53359.788	100856.424
		(1057)		MCS-II	18334.109	26814.257	38372.587	53359.874	100856.508
				Abs diff	0.000	0.085	0.003	0.086	0.084
				Rel diff	0.000	0.000	0.000	0.000	0.000
			Area	β-decomp	12517.240	7707.435	7095.114	7971.568	11192.575
				MCS-II	12517.557	7708.081	7094.829	7967.044	11188.684
				Abs diff	0.317	0.645	0.285	4.524	3.891
				Rel diff	0.003	0.008	0.004	0.057	0.035
			Time (sec)	βc + V + A	0.080	0.060	0.040	0.030	0.030
2	40	1fa8	Vol	β-decomp	33805.278	47698.544	67702.976	92974.529	166341.756
		(2000)		MCS-II	33805.341	47698.366	67703.005	92974.728	166341.756
				Abs diff	0.063	0.178	0.029	0.199	0.000
				Rel diff	0.000	0.000	0.000	0.000	0.000
			Area	β-decomp	20884.706	12994.867	12385.876	13017.280	16645.748
				MCS-II	20884.925	12995.486	12386.894	13020.522	16650.688
				Abs diff	0.219	0.619	1.018	3.242	4.940
				Rel diff	0.001	0.005	0.008	0.025	0.030
			Time (sec)	βc + V + A	0.270	0.200	0.130	0.090	0.060
3	60	1rh9	Vol	β-decomp	49777.805	67194.095	89562.978	117429.741	198039.148
		(2970)		MCS-II	49777.866	67194.116	89562.921	117429.901	198038.976
				Abs diff	0.061	0.021	0.057	0.160	0.172
				Rel diff	0.000	0.000	0.000	0.000	0.000
			Area	β-decomp	28978.563	14971.262	13686.486	14358.715	18180.924
				MCS-II	28976.581	14967.873	13682.357	14353.618	18165.279
				Abs diff	1.982	3.388	4.129	5.097	15.645
				Rel diff	0.007	0.023	0.030	0.036	0.086
			Time (sec)	βc + V + A	0.480	0.320	0.180	0.140	0.090

(*Continued*)

Table B1. (Continued)

					Offset amount (δ)				
#′	#	ID	Offset	Analysis	0.5	1.4	3.0	5.0	10.0
4	80	1lbw (3930)	Vol	β-decomp	65694.864	88752.164	118258.387	153520.723	252442.099
				MCS-II	65694.806	88752.013	118258.569	153520.699	252441.823
				Abs diff	0.058	0.151	0.182	0.025	0.276
				Rel diff	0.000	0.000	0.000	0.000	0.000
			Area	β-decomp	38338.852	19758.284	17743.607	17923.232	21998.187
				MCS-II	38335.914	19758.959	17748.089	17930.073	22004.894
				Abs diff	2.938	0.675	4.482	6.841	6.707
				Rel diff	0.008	0.003	0.025	0.038	0.030
			Time (sec)	βc + V + A	0.730	0.510	0.280	0.200	0.140
5	*100	1rf8 (5071)	Vol	β-decomp	47594.566	66021.662	95332.631	133601.144	240741.939
				MCS-II	47594.880	66021.275	95332.783	133601.766	240741.809
				Abs diff	0.314	0.387	0.152	0.622	0.130
				Rel diff	0.001	0.001	0.000	0.000	0.000
			Area	β-decomp	24361.757	18634.544	18527.257	19742.939	23679.163
				MCS-II	24358.076	18634.645	18528.667	19761.702	23694.828
				Abs diff	3.681	0.100	1.411	18.763	15.665
				Rel diff	0.015	0.001	0.008	0.095	0.066
			Time (sec)	βc + V + A	0.950	0.590	0.320	0.230	0.160

10^9 random points were used. "Abs diff" denotes the absolute difference $|V^O_{DS} - V^O_{MCS-II}|$ and "Rel diff" denotes the relative difference $\frac{|V^O_{DS} - V^O_{MCS-II}|}{V^O_{MCS-II}} \times 100$. "$\beta$c + V + A" denotes the computation of the beta-complex, the volume, and the area.

Acknowledgments

The authors thank the anonymous reviewers who helped to improve the quality of this article significantly.

Keywords: molecular surface • molecular volume • molecular area • van der Waals volume • van der Waals area • solvent accessible surface • Lee-Richards (accessible) surface • accessible volume • accessible area • offset surface • offset-volume • offset-area • Voronoi diagram of spheres • quasi-triangulation • beta-complex • beta-shape

How to cite this article: D.-S. Kim, J. Ryu, H. Shin, Y. Cho, *J. Comput. Chem.* **2012**, *33*, 1252–1273. DOI: 10.1002/jcc.22956

Additional Supporting Information may be found in the online version of this article.

[1] F. M. Richards, *J. Mol. Biol.* **1974**, *82*, 1.

[2] D.-S. Kim, C.-I. Won, J. Bhak, *J. Biomol. Struct. Dyn.* **2010**, *28*, 277.

[3] L. R. Dodd, D. N. Theodorou, *Mol. Phys.* **1991**, *72*, 1313.

[4] RCSB Protein Data Bank, 2009, Available at http://www.rcsb.org/pdb/. Accessed on September 2010.

[5] Voronoi Diagram Research Center, 2011, Available at http://voronoi.hanyang.ac.kr/.

[6] A. Shrake, J. A. Rupley, *J. Mol. Biol.* **1973**, *79*, 351.

[7] R. Abagyan, M. Totrov, D. Kuznetsov, *J. Comput. Chem.* **1994**, *15*, 488.

[8] A. Gavezzotti, *J. Am. Chem. Soc.* **1983**, *105*, 5220.

[9] J. Higo, N. Go, *J. Comput. Chem.* **1989**, *10*, 376.

[10] H. R. Karfunkel, V. Eyraud, *J. Comput. Chem.* **1989**, *10*, 628.

[11] E. Silla, F. Villar, O. Nilsson, J. L. Pascual-Ahuir, O. Tapia, *J. Mol. Graph.* **1990**, *8*, 168.

[12] E. Silla, I. Tunon, J. L. Pascual-Ahuir, *J. Comput. Chem.* **1991**, *12*, 1077.

[13] M. L. Connolly, *J. Am. Chem. Soc.* **1985**, *107*, 1118.

[14] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, M. Scharf, *J. Comput. Chem.* **1995**, *16*, 273.

[15] K. Rother, P. W. Hildebrand, A. Goede, B. Gruening, R. Preissner, *Nucleic Acids Res.* **2009**, *37*, 393.

[16] M. S. Till, G. M. Ullmann, *J. Mol. Model.* **2010**, *16*, 419.

[17] J. D. Bernal, J. L. Finney, *Discuss. Faraday Soc.* **1967**, *43*, 62.

[18] B. J. Gellatly, J. L. Finney, *J. Mol. Biol.*, **1982**, *161*, 305.

[19] M. Gerstein, J. Tsai, M. Levitt, *J. Mol. Biol.* **1995**, *249*, 955.

[20] A. Goede, R. Preissner, C. Frömmel, *J. Comput. Chem.* **1997**, *18*, 1113.

[21] H.-M. Will, Fast and Efficient Computation of Additively Weighted Voronoi Cells for Applications in Molecular Biology. In Proceedings of the 6th Skandinavian Workshop on Algorithm Theory, S. Arnborg, L. Ivansson, Eds., Vol. 1432 of Lecture Notes in Computer Science, **1998**; pp. 310–321.

[22] D.-S. Kim, Y. Cho, D. Kim, *Comput.-Aid. Des.* **2005**, *37*, 1412.

[23] D. Kim, D.-S. Kim, *Comput.-Aid. Des.* **2006**, *38*, 417.

[24] H. Edelsbrunner, E. P. Mücke, *ACM Trans. Graph.* **1994**, *13*, 43.

[25] H. Edelsbrunner, Weighted alpha shapes, Technical Report UIUCDCS-R-92-1760, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, **1992**.

[26] D.-S. Kim, J. Seo, D. Kim, J. Ryu, C.-H. Cho, *Comput.-Aid. Des.* **2006**, *38*, 1179.

[27] D.-S. Kim, Y. Cho, K. Sugihara, J. Ryu, D. Kim, *Comput.-Aid. Des.* **2010**, *42*, 911.

[28] M. L. Connolly, *J. Appl. Crystallogr.* **1983**, *16*, 548.

[29] G. Perrot, B. Cheng, K. Gibson, J. Vila, K. Palmer, A. Nayeem, B. Maigret, H. Scheraga, *J. Comput. Chem.* **1992**, *13*, 1.

[30] D. Avis, B. K. Bhattacharya, H. Imai, *Vis. Comput.* **1988**, *3*, 323.

[31] R. Fraczkiewicz, W. Braun, *J. Comput. Chem.* **1998**, *19*, 319.

[32] M. Irisa, *Comput. Phys. Commun.* **1996**, *98*, 317.

[33] B. J. McConkey, V. Sobolev, M. Edelman, *Bioinformatics* **2002**, *18*, 1365.

[34] A. Fabri, G.-J. Giezeman, L. Kettner, S. Schirra, S. Schönherr, *Soft.: Pract. Exp.* **2000**, *30*, 1167.

[35] CGAL Library Homepage, http://www.cgal.org/. Accessed on August 2009.

[36] F. Cazals, H. Kanhere, S. Loriot, Computing the Volume of a Union of Balls: A Certified Algorithm Technical Report, INRIA Sophia Antipolis, August, 2009.

[37] F. Cazals, H. Kanhere, S. Loriot, *ACM Trans. Math. Soft.* **2011**, *38*, 1.
[38] K. W. Kratky, *J. Phys. A: Math. Gen.* **1978**, *11*, 1017.
[39] K. D. Gibson, H. A. Scheraga, *Mol. Phys.* **1987**, *62*, 1247.
[40] R. Pavani, G. Ranghino, *Comput. Chem.* **1982**, *6*, 133.
[41] L. S. Chkhartishvili, *Math. Notes* **2001**, *69*, 421.
[42] D. Q. Naiman, H. P. Wynn, *Ann. Stat.* **1992**, *20*, 43.
[43] M. Petitjean, *J. Comput. Chem.* **1994**, *15*, 507.
[44] H. Edelsbrunner, *Discrete Comput. Geom.* **1995**, *13*, 415.
[45] D. Attali, H. Edelsbrunner, In Proceedings of the 21st Annual Symposium on Computational Geometry (SoCG'06), Pisa, Italy, **2006**; pp. 247–254.
[46] H. Edelsbrunner, M. Facello, P. Fu, J. Liang, In Proceedings of the 28th Annual Hawaii International Conference on System Sciences (HICSS'95), **1995**; pp. 256–264.
[47] T. Nishida, K. Sugihara, In Proceedings of the 5th International Symposium on Voronoi diagrams in Science and Engineering, **2008**; pp. 22–28.
[48] F. Aurenhammer, *SIAM J. Comput.* **1987**, *16*, 78.
[49] N. N. Medvedev, *Doklady Phys. Chem.* **1994**, *337*, 157.
[50] S. V. Anishchik, N. N. Medvedev, *Phys. Rev. Lett.* **1995**, *75*, 4314.
[51] H.-M. Will, Computation of Additively Weighted Voronoi Cells for Applications in Molecular Biology, PhD thesis, Swiss Federal Institute of Technology, Zurich, 1999.
[52] J.-D. Boissonnat, M. I. Karavelas, In Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms, **2003**; pp. 305–312.
[53] M. Gavrilova, Proximity and Applications in General Metrics, PhD thesis, Department of Computer Science, The University of Calgary, Calgary, Canada, **1998**.
[54] M. Gavrilova, J. Rokne, *Comput. Aid. Geom. Des.* **2003**, *20*, 231.
[55] V. A. Luchnikov, N. N. Medvedev, L. Oger, J.-P. Troadec, *Phys. Rev. E* **1999**, *59*, 7205.
[56] D.-S. Kim, Y. Cho, D. Kim, In Proceedings of the 16th Canadian Conference on Computational Geometry, **2004**; pp. 176–179.
[57] N. N. Medvedev, V. P. Voloshin, V. A. Luchnikov, M. L. Gavrilova, *J. Comput. Chem.* **2006**, *27*, 1676.
[58] D.-S. Kim, D. Kim, Y. Cho, K. Sugihara, *Comput.-Aid. Des.* **2006**, *38*, 808.
[59] D.-S. Kim, Y. Cho, K. Sugihara, *Comput.-Aid. Des.* **2010**, *42*, 874.
[60] D.-S. Kim, J.-K. Kim, Y. Cho, C.-M. Kim, *Comput.-Aid. Des.* **2012**, *44*, 85.
[61] D.-S. Kim, Y. Cho, J. Ryu, C.-M. Kim, *Comput.-Aid. Des.* **2010**, *42*, 795.
[62] J. Harris, H. Stocker, J. W. Harris, Handbook of Mathematics and Computational Science; Springer-Verlag New York Inc, New York, **1998**.
[63] R. PDB, RCSB PDB 2008 annual Technical report, RCSB PDB, **2008**.
[64] S. Dutta, K. Burkhardt, J. Young, G. J. Swaminathan, T. Matsuura, K. Henrick, H. Nakamura, H. M. Berman, *Mol. Biotechnol.* **2009**, *42*, 1.
[65] M.-Z. Shao, N. Badler, Spherical sampling by archimedes' theorem, Technical Report, University of Pennsylvania, **1996**.
[66] P. Glasserman, Monte Carlo Methods in Financial Engineering, Springer-Verlag New York Inc, New York, **2000**.

Received: 7 March 2011
Revised: 24 December 2011
Accepted: 28 January 2012
Published online on 7 March 2012