

The Geometry of Biomolecular Solvation

HERBERT EDELSBRUNNER AND PATRICE KOEHL

ABSTRACT. Years of research in biology have established that all cellular functions are deeply connected to the shape and dynamics of their molecular actors. As a response, structural molecular biology has emerged as a new line of experimental research focused on revealing the structure of biomolecules. The analysis of these structures has led to the development of computational biology, whose aim is to predict from molecular simulation properties inaccessible to experimental probes.

Here we focus on the representation of biomolecules used in these simulations, and in particular on the hard sphere models. We review how the geometry of the union of such spheres is used to model their interactions with their environment, and how it has been included in simulations of molecular dynamics.

In parallel, we review our own developments in mathematics and computer science on understanding the geometry of unions of balls, and their applications in molecular simulation.

1. Introduction

The molecular basis of life rests on the activity of biological macro-molecules, mostly nucleic acids and proteins. A perhaps surprising finding that crystallized over the last handful of decades is that geometric reasoning plays a major role in our attempt to understand these activities. In this paper, we address this connection between biology and geometry, focusing on hard sphere models of biomolecules.

The biomolecular revolution. Most living organisms are complex assemblies of cells, the building blocks for life. Each cell can be seen as a small chemical factory, involving thousands of different players with a large range of size and function. Among them, biological macro-molecules hold a special place. These usually large molecules serve as storage for the genetic information (the

Keywords: Molecular simulations, implicit solvent models, space-filling diagrams, spheres, balls, surface area, volume, derivatives.

Research of the two authors is partially supported by NSF under grant CCR-00-86013.

nucleic acids such as DNA and RNA), and as key actors of cellular functions (the proteins). Biochemistry, the field that studies these biomolecules, is currently experiencing a major revolution. In hope of deciphering the rules that define cellular functions, large scale experimental projects are performed as collaborative efforts involving many laboratories in many countries. The main aims of these projects are to provide maps of the genetic information of different organisms (the *genome projects*), to derive as much structural information as possible on the products of the corresponding genes (the *structural genomics projects*), and to relate these genes to the function of their products, usually deduced from their structure (the *functional genomics projects*). The success of these projects is completely changing the landscape of research in biology. As of February 2004, more than 170 whole genomes have been sequenced, corresponding to a database of over a million gene sequences. The need to store this data efficiently and to analyze its contents has led to the emergence of a collaborative effort between computer science and biology, referred to as *bio-informatics*. In parallel, the repository of biomolecular structures [Bernstein et al. 1977; Berman et al. 2000] contains more than 24,000 structures of proteins and nucleic acids. The similar need to organize and analyze the structural information contained in this database is leading to the emergence of another partnership between computer science and biology, namely *biogeometry*.

Significance of shape. Molecular structure or shape and chemical reactivity are highly correlated as the latter depends on the positions of the nuclei and electrons within the molecule. Indeed, chemists have long used three-dimensional plastic and metal models to understand the many subtle effects of structure on reactivity and have invested in experimentally determining the structure of important molecules. The same applies to biochemistry, where structural genomics projects are based on the premise that the structure of biomolecules implies their function. This premise rests on a number of specific and quantifiable correlations:

- enzymes fold into unique structures and the three-dimensional arrangement of their side-chains determines their catalytic activity;
- there is theoretical evidence that the mechanisms underlying protein complex formation depend on the shapes of the biomolecules involved [Levy et al. 2004];
- the folding rate of many small proteins correlates with a gross topological parameter that quantifies the difference between distance in space and along the main-chain [Plaxco et al. 1998; Alm and Baker 1999; Muñoz and Eaton 1999; Alm et al. 2002].

There is also evidence that the geometry of a protein plays a major role in defining its tolerance to mutation [Koehl and Levitt 2002]. We note in passing that structural biologists often refer to the ‘topology’ of a biomolecule when they mean the ‘geometry’ or ‘shape’ of the same. A common concrete model

representing this shape is a union of balls, in which each ball corresponds to an atom. Properties of the biomolecule are then expressed in terms of properties of the union. For example, the potential active sites are detected as cavities [Liang et al. 1998c; Edelsbrunner et al. 1998; Liang et al. 1998b] and the interaction with the environment is quantified through the surface area and/or volume of the union of balls [Eisenberg and McLachlan 1986; Ooi et al. 1987; Liang et al. 1998a]. In what follows, we discuss in detail the geometric properties of union of balls, and relate them to the physical properties of the biomolecules they represent.

Outline. Section 2 describes biomolecules, and surveys their different levels of representation, focusing on the hard sphere models used in nearly all molecular simulation. Section 3 describes the relationship between the geometry of a biomolecule and its energetics. Section 4 surveys analytical and approximate methods used in biomolecular simulations for computing the area and volume of a molecule, and their derivatives with respect to the atomic coordinates. Section 5 develops the mathematical background needed to give compact formulas for geometric measurements. Section 6 discusses implementations of these formulas and presents experimental results. Section 7 concludes the paper with a discussion of future research directions.

2. Biomolecules

Following the Greek philosopher Democritus, who proclaimed that all matter is an assemblage of atoms, we can build a hierarchy that relates life to atoms. All living organisms can be described as arrangements of cells, the smallest units capable of carrying functions important for life. Cells can be divided into organelles, which are themselves assemblies of biomolecules. These biomolecules are usually polymers of smaller subunits, whose atomic structures are known from standard chemistry. There are many remarkable aspects to this hierarchy, one of them being that it is ubiquitous to all life forms, from unicellular organisms to complex multicellular species like us. Unraveling the secrets behind this hierarchy has become one of the major challenges of the twentieth and now twenty-first centuries. While physics and chemistry have provided significant insight into the structure of the atoms and their arrangements in small chemical structures, the focus now is set on understanding the structure and function of biomolecules, mainly nucleic acids and proteins. Our presentation of these molecules follow the general dogma in biology that states that the genetic information contained in DNA is first transcribed to RNA molecules which are then translated into proteins.

DNA. Deoxyribonucleic acid is a long polymer built from four different building blocks, the nucleotides. The sequence in which the nucleotides are arranged contains the entire information required to describe cells and their functions.

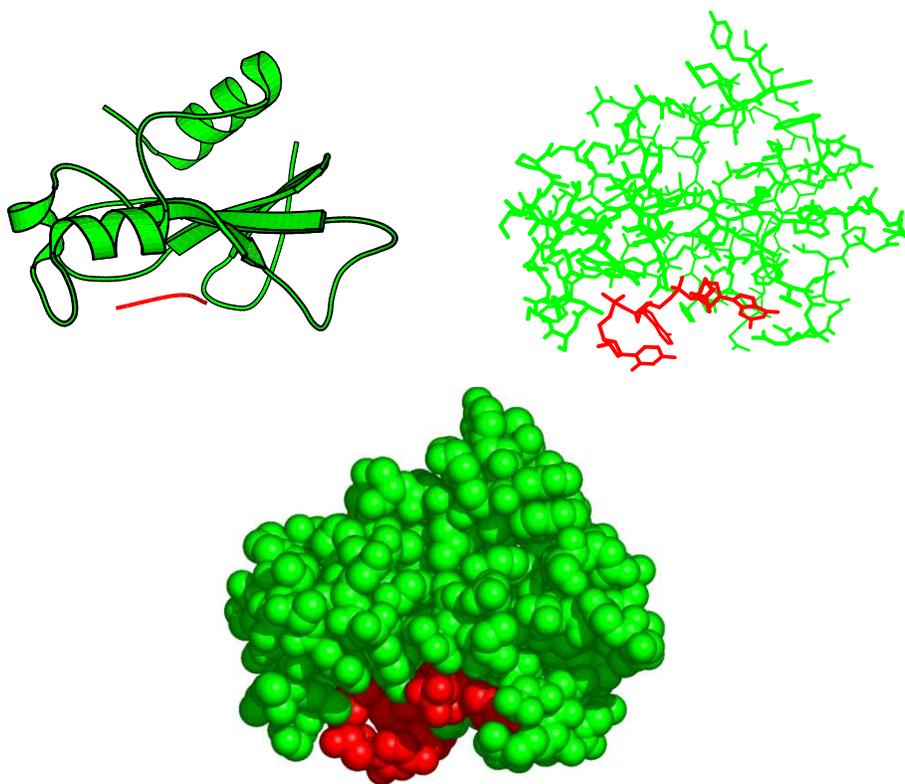


Figure 1. Visualizing protein-ligand interaction. Barnase is a small protein of 110 residues which has an endonuclease activity—it is able to cleave DNA fragments. Here we show the complex it forms with the small DNA fragment d(CGAC) [Buckle and Fersht 1994], using three different types of visualization. The coordinates are taken from the PDB file 1BRN. The protein is shown in green, and the DNA fragment in red.

Top left: Cartoon. This representation provides a high level view of the local organization of the protein in secondary structures, shown as idealized helices and strands. The DNA is shown as a short rod. This view highlights the position of the binding site where the DNA sits.

Top right: Skeletal model. This representation uses lines to represent bonds; atoms are located at their endpoints where the lines meet. It emphasizes the chemical nature of both molecules: for example, the four aromatic rings of the nucleotides of the DNA molecule are clearly visible.

Bottom: Space-filling diagram. Atoms are represented as balls centered at the atoms, with radii equal to the van der Waals radii of the atoms. This representation shows the tight binding between the protein and the ligand, that was not obvious from the other diagrams. Each of the representations is complementary to the others, and usually the biochemist uses all three when studying a protein, alone or, as illustrated here, in interaction with a ligand. The top panels were drawn using MOLSCRIPT [Kraulis 1991] and the bottom one with Pymol (<http://www.pymol.org>).

Despite this essential role in cellular functions, DNA molecules adopt surprisingly simple structures. Each nucleotide contains two parts, a backbone consisting of a deoxyribose and a phosphate, and an aromatic base, of which there are four types: adenine (A), thymine (T), guanine (G) and cytosine (C). The nucleotides are capable of being linked together to form a long chain, called a *strand*. Cells contain strands of DNA in pairs that are exact mirrors of each other. When correctly aligned, A pairs with T, G pairs with C, and the two strands form a double helix [Watson and Crick 1953]. The geometry of this helix is surprisingly uniform, with only small, albeit important, structural differences between regions of different sequences. The order in which the nucleotides appear in one DNA strand defines its sequence. Some stretches of the sequence contain information that can be translated first into an RNA molecule and then into a protein. These stretches are called *genes*; the ensemble of all genes of an organism constitutes its *genome* or *genetic information*. The remainder is *junk DNA*, which is assumed to correspond to fragments of genes that have been lost over the course of evolution. The DNA strands can stretch for millions of nucleotides. The size of the strands, as well as the fraction of junk DNA vary greatly between organisms and do not necessarily reflect differences in the complexity of the organisms. For example, the wheat genome contains approximately $1.6 \cdot 10^{10}$ bases, which is close to five times the size of the human genome. For a complete list of the genomes, see <http://wit.integratedgenomics.com/GOLD/> [Bernal et al. 2001]. The whole DNA molecules of more than 170 organisms have been sequenced in the existing genome projects, and many others are underway. There are more than a million genes that have been extracted from the DNA sequences and are collected in databases; see <http://www.ebi.ac.uk/embl>.

RNA. Ribonucleic acid molecules are very similar to DNA, being formed as sequences of four types of nucleotides, namely A, G, C, and uracil (U), which is a derivative of thymine. The sugar in the nucleotides of RNA is a ribose, which includes an extra oxygen compared to deoxyribose. The presence of this bulky extra oxygen prevents the formation of long and stable double helices. The single-stranded RNA can adopt a large variety of conformations, which remain difficult to predict based on its sequence. Interestingly, RNA is considered an essential molecule in the early steps of the origin of life. It is generally accepted now that before the appearance of living cells, the assemblies of self-replicating molecules were RNAs. In this early world, a single type of molecule performed both the function of active agents and the repository of its own description [Gilbert 1986; Gesteland and Atkins 1993; Cech 1993]. The activity of the RNA was related to its three-dimensional shape, while the coding corresponded to its linear sequence. This single molecule world had limitations since any modification of the RNA meant to improve its catalytic function could lead to a loss of its coding capabilities. Cellular life has evolved from this primary world by separating the two functions. RNA molecules now mainly serve as

templates that are used to synthesize the active molecules, namely the proteins. The information needed to synthesize the RNA is read from the genes coded by the DNA. It is assumed that DNA molecules evolved as a more stable, and consequently more reliable form of RNAs for storage purpose.

Proteins. While all biomolecules play an important part in life, there is something special about proteins, which are the products of the information contained in the genes. They are the active elements of life whose chemical activities regulate all cellular activities. According to Jacques Monod, it is in the protein that lies the secret of life: “C’est à ce niveau d’organisation chimique que gît, s’il y en a un, le secret de la vie” [Monod 1973]. As a consequence, studies of their sequence and structure occupy a central role in biology.

Proteins are heteropolymer chains of amino acids, often referred to as *residues*. This term comes from chemistry and describes the material found at the bottom of a reaction tube once a protein has been cut into pieces in order to determine its composition. There are twenty types of amino acids, which share a common *backbone* and are distinguished by their chemically diverse *side-chains*, which range in size from a single hydrogen atom to large aromatic rings and can be charged or include only nonpolar saturated hydrocarbons; see Table 1. The order

Type	Amino acids
nonpolar	glycine, alanine, valine, leucine, isoleucine, proline, methionine, tryptophan, phenylalanine
polar (neutral)	serine, threonine, asparagine, glutamine, cysteine, tyrosine
polar (acidic)	aspartic acid, glutamic acid
polar (basic)	lysine, arginine, histidine

Table 1. Classification of the 20 amino acids according to the chemical properties of their side-chains [Timberlake 1992]. Nonpolar amino acids do not have concentration of electric charges and are usually not soluble in water. Polar amino acids carry local concentration of charges, and are either globally neutral, negatively charged (acidic), or positively charged (basic). Acidic and basic amino acids are classically referred to as electron acceptors and electron donors, respectively, which can associate to form salt bridges in proteins.

in which amino acids appear defines the *primary sequence* of the protein. In its native environment, the polypeptide chain adopts a unique three-dimensional shape, referred to as the *tertiary* or *native structure* of the protein. In this structure, nonpolar amino acids have a tendency to re-group and form the core, while polar amino acids remain accessible to the solvent. The backbones are connected in sequence forming the protein *main-chain*, which frequently adopts canonical local shapes or *secondary structures*, such as α -helices and β -strands.

The former is a right handed helix with 3.6 amino acids per turn, while the latter is an approximately planar layout of the backbone. In the tertiary structure, β -strands are usually paired in parallel or anti-parallel arrangements, to form β -sheets. On average, the protein main-chain consists of about 25% in α -helix formation, 25% in β -strands, with the rest adopting less regular structural arrangements [Brooks et al. 1988]. From the seminal work of Anfinsen [1973], we know that the sequence fully determines the three-dimensional structure of the protein, which itself defines its function. While the key to the decoding of the information contained in genes was found more than fifty years ago (the genetic code), we have not yet found the rules that relate a protein sequence to its structure [Koehl and Levitt 1999; Baker and Sali 2001]. Our knowledge of protein structure therefore comes from years of experimental studies, either using X-ray crystallography or NMR spectroscopy. The first protein structures to be solved were those of hemoglobin and myoglobin [Kendrew et al. 1960; Perutz et al. 1960]. Currently, there are more than 16,000 protein structures in the database of biomolecular structures [Bernstein et al. 1977; Berman et al. 2000]; see <http://www.rcsb.org>.

Visualization. The need for visualizing biomolecules is based on the early understanding that their shape determines their function. Early crystallographers who studied proteins and nucleic acids could not rely—as it is common nowadays—on computers and computer graphics programs for representation and analysis. They had developed a large array of finely crafted physical models that allowed them to have a feeling for these molecules. These models, usually made out of painted wood, plastic, rubber and/or metal were designed to highlight different properties of the molecule under study. In the *space-filling models*, such as CPK [Corey and Pauling 1953; Koltun 1965], atoms are represented as spheres, whose radii are the atoms' van der Waals radii. They provide a volumetric representation of the biomolecules, and are useful to detect cavities and pockets that are potential active sites. In the *skeletal models*, chemical bonds are represented by rods, whose junctions define the position of the atoms. These models were used for example in [Kendrew et al. 1960], which studied myoglobin. They are useful to the chemists by highlighting the chemical reactivity of the biomolecules and, consequently, their potential activity. With the introduction of computer graphics to structural biology, the principles of these models have been translated into software such that molecules could be visualized on the computer screen. Figure 1 shows examples of computer visualizations of a protein-DNA interaction, including space-filling and skeletal representations.

3. Biomolecular Modeling

While the structural studies provide the necessary data on biomolecules, the key to their success lies in unraveling the connection between structure and

function. A survey of the many modeling initiatives motivated by this question is beyond the scope of this paper; detailed descriptions of biomolecular simulation techniques and their applications can be found in [Leach 2001; Becker et al. 2001]. We shall focus here on those in which geometry plays an essential role, mainly in the definition and computation of the energy of the biomolecule.

The apparition of computers, and the rapid increase of their power has given hope that theoretical methods can play a significant role in biochemistry. Computer simulations are expected to predict molecular properties that are inaccessible to experimental probes, as well as how these properties are affected by a change in the composition of a molecular system. For example, thermodynamics and kinetics play an important role in most functions of proteins. Proteins have to fold into a stable conformation in order to be active. Improper folding leads to inactive proteins that can accumulate and lead to disease (such as the prion proteins). Many proteins also adopt slightly different conformations in different environments. The cooperative rearrangement of hemoglobin upon binding of oxygen, for example, is essential for oxygen transport and release [Perutz 1990]. Predicting the equilibrium conformation of a protein in solution remains a "holy grail" in structural biology. In addition, while a few experimental probes exist to monitor protein dynamics events, such as hydrogen exchange experiments in NMR and small angle scattering of x-rays or neutrons, they remain elusive mainly because of the huge hierarchy of time-scale they involve. Biomolecular simulations have been designed to solve some of these problems. In particular, their aims are to describe the thermodynamic equilibrium properties of the system under study, through sampling of its free energy surface, as well as its dynamical properties.

Energy function. The state of a biomolecule is usually described in terms of its energy landscape. The native state corresponds to a large basin in this landscape, and it is mostly the structure of this basin that is of interest. Theoretically, the laws of quantum mechanics completely determine the wave function of any given molecule, and, in principle, we can compute the energy eigenvalues by solving Schrödinger's equation. In practice, however, only the simplest systems such as the hydrogen atom have an exact, explicit solution to this equation and modelers of large molecular systems must rely on approximations. Simulations of biomolecules are based on a space-filling representation of the molecule, in which the atoms are modeled by hard spheres that interact through empirical forces. A typical, semi-empirical energy function used in classical molecular simulation has the form

$$U = \sum_b k_b (r_b - r_b^0)^2 + \sum_b k_a (\theta_a - \theta_a^0)^2 + \sum_t k_t (1 + \cos n(\phi_t - \phi_t^0)) \\ + \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right)$$

[Levitt et al. 1995; Liwo et al. 1997a; 1997b; MacKerell et al. 1998; Kaminski et al. 2001; Price and Brooks 2002]. The terms in the first three sums represent bonded interactions: covalent bonds, valence angles, and torsions around bonds. The two terms in the last sum represent nonbonded interactions: a Lennard-Jones potential for van der Waals forces and the Coulomb potential for electrostatics. This sum usually excludes pairs of atoms separated by one, or two covalent bonds. The force constants, k , the minima, r^0 , θ^0 and ϕ^0 , the Lennard Jones parameters, A and B , and the atomic charges q define the force field. They are derived from data on small organic molecules, from both experiments and *ab initio* quantum calculations.

Note that U given above corresponds to the *internal energy* of the molecule, while we really need its *free energy* to describe its thermodynamic state. In thermodynamics, the term *free energy* denotes the total amount of energy in a system which can be converted to work. For a molecule, "work" is the transfer of energy related to organized motion. The free energy F is the difference between the internal energy of the molecule, and its *entropy*, where the entropy is a measure of disorder:

$$F = U - TS,$$

where T is the temperature of the system. Ideally, F is minimum when U is minimum and S is maximum. These two conditions however cannot be satisfied simultaneously by a molecule: U is minimum when there are many favorable contacts, leading to a single compact conformation for the molecule, while S is maximum when there is no privileged conformation for the molecule. In general, the thermodynamic equilibrium is reached through a compromise between these two terms. To get an estimate of the free energy of a molecule, we need to compute its internal energy, and sample the conformational space it can access. This sampling is performed through simulations, which are discussed below.

Simulation algorithms. There are three main types of algorithms used in this field, which we now describe.

Molecular dynamics simulations proceed by solving the classical equations of motions for the positions, velocities and accelerations of all atoms and molecules of the system under study. A state of the system is either described in cartesian or internal coordinates and the solution is computed numerically. In early work, macromolecules were simulated *in vacuo*, and only heavy (no hydrogen) atoms were included [McCammon et al. 1977]. This has changed as modern computers are now sufficiently powerful to simulate biomolecules in atomic detail using all-atom representations [Levitt and Sharon 1988]. The strengths of molecular dynamics are that it efficiently samples the states accessible to a system around its energy minimum, and that it provides kinetic data on the transitions between these states [Cheatham and Kollman 2000; Karplus and McCammon 2002]. The weakness of molecular dynamics is an inability to access long time-scales (on

the order of one microsecond or more even for small biomolecules [Duan and Kollman 1998].)

Monte Carlo techniques applied to biomolecular studies use stochastic moves, corresponding to rotation, translation, insertion or deletion of whole molecules, to sample the conformational space available to the molecule under study, and to calculate ensemble averages of physical or geometric quantities of interest, such as energy, or the fluctuation of some specific inter-atomic distances. In the limit of long Monte Carlo simulations, these ensemble averages correspond to thermodynamics equilibrium properties. A strength of Monte Carlo simulations is that they can be adapted to explore unfavorable regions of the energy landscape. This has been used to sample conformations of small simplified models of proteins, yielding a full characterization of the thermodynamics of their folding process [Hao and Scheraga 1994a; Hao and Scheraga 1994b].

A *molecular mechanics* study is not really a simulation as such, rather a mechanical investigation of the properties of one or more molecules. A good example would be finding the minimum of the potential energy U of a molecule. Note that U does not include entropic effects. Thus, the conformation of a molecule obtained through minimization of U does not necessarily correspond to the thermodynamic equilibrium state, which corresponds to the minimum of the free energy.

Protein solvation. Soluble biomolecules adopt their stable conformation in water, and are unfolded in the gas phase. It is therefore essential to account for water in any modeling experiment. Molecular dynamics simulation that include a large number of solvent molecules are the state of the art in this field, but they are inefficient as most of the computing time is spent on updating the position of the water molecule. It should further be noted that it is not always possible to account for the interaction with the solvent explicitly. For example, energy minimization of a system including both a protein and water molecules does not account for the entropy of water, which would behave like ice with respect to the protein. An alternative approach takes the effect of the solvent implicitly into account. In such an implicit solvent model, the effects of water is included in an effective solvation potential, $W = W_{\text{elec}} + W_{\text{np}}$, in which the first term accounts for the molecule-solvent electrostatics polarization, and the second for the molecule-solvent van der Waals interactions and for the formation of a cavity in the solvent. There is a large body of work that focuses on computing W_{elec} . A survey of the corresponding models is beyond the scope of this paper and we refer the reader to the excellent review [Simonson 2003] for more information.

Here we focus on computing W_{np} , the nonpolar effect of water on the biomolecule, sometimes referred to as the *hydrophobic effect*. Biomolecules contain both hydrophilic and hydrophobic parts. In their folded states, the hydrophilic parts are usually at the surface, where they can interact with water, and the

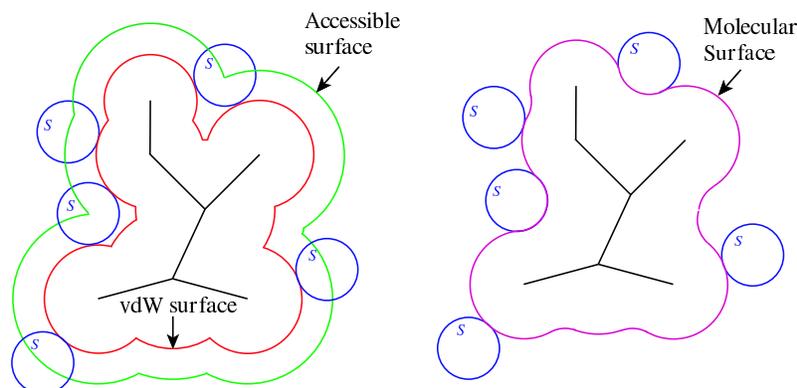


Figure 2. Different notions of protein surface. The *van der Waals surface* of a molecule (shown in red) is the surface of the union of balls representing all atoms, with radii set to the van der Waals radii. The *accessible surface* of the same molecule (shown in green) is the surface generated by the center of a solvent sphere (marked S) rolling on the van der Waals surface. The radius of the solvent sphere is usually set to 1.4 Å, the approximate radius of a water molecule. The accessible surface is also the obtained after expanding the radius of the atomic spheres by the radius of the solvent sphere. The *molecular surface* (shown in magenta) is the envelope generated by the rolling sphere. It differs from the van der Waals surface by covering portions of the volume inaccessible to the rolling sphere.

hydrophobic parts are buried in the interior, where they form an “oil drop with a polar coat” [Kauzmann 1959].

Quantifying the hydrophobic effect. In order to quantify the hydrophobic effect, Lee and Richards introduced the concept of the solvent-accessible surface [Lee and Richards 1971], illustrated in Figure 2. They computed the accessible area of each atom in both the folded and extended state of a protein, and found that the decrease in accessible area between the two states is greater for hydrophobic than for hydrophilic atoms. These ideas were further refined by Eisenberg and McLachlan [1986], who introduced the concept of a solvation free energy, computed as a weighted sum of the accessible areas A_i of all atoms i of the biomolecule:

$$W_{\text{np}} = \sum_i \alpha_i A_i,$$

where α_i is the atomic solvation parameter. It is not clear, however, which surface area should be used to compute the solvation energy [Wood and Thompson 1990; Tunon et al. 1992; Simonson and Brünger 1994]. There is also some evidence that for small solute, the hydrophobic term W_{np} is not proportional to the surface area [Simonson and Brünger 1994], but rather to the solvent excluded volume of the molecule [Lum et al. 1999]. A volume-dependent solvation term was originally introduced by Gibson and Scheraga [1967] as the hydration shell

model. Note that the ambiguity in the choice of the definition of the surface of a protein extends to the choice of its volume definition. Within this debate on the exact form of the solvation energy, there is however a consensus that it depends on the geometry of the biomolecule under study. Inclusion of W_{np} in a molecular simulation therefore requires the calculation of accurate surface areas and volumes. If the simulations rely on minimization, or integrate the equations of motion, the derivatives of the solvation energy are also needed. The calculation of second derivatives is also of interest in studying the normal modes of a biomolecule in a continuum solvent.

4. Computing Volumes and Areas

In this section, we review existing approaches to computing the surface area and/or volume of a biomolecule represented as a union of balls. The original approach of Lee and Richards [1971] computed the accessible surface area by first cutting the molecule with a set of parallel planes. The intersection of a plane with an atomic ball, if it exists, is a circle which can be partitioned into accessible arcs on the boundary and occluded arcs in the interior of the union. The accessible surface area of atom i is the sum of the contributions of all its accessible arcs, computed approximately as the product of the arc length and the spacing between the plane. This method was originally implemented in the program ACCESS [Lee and Richards 1971] and later in NACCESS (<http://wolf.bms.umist.ac.uk/naccess/>). Shrake and Rupley [1973] refined Lee and Richards' method and proposed a Monte Carlo numerical integration of the accessible surface area. Their method placed 92 points on each atomic sphere, and determined which points were accessible to solvent (not inside any other sphere). Efficient implementations of this method include applications of look-up tables [Legrand and Merz 1993], of vectorized algorithm [Wang and Levinthal 1991] and of parallel algorithms [Futamura et al. 2004]. Similar numerical methods have been developed for computing the volume of a union of balls [Rowlinson 1963; Pavani and Raghino 1982; Gavezzotti 1983].

The surface area and/or volume computed by numerical integration over a set of points, even if closely spaced, is not accurate and cannot be readily differentiated. To improve upon the numerical methods, analytical approximations to the accessible surface area have been developed, which either treat multiple overlapping balls probabilistically [Wodak and Janin 1980; Hasel et al. 1988; Cavallo et al. 2003] or ignore them altogether [Street and Mayo 1998; Weiser et al. 1999a]. Better analytical methods describe the molecule as a union of pieces of balls, each defined by their center, radius, and arcs forming their boundary, and subsequently apply analytical geometry to compute the surface area and volume [Richmond 1984; Connolly 1985; Dodd and Theodorou 1991; Petitjean 1994; Iriša 1996]. Pavani and Raghino [1982] proposed a method for computing the volume of a molecule by inclusion-exclusion. In their implementation, only

intersections of up to three balls were considered. Petitjean however noticed that practical situations for proteins frequently involve simultaneous overlaps of up to six balls [Petitjean 1994]. Subsequently, Pavani and Ranghino's idea was generalized to any number of simultaneous overlaps by Gibson and Scheraga [Gibson and Scheraga 1987] and by Petitjean [Petitjean 1994], applying a theorem that states that higher-order overlaps can always be reduced to lower-order overlaps [Kratky 1978]. Doing the reduction correctly remains however computationally difficult and expensive. The Alpha Shape Theory solves this problem using Delaunay triangulations and their filtrations, as described by Edelsbrunner [Edelsbrunner 1995]. It will be presented in greater detail in the next section.

The distinction between approximate and exact computation also applies to existing methods for computing the derivatives of the volume and surface area of a molecule with respect to its atomic coordinates [Kundrot et al. 1991; Gogonea and Osawa 1994; Gogonea and Osawa 1995; Cossi et al. 1996]. In the case of the derivatives of the surface area, computationally efficient methods were implemented in the MSEED software by Perrot et al. [1992] and in the SASAD software by Sridharan et al. [1994]. All these methods introduce approximations to deal with singularities caused by numerical errors or by discontinuities in the derivatives [Gogonea and Osawa 1995]. There is also an inherent difficulty in using a potential based on surface area or volume in biomolecular simulations. Although the area and volume are continuous in the position of the atoms, their derivatives are not. This problem of discontinuities was studied in more details for surface area calculation [Perrot et al. 1992; Wawak et al. 1994].

The complexity of the computation of the area and volume of a union of balls, the problems of singularities encountered when computing their derivatives, and the inherent existence of discontinuities have led to the development of alternative geometric representations of molecules. Here we mention the Gaussian description of molecular shape, that allows for easy analytical computation of surface area, volume and derivatives [Grant and Pickup 1995; Weiser et al. 1999b], and the molecular skin, which will be described in the next section.

5. Alpha Shape Theory

In this section, we discuss in some detail the inclusion-exclusion approach to computing area, volume, and their derivatives. It is based on the concept of alpha complexes [Edelsbrunner et al. 1983; Edelsbrunner and Mücke 1994], which are sub-complexes of the Delaunay triangulation [Delaunay 1934] of a set of spheres.

Voronoi decomposition and dual complex. Consider a finite set of spheres S_i with centers $z_i \in \mathbb{R}^3$ and radii $r_i \in \mathbb{R}$ and let B_i be the ball bounded by S_i . To allow for varying radii, we measure square distance of a point x from S_i using $\pi_i(x) = \|x - z_i\|^2 - r_i^2$. The *Voronoi region* of S_i consists of all points

x at least as close to S_i as to any other sphere: $V_i = \{x \in \mathbb{R}^3 \mid \pi_i(x) \leq \pi_j(x)\}$. As illustrated in Figure 3, the Voronoi region of S_i is a convex polyhedron obtained as the common intersection of finitely many closed half-spaces, one per sphere $S_j \neq S_i$. If S_i and S_j intersect in a circle then the plane bounding the corresponding half-space passes through that circle. It follows that the Voronoi regions decompose the union of balls B_i into convex regions of the form $B_i \cap V_i$. The boundary of each such region consists of spherical patches on S_i and planar patches on the boundary of V_i . The spherical patches separate the inside from the outside and the planar patches decompose the inside of the union. The

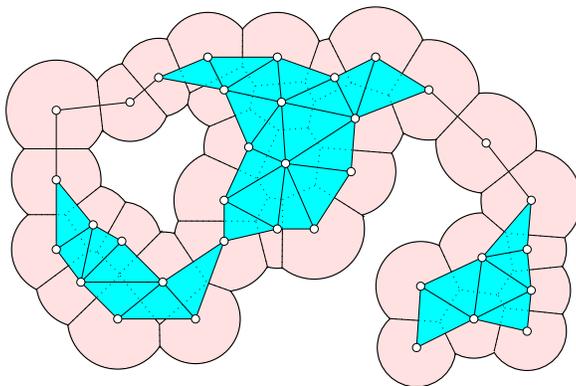


Figure 3. Voronoi decomposition and dual complex. Given a finite set of disks, the Voronoi diagram decomposes the plane into regions in which one circle minimizes the square distance measured as $\|x - z_i\|^2 - r_i^2$. In the drawing, we restrict the Voronoi diagram to within the portion of the plane covered by the disks and get a decomposition of the union into convex regions. The dual Delaunay triangulation is obtained by drawing edges between circle centers of neighboring Voronoi regions. To draw the dual complex of the disks we limit ourselves to edges and triangles between centers whose corresponding restricted Voronoi regions have a nonempty common intersection.

Delaunay triangulation is the dual of the Voronoi diagram, obtained by drawing an edge between the centers of S_i and S_j if the two corresponding Voronoi regions share a common face. Furthermore, we draw a triangle connecting z_i , z_j and z_k if V_i , V_j and V_k intersect in a common line segment, and we draw a tetrahedron connecting z_i , z_j , z_k and z_ℓ if V_i , V_j , V_k and V_ℓ meet at a common point. Assuming general position of the spheres, there are no other cases to be considered. We refer to this as the *generic case* but hasten to mention that because of limited precision it is rare in practice. Nevertheless, we can simulate a perturbation in our algorithm [Edelsbrunner and Mücke 1990], which is an effective method to consistently unfold potentially complicated degenerate cases to nondegenerate ones.

Suppose we limit the construction of the dual triangulation to within the union of balls, as illustrated in Figure 3. In other words, we draw a dual edge between z_i and z_j only if $B_i \cap V_i$ and $B_j \cap V_j$ share a common face, and similarly for triangles and tetrahedra. The result is a sub-complex of the Delaunay triangulation which we refer to as the *dual complex* $K = K_0$ of the set of spheres. For various reasons, including the definition of pockets in biomolecules [Edelsbrunner et al. 1998], it is useful to alter the spheres by increasing or decreasing their radii. We do this in a way that leaves the Voronoi diagram invariant. Modeling growth with a positive real number and shrinkage with a positive real multiple of the imaginary unit, both denoted as α , we obtain a real number α^2 that may be positive or negative. For each i let $S_i(\alpha)$ be the sphere with center z_i and radius $\sqrt{r_i^2 + \alpha^2}$. Interpreting spheres with imaginary radii as empty, the *alpha complex* K_α of the spheres S_i is the dual complex of the spheres $S_i(\alpha)$. If we increase α^2 continuously from $-\infty$ to $+\infty$ we get a continuous nested sequence of unions of balls and a discrete nested sequence of alpha complexes.

Area and volume formulas. A simplex τ in the dual complex can be interpreted abstractly as a collection of balls, one ball if it is a vertex, two if it is an edge, etc. In this interpretation, the dual complex is a system of sets of balls, and because every face of a simplex in K also belongs to K , this system is closed under containment. It now makes sense to write $\text{vol} \bigcap \tau$ for the volume of the intersection of the balls in τ . This is the kind of term we would see in an inclusion-exclusion formula for the volume of the union of balls, $\bigcup_i B_i$. As proved in [Edelsbrunner 1995], the inclusion-exclusion formula that corresponds to the dual complex gives indeed the correct volume.

VOLUME THEOREM:

$$\text{vol} \bigcup_i B_i = \sum_{\tau \in K} (-1)^{\dim \tau} \text{vol} \bigcap \tau.$$

Here $\dim \tau = \text{card} \tau - 1$ is the dimension of the simplex. This result overcomes past difficulties by implicitly reducing higher-order to lower-order overlaps. An added advantage of this formula is that the balls in each term form a unique geometric configuration so that the analytic calculation of the volume can be done without case analysis. Specifically, the balls in a simplex $\tau \in K$ are *independent* in the sense that for every face $v \subseteq \tau$ there exists a point that lies inside all balls that belong to v and outside all balls that belong to τ but not to v .

A similar formula can be derived for the area of the boundary of the union of balls. One way to arrive at this formula is to consider a sphere S_i and to observe that its contribution is the area of the entire sphere, $4\pi r_i^2$, minus the portion covered by caps of the form $S_i \cap B_j$, for $j \neq i$. The configuration of caps on S_i is but a spherical version of a configuration of disks, and computing its area is the same problem as computing the volume of a set of balls, only one dimension lower. To express that area as an alternating sum we need its dual complex,

but this is nothing other than the link of S_i in K , consisting of all simplices v that do not contain B_i but are faces of simplices that contain B_i : $B_i \notin v$ and $v \cup \{B_i\} \in K$. Specifically, the area contribution of S_i is the area of the sphere minus the sum of $(-1)^{\dim v} \text{area}(S_i \cap \bigcap v)$. We collect all these contributions and combine terms to get the final result.

AREA THEOREM:

$$\text{area} \bigcup_i B_i = \sum_{\tau \in K} (-1)^{\dim \tau} \text{area} \bigcap \tau.$$

We see that the principle of inclusion-exclusion is quite versatile, which is important for applications in which we might want to measure aspects of the union of balls that are similar to but different from its volume and surface area. Examples are

- the total length of arcs in the boundary;
- voids of empty space surrounded by the union;
- weighted versions of the above.

Of the three extensions, the least obvious is how to measure voids. The other two are needed to express the derivative of the weighted volume and area, which are discussed next.

Area and volume derivatives. We are interested in the derivatives of the area and the volume of a union of n balls with respect to their positions in space. Since we keep the radii fixed, we may specify the configuration by the vector $\mathbf{z} \in \mathbb{R}^{3n}$ of center coordinates. The area thus becomes a function $f : \mathbb{R}^{3n} \rightarrow \mathbb{R}$, and similar for the volume. The *derivative* of f at \mathbf{z} is the best linear approximation at that configuration, $Df_{\mathbf{z}} : \mathbb{R}^{3n} \rightarrow \mathbb{R}$. This linear function is completely specified by the gradient $\mathbf{a} = \nabla f(\mathbf{z})$, namely

$$Df_{\mathbf{z}}(\mathbf{t}) = \langle \mathbf{a}, \mathbf{t} \rangle,$$

in which $\mathbf{t} \in \mathbb{R}^{3n}$ is the motion vector. In [Edelsbrunner and Koehl 2003; Bryant et al. 2004] we gave formulas for the derivatives by specifying the gradient in terms of simple parameters readily computable from the input spheres. To state the result for the area, let $\zeta_{ij} = \|z_i - z_j\|$ be the distance between the two centers and write $u_{ij} = (z_i - z_j)/\zeta_{ij}$ for the unit vector in the direction of the connecting line. For each $k \neq i, j$ let

$$w_{ijk} = u_{ik} - \langle u_{ik}, u_{ij} \rangle \cdot u_{ij}$$

be the component of u_{ik} normal to u_{ij} , and let $u_{ijk} = w_{ijk}/\|w_{ijk}\|$ be the unit vector in that normal direction. Finally, let r_{ijk} be half the distance between the two points at which the spheres S_i , S_j , and S_k meet. For completeness, we state the result for the case in which the area contribution is weighted by the constant α_i , the corresponding atomic solvation parameter.

WEIGHTED AREA DERIVATIVE THEOREM: *The gradient $\mathbf{a} \in \mathbb{R}^{3n}$ of the weighted area derivative at a configuration of balls $\mathbf{z} \in \mathbb{R}^{3n}$ is*

$$\begin{bmatrix} \mathbf{a}_{3i+1} \\ \mathbf{a}_{3i+2} \\ \mathbf{a}_{3i+3} \end{bmatrix} = \sum_j \left(s_{ij} \cdot \mathbf{a}_{ij} + \sum_k b_{ijk} \cdot \mathbf{a}_{ijk} \right),$$

$$\mathbf{a}_{ij} = \pi \left((\alpha_i r_i + \alpha_j r_j) - (\alpha_i r_i - \alpha_j r_j) \frac{r_i^2 - r_j^2}{\zeta_{ij}^2} \right) \cdot \mathbf{u}_{ij},$$

$$\mathbf{a}_{ijk} = 2r_{ijk} \frac{\alpha_i r_i - \alpha_j r_j}{\zeta_{ij}} \cdot \mathbf{u}_{ijk},$$

for $0 \leq i < n$. The sums are over all boundary edges $z_i z_j$ and their triangles $z_i z_j z_k$ in K .

The geometrically interesting terms in the formula are s_{ij} , the fraction of the circle $S_i \cap S_j$ that belongs to the boundary of the union, and b_{ijk} , the fraction of the line segment connecting the point pair $S_i \cap S_j \cap S_k$ that belongs to the Voronoi segment $V_i \cap V_j \cap V_k$. A remarkable aspect of the formula is the existence of terms that depend on three rather than just two spheres. These terms vanish in the unweighted case if all radii are the same. We can reuse some of the notation to state the result for the volume. We again state the result for the case in which the volume of $B_i \cap V_i$ is weighted by the constant α_i .

WEIGHTED VOLUME DERIVATIVE THEOREM: *The gradient $\mathbf{v} \in \mathbb{R}^{3n}$ of the weighted volume derivative of a configuration of balls $\mathbf{z} \in \mathbb{R}^{3n}$ is*

$$\begin{bmatrix} \mathbf{v}_{3i+1} \\ \mathbf{v}_{3i+2} \\ \mathbf{v}_{3i+3} \end{bmatrix} = \sum_j b_{ij} r_{ij}^2 \pi (y_{ij} \cdot \mathbf{u}_{ij} + x_{ij} \cdot \mathbf{v}_{ij}),$$

$$y_{ij} = \frac{\alpha_i + \alpha_j}{2} + \frac{(\alpha_j - \alpha_i)(r_i^2 - r_j^2)}{2\zeta_{ij}^2},$$

$$x_{ij} = \frac{2(\alpha_i - \alpha_j)}{3\zeta_{ij}},$$

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

Here r_{ij} is the radius of the disk spanned by the circle $S_i \cap S_j$ and b_{ij} is the fraction of this disk that belongs to the corresponding Voronoi polygon, $V_i \cap V_j$. The most interesting term in this formula is the average vector \mathbf{v}_{ij} from the center of the disk to the boundary of its intersection with the Voronoi polygon. In computing the average, we weight each point on this boundary by the area of the infinitesimal triangle it defines with the center. This vector is used to express the gain and loss of weighted volume as the disk rotates and trades off contributions of the two balls it separates. In the unweighted case, we gain as much as we lose which explains why x_{ij} vanishes and thus cancels any effect \mathbf{v}_{ij} would have.

Continuity of the derivative. If considered over all configurations, the derivative of f is a function $Df : \mathbb{R}^{3n} \times \mathbb{R}^{3n} \rightarrow \mathbb{R}$. As described earlier, for each state $\mathbf{z} \in \mathbb{R}^{3n}$, this is a linear function $\mathbb{R}^{3n} \rightarrow \mathbb{R}$ completely specified by the gradient at \mathbf{z} . It is convenient to introduce another function $\nabla f : \mathbb{R}^{3n} \rightarrow \mathbb{R}^{3n}$ such that $Df(\mathbf{z}, \mathbf{t}) = \langle \nabla f(\mathbf{z}), \mathbf{t} \rangle$. For the purpose of simulating molecular motion, it is important that ∇f be continuous, at least mostly, and if there are discontinuities, that we are able to recognize and predict them. Unfortunately, the derivatives of the weighted area and the weighted volume are both not everywhere continuous. The good news is that the formulas in the two Derivative Theorems permit a complete analysis.

Interestingly, a configuration at which ∇f is not continuous is necessarily a configuration at which the dual complex is ambiguous, and this is true for the area and the volume. For example, the area derivative has a discontinuity at configurations that contain two spheres touching in a point that belongs to the boundary of the union. The set of configurations \mathbf{z} that contain such spheres is a $(3n-1)$ -dimensional subset of \mathbb{R}^{3n} . In contrast, the volume derivative has discontinuities only at configurations that contain two equal spheres or three spheres that meet in a common circle, both in the weighted and the unweighted case. The set of such configurations is a $(3n-3)$ -dimensional subset of \mathbb{R}^{3n} . A molecular dynamics simulation has to do extra work to compensate for the missing information whenever it runs into a discontinuity of the derivative [Carver 1978; Gear and Østerby 1984]. This occurs less often for the volume than for the area, firstly because the dimension of such configurations is less and secondly because the specific structure of these configurations makes them physically unlikely.

Voids and pockets. A *void* V is a maximal connected subset of space that is disjoint from and completely surrounded by the union of balls. Its surface area is easily computed by identifying the sphere patches on the boundary of the union that also bound the void. It helps to know that there is a deformation retraction from $\bigcup_i B_i$ to the dual complex [Edelsbrunner 1995]. Similarly, there is a corresponding void in K represented by a connected set of simplices in the Delaunay triangulation, that do not belong to K . This set U is open and its boundary (the simplices added by closure) forms what one may call the dual complex of the boundary of V . We use normalized angles to select the relevant portions of the intersections of balls. To define this concept, let v be a face of a simplex τ and consider a sufficiently small sphere in the affine hull of τ whose center is in the interior of v . The *normalized angle* $\varphi_{v,\tau}$ is the fraction of the sphere contained in τ . For example, if τ is a tetrahedron then we get the solid angle at a vertex, the dihedral angle at an edge, and $\frac{1}{2}$ at a triangle.

VOID AREA THEOREM:

$$\text{area } V = \sum_{v \subseteq \tau} (-1)^{\dim v} \varphi_{v,\tau} \text{area } \bigcap v.$$

The sum is over all faces $v \in K$ of simplices $\tau \in U$.

The correctness of the formula is not immediate and relies on an identity for simplices proved in [Edelsbrunner 1995]. Similarly, we can use U to compute the volume of V .

VOID VOLUME THEOREM:

$$\text{vol } V = \text{vol } U - \sum_{v \subseteq \tau} (-1)^{\dim v} \varphi_{v,\tau} \text{vol } \bigcap v.$$

The sum is over all faces $v \in K$ of simplices $\tau \in U$.

Here, $\text{vol } U$ is simply the sum of volumes of the tetrahedra of U . There are similar angle-weighted formulas for the entire union of balls. It would be interesting to generalize the Void Area and Volume Theorems to pockets as defined in [Edelsbrunner et al. 1998]. In contrast to a void, a *pocket* is not completely surrounded but connected to the outside through narrow channels. Again we have a corresponding set of simplices in the Delaunay triangulation that do not belong to the dual complex, but this set is partially closed at the places the pocket connects to the outside. The inclusion-exclusion formulas still apply, but there are cases in which the cancellation of terms near the connecting channel is not complete and leads to slightly incorrect measurements.

Alternative geometric representations. The sensitivity of simulation software to discontinuities in the derivative suggests that we approximate the surface area by another function. For example, we may use a *shell representation* and approximate area by the volume in that shell. This can be done with uniform thickness everywhere, or with variable thickness that depends on the radii, such as $\bigcup_i B_i(\varepsilon) - \bigcup_i B_i(-\varepsilon)$, where the small positive ε affects the radii as formulated in the definition of the alpha complex. The latter lends itself to fast computation because both the outer and the inner union have their dual complex in the same Delaunay triangulation and measuring both takes barely more time than measuring one. Another alternative to the union of balls is the *molecular surface* explained in Figure 2. Here we roll a sphere with fixed radius r about a union of balls. The rolling motion is captured by the boundary of another union in which all balls grow by r in radius. For each patch, arc, and vertex in this boundary the molecular surface contains a (smaller) sphere patch, a torus patch, and a (reversed) sphere patch. We can therefore collect all patches of the molecular surface using the dual complex of the grown balls and get the surface area by accumulation. At rare occasions, the patches form self-intersections which leads to slightly incorrect measurements. Computing these self-intersections can be rather involved analytically [Bajaj et al. 1997]. A similar alternative to unions of balls is the *molecular skin* as defined in [Edelsbrunner 1999]. Instead of torus patches, it uses hyperboloids of one and two sheets to blend between the spheres; see Figure 4. The surface is decomposed into simple patches by a mix of the

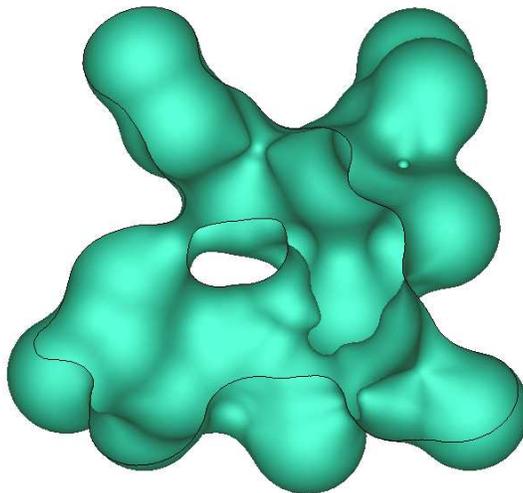


Figure 4. Molecular skin in cut-away view. Half the surface of a small molecule of about forty atoms.

Voronoi diagram and the Delaunay triangulation. These patches are free of self-intersections and the area can be computed by accumulation, as before but without running the risk of making mistakes. At this time, there is no complete analysis of the volume and area derivatives available, neither for the molecular surface nor the molecular skin.

6. Algorithm and Implementation

We have written a new version of the Alpha Shape software [Edelsbrunner and Mücke 1994], specific to molecular simulation applications, implementing the weighted surface area, the weighted volume, and the derivatives of both. The software is distributed as an Open Source program under the name AlphaVol at <http://biogeometry.duke.edu/software/proshape>.

Overview. The software takes as input a set of spheres S_i in \mathbb{R}^3 , each specified by the coordinates of its center z_i and its radius r_i . Such a set representing a protein can for example be extracted from the corresponding PDB file using one of several standard sets of van der Waals radii. The computation is performed through three successive tasks:

1. Construct the Delaunay triangulation.
2. Extract the dual complex.
3. Measure the union using inclusion-exclusion.

The main difference to the old Alpha Shapes software is the speed resulting from an improvement of all steps by about two orders of magnitude. We achieve this

through careful redesign of low-level computations (determinants in Task 1 and term management in Task 3) and the limitation in scope (dual complex instead of filtration of alpha complexes in Task 2). We review all three steps, focusing on nonobvious implementation details that have an impact on the correctness and running time of the software.

Delaunay triangulation. Our implementation of the Delaunay triangulation is based on the randomized incremental algorithm described in [Edelsbrunner and Shah 1996]. Following the paper's recommendation, we use a minimalist approach to storing the triangulation in a linear array of tetrahedra. For each tetrahedron, we store the indices of its four vertices, the indices of the four neighboring tetrahedra, a label, and the position of the opposite vertex in the vertex list of each neighboring tetrahedron. For each vertex we use four double-precision real numbers for the coordinates and the radius of the corresponding sphere. The triangles and edges are implicit in this representation.

The triangulation is constructed incrementally, by adding one sphere at a time. Before starting the construction, we re-index such that S_1, S_2, \dots, S_n is a random permutation of the input spheres. To reduce the number of cases, we choose four additional spheres with their centers at infinity so that all input spheres are contained in the tetrahedron they define. Let D_i be the Delaunay triangulation of the four spheres at infinity together with S_1, S_2, \dots, S_i . The algorithm proceeds by iterating three steps:

For $i = 1$ to n ,

- 1.1. find the tetrahedron $\tau \in D_{i-1}$ that contains z_i ;
- 1.2. add z_i to decompose τ into four tetrahedra;
- 1.3. flip locally non-Delaunay triangles in the link of z_i .

Step 1.1 is implemented using the jump-and-walk technique proposed by Mücke et al. [1999]. Here we choose a small random sample of the vertices in the current triangulation and walk from the vertex closest to z_i to τ . In this walk, we repeatedly test whether z_i is inside a tetrahedron v and whether v remains in the current Delaunay triangulation. These tests are decided by computing the signs of the determinants of four 4-by-4 matrices, which place z_i relative to the faces of v , and the sign of one 5-by-5 matrix. By noticing that any two of the 4-by-4 matrices share three rows (corresponding to z_i and the vertices of a shared edge) we find that 28 multiplications suffice to compute all five determinants. In Step 1.2, the sphere S_i is sometimes discarded without decomposing τ , namely when its Voronoi region is empty. This usually does not happen for molecular data. A flip in Step 1.3 replaces two tetrahedra by three or three by two. We are also prepared to remove a sphere by replacing four tetrahedra by one, but this again is usually not necessary for molecular data. The fact that any arbitrary ordering of the flips will successfully repair the Delaunay triangulation is nontrivial but has been established in [Edelsbrunner and Shah 1996]. The numerical tests needed to

decide which flips to make compute again signs of determinants of 4-by-4 and 5-by-5 matrices. As before, we save time by recognizing common rows and reusing partial results in the form of shared minors. An important ingredient in this context is the treatment of singularities. Inexact versions of the numerical tests are vulnerable to roundoff errors and can lead to wrong output. Following work in computational geometry [Fortune and VanWyk 1996], we implemented both tests using a so-called floating-point filter that first evaluates the tests approximately, using floating-points arithmetic, and if the results cannot be trusted, switches to exact arithmetic. As a side-benefit, we can now correctly recognize degenerate cases and use a simulated perturbation to consistently reduce them to general cases [Edelsbrunner and Mücke 1990].

Dual complex. Given the Delaunay triangulation D of the input spheres, we construct the dual complex $K \subseteq D$ by labeling the Delaunay simplices. Specifically, for each simplex $\tau \in D$ there is a threshold α_τ such that $\tau \in K_\alpha$ iff $\alpha_\tau^2 \leq \alpha^2$. Hence τ belongs to the dual complex iff $\alpha_\tau^2 \leq 0$. We call τ a *critical* simplex if α_τ separates the case in which the balls $B_i(\alpha)$ defining τ have an empty common intersection from the case in which they have a nonempty common intersection. These simplices are characterized by the fact that all other balls are further than orthogonal from the smallest sphere orthogonal to all balls B_i defining τ . (Two balls B_i and B_j of centers z_i and z_j and radii r_i and r_j , respectively, are orthogonal iff $\|z_i - z_j\|^2 = r_i^2 + r_j^2$.) All other simplices are *regular* and need a critical simplex they are face of to be included in the dual complex. To label the Delaunay simplices, we therefore need to be able to recognize critical simplices and to decide the signs of their square thresholds. Both tests can be expressed in terms of the signs of the determinants of small matrices whose entries are center coordinates and square radii of the input spheres. Detailed expressions for these tests can be found in [Edelsbrunner 1992; Edelsbrunner and Mücke 1994].

We evaluate these tests with the same care for singularities and numerical uncertainties as used in the construction of the Delaunay triangulation. Specifically, we apply filters and repeat the computation in exact arithmetic unless we can be sure that the initial floating-point computation gives the correct sign.

Weighted surface area and volume. We compute the weighted volume of a union of balls using the Volume Theorem in Section 4. The weights are worked into the formula by decomposing each term, $\text{vol} \cap \tau$, into $\dim \tau + 1$ terms using the bisector planes also used in the Voronoi diagram. This decomposition is natural since it is the easiest way to compute the volume of $\cap \tau$ in the first place, even in the unweighted case.

We could do the same for the weighted area, effectively reducing the formula in the Area Theorem further to an alternating sum in which every term is the area of the intersection of a sphere with up to three half-spaces. Simple analytic formulas for the area of such an intersection can be found in [Edelsbrunner and Fu 1994]. We choose an alternative path deriving a similar formula (yielding the

same result) from the angle-weighted formula given in the Void Area Theorem. An adaptation of this formula to an entire union of balls gives

$$\text{area} \bigcup_i B_i = \sum_v (-1)^{\dim v} \varphi_v \text{area} \bigcap v,$$

where the sum is over all simplices v in the boundary of K and φ_v is the normalized angle around v not covered by simplices that contain v as a face. As before, we further decompose each term into the intersection of a sphere and a small number of half-spaces. The above sum is usually shorter than that in the straight Area Theorem, which has a term for every simplex in the dual complex. Another difference is that each term is the intersection of at most three balls as opposed to at most four in the Area Theorem. The two differences compensate for the extra effort of computing normalized angles and more, leading to code that for proteins is about twice as fast as that based on the straight Area Theorem.

Derivatives. We now explain how we compute the geometric ingredients in the two Derivative Theorems stated in Section 5. For the area derivative, these are the fractions s_{ij} and b_{ijk} . Both can be computed using inclusion-exclusion over links inside the dual complex. Recall that s_{ij} is the fraction of the circle $S_i \cap S_j$ that belongs to the boundary of the union of balls. Equivalently, it is the fraction of the circle not covered by arcs of the form $S_i \cap S_j \cap B_k$. We may interpret these arcs as one-dimensional balls and measure their union using inclusion-exclusion, not unlike the formula in the Volume Theorem. We find the same symmetry in dimension in the corresponding combinatorial complexes. Specifically, the (one-dimensional) dual complex of the arcs is isomorphic to the link of the edge $z_i z_j$ in the dual complex of the balls. The link of this edge in the Delaunay triangulation is a cycle and in K is a sub-complex consisting of vertices z_k and edges $z_k z_\ell$. Writing s_{ij}^k and $s_{ij}^{k\ell}$ for the fractions of the circle inside B_k and inside $B_k \cap B_\ell$, we have

$$s_{ij} = 1 - \sum_k s_{ij}^k + \sum_{k,\ell} s_{ij}^{k\ell},$$

where the sums range over the link of the edge $z_i z_j$ in K . The computation of b_{ijk} is similar but simpler because the dimension of the link of a triangle is only zero, consisting of at most two vertices. Consider the line segment connecting the two points at which S_i , S_j and S_k meet and note that all points x on this line segment have the same distance to the three spheres: $\pi_i(x) = \pi_j(x) = \pi_k(x)$. Writing b_{ijk}^ℓ for the fraction of points x whose square distance from S_ℓ is less than from the three defining spheres we get $b_{ijk} = 1 - \sum_\ell b_{ijk}^\ell$, where the sum is over the vertices z_ℓ in the link of the triangle. For further details refer to [Bryant et al. 2004]. The same quantity but one dimension higher appears in the volume derivative. Specifically, b_{ij} is the fraction of the disk B_{ij} spanned by the circle $S_i \cap S_j$ that belongs to the corresponding Voronoi polygon. Let B_{ij}^k

be the subset of points x in this disk whose square distance to S_k is less than to the two defining spheres: $\pi_k(x) < \pi_i(x) = \pi_j(x)$. Similarly, let $B_{ij}^{k\ell} = B_{ij}^k \cap B_{ij}^\ell$ and write b_{ij}^k and $b_{ij}^{k\ell}$ for the respective fractions of the disk they define. Then $b_{ij} = 1 - \sum_k b_{ij}^k + \sum_{k,\ell} b_{ij}^{k\ell}$. Finally consider the average vector v_{ij} from the center of the disk to the boundary of its intersection with the Voronoi polygon. Its computation follows the same pattern of inclusion-exclusion over the link of the edge, $v_{ij} = 0 - \sum_k v_{ij}^k + \sum_{k,\ell} v_{ij}^{k\ell}$, where v_{ij}^k is the average vectors to the arc minus the average vector to the line segment in the boundary of B_{ij}^k , and similarly $v_{ij}^{k\ell}$ is the difference between the two average vectors of $B_{ij}^{k\ell}$. For further details refer to [Edelsbrunner and Koehl 2003].

Performance. We discuss the actual performance of AlphaVol. We have computed the weighted surface areas and volumes, as well as their derivatives with respect to atomic coordinates, of 2,868 proteins varying in size from 17 to 500 residues. These proteins contain between 124 and 4,063 atoms. Computing times for AlphaVol on an Intel 1600 MHz Pentium IV computer are shown in Figure 5.

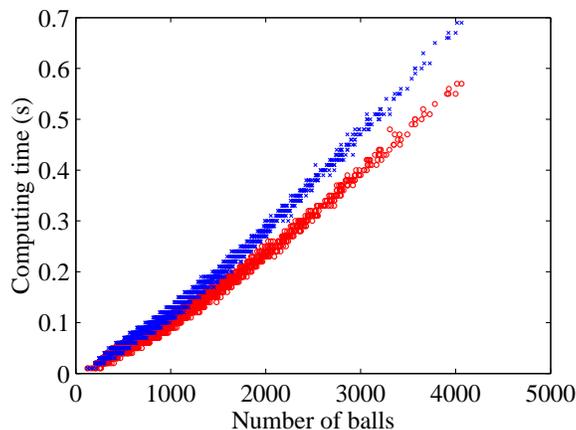


Figure 5. Performance of AlphaVol. The running time (in seconds) required by AlphaVol to compute the weighted volume and weighted surface area of a protein, with (x) and without (o) derivative is plotted against the number of atoms of the protein. The running times are measured on an Intel 1600 MHz Pentium IV computer, running Linux. AlphaVol is written in Fortran, and was compiled using ifc, the Intel Fortran compiler for Linux.

As described above, AlphaVol first computes the Delaunay triangulation of the n input spheres. Although in the worst case this takes quadratic time for constructing a quadratic number of simplices, for protein data the running time is typically $O(n \log n)$ for constructing $O(n)$ simplices. The time for constructing the dual complex and measuring the union of balls is linear in the number of simplices in the Delaunay triangulation and therefore typically in $O(n)$. The experimentally observed total running time of AlphaVol is compatible with a

complexity of $O(n \log n)$, both with and without derivatives, for up to 4,000 balls. Approximately 45% of the total running time is spent on the Delaunay triangulation, 10% on the dual complex, and 45% on the weighted area and volume. Computing the derivatives of both adds another 20%.

Applications. AlphaVol exists as a stand-alone program that can be used to compute the solvation energy of a biomolecule. We have also inserted AlphaVol into the molecular dynamics software ENCAD [Levitt et al. 1995] and GROMACS [Lindahl et al. 2001], but it is too early to say anything about the corresponding results. Recall from Section 3 that AlphaVol accounts for the nonpolar effect of water on a biomolecule, W_{np} , which is only one element of the effective solvation potential W to be used in simulations with implicit solvent. While there is a large body of work on computing the other part, W_{elec} [Simonson 2003], there is not yet any consensus on the model to be used for simulation. We have recently started a project on this specific problem.

7. Discussion

The Alpha Shape Theory with the two Derivative Theorems provides a fast, accurate and robust method for computing the interaction of water with a biomolecule in an implicit solvent model. To our knowledge, the corresponding software, AlphaVol, is the only program that deals explicitly with the problem of discontinuities of the derivatives, which are detected as singularities in the construction of the dual complex [Edelsbrunner and Koehl 2003; Bryant et al. 2004]. We conclude this paper with a short discussion of two immediate applications of this work.

Macro-molecular machinery. Recent advances in structural biology have produced an abundance of data on large macro-molecular complexes; see for example the myosin motors at <http://www.proweb.org/myosin/index.html>, the RNA polymerase transcription complexes [Cramer et al. 2001; Bushnell and Kornberg 2003], and the ribosome complexes [Wimberly et al. 2000; Yusupov et al. 2001; Ban et al. 2002]. Modeling the dynamics of such large systems is as important as modeling smaller proteins. It becomes impractical, however, to consider all atoms of the molecular machinery, and we need to introduce approximations that consider the system at coarser levels of detail. One possible approach is to represent the macro-molecular complex with a small number of spheres, supplemented with a model for their interactions that captures the physics of the underlying atomic model. These interactions will include an internal potential, and a potential to account for the solvent environment of the system. We expect the latter to resemble the solvation potential described in Section 2, in which the software AlphaVol will prove useful.

Normal modes. Collective motions in which substantial parts move as units relative to the rest play an important role in defining the function of a biomolecule. Examples include domain motions during catalytic activities (e.g. citrate synthase [Remington et al. 1982]), as well as the transition from one conformation to another for proteins that have more than one functionally distinct state. These processes involve the correlated motion of many atoms and are slower than local vibrations. They are difficult and costly to detect using classical molecular dynamics simulations, which motivates the use of normal modes dynamics as an alternative approach to detecting these collective motions [Go et al. 1983; Levitt et al. 1983; Brooks and Karplus 1983]. The normal modes are found by assuming that the potential energy can be approximated as a quadratic function of its variables and solving an eigenvalue problem to give a closed analytical description of the motion. The eigenvalues give the frequencies of the modes and the eigenvectors give the details of the corresponding motions. At a local minimum, the quadratic approximation is obtained by a Taylor expansion to the second order of the total potential energy. Computing normal modes therefore requires computing the second derivatives of the energy function. However, it is difficult to define a meaningful energy minimum for a system involving a large biomolecule in the midst of small water molecules since their geometric and physical properties are so different. We believe that this difficulty can be circumvented by using an implicit solvent model. Computing the Taylor expansion of the energy function including an implicit solvent model would then require the second derivatives of the weighted surface area and/or volume of the biomolecule. We have recently applied the mathematical tools described in this paper to derive formulas for both (manuscript in preparation).

References

- [Alm and Baker 1999] E. Alm and D. Baker, “Prediction of protein-folding mechanisms from free energy landscapes derived from native structures”, *Proc. Natl. Acad. Sci. (USA)* **96** (1999), 11305–11310.
- [Alm et al. 2002] E. Alm, A. V. Morozov, T. Kortemme, and D. Baker, “Simple physical models connect theory and experiments in protein folding kinetics”, *J. Mol. Biol.* **322** (2002), 463–476.
- [Anfinsen 1973] C. B. Anfinsen, “Principles that govern protein folding”, *Science* **181** (1973), 223–230.
- [Bajaj et al. 1997] C. Bajaj, H. Y. Lee, R. Merkert, and V. Pascucci, “NURBS based B-rep models from macromolecules and their properties”, pp. 217–228 in *Proc. 4th Sympos. Solid Modeling Appl.*, 1997.
- [Baker and Sali 2001] D. Baker and A. Sali, “Protein structure prediction and structural genomics”, *Science* **294** (2001), 93–96.
- [Ban et al. 2002] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, “The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution”, *Science* **289** (2002), 905–920.

- [Becker et al. 2001] O. M. Becker, A. D. Mckerell, B. Roux, and M. Watanabe (editors), *Computational biochemistry and biophysics*, edited by O. M. Becker et al., Marcel Dekker Inc., New York, 2001.
- [Berman et al. 2000] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, et al., “The Protein Data Bank”, *Nucl. Acids. Res.* **28** (2000), 235–242.
- [Bernal et al. 2001] A. Bernal, U. Ear, and N. Kyrpides, “Genomes OnLine Database (GOLD): a monitor of genome projects world-wide”, *Nucl. Acids. Res.* **29** (2001), 126–127.
- [Bernstein et al. 1977] F. C. Bernstein, T. F. Koetzle, G. William, D. J. Meyer, M. D. Brice, J. R. Rodgers, et al., “The protein databank: a computer-based archival file for macromolecular structures”, *J. Mol. Biol.* **112** (1977), 535–542.
- [Brooks and Karplus 1983] B. R. Brooks and M. Karplus, “Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor”, *Proc. Natl. Acad. Sci. (USA)* **80** (1983), 3696–3700.
- [Brooks et al. 1988] C. Brooks, M. Karplus, and M. Pettitt, “Proteins: a theoretical perspective of dynamics, structure and thermodynamics”, *Adv. Chem. Phys.* **71** (1988), 1–259.
- [Bryant et al. 2004] R. Bryant, H. Edelsbrunner, P. Koehl, and M. Levitt, “The area derivative of a space-filling diagram”, *Discrete Comput. Geom.* **32** (2004), 293–308.
- [Buckle and Fersht 1994] A. M. Buckle and A. R. Fersht, “Subsite binding in an RNase: structure of a barnase-tetranucleotide complex at 1.76 angstrom resolution”, *Biochemistry* **33** (1994), 1644–1653.
- [Bushnell and Kornberg 2003] D. A. Bushnell and R. D. Kornberg, “Complete, 12-subunit RNA polymerase II at 4.1-angstrom resolution: implications for the initiation of transcription”, *Proc. Natl. Acad. Sci. (USA)* **100** (2003), 6969–6973.
- [Carver 1978] M. B. Carver, “Efficient integration over discontinuities in ordinary differential equation simulators”, *Math. Comput. Simul.* **20** (1978), 190–196.
- [Cavallo et al. 2003] L. Cavallo, J. Kleijung, and F. Fraternali, “POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level”, *Nucl. Acids. Res.* **31** (2003), 3364–3366.
- [Cech 1993] T. R. Cech, “The efficiency and versatility of catalytic RNA: implications for an RNA world”, *Gene* **135** (1993), 33–36.
- [Cheatham and Kollman 2000] T. E. Cheatham and P. A. Kollman, “Molecular dynamics simulation of nucleic acids”, *Ann. Rev. Phys. Chem.* **51** (2000), 435–471.
- [Connolly 1985] M. L. Connolly, “Computation of molecular volume”, *J. Am. Chem. Soc.* **107** (1985), 1118–1124.
- [Corey and Pauling 1953] R. B. Corey and L. Pauling, “Molecular models of amino acids, peptides and proteins”, *Rev. Sci. Instr.* **24** (1953), 621–627.
- [Cossi et al. 1996] M. Cossi, B. Mennucci, and R. Cammi, “Analytical first derivatives of molecular surfaces with respect to nuclear coordinates”, *J. Comp. Chem.* **17** (1996), 57–73.
- [Cramer et al. 2001] P. Cramer, D. A. Bushnell, and R. D. Kornberg, “Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution”, *Science* **292** (2001), 1863–1876.

- [Delaunay 1934] B. Delaunay, “Sur la sphère vide”, *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk* **7** (1934), 793–800.
- [Dodd and Theodorou 1991] L. R. Dodd and D. N. Theodorou, “Analytical treatment of the volume and surface area of molecules formed by an arbitrary collection of unequal spheres intersected by planes”, *Mol. Phys.* **72** (1991), 1313–1345.
- [Duan and Kollman 1998] Y. Duan and P. A. Kollman, “Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution”, *Science* **282** (1998), 740 – 744.
- [Edelsbrunner 1992] H. Edelsbrunner, “Weighted alpha shapes”, Technical Report UIUC-CS-R-92-1760, Comput. Sci. Dept., Univ. Illinois, Urbana, Illinois, 1992.
- [Edelsbrunner 1995] H. Edelsbrunner, “The union of balls and its dual shape”, *Discrete Comput. Geom.* **13** (1995), 415–440.
- [Edelsbrunner 1999] H. Edelsbrunner, “Deformable smooth surface design”, *Discrete Comput. Geom.* **21** (1999), 87–115.
- [Edelsbrunner and Fu 1994] H. Edelsbrunner and P. Fu, “Measuring space filling diagrams and voids”, Technical Report UIUC-BI-MB-94-01, Beckman Inst., Univ. Illinois, Urbana, Illinois, 1994.
- [Edelsbrunner and Koehl 2003] H. Edelsbrunner and P. Koehl, “The weighted-volume derivative of a space-filling diagram”, *Proc. Natl. Acad. Sci. (USA)* **100** (2003), 2203–2208.
- [Edelsbrunner and Mücke 1990] H. Edelsbrunner and E. P. Mücke, “Simulation of simplicity: a technique to cope with degenerate cases in geometric algorithms”, *ACM Trans. Graphics* **9** (1990), 66–104.
- [Edelsbrunner and Mücke 1994] H. Edelsbrunner and E. P. Mücke, “Three-dimensional alpha shapes”, *ACM Trans. Graphics* **13** (1994), 43–72.
- [Edelsbrunner and Shah 1996] H. Edelsbrunner and N. R. Shah, “Incremental topological flipping works for regular triangulations”, *Algorithmica* **15** (1996), 223–241.
- [Edelsbrunner et al. 1983] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel, “On the shape of a set of points in the plane”, *IEEE Trans. Inform. Theory* **IT-29** (1983), 551–559.
- [Edelsbrunner et al. 1998] H. Edelsbrunner, M. A. Facello, and J. Liang, “On the definition and construction of pockets in macromolecules”, *Discrete Appl. Math.* **88** (1998), 83–102.
- [Eisenberg and McLachlan 1986] D. Eisenberg and A. D. McLachlan, “Solvation energy in protein folding and binding”, *Nature (London)* **319** (1986), 199–203.
- [Fortune and VanWyk 1996] S. Fortune and C. J. VanWyk, “Static analysis yields efficient exact integer arithmetic for computational geometry”, *ACM Trans. Graph.* **15** (1996), 223–248.
- [Futamura et al. 2004] N. Futamura, S. Alura, D. Ranjan, and B. Hariharan, “Efficient parallel algorithms for solvent accessible surface area of proteins”, *IEEE Trans. Parallel Dist. Syst.* **13** (2004), 544–555.
- [Gavezzotti 1983] A. Gavezzotti, “The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity”, *J. Am. Chem. Soc.* **105** (1983), 5220–5225.

- [Gear and Østerby 1984] C. W. Gear and O. Østerby, "Solving ordinary differential equations with discontinuities", *ACM Trans. Math. Softw.* **10** (1984), 23–24.
- [Gesteland and Atkins 1993] R. F. Gesteland and J. A. Atkins, *The RNA world: the nature of modern RNA suggests a prebiotic RNA world*, Cold Spring Harbor Laboratory Press, Plainview, NY, 1993.
- [Gibson and Scheraga 1967] K. D. Gibson and H. A. Scheraga, "Minimization of polypeptide energy, I: Preliminary structures of bovine pancreatic ribonuclease S-peptide", *Proc. Natl. Acad. Sci. (USA)* **58** (1967), 420–427.
- [Gibson and Scheraga 1987] K. D. Gibson and H. A. Scheraga, "Exact calculation of the volume and surface-area of fused hard-sphere molecules with unequal atomic radii", *Mol. Phys.* **62** (1987), 1247–1265.
- [Gilbert 1986] W. Gilbert, "The RNA world", *Nature* **319** (1986), 618.
- [Go et al. 1983] N. Go, T. Noguti, and T. Nishikawa, "Dynamics of a small globular protein in terms of low-frequency vibrational modes", *Proc. Natl. Acad. Sci. (USA)* **80** (1983), 3696–3700.
- [Gogonea and Osawa 1994] V. Gogonea and E. Osawa, "Implementation of solvent effect in molecular mechanics, 3: The first and second-order analytical derivatives of excluded volume.", *J. Mol. Struct. (Theochem)* **311** (1994), 305–324.
- [Gogonea and Osawa 1995] V. Gogonea and E. Osawa, "An improved algorithm for the analytical computation of solvent-excluded volume: the treatment of singularities in solvent-accessible surface-area and volume functions", *J. Comp. Chem.* **16** (1995), 817–842.
- [Grant and Pickup 1995] J. A. Grant and B. T. Pickup, "A Gaussian description of molecular shape", *J. Phys. Chem.* **99** (1995), 3503–3510.
- [Hao and Scheraga 1994a] M. H. Hao and H. A. Scheraga, "Monte Carlo Simulation of a first order transition for protein folding", *J. Phys. Chem.* **98** (1994), 4940–4948.
- [Hao and Scheraga 1994b] M. H. Hao and H. A. Scheraga, "Statistical thermodynamics of protein folding – sequence dependence", *J. Phys. Chem.* **98** (1994), 9882–9893.
- [Hasel et al. 1988] W. Hasel, T. F. Hendrikson, and W. C. Still, "A rapid approximation to the solvent accessible surface areas of atoms", *Tetrahed. Comp. Method.* **1** (1988), 103–106.
- [Irisa 1996] M. Irisa, "An elegant algorithm of the analytical calculation for the volume of fused spheres with different radii", *Comp. Phys. Comm.* **98** (1996), 317–338.
- [Kaminski et al. 2001] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides", *J. Phys. Chem. B.* **105** (2001), 6474–6487.
- [Karplus and McCammon 2002] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules", *Nature Struct. Biol.* **9** (2002), 646–652.
- [Kauzmann 1959] W. Kauzmann, "Some factors in the interpretation of protein denaturation", *Adv. Protein Chem.* **14** (1959), 1–63.
- [Kendrew et al. 1960] J. Kendrew, R. Dickerson, B. Strandberg, R. Hart, D. Davies, and D. Phillips, "Structure of myoglobin: a three dimensional Fourier synthesis at 2 angstrom resolution", *Nature* **185** (1960), 422–427.

- [Koehl and Levitt 1999] P. Koehl and M. Levitt, “A brighter future for protein structure prediction”, *Nature Struct. Biol.* **6** (1999), 108–111.
- [Koehl and Levitt 2002] P. Koehl and M. Levitt, “Protein topology and stability defines the space of allowed sequences”, *Proc. Natl. Acad. Sci. (USA)* **99** (2002), 1280–1285.
- [Koltun 1965] W. L. Koltun, “Precision space-filling atomic models”, *Biopolymers* **3** (1965), 665–679.
- [Kratky 1978] K. W. Kratky, “Area of intersection of n equal circular disks”, *J. Phys. A.: Math. Gen.* **11** (1978), 1017–1024.
- [Kraulis 1991] P. J. Kraulis, “MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures”, *J. Appl. Crystallo.* **24** (1991), 946–950.
- [Kundrot et al. 1991] C. E. Kundrot, J. W. Ponder, and F. M. Richards, “Algorithms for calculating excluded volume and its derivatives as a function of molecular-conformation and their use in energy minimization”, *J. Comp. Chem.* **12** (1991), 402–409.
- [Leach 2001] A. R. Leach, *Molecular modelling: principles and applications*, 2nd ed., Prentice Hall, 2001.
- [Lee and Richards 1971] B. Lee and F. M. Richards, “Interpretation of protein structures: estimation of static accessibility”, *J. Mol. Biol.* **55** (1971), 379–400.
- [Legrand and Merz 1993] S. M. Legrand and K. M. Merz, “Rapid approximation to molecular-surface area via the use of boolean logic and look-up tables”, *J. Comp. Chem.* **14** (1993), 349–352.
- [Levitt and Sharon 1988] M. Levitt and R. Sharon, “Accurate simulation of protein dynamics in solution”, *Proc. Natl. Acad. Sci. (USA)* **85** (1988), 7557–7561.
- [Levitt et al. 1983] M. Levitt, C. Sander, and P. S. Stern, “Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme”, *J. Mol. Biol.* **181** (1983), 423–447.
- [Levitt et al. 1995] M. Levitt, M. Hirshberg, R. Sharon, and V. Daggett, “Potential-energy function and parameters for simulations of the molecular-dynamics of proteins and nucleic-acids in solution”, *Comp. Phys. Comm.* **91** (1995), 215–231.
- [Levy et al. 2004] Y. Levy, P. G. Wolynes, and J. N. Onuchic, “Protein topology determines binding mechanism”, *Proc. Natl. Acad. Sci. (USA)* **101** (2004), 511–516.
- [Liang et al. 1998a] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramaniam, “Analytical shape computation of macromolecules, I: Molecular area and volume through alpha shape”, *Proteins: Struct. Func. Genet.* **33** (1998), 1–17.
- [Liang et al. 1998b] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramaniam, “Analytical shape computation of macromolecules, II: Inaccessible cavities in proteins”, *Proteins: Struct. Func. Genet.* **33** (1998), 18–29.
- [Liang et al. 1998c] J. Liang, H. Edelsbrunner, and C. Woodward, “Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design”, *Prot. Sci.* **7** (1998), 1884–1897.
- [Lindahl et al. 2001] E. Lindahl, B. Hess, and D. van der Spoel, “GROMACS 3.0: a package for molecular simulation and trajectory analysis”, *J. Molec. Mod.* **7** (2001), 306–317.

- [Liwo et al. 1997a] A. Liwo, S. Oldziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, “A united-residue force field for off-lattice protein-structure simulations, 1: Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data”, *J. Comp. Chem.* **18** (1997), 849–873.
- [Liwo et al. 1997b] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, S. Oldziej, and H. A. Scheraga, “A united-residue force field for off-lattice protein-structure simulations, 2: Parameterization of short-range interactions and determination of weights of energy terms by Z-score optimization”, *J. Comp. Chem.* **18** (1997), 874–887.
- [Lum et al. 1999] K. Lum, D. Chandler, and J. D. Weeks, “Hydrophobicity at small and large length scales”, *J. Phys. Chem. B.* **103** (1999), 4570–4577.
- [MacKerell et al. 1998] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, et al., “All-atom empirical potential for molecular modeling and dynamics studies of proteins”, *J. Phys. Chem. B.* **102** (1998), 3586–3616.
- [McCammon et al. 1977] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins”, *Nature* **267** (1977), 585–590.
- [Monod 1973] J. Monod, *Le hasard et la nécessité*, Le Seuil, Paris, France, 1973.
- [Mücke et al. 1999] E. P. Mücke, I. Saias, and B. Zhu, “Fast randomized point location without preprocessing in two- and three-dimensional Delaunay triangulations”, *Comput. Geom.: Theory Appl.* **12** (1999), 63–83.
- [Muñoz and Eaton 1999] V. Muñoz and W. A. Eaton, “A simple model for calculating the kinetics of protein folding from three-dimensional structures”, *Proc. Natl. Acad. Sci. (USA)* **96** (1999), 11311–11316.
- [Ooi et al. 1987] T. Ooi, M. Oobatake, G. Nemethy, and H. A. Scheraga, “Accessible surface-areas as a measure of the thermodynamic parameters of hydration of peptides”, *Proc. Natl. Acad. Sci. (USA)* **84** (1987), 3086–3090.
- [Pavani and Ranghino 1982] R. Pavani and G. Ranghino, “A method to compute the volume of a molecule”, *Computers and Chemistry* **6** (1982), 133–135.
- [Perrot et al. 1992] G. Perrot, B. Cheng, K. D. Gibson, J. Vila, K. A. Palmer, A. Nayeem, et al., “MSEED: a program for the rapid analytical determination of accessible surface-areas and their derivatives”, *J. Comp. Chem.* **13** (1992), 1–11.
- [Perutz 1990] M. F. Perutz, “Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide”, *Annu. Rev. Physiol.* **52** (1990), 1–25.
- [Perutz et al. 1960] M. Perutz, M. Rossmann, A. Cullis, G. Muirhead, G. Will, and A. North, “Structure of hemoglobin: a three-dimensional Fourier synthesis at 5.5 angstrom resolution, obtained by X-ray analysis”, *Nature* **185** (1960), 416–422.
- [Petitjean 1994] M. Petitjean, “On the analytical calculation of van-der-Waals surfaces and volumes: some numerical aspects”, *J. Comp. Chem.* **15** (1994), 507–523.
- [Plaxco et al. 1998] K. W. Plaxco, K. T. Simons, and D. Baker, “Contact order, transition state placement and the refolding rates of single domain proteins”, *J. Mol. Biol.* **277** (1998), 985–994.
- [Price and Brooks 2002] D. J. Price and C. L. Brooks, “Modern protein force fields behave comparably in molecular dynamics simulations”, *J. Comp. Chem.* **23** (2002), 1045–1057.

- [Remington et al. 1982] S. Remington, G. Weigand, and R. Huber, "Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 angstrom resolution", *J. Mol. Biol.* **158** (1982), 111–152.
- [Richmond 1984] T. J. Richmond, "Solvent accessible surface-area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect", *J. Mol. Biol.* **178** (1984), 63–89.
- [Rowlinson 1963] J. S. Rowlinson, "The triplet distribution function in a fluid of hard spheres", *Mol. Phys.* **6** (1963), 517–524.
- [Shrake and Rupley 1973] A. Shrake and J. A. Rupley, "Environment and exposure to solvent of protein atoms: lysozyme and insulin", *J. Mol. Biol.* **79** (1973), 351–371.
- [Simonson 2003] T. Simonson, "Electrostatics and dynamics of proteins", *Rep. Prog. Phys.* **66** (2003), 737–787.
- [Simonson and Brünger 1994] T. Simonson and A. T. Brünger, "Solvation free-energies estimated from macroscopic continuum theory: an accuracy assessment", *J. Phys. Chem.* **98** (1994), 4683–4694.
- [Sridharan et al. 1994] S. Sridharan, A. Nicholls, and K. A. Sharp, "A rapid method for calculating derivatives of solvent accessible surface areas of molecules", *J. Comp. Chem.* **16** (1994), 1038–1044.
- [Street and Mayo 1998] A. G. Street and S. L. Mayo, "Pairwise calculation of protein solvent-accessible surface areas", *Folding & Design* **3** (1998), 253–258.
- [Timberlake 1992] K. C. Timberlake, *Chemistry*, 5th ed., Harper Collins, New York, 1992.
- [Tunon et al. 1992] I. Tunon, E. Silla, and J. L. Pascual-Ahuir, "Molecular-surface area and hydrophobic effect", *Protein Eng.* **5** (1992), 715–716.
- [Wang and Levinthal 1991] H. Wang and C. Levinthal, "A vectorized algorithm for calculating the accessible surface area of macromolecules", *J. Comp. Chem.* **12** (1991), 868–871.
- [Watson and Crick 1953] J. D. Watson and F. H. C. Crick, "A Structure for Deoxyribose Nucleic Acid", *Nature* **171** (1953), 737–738.
- [Wawak et al. 1994] R. J. Wawak, K. D. Gibson, and H. A. Scheraga, "Gradient discontinuities in calculations involving molecular-surface area", *J. Math. Chem.* **15** (1994), 207–232.
- [Weiser et al. 1999a] J. Weiser, P. S. Shenkin, and W. C. Still, "Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)", *J. Comp. Chem.* **20** (1999), 217–230.
- [Weiser et al. 1999b] J. Weiser, P. S. Shenkin, and W. C. Still, "Optimization of Gaussian surface calculations and extension to solvent accessible surface areas", *J. Comp. Chem.* **20** (1999), 688–703.
- [Wimberly et al. 2000] B. T. Wimberly, D. E. Brodersen, W. M. Clemons Jr., R. J. Morgan-Warren, A. P. Carter, C. Vornrhein, et al., "Structure of the 30S ribosomal subunit", *Nature* **407** (2000), 327–339.
- [Wodak and Janin 1980] S. J. Wodak and J. Janin, "Analytical approximation to the accessible surface-area of proteins", *Proc. Natl. Acad. Sci. (USA)* **77** (1980), 1736–1740.

- [Wood and Thompson 1990] R. H. Wood and P. T. Thompson, “Differences between pair and bulk hydrophobic interactions”, *Proc. Natl. Acad. Sci. (USA)* **87** (1990), 946–949.
- [Yusupov et al. 2001] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. D. Cate, and H. F. Noller, “Crystal structure of the ribosome at 5.5 angstrom resolution”, *Science* **292** (2001), 883–896.

HERBERT EDELSBRUNNER
DEPARTMENT OF COMPUTER SCIENCE
DUKE UNIVERSITY
DURHAM, NC 27708
UNITED STATES
edels@cs.duke.edu

PATRICE KOEHL
DEPARTMENT OF COMPUTER SCIENCE AND GENOME CENTER
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616
UNITED STATES
koehl@cs.ucdavis.edu

