PROTEINS: Structure, Function, and Genetics 33:18–29 (1998)

Analytical Shape Computation of Macromolecules: II. Inaccessible Cavities in Proteins

Jie Liang,[1,2] Herbert Edelsbrunner,[2] Ping Fu,[1] Pamidighantam V. Sudhakar,[1] and Shankar Subramaniam[1,3]*

[1]*National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois*
[2]*Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois*
[3]*Beckman Institute for Advanced Science and Technology, Departments of Biochemistry, Molecular & Integrative Physiology, and Chemical Engineering, Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois*

ABSTRACT **The structures of proteins are well-packed, yet they contain numerous cavities which play key roles in accommodating small molecules, or enabling conformational changes. From high-resolution structures it is possible to identify these cavities. We have developed a precise algorithm based on alpha shapes for measuring space-filling-based molecular models (such as van der Waals, solvent accessible, and molecular surface descriptions). We applied this method for accurate computation of the surface area and volume of cavities in several proteins. In addition, all of the atoms/residues lining the cavities are identified. We use this method to study the structure and the stability of proteins, as well as to locate cavities that could contain structural water molecules in the proton transport pathway in the membrane protein bacteriorhodopsin. Proteins 33:18–29, 1998.** © 1998 Wiley-Liss, Inc.

Key words: molecular cavities; packing defects; Delaunay complex; alpha shape; structural solvent in proteins

INTRODUCTION

Native protein structures are well-packed and a large degree of the stabilization comes from packing interactions. Nevertheless, proteins have cavities, which either accommodate prosthetic groups or functional water molecules, or serve the function of allowing conformational flexibility. Extensive mutagenesis studies on proteins show that the alteration of cavity shape and size influences both the thermodynamic stability[1,2,3] and in some cases the function of the protein.[4] It is therefore important to calculate accurately the shapes and sizes of cavities in proteins.

Lee and Richards[5] used a modified version of their solvent accessibility area and volume computation method to calculate cavities in myoglobin and in other proteins. Using a modified version of the Shrake and Rupley algorithm, Rashin et al.[6] calculated cavities in a set of proteins. They analyzed their results to probe the immediate vicinity of occupied and empty cavities. They concluded that cavities lined by polar residues accommodate one or more water molecules and these water molecules form up to three hydrogen bonds with the amino acids lining the cavity. They also estimated that the energetic cost of creating a cavity is about 1kT/10 Å. More recently, Hubbard et al.[7,8] did an exhaustive analysis of cavities in 121 high-resolution protein structures. They find that cavities that do not contain ordered water molecules are surrounded by hydrophobic sidechains of residues in well-defined secondary structures, while cavities containing one or more water molecules are surrounded by coil regions. Experiments using water-sensitive two-dimensional heteronuclear NMR techniques show that cavities in proteins could contain disordered water molecules not apparent in the X-ray structural data.[9] Recently, a method for quantitative measurement of the hydrophilicity of cavities was developed by Zhang and Hermans,[10] along with the free energy analysis of introducing buried waters in protein cavities. Using this method, hydrated cavities were shown to be distinguishable from empty cavities by a threshold value of the water–protein interaction energy of −12 kcal/mol.

Cavities also play an important role in protein–protein interfaces. Interdomain contacts in proteins are often mediated by structurally well-defined interfacial water molecules isolated from bulk solvent. In addition, allosteric transformations in proteins are

Grant sponsor: NSF; Grant numbers: ASC 94-04900, ASC 92-00301, CCR 91-18874, ASC 89-02829, MCB 92-19619, and DBI 94-04223; Grant sponsor: ONR; Grant number: N00014-95-1-0692.

Jie Liang's current address is Department of Cheminformatics, SmithKline Beecham Pharmaceuticals, PO Box 1539, UW2940, 709 Swedeland Road, King of Prussia, PA 19406.

*Correspondence to: Shankar Subramaniam, Beckman Institute for Advanced Science and Technology, Departments of Biochemistry, Molecular & Integrative Physiology, and Chemical Engineering, Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801. E-mail: shankar@ncsa.uiuc.edu

Received 3 September 1996; Accepted 7 April 1998

© 1998 WILEY-LISS, INC.

associated with changes in domain contacts and concomitantly with sizes and shapes of interfacial cavities. Furthermore, water-filled cavities play the role of modulating pK_a values of acidic and basic residues surrounding the cavities. In the absence of high-resolution structural information capable of resolving all the water molecules in protein cavities, it will be extremely useful to develop accurate and fast computational methods for quantitatively calculating the shapes and sizes of these cavities.

In this article we present an algorithm for accurately computing the location, shape, and size of internal cavities in proteins. The algorithm is based on the notions of alpha shapes and weighted Voronoi dissections of space-filling-based molecular models, such as the van der Waals, solvent accessible, and molecular surface models. The rest of the article consists of two parts. The first part presents the alpha shape method and algorithmic details; the second part discusses applications of this method to a set of proteins and compares the results of this study with the results from experiments. The method is also applied to study the role cavities play in the stability of the protein ribonuclease S, the conformational flexibility in the oxygen transport pathway in myoglobin, and the proton pathway in bacteriorhodopsin. Our goal is to highlight the specificity of these cavities and correlate their presence with experimental data.

MODELING, COMPUTING, AND MEASURING INACCESSIBLE CAVITIES

Geometric Models

Space-filling diagrams are widely used geometric models for macromolecules such as proteins and nucleic acids. They model a molecule as the union of many fused spherical balls in three-dimensional space. Each ball represents an atom by adopting the spatial location and an appropriate van der Waals radius of the atom. The van der Waals or VW model is the union of these spherical balls. Two related but different models are often used to describe the interaction of the molecule with a solvent. The solvent is treated as a sphere of appropriate radius and is rolled around the van der Waals surface of the protein. The center of the solvent sphere sweeps out the surface of the solvent accessible or SA model, while the front of the solvent sphere defines the molecular surface or MS model (see References 5,11,12).

Recently, a combinatorial model of a molecule in terms of an associated geometric construct called dual complex has been proposed. It is based on the weighted Voronoi diagram, the Delaunay complex, and the theory of alpha shapes that have been developed in the computational geometry community.[13] For the VW model of the molecule, the Voronoi diagram of the atom balls divides the space into Voronoi regions, one per atom. A Voronoi region is generated by an atom, and consists of the part of space closest to this atom.[14] Adjacent Voronoi regions are separated by the radical plane of their balls.

The Delaunay complex is a geometric construct that can be derived from the Voronoi diagram by the following direct translation. The center of an atom ball with a Voronoi region becomes a vertex in the Delaunay complex. If two Voronoi regions share a common facet, then the edge connecting the centers of the two corresponding atom balls is in the complex. If three Voronoi regions share a common edge, then the triangle spanned by the three ball centers is in the complex. Finally, if four Voronoi regions share a common point, then the tetrahedron spanned by the four ball centers is in the complex. We have thus accounted for all possible intersection patterns among Voronoi regions, since in three-dimensional space there can be no more than four Voronoi regions that meet. The vertexes, edges, triangles, and tetrahedra are the four basic types of elements, called simplices.

The Voronoi regions decompose the VW molecule into convex cells that are either disjoint or overlap along common boundary pieces. Each convex cell contains one atom. The dual complex (or the 0-alpha complex) consists of the four different types of simplices described above. The vertices are the centers of the atom balls. We add an edge connecting two atom centers to the dual complex if their two convex cells overlap along a common face. Similarly, we add a triangle spanned by three centers if their convex cells share a common edge. Finally, we add a tetrahedron spanned by four centers if their convex cells share a common point. (See Reference 15 for a complete and rigorous treatment of dual complexes and to the companion paper[16] for an intuitive description concerning biology applications.)

Most importantly, the dual complex faithfully represents geometric and topological properties of the molecule as represented by the VW model. Since the SA model is also a union of balls, we can repeat the same construction after incrementing the radii of all atoms by the radius of the solvent ball, and obtain a dual complex representing the properties of this enlarged model.

Representing Inaccessible Cavities

We consider the SA surface area model of a molecule. The complement space is the part of three-dimensional space not covered by any atom ball in that model. It may consist of several components, and each component except the infinitely large one, namely the "outside," is an inaccessible cavity, also referred to as a void. Since a void is disconnected from the outside component, solvent trapped within cannot escape to the outside or infinity if there is no dynamic structural fluctuation of the molecule. A proven mathematical fact is that for every void of the SA model of the molecule, there is a unique corresponding void in the dual complex, and the SA

molecular void is contained within the complex void.[15] Therefore, we first identify and compute the voids in the dual complex, and then (conceptually) map them into voids in the molecules.

A void in the dual complex is triangulated by a set of Delaunay tetrahedra and can therefore be represented by them. These tetrahedra do not belong to the dual complex. The surface of the void can be represented by a subset of triangles from the triangular faces of these tetrahedra. Multiple sets of selected Delaunay tetrahedra can be used to represent the whole collection of voids in a molecule, with each set of tetrahedra corresponding to one void. A union-find data structure is used to implement the set system. Our algorithm identifies all Delaunay tetrahedra that do not belong to the dual complex, and collects them in the union-find system. Every triangular interface between two such tetrahedra will then trigger a merge operation so the two tetrahedra are labeled as members of the same void in the dual complex. More details of the union-find system and the algorithm are described in the Appendix.

Computing Volume and Area of Molecular Void

The volume and area of the molecular void can be computed from the corresponding void in the dual complex. The same short inclusion-exclusion formula described in the companion article is applied. Essentially, we first compute the volume of the voids in the dual complex, which is the sum of the individual tetrahedron volume. We then subtract the volume of a sector of the atom ball (which is the part of the atom contained within a corner of the tetrahedron), since this atom reaches into the void tetrahedron. Because the reached-in portions of atom balls have two overlaps, we add back corresponding wedges of the intersections of two balls (lenses) contained within the tetrahedron. For three overlaps, we subtract halves of intersections of three balls. We use a similar strategy to compute the area of the molecular void.

The volume and area computations are performed by the software module VOLBL, which is part of the standard distribution of the alpha shapes software. The entire computation is analytical, and numerical errors are introduced only through imprecise floating-point arithmetic and mathematical function evaluation, such as the square-root and the inverse of the cosine. A recent version of VOLBL extends the volume and area computation to the MS model, albeit in some circumstances the model is ambiguous and it is not entirely clear what its volume and area should mean.

We conclude this section by mentioning that our software is robust and complete, i.e., no inaccessible cavities are missed no matter how small. This is in contrast to other available software, which typically have difficulties with high degree of overlap, with surface arcs that form complete circles, etc.[17] Three-dimensional pictures of cavities bounded by spherical patches can be generated by software based on the structure of alpha complexes.[18]

RESULTS AND DISCUSSION

We first compare the number and size of cavities for a set of proteins investigated previously.[6] We then compute cavities in three other proteins where the presence of solvent and possible roles of cavities in protein function are exemplified: myoglobin, where xenon binds in its hydrophobic cavities; ribonuclease (native and modified proteins), where isolated cavities and protein packing changes when proteins are modified; and bacteriorhodopsin, where sites for structural water molecules may play important roles in proton pumping. All structures are taken from the Brookhaven Protein Data Bank,[19] and all solvent water molecules are explicitly removed. Heteroatoms are included: the heme group in myoglobin and retinal in bacteriorhodopsin. A probe of radius 1.2 Å for water is used in our calculations, except for bacteriorhodopsin, where a radius of 1.4 Å is used.

Inaccessible Cavities and Comparison With Previous Results

A list of proteins (the set computed by Rashin et al.[6]), together with the number of cavities, total cavity SA area/volume, and total cavity MS area/volume are given in (Table I). Rashin et al. used an approximate algorithm to compute cavities. The overall results in computed SA area of the two methods are comparable. However, the MS volumes differ significantly. We note that a Monte Carlo-based point-sampling method was used by Rashin et al. to estimate the MS volume. Pronounced differences are observed for 1eca, 2ptn and 8tln (3tln used by Rashin) molecules. In general, VOLBL finds more cavities for proteins (except 2cha and 5mbn), indicating the accuracy and sensitivity of the method. VOLBL-computed cavities show large differences from the results of Hubbard et al.,[8] where a probe radius of 1.4 Å was used.

Myoglobin: Ligand-Binding Cavities

Myoglobin is an oxygen transport protein found in skeletal muscles. Diffusion and binding of oxygen to the heme iron have been the subject of numerous investigations.[20,4,21] The X-ray structure of myoglobin shows no direct pathway for the transport of oxygen from the exterior to the heme. It has been postulated that protein conformational changes occur concomitantly with oxygen diffusion.[22] These conformational changes are related to atomic interactions and structural changes such as residue sidechain movement and cavity formation. Cavities in myoglobin have been explored experimentally through binding of the noble gas atom xenon. X-ray structural analysis shows that xenon binding cavi-

TABLE I. Cavities Computed for the Listed Proteins Using Alpha Shape Method (VOLBL); Comparison With Results From Rashin et al.[6†]

	# of cavities		Area			Volume		
Protein	VOLBL	Rashin	SA	MS	Rashin (SA)	SA	MS	Rashin (MS)
1eca	10	9	46.4	435.8	69	6.4	252.5	401
1nxb	3	0	0.4	62.5	0	0.0	26.7	0
2act	20	21	140.4	958.2	130	35.0	611.2	449
2cha	23	26	132.2	1079.7	120	20.4	647.6	571
2lyz	12	8	58.7	498.7	53	9.0	297.2	190
2ptn	19	13	175.4	1123.0	168	31.5	702.0	494
2sn3	2	2	6.0	80.2	6	0.4	44.4	32
3cyt	8	5	3.1	211.0	2	0.1	97.1	34
3rn3	4	5	1.3	94.9	3	0.1	42.7	41
4pti	2	2	23.4	132.7	20	3.7	87.3	70
5mbn	17	23	95.8	877.0	85	14.1	503.4	391
8tln	42	30	163.4	1777.8	117	22.2	987.0	528

†Areas are in Å^2 and volumes in Å^3. SA represents the solvent accessible model and MS represents the molecular surface model. Rashin et al. used structures from earlier PDB entries 1sn3, 1rn3, 2mbn, and 3tln.

ties are surrounded by relatively nonpolar residues.[23] Four known xenon binding sites are seen in the X-ray structure with varying degrees of occupancy. One site is found in a cavity near the proximal histidine and a pyrrole group of the heme.[24] A second site is located near the corner of G-H and A-B helices close to the external surface.[25] These two xenon binding sites are also observed in ^{129}Xe NMR experiments on met-myoglobin in solution.[26] A third xenon binding site is located at the corner of the E-F helix and H helix and is close to the surface. A fourth site is on the distal side of the heme and is close to the oxygen binding site.

Computation of cavities in myoglobin would provide insights into the locations of hydrophobic sites where ligands (such as xenon) can bind, as well as help to identify cavities that can aid in diffusion of the ligands. The cavities in myoglobin computed from alpha-shape program VOLBL are shown in the Figure 1. We obtain a total of 17 inaccessible cavities in myoglobin. Rashin et al., using the same size probe, obtained 23 cavities using a modified Shrake and Rupley procedure.

In our result, cavities 1, 2, 4, and 7 correspond to the xenon binding sites. The volume contained in each cavity is listed, along with the identity of surrounding residues, in Table II (information at the level of atomic detail is computed by our software, but these are omitted for the sake of brevity). The xenon binding site I corresponds to cavity 7, which has the highest occupancy and the lowest B factor. Site II corresponds to cavity 4, site III to cavity 1, and site IV to cavity 2. Of the four xenon binding cavities sites, site I has the smallest volume and surface area. Site II has the largest B factor (49.6 Å^2) and sites III and IV have large volumes and areas. The sizes of the computed cavities correlate well with results of the NMR experiments. Tightly bound xenon atoms with slow off-rates (1×10^{-5} M/s) and sensitivity to the paramagnetic iron atom. These are the xenon atoms in sites I and II, which have small volume and are close to the heme prosthetic group. Xenon atoms with larger off-rates are found in the larger cavities III and IV. Residue F138 is close to site III but has no direct contact with site I. Free energy perturbation molecular dynamics simulations[27] suggest that, apart from the immediate residues surrounding xenon site I, residue F138 also experiences a large movement. This suggests that site II located between sites I and III may mediate the changes in F138.

The oxygen binding region is also apparent in our cavity computations. Cavity 5 is flanked by heme on one side, and the distal histidine H64 and hydrophobic residues L29, L32, F43, and I107 on another side. Cavity 12 is flanked by residues L29, F43, and H64 (also shared by cavity 5). These cavities form a possible pathway through which oxygen diffuses to the iron-binding site. Cavity 5 is also close to the xenon binding site IV. This suggests that the structure around the oxygen binding pocket is flexible. There are other cavities (cavities 9 and cavity 14) in addition to cavity 12 along the path toward the CD loop. These cavities could also play a role in providing a pathway for oxygen diffusion into myoglobin from the exterior.

Ribonuclease: Cavities and Packing Interactions in Proteins

Packing interactions in proteins play an important role for protein folding and protein stability. The role of packing was examined in ribonuclease S formed by the S-protein–S-peptide complex.[28] The S-peptide is obtained by subtilisin cleavage of the ribonuclease S. The S-peptide and S-protein associate to form native-like ribonuclease-S structure. Methionine 13 on the S-peptide packs into a hydrophobic cluster in the ribonuclease S native structure. By replacing

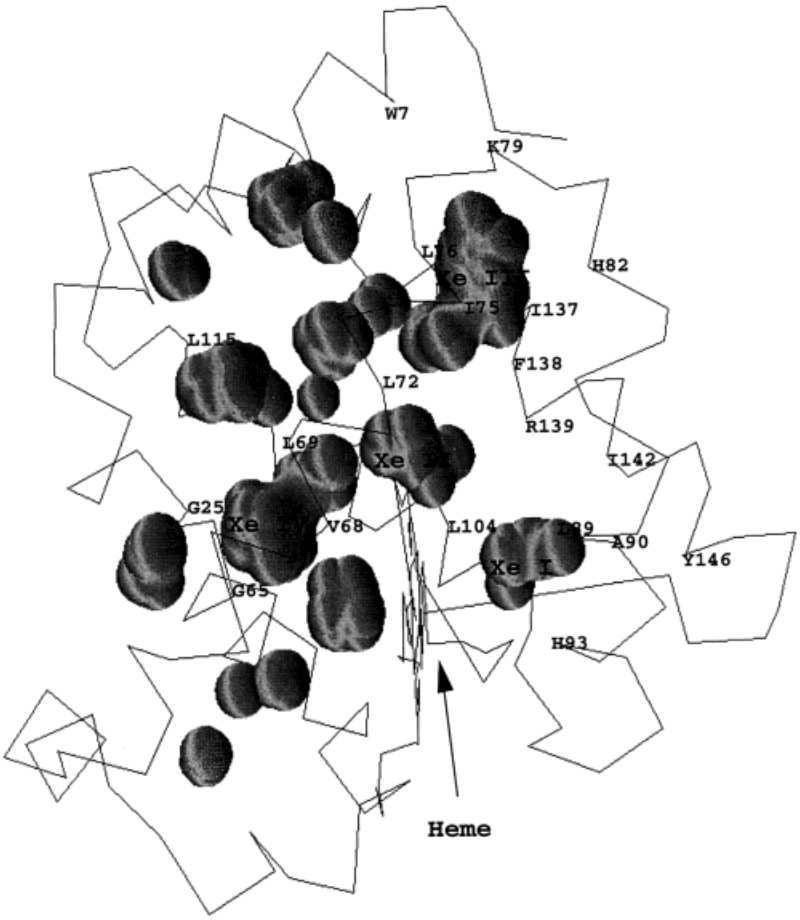

Fig. 1. Inaccessible cavities in myoglobin (5mbn). Xenon binding sites are labeled.

methionine in position 13 with other amino acids, Richards and coworkers[28] investigated the role of hydrophobic interactions in maintaining the stability of the protein. The crystallographic and thermodynamic analysis of the modified proteins does not show good correlation with overall biophysical parameters such as polarizability, volume, and charge. A key problem in examining the packing interactions is the presence of small cavities arising in the modified proteins, some of which accommodate one or more water molecules. Varadarajan and Richards[28] estimated the volume changes of the cavities by subtracting mean Voronoi volumes from residue Voronoi volume at position 13 for all of the mutants.

We computed the size and shape of cavities in the native and modified S-protein–S-peptide complexes in order to examine the changes in packing interactions arising from these small cavities. The number and sizes of cavities computed for these proteins are listed in Tables III–IX. In the wild-type Rnase (2rns), where M is at position 13, four cavities are found. All of them are relatively small (none exceeds 23 $Å^2$ in MS area and 10.1 $Å^3$ in MS volume).

Structures of the modified proteins show significant differences in the number of cavities and their sizes from those in the native protein. The modified protein M13L (1rbh) has five cavities, one of which (cavity 1) is in the vicinity of residue 13 (Table IV). This cavity also exists in the native protein. However, its size increases for 10.1 $Å^3$ (MS volume) in the native to 28.4 $Å^3$ in the modified structure. This suggests that the branched sidechain of leucine introduces significant structural changes. On the other hand, $\Delta\Delta G$ for the M13L mutant is close to zero. It has been suggested that changes in enthalpic contributions are compensated by changes in entropic contributions.[28]

In the modified protein M13A (1rbc) (see Table V), cavity 1 surrounding residue 13 is present, as in the native protein. It is enlarged due to the replacement by the small sidechain in alanine. This modified protein also displays a change in cavity 5 (cavity 3 in the native protein), which is surrounded now by residues D121 and A109, in addition to residue H119 present in the native protein (the native protein has residues A4, V118, and H119 lining this cavity). Several additional cavities are also formed. Varadarajan and Richards[28] observe the presence of a water molecule in the M13A protein structure.

TABLE II. Inaccessible Cavities Computed Using VOLBL for Myoglobin (PDB Entry 5mbn)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	105.2	31.5	6.4	144.6	W7 L72 I75 L76 K79 G80 H82 A134 L135 F138 L137
2	80.1	20.1	3.2	132.5	G25 I28 L29 G65 V68 L69 L72 I107 I111 Hem130
3	41.9	9.4	1.0	67.7	W14 V17 H24 I28 L69 I111 L115
4	48.1	9.4	0.9	82.9	L104 I107 S108 I111 L135 F138 Hem130 R139
5	42.8	10.1	1.4	65.0	L29 L32 F43 H64 V68 I107 Hem130
6	33.0	5.5	0.5	65.4	L9 V10 V13 F123 A127 A130 M131
7	25.6	2.4	0.1	49.9	L89 A90 H93 L104 I142 Y146
8	25.9	4.2	0.4	44.3	W14 L72 L76 I111 M131 L135
9	21.3	1.9	0.1	42.3	A22 G25 Q26 K62 G65
10	11.4	0.3	0.0	24.9	F33 L40 L49 M55
11	14.0	0.4	0.0	29.9	L76 M131 A134 L135
12	10.8	0.2	0.0	26.5	L29 F33 F43 F46 H64
13	11.9	0.4	0.0	25.7	V13 V17 L115 H119
14	7.3	0.0	0.0	18.2	L29 F33 F46 L61
15	8.8	0.0	0.0	20.6	L11 W14 L76 K77
16	7.4	0.0	0.0	18.4	H93 I99 L104 Y146
17	8.1	0.0	0.0	20.0	E109 I112

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS themolecular surface model. Residues contributing to the cavities are also listed.

TABLE III. Inaccessible Cavities Computed Using VOLBL for Ribonuclease (PDB Entry 2rns)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	0.1	22.8	0.0	10.1	8F 47V 54V 106I 120F
2	0.0	20.4	0.0	8.6	14D 29M 33R 46F
3	0.0	22.3	0.0	8.9	4A 118V 119H
4	0.0	18.2	0.0	7.3	55Q 115Y 116V

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

TABLE IV. Inaccessible Cavities Computed Using VOLBL for Modified Ribonuclease S Protein M13L (PDB Entry 1rbh)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	5.3	48.8	0.6	28.4	F8 H12 L13 V47 V54
2	0.6	27.4	0.0	13.0	N34 L35 K37 D38
3	0.0	21.1	0.0	9.0	F8 V47 I106 F120
4	0.0	20.9	0.0	8.9	F8 V47 V54 I106
5	0.1	31.4	0.0	11.5	Q55 Y115 V116

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

TABLE V. Inaccessible Cavities Computed Using VOLBL for Modified Ribonuclease S Protein M13A (PDB Entry 1rbc)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	25.2	134.9	2.6	83.5	F8 H12 A13 D14 S15 V47 E49 L51 V54
2	0.9	30.6	0.0	14.8	N34 L35 K37 D38
3	0.1	21.8	0.0	9.5	F8 E9 A13 L51
4	0.0	18.2	0.0	7.3	F8 V47 V54 I106
5	0.0	19.4	0.0	7.9	A109 H119 D121

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

TABLE VI. Inaccessible Cavities Computed Using VOLBL for Modified Ribonuclease S Protein M13F (PDB Entry 1rbe)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	7.7	66.0	0.8	38.1	S21 N27 M30 K31 T36 C95 Y97
2	5.3	50.1	0.6	29.6	F8 H12 F13 V47 V54
3	0.2	25.0	0.0	11.5	F8 V47 V54 I106 F120
4	0.1	20.7	0.0	8.8	T45 I81 I106 F120 S123
5	0.1	29.4	0.0	11.7	S21 K31 T36 C95
6	0.2	23.3	0.0	10.4	D83 K98 T100
7	0.0	21.6	0.0	8.8	V54 I106 V108

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

Modified protein M13F (1rbe) has seven cavities, three adjacent to residue 13, as listed in Table VI. This is rather counterintuitive, since the bulkier phenylalanine replacement introduces more cavities. It has been suggested that phenylalanine substitution induces conformational changes in its vicinity.[28] This changes both the number of cavities as well as residue make-up of the cavity walls.

In Table VII, a similar effect is observed in M13G, where, contrary to intuition, there are fewer cavities. The total cavity volume, however, is increased in this modified protein compared to the native protein.

Table VIII shows that the more isochoric substitution M13I (1rbg) leads to very small changes in the cavity sizes as compared to the native protein. Protein M13V (1rbi) has an additional cavity (volume 32.2 Å^3), although not in the immediate vicinity of residue 13 (Table IX). Such new cavity formation could affect protein stability.

The variable volumes of cavities formed in these modified proteins show that replacement at residue

TABLE VII. Inaccessible Cavities Computed Using VOLBL for Modified Ribonuclease S Protein M13G (PDB Entry 1rbf)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	13.8	69.0	2.9	49.0	S21 N27 M30 K31 T36 C95 Y97
2	0.4	26.2	0.0	12.1	N34 L35 K37 D38
3	0.0	20.4	0.0	8.5	T45 I106 F120 S123

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

TABLE VIII. Inaccessible Cavities Computed Using VOLBL for Modified Ribonuclease S Protein M13I (PDB Entry 1rbg)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	1.1	36.0	0.0	17.1	N34 L35 K37 D38 R39
2	0.7	28.2	0.0	13.6	F8 V47 V54 I106 F120
3	0.0	18.3	0.0	7.4	F8 H12 I13 V47
4	0.0	18.6	0.0	7.5	S21 K31 P93 C95
5	0.0	18.7	0.0	7.5	Q55 Y115 V116

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

TABLE IX. Inaccessible Cavities Computed Using VOLBL for Modified Ribonuclease S Protein M13V (PDB Entry 1rbi)†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	4.8	55.7	0.4	32.1	S21 N27 M30 K31 T36 C95 Y97
2	1.3	32.1	0.1	16.4	F8 H12 V13 V47 V54
3	0.1	23.8	0.0	10.7	F8 V47 V54 I106 F120
4	0.5	26.7	0.0	12.4	N34 L35 K37 D38
5	0.0	21.8	0.0	9.4	S75 S77 M79

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

13 leads to subtle changes in the conformation and packing of the protein structure. While relative deviations from the native structure may not be significant, modified proteins may contain cavities which differ both in size and location. Phenomenological thermodynamic analysis of protein stability due to mutations should, therefore, take into account these changes in cavities. The changes in cavities may also be counterintuitive to size-based arguments. This is evidenced by the fact that mutation of methionine to glycine results in elimination of a cavity. Lacking a sidechain, glycine provides the flexibility necessary for a conformational change in the protein backbone.

TABLE X. Inaccessible Cavities Computed by VOLBL for Membrane Protein Bacteriorhodopsin 2brd†

Cavity	Area SA	Area MS	Volume SA	Volume MS	Contributing residues
1	40.3	146.2	10.5	134.0	L97 A98 V101 D102 T107 L111 V151 L152 F154 G155 F171
2	13.9	98.7	2.42	71.2	L13 Y57 M60 I78 R82 T205 F208 M209 D212
3	10.6	101.0	1.31	71.0	L93 L174 V177 T178 W182 A215 F219 I222
4	9.29	73.4	1.25	51.9	Y79 R82 Y83 W86 W189 S193 E194 F208
5	5.79	58.3	0.80	39.9	F42 L100 L223 S226 R227
6	3.70	50.3	0.40	32.4	F27 F42 Y43 T46 L223 R227
7	6.13	71.6	0.56	48.6	M20 P50 A53 V213 K216 V217
8	4.14	53.7	0.49	34.8	S59 Y64 P77 W80 A81
9	3.71	63.6	0.21	35.8	P70 F71 G72 E74 Y79 E194
10	1.81	47.1	0.09	28.5	G122 W137 W138 W189 Ret(C3)
11	2.38	47.8	0.17	28.1	T46 V49 P50 L93 K216 F219
12	2.21	50.7	0.13	28.9	L111 A114 Y147 I148 V151
13	0.31	31.7	0.008	16.6	I45 T46 V49 L93 D96 F219
14	0.82	41.4	0.03	22.4	Y57 M60 R82 D85 W86 D212
15	0.84	46.7	0.019	21.9	V188 I191 G192 V199 L207
16	0.62	34.9	0.025	19.0	W12 L15 L206 M209
17	0.02	26.3	0.000	12.7	L152 F171 L174 R175
18	0.01	26.6	0.000	12.7	N76 Y79 L127

†Areas are in Å^2 and volumes are in Å^3. SA represents the solvent accessible model and MS the molecular surface model. Residues contributing to the cavities are also listed.

Bacteriorhodopsin: Cavities and Proton Pathway

A 3.5 Å resolution structure of the ground state of the light-driven proton pump bacteriorhodopsin (bR) has been obtained by Henderson and coworkers from electron microscopy studies.[29,30] The structure reveals the presence of a putative channel lined by a set of polar residues and the Schiff base, through which a proton can be translocated across the membrane. The role played by several of the proposed channel residues has been confirmed by mutagenesis experiments.[31–36] FTIR and fluorescence label-

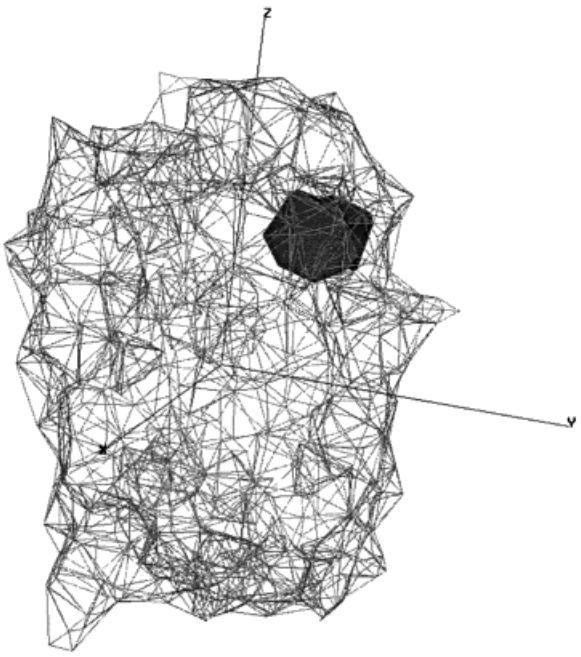

Fig. 2. Alpha shape representation of cavity 1 in bacteriorhodopsin (2brd). It is larger than the molecular cavity defined by the molecular surface.

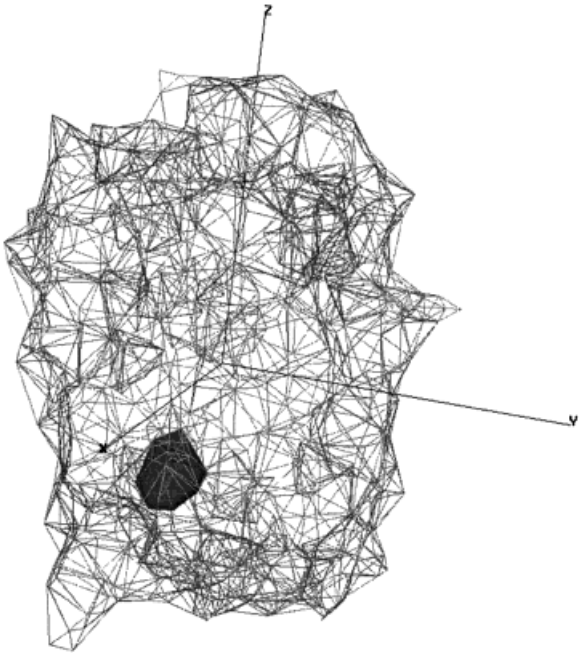

Fig. 3. Alpha shape representation of cavity 2 in bacteriorhodopsin (2brd).

ing experiments have shown that the immediate environment around a number of key residues in the putative proton transfer pathway change during the photocycle.[37] It has been postulated that a select number of water molecules in the pathway play an important role in proton transport.[38–44] A quantitative study of cavities which are large enough to accommodate one or more water molecules in the bR structure would shed light on possible interactions between the water molecules and the residues in the pathway, as well as environmental changes around key residues during the photocycle. We compute below the number and sizes of cavities in the ground state structure of bR. These are calculations from a recent structure whose coordinates were provided by Henderson and coworkers[29,30] (2brd in PDB files).

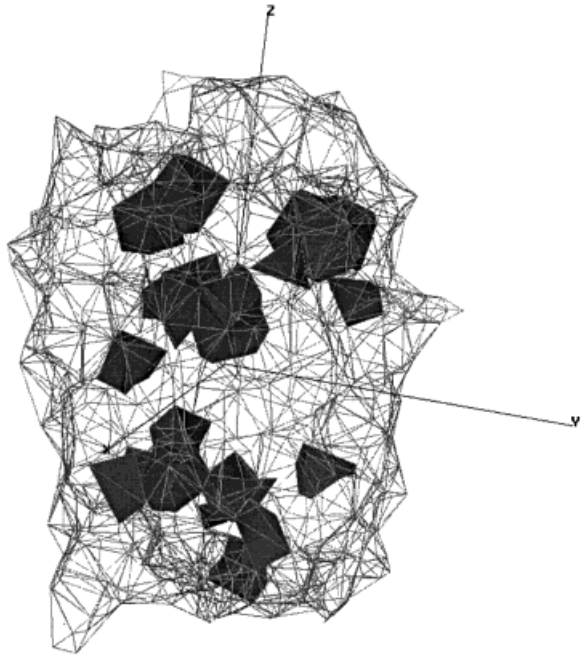

Fig. 4. Alpha shape representations of all cavities in bacteriorhodopsin.

We use 1.4 Å as the probe size for a water molecule. All membrane lipid molecules are removed for the calculation. The sizes of the computed cavities, along with the flanking residues, are presented in Table X. There are 18 cavities in bR which are inaccessible to bulk solvent and can accommodate one or more water molecules. The largest of these cavities (1 and 2) are at the cytoplasmic and extracellular ends of the protein, respectively. They are capable of accommodating 2–4 water molecules, which can serve as the proton transfer link to the bulk electrolyte. In Figures 2, 3, 4, and 5 we present a model of the protein along with the cavities. As shown in the figures, the largest cavity (cavity 1), which lines the proton channel from the cytoplasmic interior, is encompassed between helices C and G, with a few contacts from residues on helices B and F. It is in the vicinity of cavity 13, where C_β from Asp 96 is part of the cavity wall. The hydrophobic environment around Asp 96 in the ground state has been postulated on the basis of numerous experiments

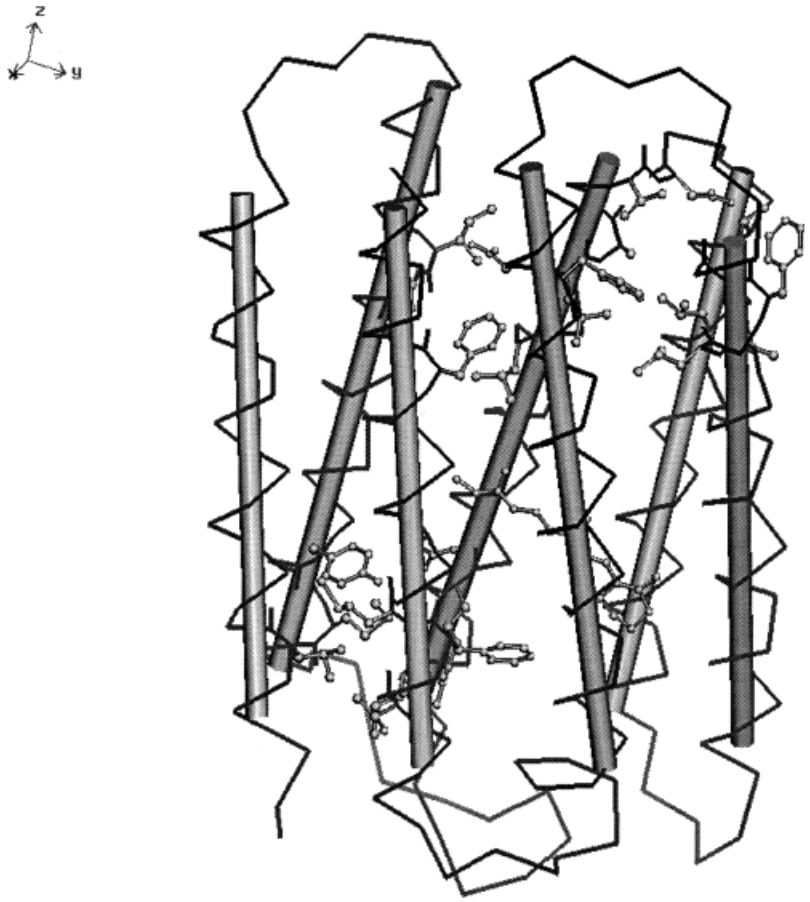

Fig. 5. The alpha carbons of the seven helices and the retinal, together with residue sidechains which contribute to cavities 1, 2, and 13.

and it has been suggested that the pK_a of Asp 96 in the bR ground state is as high as 13.[45] Our calculations reveal that wall atoms of both cavity 1 and cavity 13 are very hydrophobic: 18 out of the 19 wall atoms of cavity 1 are carbon atoms (except one oxygen atom from Val 151); whereas all six atoms on the wall of cavity 13 are carbon atoms. We display the sidechains contributing to cavities 1, 2, and 13 in Figure 5. Recent experiments, including the structure of the M intermediate,[46–48] show that during the photocycle the movement of the F helix changes the solvation environment around Asp 96. We postulate that the inaccessible and hydrophobic cavities 1 and 13 in the ground state becomes exposed to the cytosol in this process, thus exposing Asp96 and bulk water and altering its pK_a. This enables deprotonation of Asp 96 and translocation of the proton to the Schiff base.

CONCLUSIONS

We have presented an accurate method for computing inaccessible cavities in proteins using the alpha shape method. Molecular and solvent-accessible surface areas and volumes for the cavities can be computed using our method. In addition to computing total areas and volumes, our method also provides information on atoms and residues which contribute to the cavities. Precise metrics computed by our method can be used for probing protein structure and stability, as well as for protein engineering. Further, the computed cavities also reveal potential water-binding sites.

We applied our method to obtain cavities in previously studied proteins. The power of the method is illustrated by three key applications. In myoglobin, our computation located all of the cavities which bind xenon. In ribonuclease S-protein–S-peptide complex, we analyzed the role cavities play in engineered protein stability. In bacteriorhodopsin, we show that the two largest cavities line the putative channel involved in proton transport. The precise computation of cavities can be quantitated further to analyze protein thermodynamics and this study is in progress. Effects of buried surface area/volume in multidomain proteins and protein–protein complexes is also underway.

Currently, with our software, VOLBL, we do not store computed area/volume for pair and triple atoms intersection. Such information would be useful for further detailed analysis of polar–polar, polar–nonpolar, and nonpolar–nonpolar interactions. Imple-

mentations for accessing such information is also in progress.

ACKNOWLEDGMENTS

The software for constructing weighted Delaunay complexes and the alpha shape filters was written by Ernst Mücke and Michael Facello. We thank them for creating reliable and robust software so that we can build on their results. The authors thank NSF Meta Center Allocation for providing computational resources. The software VOLBL is part of the standard distribution of the alpha shapes software, and is available at http://alpha.ncsa.uiuc.edu/alpha.

REFERENCES

1. Kellis, J.T.J., Nyberg, K., and Fersht, A.R. Energetics of complementary side-chain packing in a protein hydrophobic core. Biochemistry 28:4914–4922, 1989.
2. Eriksson, A.E., Baase, W.A., Wozniak, J.A., Mathews, B.M. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature 355:371–373, 1992.
3. Erikson, A.E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Baldwin, E.P., Mathews, B.M. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183, 1992.
4. Lambright, D.G., Balasubramanian, S., Decatur, S.M. Anatomy and dynamics of a ligand-binding pathway in myoglobin: The roles of residues 45, 60, 64, and 68. Biochemistry 33:5518–5525, 1994.
5. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
6. Rashin, A., Iofin, M., Honig, B. Internal cavities and buried waters in globular proteins. Biochemistry 25:3619–3625, 1986.
7. Hubbard, S.J., Gross, K.-H., Argos, P. Intramolecular cavities in globular proteins. Protein Eng. 7:613–626, 1994.
8. Hubbard, S.J., Argos, P. Cavities and packing at protein interfaces. Protein Sci. 3:2194–2206, 1994.
9. Ernst, J.A., Clubb, R.T., Zhou, H.X., Gronenborn, A.M., Clore, G.M. Demonstration of positionally disordered water within a protein hydrophobic cavity by nmr. Science 267:1813–1817, 1995.
10. Zhang, L., Hermans, J. Hydrophilicity of cavities in proteins. Proteins 24:433–438, 1996
11. Richards, F.M. Areas, volumes, packing, and protein structures. Annu. Rev. Biophys. Bioeng. 6:151–176, 1977.
12. Connolly, T. Analytical molecular surface calculation. J. Appl. Cryst. 16:548–558, 1983.
13. Edelsbrunner, H., Mücke, E. Three-dimensional alpha shapes. ACM Trans. Graphics 13:43–72, 1994.
14. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire: Sur quelques propriétés de formes quadratiques positives parfaites. Journal für die Reine und Angewandte Mathematik 133:97–178, 1907 (in French).
15. Edelsbrunner, H. The union of balls and its dual shape. Discrete Comput. Geom. 13:415–440, 1995.
16. Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P., Subramaniam, S. Analytic shape computation of macromolecules. I. Molecular area and volume through alpha shape. Proteins 32: 1998.
17. Perrot, G., Cheng, B., Gilson, K., Palmer, K., Nayeem, A., Maigret, B., Scheraga, H. MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives. J. Comp. Chem. 13:1–11, 1992.
18. Akkiraju, N., Edelsbrunner, H. Triangulating the surface of a molecule. Discrete Appl. Math. 71:5–54, 1996.
19. Bernatein, F.C., Koetzle, T.F.W., B., G.J., Meyer, E.J., Brice, M.D., Rogers, J.R., Kennerd, O., Shimanouchi, T. The protein data bank: A computer based archival file for macromolecular structures. J. Mol. Biol. 112:532–542, 1977.
20. Ansari, A., Jones, C.M., Henry, E.R. Conformational relaxation and ligand binding in myoglobin. Biochemistry 33: 5128–5145, 1994.
21. Nienhaus, G.U., Mourant, J.R., Chu, K. Ligand binding to heme proteins: The effect of light on ligand binding in myoglobin. Biochemistry 33:13413–13430, 1994.
22. Elber, R., Karplus, M. Enhanced sampling in molecular dynamics—Use of the time dependent hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. J. Amer. Chem. Soc. 112:9161–9175, 1990.
23. Tilton, R.F., Kuntz, I.D., Petsko, G.A. Cavities in proteins: Structure of a myoglobin-xenon complex solved to 1.9 Å. Biochemistry 23:2849–2857, 1984.
24. Schoenborn, B.P. Binding of xenon to horse haemoglobin. Nature (London) 208:760–762, 1965.
25. Schoenborn, B.P. Structure of alkaline metmyoglobin-xenon complex. J. Mol. Biol. 45:297–303, 1969.
26. Tilton, Jr., R.F., Kuntz, I.D. Jr. Nuclear magnetic resonance stidues of xenon-129 with myoglobin and hemoglobin. Biochemistry 21:6850–6857, 1982.
27. Hermans, J., Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics. Israel J. Chem. 27:225–227, 1986.
28. Varadarajan, R., Richards, F. Crystallographic structures of ribonuclease s variants with nonpolar substitution at position 13: Packing and cavities. Biochemistry 31:12315–12327, 1992.
29. Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., Downing, K.H. Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy. J. Mol. Biol 213:899–929, 1990.
30. Grigorieff, N., Ceska, T., Downing, K., Baldwin, J., Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. J. Mol. Biol. 259:393–421, 1996.
31. Mogi, T., Stern, L.J., Hackett, N.R., Khorana, H.G. Bacteriorhodopsin mutants containing single tyrosine to phenylalanine substitution are all active in proton translocation. Proc. Natl. Acad. Sci. USA 85:5595–5599, 1987.
32. Mogi, T., Stern, L.J., Marti, T., Chao, B.H., Khorana, H.G. Aspartic acid substitutions affect proton translocation by bacteriorhodopsin. Proc. Natl. Acad. Sci. USA 85:4148–4152, 1988.
33. Stern, L.J., Khorana, H.G. Structure-function studies on bacteriorhodopsin. Individual substitutions of Arg residues by Gln affect chromophore formation, photocylce and proton translocation. J. Biol. Chem. 264:14202–14208, 1989.
34. Greenlagh, D.A.C., Altenbach, C., Hubbel, W.L., Khorana, H.G. Location of Arg-82, Asp-85, and Asp-96 in helix C of bacteriorhodopsin relative to the aquesous boundaries. Proc. Natl. Acad. Sci. USA 88:8626–8630, 1991.
35. Marti, T., Otto, H., Mogi, T., Rosselet, S.J., Heyn, M.P., Khorana, H.G. Bacteriorhodopsin mutants containing single substitution of seine or threonine residues are all active in proton translocation. J. Biol. Chem. 266:6919–6927, 1991.
36. Rath, P., Marti, T., Sonar, S., Khorana, H.G., Rothschild, K.J. Hydrogen bonding interactions with the Schiff base of bacteriorhodopsin. Resonance Raman spectroscopy of the mutants D85N and D85A. J. Biol. Chem. 268:17742–17749, 1993.
37. leCoutre, J., Tittor, J., Oesterhelt, D. Experimental evidence for hydrogen-bonded network proton transfer in bacteriorhodopsin shown by Fourier-transform infrared spectroscopy using azide as catalyst. Proc. Natl. Acad. Sci. USA 92:4962–4966, 1995.
38. Hildebrandt, P., Stockburger, M. Role of water in bacteriorhodopsin's chromophore: Resonance Raman study. Biochemistry 23:5539–5548, 1984.
39. Harbison, G.S., Roberts, J.E., Herzfeld, J., Griffin, R.G. Solid state NMR detection of proton exchange between the bacteriorhodopsin Schiff base and bulk water. J. Am. Chem. Soc. 110:7221–7227, 1988.
40. Cao, Y., Varo, G., Chang, M., Ni, B., Needleman, R., Lanyi, J. Water is required for proton transfer from Asp96 to the bacteriorhodopsin Schiff base. Biochemistry 30:10972–10979, 1991.

41. Rothschild, K.J., He, Y.-W., Sonar, S., Marti, T., Khorana, H.G. Vibrational spectroscopy of bacteriorhodopsin mutants. Evidence that Thr-89 and Thr-89 form part of a transient network of hydrogen bonds. J. Biol. Chem. 267: 1615–1622, 1992.
42. Deng, H., Callender, R., Ebrey, T. Evidence for a bound water molecule next to the retinal Schiff base in bacteriorhodopsin and rhodopsin: A resonance Raman study of the Schiff base hydrogen/deuterium exchange. Biophys. J. 66: 1129–1136, 1994.
43. Humphrey, W., Loguno, I., Schulten, K., Sheves, M. Molecular dynamics study of bacteriorhodopsin and artificial pigments. Biochemistry 33:3668–3678, 1994.
44. Balashov, S.P., Imasheva, E.S., Govindjee, R., Ebrey, T.G. Titration of aspartate-85 in bacteriorhodopsin: What it says about chromophore isomerization and proton release. Biophys. J. 70:473–481, 1996.
45. Sheves, M., Albeck, A., Friedman, N., Ottolenghi, M. Controlling the pK_a of the bacteriorhodopsin Schiff base by use of artificial retinal analogue. Proc. Natl. Acad. Sci. USA 86:3262–3266, 1986.
46. Subramaniam, S., Gerstein, M., Oesterhelt, D., Henderson, R. Electron diffraction analysis of structural changes in photocyle of bacteriorhodopsin. EMBO J. 12:1–8, 1993.
47. Varo, G., Lanyi, J.K. Effect of hydrostatic pressure on the kinetics reveal a volume increase in the bacteriorhodopsin photocyle. Biochem. 34:12161–12169, 1995.
48. Varo, G., Needleman, R., Lanyi, J.K. Protein structural change at the cytoplasmic surface as the cause of cooperativity in the bacteriorhodopsin photocycle. Biophys. J. 70: 461–467, 1996.
49. Cormen, T., Leiserson, C., Rivest, R. "Introduction to Algorithms." Cambridge, MA: MIT Press, 1990.
50. Edelsbrunner, H., Fu, P. Measuring space filling diagrams and voids. Rept. UIUC-BI-MB-94-01, Molecular Biophysics Group, Beckman Inst. Univ. Illinois, Urbana, IL, 1994.

APPENDIX

Filter of Delaunay Simplices

We repeat the steps in constructing the dual complex, as described earlier, by connecting atom centers to obtain simplices other than vertices: edges, triangles, and tetrahedra. However, we perform the construction of simplices directly for the Voronoi regions, rather than the Voronoi regions clipped to be within the VW molecule, we get the Delaunay triangulation or complex. The dual complex of the Voronoi dissection of the molecule is a subset of the Delaunay complex, as it is obtained by the same process but with a smaller number of regions with less overlap.

Indeed, we can simultaneously grow all balls in such a way that their Voronoi regions remain invariant and maintain the growing dual complex. This way we get a nested sequence of dual complexes, which are all subsets of the Delaunay complex. The last complex in the sequence is the Delaunay complex itself. To formalize this idea, we parameterize the growth of balls by a real number α and refer to the dual complex at time α as the α-complex. The 0-complex is the dual complex of the VW or the SA model, depending on the choice of radii. Negative values of α correspond to shrinking the balls. Although growing the balls is a continuous process, we get a finite sequence of alpha complexes:

$$\kappa_0, \kappa_1, \ldots, \kappa_m.$$

The first complex, κ_0, is the empty complex and the last complex, κ_m, is the Delaunay complex. We use this sequence of alpha complexes to sort the simplices in the Delaunay complex. For each Delaunay simplex, σ, there is a unique rank $r = r(\sigma)$ such that $\sigma \in \kappa_r$ and $\sigma \in \kappa_{r-1}$. The filter is an ordering of the simplices,

$$\sigma_0, \sigma_1, \ldots, \sigma_n,$$

so that the rank increases from left to right, that is, $r(\sigma_j) \leq r(\sigma_{j+1})$ for all j. It can happen that two simplices have the same rank, in which case we order vertex before edge before triangle before tetrahedron. The relevance of the filter will become clear shortly when we discuss the computation of inaccessible cavities in a molecule.

Representing Inaccessible Cavities

Consider the sequence of alpha complexes: κ_0, $\kappa_1, \ldots \kappa_m$, and let I be the index so that κ_I is the dual complex of the SA model. Let J be the largest index of any simplex in the filter with rank $r(\sigma_J) = I$. Hence σ_0 through σ_J belong to κ_1 and σ_{J+1} through σ_n do not belong to κ_n. We can take advantage of the fact that the simplices in the dual complex of the SA model are readily available as a prefix of the filter. All we need is the homotopy equivalence result proved in Reference 15: for every void of the SA model there is a unique corresponding void of κ_I which contains the SA void.

A void in κ_I is triangulated by Delaunay simplices that are not in κ_I. We use this set to represent the void. The set consists of tetrahedra, of triangles connecting tetrahedra, and of edges connecting triangles. The set is completely determined by its tetrahedra, and a triangle or edge σ_j belongs to the set if and only if $j > J$ and σ_j is a face of at least one tetrahedron in the set.

Constructing Inaccessible Cavities

We describe an algorithm that constructs sets of tetrahedra, where each set corresponds to a void of κ_I. The sets are stored in a union-find system, which is a standard data structure for maintaining a system of disjoint sets. Any tetrahedron can belong to at most one set or void. The set system is manipulated through a sequence of the following types of operations:

ADD(σ): Add the tetrahedron σ as the only member of a new set to the system.

FIND(σ): Determine and return the set that contains the tetrahedron σ.

UNION(U,V): Let $U \neq V$ be two sets in the system and replace them by the union, $V \cup V$.

Specific implementations of the union-find system can be found in most algorithm texts (see e.g., Reference 49). Our implementation takes time O(n log n) for a sequence of n operations. There are other implementations that are somewhat faster, but they

represent sets by trees, which is less convenient than the lists used in our implementation.

We return to the main problem, namely, constructing the voids. This is done by traversing the filter backwards, from σ_n down to σ_{J+1}. The tetrahedra among these simplices are collected in the sets of the union-find system. The triangles are used to trigger the union of sets. There is one special set in the system that collects all tetrahedra outside κ_I. This set does not represent any void but rather the outside component of the complement. We initialize the special set to $\{\sigma_\infty\}$, and we think of σ_∞ as the outside or complement of the Delaunay complex.

```
Initialize the set system to contain only the special
set: {{σ∞}};
for j := n down to J + 1 do
    if σj is a tetrahedron then
      ADD(σj)
    elseif σj is a triangle then
      determine the two tetrahedra σ, σ′ that share
      σj (one can be σ∞;
      U := FIND(σ); V := FIND(σ′);
      if U ≠ V then UNION(U,V) endif;
    endif
endfor.
```

For the correctness of the algorithm it is important that when a triangle σ_j is encountered, its two tetrahedra, σ and σ', have already been added to the system. The construction of the filter guarantees that this is indeed always the case. After running the algorithm, each set but the special one represents a void of κ_I. Next we show how such a set of tetrahedra can be used to compute the volume and the surface area of the corresponding void in the SA model.

Computing Volume

The method for computing the volume of the SA void represented by a set V of Delaunay tetrahedra is based on an inclusion-exclusion formula proved in Edelsbrunner.[15] We just explain the method. The volume of the SA void is

$$\text{vol} = \text{vol}_0 - \text{vol}_1 + \text{vol}_2 - \text{vol}_3$$

where the four terms accumulate the volume of four different types of geometric objects: tetrahedra, sectors of balls, wedges of intersections of two balls, and halves of intersections of three balls.

$\text{vol}_0 = \Sigma$ vol (σ), where vol(σ) is the volume of the tetrahedron σ and the sum extends over all $\sigma \in$ V.

$\text{vol}_1 = \Sigma\, \varphi_{\upsilon,\sigma} \cdot \text{vol}\,(b_\upsilon)$, where $\sigma \in$ V, $\in \nu\, \kappa_I$ is a vertex of σ, and b_υ is the ball with center υ. vol(b_υ) is the volume of b_υ and $\varphi_{\upsilon,\sigma}$ is the solid angle at υ inside σ. The sum extends over all $\sigma \in$ V and over all vertices $\upsilon \in \kappa_I$ of σ.

$\text{vol}_2 = \Sigma\, \varphi_{\epsilon,\sigma} \cdot \text{vol}\,(v_\upsilon \cap b_\mu)$, where $\sigma \in$ V, $\epsilon \in \kappa_I$ is an edge of σ, and υ and μ are the endpoints of ϵ, vol ($b_\upsilon \cap b_\mu$) is the volume of the intersection of the two balls, and $\varphi_{\epsilon,\sigma}$ is the dihedral angle at ϵ inside σ. The sum extends over all $\sigma \in$ V and over all edges $\epsilon \in \kappa_I$ of σ.

$\text{vol}_3 = \Sigma\, \frac{1}{2} \cdot \text{vol}\,(b_\upsilon \cap b_\mu \cap b_\lambda)$, where υ, μ, λ are vertices of a tetrahedron $\sigma \in$ V that span a triangle in $\kappa_I \cdot$ vol ($b_\upsilon \cap b_\mu \cap b_\lambda$) is the volume of the triple intersection.

All angles, whether solid or dihedral, are measured in revolutions, so 0 is the empty and 1 is the full angle. The above expression for the volume of a void is fairly natural and can easily be seen to be correct if the balls lining the void overlap at most in triplets. As proved in Edelsbrunner,[15] the expression is exact no matter how big the balls are and how many of them form common overlaps.

Computing Area

The above method for volume computation extends to area. The formula for the area of the SA void represented by the set V of tetrahedra is

$$\text{area} = \text{area}_1 - \text{area}_2 + \text{area}_3.$$

Again, we look at sectors of balls, wedges of intersections of two balls, and halves of intersections of three balls.

$\text{area}_1 = \Sigma\, \varphi_{\upsilon,\sigma} \cdot \text{area}\,(b_\upsilon)$, where the notation is the same as for vol_1, except that area (b_υ) is the surface area of ball b_υ.

$\text{area}_2 = \Sigma\, \varphi_{\epsilon,\sigma} \cdot \text{area}\,(b_\upsilon \cap b_\upsilon)$, where the notation is the same as in vol_2, except that area ($b_\upsilon \cap b_\mu$) denotes the surface area of the intersection of the two balls.

$\text{area}_3 = \Sigma\, \frac{1}{2} \cdot \text{area}\,(b_\upsilon \cap b_\mu \cap b_\lambda)$, where the notation is the same as in vol_3, except that area ($b_\upsilon \cap b_\mu \cap b_\lambda$) is the surface area of the common intersection of the three balls.

This method works for arbitrary arrangements of spherical balls and takes no advantage of properties of molecules, such as fairly uniform ball size and limited overlap. The above formulas reduce the computation to a small collection of primitive elements. A complete description of the analytic functions used to handle the primitive elements can be found in Reference 50. Many of the functions are the same as in Reference 12.