might be possible to show by femtosecond techniques that the rapid fluctuations shown for solvent dipolar correlations in time-dependent dielectric friction[36] or in computer simulation studies of dipolar fluids[37,38] appear as fluctuations in fluorescence maxima or intensities.

Note Added in Proof. The conversion of the initially formed excited state ($S_{1,bent}$) of bimanes (1,5-diazabicyclo[3.3.0]octa-3,6-dien-2,8-diones[39,40]) into a successor state ($S_{1,quasi-planar}$) is characterized by τ_1' in linear alkanols. The conversions involve a modest degree of charge rearrangement.[41]

Acknowledgment. E.M.K. is grateful for the facilities afforded him by Prof. A. Streitwieser, Jr., and the Chemistry Department, University of California, Berkeley, during the fall semester, 1983. Helpful discussions with Prof. Berni Alder (Berkeley-Livermore), Prof. H. L. Friedman (Stony Brook), Prof. M. Bixon (Tel-Aviv), and Prof. R. A. Marcus (California Institute Technology) are greatly appreciated. Information from Prof. J. H. Clark (Berkeley) in advance of publication was very helpful.

(37) Impey, R. W.; Madden, P. W.; McDonald, I. R. *Mol. Phys.* **1982**, *46*, 513-539.
(38) Edwards, D. M. F.; Madden, P. A. *Mol. Phys.* **1984**, *51*, 1163-1179.
(39) Kosower, E. M.; Ben-Shoshan, M.; Faust, D.; Goldberg, I. *J. Org. Chem.* **1982**, *47*, 213-221.
(40) Goldberg, I.; Kosower, E. M. *J. Phys. Chem.* **1982**, *86*, 332-335.
(41) Giniger, R.; Huppert, D.; Kosower, E. M., unpublished results.

Computation of Molecular Volume

Michael L. Connolly

Contribution from the Department of Molecular Biology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Rd., La Jolla, California 92037. *Received May 2, 1984*

Abstract: Volume is a fundamental physical property of molecules that is important in understanding their structure, function, and interactions. Present methods for computing volumes of macromolecules from crystallographically determined atomic coordinates introduce numerical errors that are, in the case of highly refined protein structures, larger than the experimental errors in the determination of the atomic coordinates. In order to obtain the maximum benefit from this high-quality experimental data, it is necessary to develop a volume-computation method whose numerical error is significantly less than the experimental error. Such a method is presented here. The molecule is modeled as a static collection of hard spheres which completely exclude a spherical probe representing a solvent molecule. van der Waals volumes are computed exactly, and solvent-excluded volumes are computed with an error of about 0.01%. The method's accuracy makes it particularly useful for comparing three-dimensional structures of a macromolecule in slightly differing conformations. Causes of such differences include temperature, oxidation state, presence of ligands, crystal form, and X-ray crystallographic refinement technique. Molecular volume changes during energy minimization, molecular dynamics simulation, and X-ray refinement can be monitored. This approach should also be of general utility in measuring the volumes of packing defects in protein interiors, ligand-binding pockets on protein surfaces, and gaps between molecules at subunit interfaces. Because the volume is defined analytically, it can be differentiated for use in energy functions.

Molecular volume is important, and it is important to measure it accurately. Volume is directly related to other physical chemical properties, such as charge, temperature, and pressure,[1] and its converse, which is density, has proved useful in studying protein tertiary structure. Density variations in different regions of the protein interior[2,3] and packing defects[4] have been identified and related to conformational fluctuations, hydrogen exchange, and the protein folding problem. These studies have also emphasized the importance of a suitable definition of the molecular surface in order to accurately measure molecular volume. It is important to measure molecular volume accurately in order to make full use of the information contained in high-resolution structural determinations of macromolecules.

Before proceeding to describe the volume computation method, it is necessary to define the terminology used in this work. The van der Waals volume is the volume occupied by the atoms when considered to be hard spheres with van der Waals radii. The solvent-excluded volume is the volume of space from which solvent is excluded by the presence of the molecule, when the solvent molecule is also modeled as a hard sphere, called the probe sphere. The interstitital volume consists of packing defects between the atoms that are too small to admit a probe sphere of a given radius. The solvent-excluded volume is the van der Waals volume plus the interstitial volume. The van der Waals volume can be considered to be a special case of the solvent-excluded volume, where the probe radius is zero. The general term, molecular volume, will be used to refer to both van der Waals and solvent-excluded volumes.

The term solvent-accessible surface will be used to refer to the smooth network of convex and reentrant surface traced by the inward-facing part of the probe sphere as it rolls over the molecule.[5,6] This surface is chemically important because it (i) forms the boundary of the solvent-excluded volume, (ii) has convex regions that are coincident with the part of the van der Waals surface that is accessible to a probe sphere, and (iii) is useful for graphically representing and analyzing the interface of macromolecules with each other and with solvent, drugs, and other small molecules. The term was originally used to refer to the surface traced out by the center of a probe sphere,[7] which is much easier to calculate but which lacks the advantages listed above. The general term, molecular surface, will also be used to refer to the solvent-accessible surface.

Two general approaches to geometric computations may be distinguished: numerical and analytical. A numerical algorithm subdivides a geometric object into a large number of small, similar units. For a three-dimensional object, these units may be cubes. The answer given is only approximate, and the amount of computation required for high accuracy is usually greater than that for an analytical method. An analytical method gives the answer

(1) Morild, E. *Adv. Protein Chem.* **1981**, *34*, 93.
(2) Richards, F. M. *J. Mol. Biol.* **1974**, *82*, 1.
(3) Kuntz, I. D.; Crippen, G. M. *Int. J. Peptide Protein Res.* **1979**, *13*, 223.
(4) Richards, F. M. *Carlsberg Res. Commun.* **1979**, *44*, 47.
(5) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(6) Connolly, M. L. *Science* **1983**, *221*, 709.
(7) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

0002-7863/85/1507-1118$01.50/0

Table I. Surface Definition Equations[a]

variable name	value
atomic coordinates	$\mathbf{a}_i$, $\mathbf{a}_j$, $\mathbf{a}_k$, ... (input)
van der Waals radii	r_i, r_j, r_k, ... (input)
probe radius	r_p (input)
$V_{ci} = (\phi_s/6)r_i^3 \sin\theta_{si}\cos^2\theta_{si}$	$d_{ij} = \lvert\mathbf{a}_j - \mathbf{a}_i\rvert$
torus axis $(\phi_s/2)[r_{ij}^2 r_p$ vector	$\mathbf{u}_{ij} = (\mathbf{a}_j - \mathbf{a}_i)/d_{ij}$
torus center	$\mathbf{t}_{ij} = {}^1/_2(\mathbf{a}_i + \mathbf{a}_j) + {}^1/_2(\mathbf{a}_j - \mathbf{a}_i)\cdot[(r_i + r_p)^2 - (r_j + r_p)^2]/d_{ij}^2$
torus radius	$r_{ij} = {}^1/_2[(r_i + r_j + 2r_p)^2 - d_{ij}^2]^{1/2}\cdot[(d_{ij}^2 - (r_i - r_j)^2]^{1/2}/d_{ij}$
base triangle angle	$\omega_{ijk} = \arccos(\mathbf{u}_{ij}\cdot\mathbf{u}_{ik})$
base plane normal vector	$\mathbf{u}_{ijk} = \mathbf{u}_{ij} \times \mathbf{u}_{ik}/\sin\omega_{ijk}$
torus base point unit vector	$\mathbf{u}_{ib} = \mathbf{u}_{ijk} \times \mathbf{u}_{ij}$
base point	$\mathbf{b}_{ijk} = \mathbf{t}_{ij} + \mathbf{u}_{ib}[\mathbf{u}_{ijk}\cdot(\mathbf{t}_{ik} - \mathbf{t}_{ij})]/\sin\omega_{ijk}$
probe height	$h_{ijk} = [(r_i + r_p)^2 - \lvert\mathbf{b}_{ijk} - \mathbf{a}_i\rvert^2]^{1/2}$
probe position	$\mathbf{p}_{ijk} = \mathbf{b}_{ijk} \pm h_{ijk}\mathbf{u}_{ijk}$
vertex	$\mathbf{v}_{pi} = (r_i\mathbf{p}_{ijk} + r_p\mathbf{a}_i)/(r_i + r_p)$
contact circle center	$\mathbf{c}_{ij} = (r_i\mathbf{t}_{ij} + r_p\mathbf{a}_i)/(r_i + r_p)$
contact circle radius	$r_c = r_{ij}r_i/(r_i + r_p)$
contact circle displacement	$d_c = \mathbf{u}_{ij}\cdot(\mathbf{c}_{ij} - \mathbf{a}_i)$
concave arc plane normal vector	$\mathbf{n}_{ijk} = (\mathbf{p}_{ijk} - \mathbf{t}_{ij}) \times \mathbf{u}_{ij}/r_{ij}$
concave triangle angle	$\beta_v = \arccos(\mathbf{n}_{ijk}\cdot\mathbf{n}_{ikj})$
convex face angle	$\alpha_v = \pi - \beta_v$
saddle wrap angle	$\phi_s = \arccos(\mathbf{n}_{ijk}\cdot\mathbf{n}_{ijl})$ when $\mathbf{n}_{ijk} \times \mathbf{n}_{ijl}\cdot\mathbf{u}_{ij} \geq 0$; $= 2\pi - \arccos(\mathbf{n}_{ijk}\cdot\mathbf{n}_{ijl})$ when $\mathbf{n}_{ijk} \times \mathbf{n}_{ijl}\cdot\mathbf{u}_{ij} < 0$
saddle width angle	$\theta_{si} = \arctan(d_c/r_c)$
euler characteristic	$\chi = 2$ − number of cycles

[a] As the probe sphere rolls around a pair of atoms, i and j, it traces out the volume of a torus, which has an axis $\mathbf{u}_{ij}$, center $\mathbf{t}_{ij}$, and radius r_{ij}. The circle of contact between the probe sphere and atom i has center $\mathbf{c}_{ij}$, radius r_c, and signed displacement d_c from the center of atom i. When the probe is simultaneously tangent to three atoms, i, j, and k, it has a center $\mathbf{p}_{ijk}$, at a height h_{ijk} above a base point $\mathbf{b}_{ijk}$, lying on the base triangle connecting the three atom centers. The contact point between the probe and atom i is called a vertex of the surface and is denoted by $\mathbf{v}_{pi}$. Concave triangles and convex faces meet at these vertices and have interior angles of β_v and α_v, respectively. The angle that a saddle face wraps around the torus axis is denoted by ϕ_s. The saddle width angle, θ_{si}, is defined in Figure 4. Also see Figure 1c,d.

Table II. Molecular Areas

face	area
convex	$A_+ = r_i^2[2\pi\chi - \sum_s \phi_s \sin\theta_{si} - \sum_v(\pi - \alpha_v)]$
saddle	$A_s = \phi_s[r_{ij}r_p(\theta_{si} + \theta_{sj}) - r_p^2(\sin\theta_{si} + \sin\theta_{sj})]$
concave	$A_- = r_p^2(\sum_v \beta_v - \pi)$

Table III. Volumes of Surface Pieces

Convex

$V_+ = {}^1/_3 r_i A_+$

Saddle

$V_{ci} = (\phi_s/6)r_i^3 \sin\theta_{si}\cos^2\theta_{si}$

$V_{si} = (\phi_s/2)[r_{ij}^2 r_p \sin\theta_{si} - r_{ij}r_p^2(\sin\theta_{si}\cos\theta_{si} + \theta_{si}) + (r_p^3/3)(\sin\theta_{si}\cos^2\theta_{si} + 2\sin\theta_{si})]$

$V_s = V_{ci} + V_{si} + V_{sj} + V_{cj}$

Concave

$A_f = {}^1/_2 d_{ij}d_{ik}\sin\omega_{ijk}$

$V_- = {}^1/_3(h_{ijk}A_f - r_p A_-)$

exactly as an equation, or a set of equations.

Methods

The starting point for computing solvent-excluded volume is to represent the solvent-accessible molecular surface and its area in an analytical fashion. This problem has been solved,[8] and the resulting definitions and equations are shown in Tables I and II. These earlier results are needed to define the variables used in the equations for the volumes (Table III).

Interior polyhedron. A polyhedron is derived from the analytical representation of the solvent-accessible surface. For each concave triangle of the surface, a flat triangle is constructed between the centers of the three atoms that the concave triangle bridges. These flat triangles define a polyhedron inside the surface.

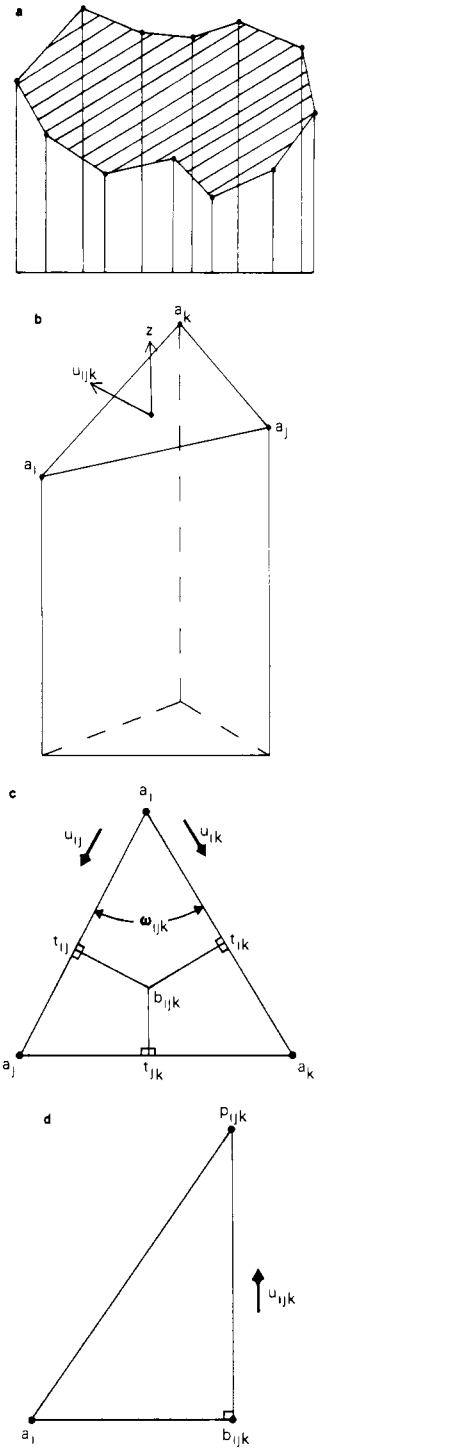

Figure 1. Interior polyhedron volume (shaded). (a) Two-dimensional representation of a polyhedron as a polygon. The sides of the polygon represent faces of the polyhedron. The horizontal line at the bottom represents the xy plane. The trapezoids between the polygon sides and the bottom line represent truncated triangular prisms between the polyhedron faces and the xy plane. The unshaded region represents cancellation between prisms with negative volumes and parts of the prisms with positive volumes. (b) Truncated triangular prism. The cosine of the angle between the triangle face normal vector and a unit vector pointing along the positive z axis is multiplied by the triangular face area to give the right-section area. (c) Triangular face connecting three solvent-accessible atoms. The two unit vectors $\mathbf{u}_{ij}$ and $\mathbf{u}_{ik}$ define the plane of the triangle. (d) The unit vector $\mathbf{u}_{ijk}$ is perpendicular to the plane of the triangle and points from the base point $\mathbf{b}_{ijk}$ to the probe sphere center $\mathbf{p}_{ijk}$.

(8) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.

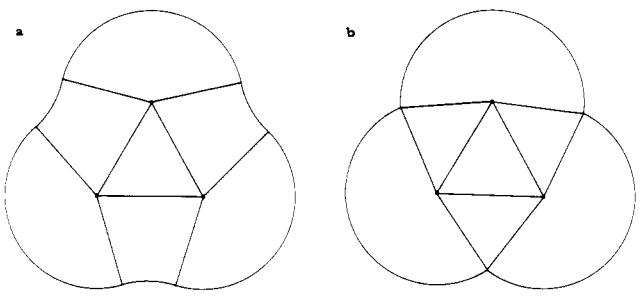

Figure 2. Two-dimensional representation of analytical partition: (a) Partition of solvent-excluded volume, (b) partition of van der Waals volume.

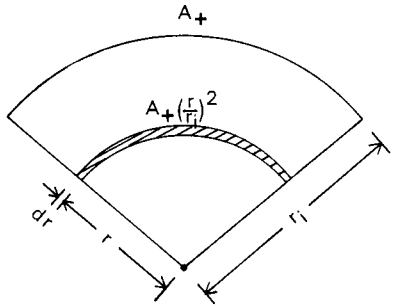

Figure 3. Two-dimensional representation of a convex piece. The atom center is at the vertex. The outer arc represents the convex face. Concentric arcs represent pieces of concentric spherical shells.

The volume of the interior polyhedron may be decomposed into truncated triangular prisms between each triangular face of the polyhedron and the xy plane (Figure 1). The volume of a truncated triangular prism is treated as a signed quantity, being the product of two signed quantities: the area of a right section of the prism and the average length of the three vertical edge of the prism.[9] A right section is the intersection of the prism with a plane perpendicular to the prism axis. The area of the right section is the product of the area of the polyhedral face, A_f (derived below), with the cosine of the angle between the normal vector to the polyhedral face, $\mathbf{u}_{ijk}$, and the prism axis. Denoting a unit vector along the prism or z axis by $\mathbf{z}$, then the area of the right section is $(\mathbf{u}_{ijk}\cdot\mathbf{z})A_f$. The average length of the three vertical edges of the prism is $(^1/_3)\mathbf{z}\cdot(\mathbf{a}_i + \mathbf{a}_j + \mathbf{a}_k)$. Summing the prism volumes over all faces of the polyhedron gives the polyhedron volume

$$V_p = \sum_f \frac{1}{3}\mathbf{z}\cdot(\mathbf{a}_i + \mathbf{a}_j + \mathbf{a}_k)(\mathbf{u}_{ijk}\cdot\mathbf{z})A_f$$

The volume outside the polyhedron but inside the surface, which will be named the surface layer, is decomposed or partitioned into a set of disjoint pieces (Figure 2). There is one piece for each face of the surface. Since there are three shapes of surface faces (convex, saddle-shaped, and concave), there are then three shapes of surface pieces. The volumes of the surface pieces may be calculated by using solid geometry and integral calculus, as described below.

Convex Piece. This computation is the simplest, being just the volume between the center of the atom and the convex face (Figure 3). This volume is the limit, as the shell thickness approaches zero, of the sum of the volumes of concentric spherical shell pieces, identical in shape but shrinking in size toward the atom center. The volume of each shell piece is equal to the convex face area, scaled by the square of the ratio of the shell and atom radii and multiplied by the infinitesimal thickness dr.

$$V_+ = \int_0^{r_i} A_+\left(\frac{r}{r_i}\right)^2 dr = \frac{A_+}{r_i^2}\int_0^{r_i} r^2\, dr = {}^1/_3 r_i A_+$$

Saddle Piece. The volume of the saddle piece (Figure 4) may be divided into four parts. Between the center of each atom and its circle of contact with the probe sphere (Table I) is a sector of cone. Between these two conical pieces is the volume of a sector of the hole inside a torus, which for ease of integration is divided into two parts lying on either side of the plane bisecting the torus.

The volume of the cone of atom i is one-third of the product of its altitude with the area of its base: $^1/_3\, d_c\pi r_c^2$. This volume must be multiplied by the fraction of a complete circle that the saddle face wraps around the torus axis: $\phi_s/(2\pi)$. This product is the volume of the conical sector next to atom i: $(\phi_s/6)\, d_c r_c^2$. From Figure 4 it can be seen that $d_c = r_i \sin\theta_{si}$ and $r_c = r_i \cos\theta_{si}$, so this volume may be written

$$V_{ci} = (\phi_s/6) r_i^3 \sin\theta_{si}\cos^2\theta_{si}$$

There is a similar term V_{cj} for the volume of the conical sector for atom j.

The volume of the hole of the torus between the two contact circles may be calculated by integral calculus. The volume is the limit, as the disk thickness approaches zero, of the sum of the volumes of circular disks of radius x and infinitesimal thickness $r_p \cos\theta\, d\theta$, where $x = r_{ij} - r_p \cos\theta$ (Figure 4). The volume of the part of the torus hole between the torus

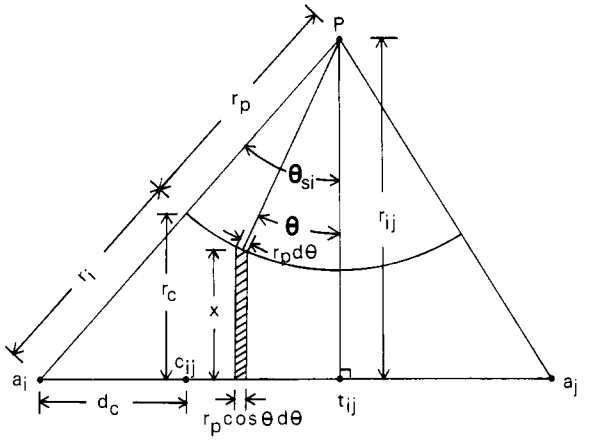

Figure 4. Cross-section of a saddle piece. This area is rotated part way around the interatomic axis to generate the saddle piece. The volume of the torus hole is computed by integrating infinitesimally thin circular disks between the contact circles.

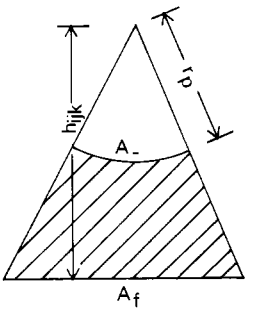

Figure 5. Two-dimensional representation of a concave piece (hatched). The probe center is at the apex. The arc represents the concave face and the line at the bottom represents the base triangle.

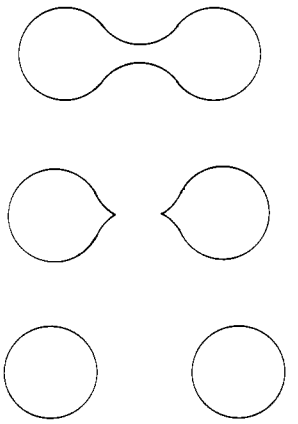

Figure 6. Two-dimensional representation of cusps. Two atoms close together share a continuous solvent-accessible surface, while two atoms separated by more than the probe diameter have distinct, spherical surfaces. In a limited intermediate range of interatomic distance, the solvent-excluded volumes extend part way toward each other, ending in cusps.

(9) Beyer, W. H. "CRC Standard Mathematical Tables"; CRC Press: Cleveland, Ohio, 1976; p 13.

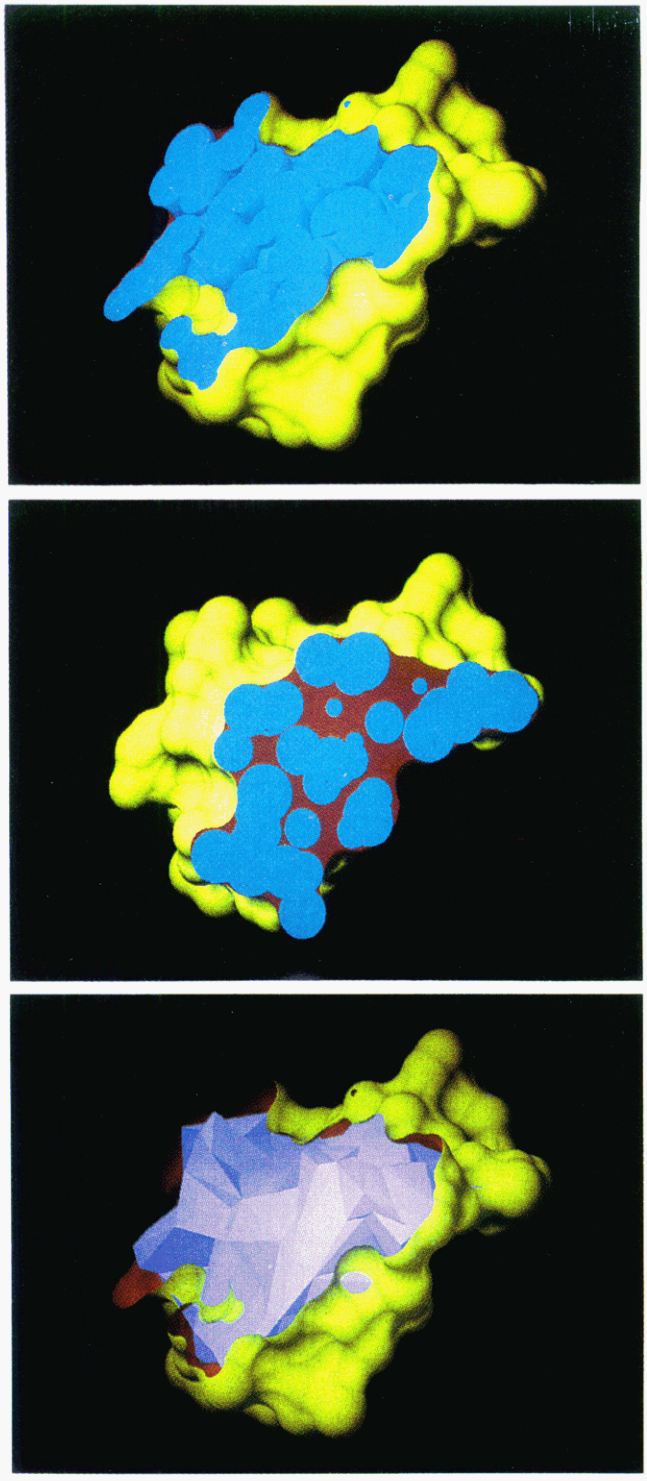

Figure 7. Crambin protein, (a, top) Cutaway view of atoms surrounded by a surface envelope. (b, middle) Cross-section of van der Waals (blue) and interstitial (red) volumes. (c, bottom) Interior polyhedron (magenta) inside clipped surface.

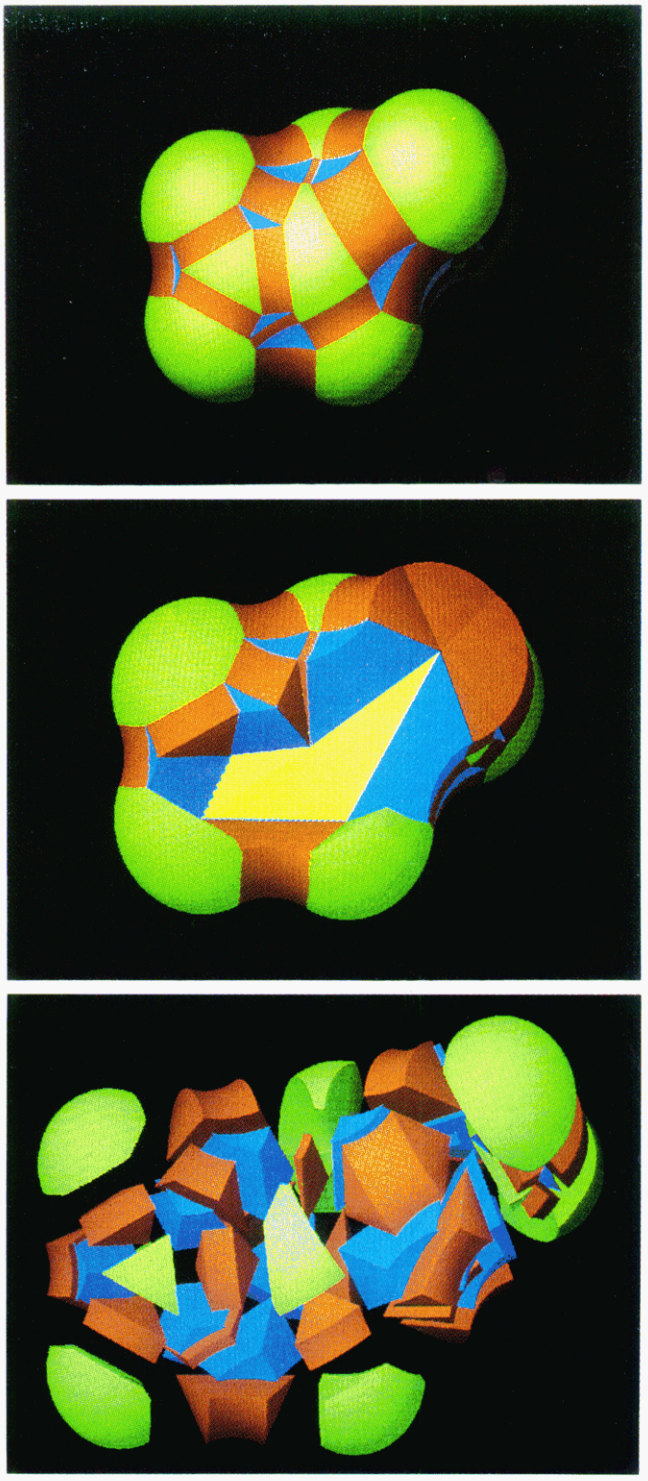

Figure 8. Asparagine amino acid. (a, top) Solvent-accessible surface: green, convex pieces; red, saddle pieces; blue, concave pieces. (b, middle) Cutaway view obtained by removing some surface pieces, not by clipping; yellow; interior polyhedron. (c, bottom) Exploded view of surface pieces. Interior polyhedron not shown.

center and the contact circle on atom i is given by

$$V = \int_0^{\theta_{si}} \pi x^2 r_p \cos\theta \, d\theta = \int_0^{\theta_{si}} \pi (r_{ij} - r_p \cos\theta)^2 r_p \cos\theta \, d\theta$$

But the saddle piece is only a sector of this hole, so the volume obtained by evaluating this integral must be multiplied by $\phi_s/(2\pi)$ to get the volume V_{si} of the toroidal sector

$$V_{si} = (\phi_s/2)[r_{ij}^2 r_p \sin\theta_{si} - r_{ij} r_p^2(\sin\theta_{si}\cos\theta_{si} + \theta_{si}) + (r_p^3/3) \times (\sin\theta_{si}\cos^2\theta_{si} + 2\sin\theta_{si})]$$

There is a similar term, V_{sj}, for atom j. The total volume of the saddle piece is then the sum of four terms

$$V_s = V_{ci} + V_{si} + V_{sj} + V_{cj}$$

Concave Piece. This volume (Figure 5) is the volume of a triangular pyramid minus the volume of a piece of a probe sphere. The base of the pyramid is the triangle between the centers of atoms i, j, and k. The area

of this flat face, A_f, is one-half the product of the lengths of two sides of the triangle times the sine of the angle between them

$$A_f = \tfrac{1}{2} d_{ij} d_{ik} \sin \omega_{ijk}$$

The volume of the pyramid with this triangle as base and the center of the probe as vertex is equal to one-third the product of the base triangle area, A_f, with the probe height, h_{ijk}. From this volume must be subtracted the volume of the piece of the probe sphere between the sphere center and the concave surface triangle. This volume may be calculated by the same method used to calculate the convex piece volume, giving one-third the sphere radius, r_p, times the concave face area A_-. The volume of the concave piece is then given by

$$V_- = \tfrac{1}{3}(h_{ijk} A_f - r_p A_-)$$

Cusps. On the rare occasions where the molecular surface intersects itself, generating cusps (Figure 6), the saddle pieces will be self-overlapping, and the concave pieces will overlap each other. The molecular volume must be corrected to account for these overlap volumes. The correction terms for simple cusps may be calculated analytically, and the sum of all the analytical cusp correction volumes is denoted by V_{ac}. Some cusps, such as those involving three or more concave pieces, are too complicated for analytical corrections, and their overlap volumes must be computed numerically. The sum of the correction volumes for these cusps is denoted by V_{nc}. The analytical and numerical cusp correction methods are complex and their detailed description is beyond the scope of this paper. Briefly described, an analytical saddle-piece correction involves starting the integration of each half of the torus hole not at the torus center but at the cusp point. An analytical concave-piece correction involves decomposing the overlap volume of two concave pieces into pyramids and pieces of spheres and cones. A numerical correction for multiply overlapping concave pieces is computed by filling the overlap volume with small cubes.

The total solvent-excluded volume of the molecule may then be written as

$$V_{mol} = V_p + \sum V_+ + \sum V_s + \sum V_- + V_{ac} + V_{nc}$$

where the sums are taken over the convex, saddle, and concave pieces, respectively. The van der Waals volume is computed as the solvent-excluded volume for a zero-radius probe.

Computer Program. A computer program (VAM) implementing this method for calculating molecular volumes, and which also calculates molecular areas, has been written. This program accepts as input the analytical molecular surface computed by the AMS program.[8] Both programs are written in Fortran 77. The AMS program uses 2–3 s/atom and the VAM program uses 0.5–0.7 s/atom of computer time on a VAX 11/750 with floating-point accelerator operating under VMS.

Input Parameters and Data. Unless otherwise stated, the surfaces and volumes in this work have been computed with use of a probe radius of 1.5 Å and van der Waals radii with implicit hydrogens taken from ref 10. The protein coordinates were obtained from the protein data bank at Brookhaven National Laboratory.[11]

Results and Discussion

Illustrations. The solvent-accessible surface of the small plant protein crambin[12] is shown in Figure 7a. This surface forms a contour surrounding the atoms of the molecule. The solvent-excluded volume can be seen to consist mainly of the van der Waals volume but also of interstitial volume between the atoms (Figure 7b). The interior polyhedron (Figure 7c) connecting the solvent-accessible atoms of crambin contains most of the molecular volume. The surface layer of an asparagine amino acid (Figure 8a) may have some of its pieces removed to reveal part of the surface of a small interior polyhedron (Figure 8b). Alternatively, the surface pieces may be exploded away from each other to show their individual shapes (Figure 8c).

Sensitivity to Parameters. The dependence of the solvent-excluded volume of crambin on the probe radius is shown in Figure 9. The van der Waals volume (4245 Å^3) is 82% of the solvent-excluded volume (5172 Å^3) for a 1.5-Å-radius probe and 76%

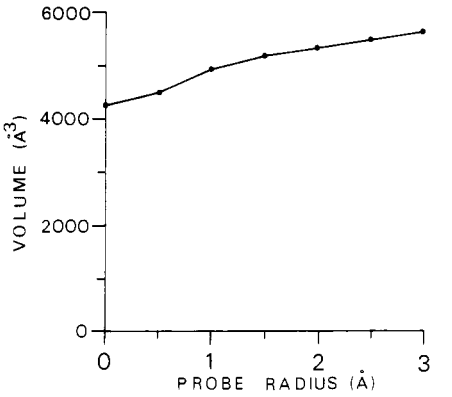

Figure 9. Solvent-excluded volume of crambin vs. probe radius.

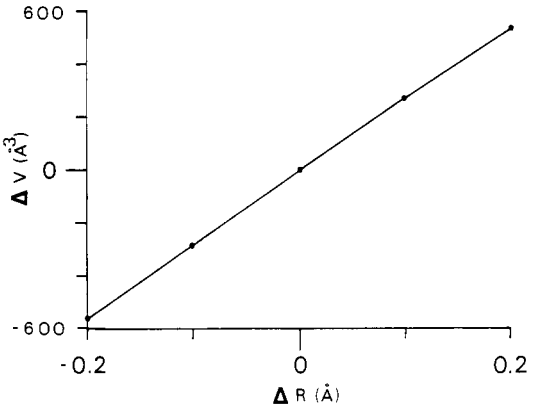

Figure 10. Change in solvent-excluded volume of crambin vs. change in atomic radii. The volume is 5172 Å^3 for a 1.5-Å-radius probe and unchanged atomic radii.

of the solvent-excluded volume (5619 Å^3) for a 3.0-Å-radius probe. The crambin volume dependence on the van der Waals radii of the atoms is shown in Figure 10. The volume changes by about 5% for each 0.1-Å change in atomic radii.

Accuracy. The extent to which the numbers computed by the analytical partition algorithm accurately reflect the physical property of molecular volume depends on both how well the hard-sphere model represents a molecule and how well the computer algorithm represents the hard-sphere model. Clearly the physical and chemical properties of a molecule are not completely accurately represented by considering the molecule to be a static collection of spheres, and the interactions with the solvent are more complex than simple steric exclusion of a spherical probe. The dynamics of molecules and the detailed nature of the solvent–solute interface are not addressed by this work. Only the geometric accuracy of the method is considered, that is, the accuracy with which the computer algorithm matches the hard-sphere model of the molecule.

An estimate of the geometric accuracy of the method may be made. Only the numerical correction term introduces an error. The accuracy of the numerical correction method for complicated cusps can be estimated from its accuracy when applied to simple cusps. In this case, the analytical correction can be used as a reference standard. For simple cusps, the numerical and analytical corrections are both computed, and these corrections differ by an average of 10%. When this is used as a guide for complicated cusps, where the volume of the numerical correction is typically 0.1% of the volume of the protein, then the protein volume would be in error by about 10% × 0.1% = 0.01%. The method is exact for van der Waals volumes, because there are no cusps.

Comparison with Previous Methods. The solvent-excluded volume has been computed by summing the volumes of small cubes contained within the solvent-accessible surface (Figure 11a). Cube methods have been developed by several researchers, most accurately by Muller[13] and Pavlov and Fedorov.[14] The latter

(10) McCammon, J. A.; Wolynes, P. G.; Karplus, M. *Biochemistry* **1977**, *18*, 927.
(11) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(12) Hendrickson, W. A.; Teeter, M. M. *Nature (London)* **1981**, *290*, 107.

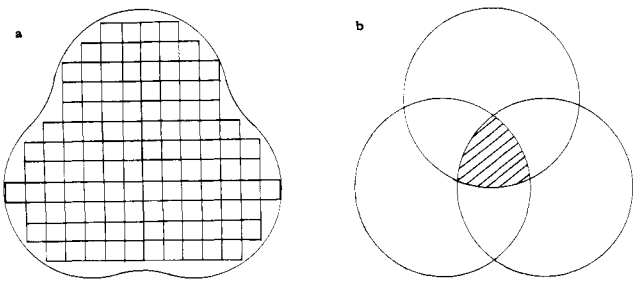

Figure 11. Two-dimensional representations of molecular volumes: (a) Solvent-excluded volume filled with small cubes, (b) van der Waals volume of three intersecting spheres. The triple overlap volume is shaded.

group has compared the volumes computed by their method to volumes calculated from experimental partial specific volumes. The average difference between the two sets of volumes was 2%.

The analytical partition method computes a volume for sperm whale myoglobin[15] of 19 793 Å^3 and a volume for T4 phage lysozyme[16] of 20571 Å^3, using the van der Waals radii of Lee and Richards[7] and a 1.5-Å-radius probe. Using the same radii, Pavlov and Fedorov[14] calculate values of 19 192 and 19 960 Å^3, respectively, which are both about 3% smaller. Their inaccuracy is caused by the fact that the smoothly curved molecular surface is approximated by the jagged boundary of the cubes. Methods for calculating solvent-excluded volumes using Voronoi polyhedra[17] have been developed and applied to proteins.[2,18-21] Again, an error is introduced because the molecular surface is represented only approximately, in this case as a polyhedron.

The van der Waals volume of a molecule is the union of the volumes of intersecting spheres representing its atoms (Figure 11b). The volume of the union of spheres may, in principle, be calculated by using the inclusion–exclusion principle.[22] This principle represents the volume of a union of objects as the sum of individual object volumes, adjusted by a series of intersection or overlap volumes. For three spheres, A, B, and C, the volume of the union is given by

$$V(A \cup B \cup C) = V(A) + V(B) + V(C) - V(A \cap B) - V(A \cap C) - V(B \cap C) + V(A \cap B \cap C)$$

Using this principle, van der Waals volumes have been calculated by Bondi,[23] who has computed pairwise overlaps analytically and ignored higher-order overlaps, and by Pavani and Ranghino,[24] who have gone one step further and computed triple-overlap volumes by numerical integration. In order to compute an exact van der Waals volume with use of the inclusion–exclusion principle, it would be necessary to compute overlap volumes of higher orders analytically, which has not been done. The analytical partition method gives union volumes directly, without considering intersections, because the pieces fit together without overlaps. It may be mentioned in passing that if in the future there develops a need to calculate intersection volumes, they may be computed from the union volumes given by the analytical partition method, using the converse or dual formulation of the inclusion–exclusion principle

$$V(A \cap B \cap C) = V(A) + V(B) + V(C) - V(A \cup B) - V(A \cup C) - V(B \cup C) + V(A \cup B \cup C)$$

In summary, previous methods for computing solvent-excluded volumes generate inaccuracies because of the crudity of their surface envelopes, while previous methods for van der Waals volumes generate inaccuracies because higher-order overlap volumes are ignored or are computed numerically rather than analytically.

Measuring Small Conformational Differences. The analytical partition method is able to compare very closely related structures, whose volumes differ by only a fraction of a percent. The volumes of two bovine pancreatic trypsin structures are compared: the diisopropylfluorophosphate-inhibited (DIP) structure[25] and the benzamidine-inhibited structure.[26] A comparison[25] between their independent refinements has shown that the average positional difference between internal main-chain atoms is 0.146 Å. Only the 1629 atoms of the protein and none of the buried water molecules or inhibitor atoms have been used in the volume calculations. The DIP-inhibited structure has a solvent-excluded volume of 27321 Å^3 and a van der Waals volume of 20903 Å^3. The benzamidine-inhibited structure has a solvent-excluded volume of 27 420 Å^3 and a van der Waals volume of 21 143 Å^3. The difference in van der Waals volumes (240 Å^3) is considerably greater than the difference in solvent-excluded volumes (99 Å^3). A structural change that would increase the van der Waals volume more than the solvent-excluded volume would be an increase in bond lengths, because there is then less overlap between bonded atoms. Indeed, a calculation shows that there is an average increase of 0.01 Å in the bond lengths of the benzamidine-inhibited structure relative to the DIP-inhibited structure. This increase could likely have resulted from different bond length parameters used in the two refinement procedures. Most of the increase in the van der Waals volume is absorbed by a decrease in the interstitial volume (141 Å^3). The solvent-excluded volume more accurately reflects the size of the protein as a whole, since it is less sensitive to stereochemical parameters.

We are now in a position to compare the magnitudes of volume variations introduced from three sources: van der Waals radii, crystallographic coordinates, and computational method. The atomic radii typically used in chemical calculations[2,7,10,14,23,24,27] vary by about 0.1 Å, which introduces a volume variability of about 5% (Figure 10). Probe radii used[6,7,13,14,18] typically vary by about 0.1 Å, which introduces a volume variability of about 1.0% (Figure 9). Differences in atomic coordinates of the two highly refined trypsin structures produce a solvent-excluded volume difference of 0.4%. The computational error is 3% for the improved cube method and 0.01% for the analytical partition method. Since the analytical partition method introduces an error that is significantly smaller than that caused by uncertainties in radii and coordinates, it does not degrade structural information. Although the volume variations caused by uncertainties in radii are greater than the difference in trypsin volumes, the volume comparison is still valid because the same radii were used for both structures.

There are many other situations where a macromolecule has had two slightly different three-dimensional structures determined. This is the case for myoglobin, which has had its structure solved at both room temperature and low temperature.[28] Both the liganded (oxy) and unliganded (deoxy) myoglobin structures[29] have been determined at high resolution (1.6 and 1.4 Å, respectively). The structures at different temperatures of a B DNA dodecamer have had their volumes compared by the analytical partition method.[6] The oxidized and reduced structures of cytochrome *c* have been solved at 1.8 and 1.5 Å, respectively.[30] The

(13) Muller, J. J. *J. Appl. Crystallogr.* **1983**, *16*, 74.
(14) Pavlov, M. Yu.; Fedorov, B. A. *Biopolymers* **1983**, *22*, 1507.
(15) Watson, H. C. *Prog. Stereochem.* **1969**, *4*, 299.
(16) Remington, S. J.; Ten Eyck, L. F.; Matthews, B. W. *Biochem. Biophys. Res. Commun.* **1977**, *75*, 265.
(17) Voronoi, G. F. *Reine Angew. Math.* **1908**, *134*, 198.
(18) Finney, J. L. *J. Mol. Biol.* **1978**, *119*, 415.
(19) Gellatly, B. J.; Finney, J. L. *J. Mol. Biol.* **1982**, *161*, 305.
(20) McCammon, J. A.; Lee, C. Y.; Northrup, S. H. *J. Am. Chem. Soc.* **1983**, *105*, 2232.
(21) Brostow, W.; Dussault, J. P.; Fox, B. *J. Comput. Phys.* **1978**, *29*, 81.
(22) Hall, M., Jr. "Combinatorial Theory"; Blaisdell: Waltham, MA., 1967; p 8.
(23) Bondi, A. "Molecular Crystals, Liquids and Glasses"; Wiley: New York, 1968.
(24) Pavani, R.; Ranghino, G. *Comput. Chem.* **1982**, *6*, 133.
(25) Chambers, J. L.; Stroud, R. M. *Acta Crystallogr., Sect. B* **1979**, *B35*, 1861.
(26) Bode, W.; Schwager, P. *J. Mol. Biol.* **1975**, *98*, 693.
(27) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(28) Hartmann, H.; Parak, F.; Steigemann, W.; Petsko, G. A.; Ponzi, D.; Frauenfelder, H. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4967.
(29) Phillips, S. E. V. *J. Mol. Biol.* **1980**, *142*, 531.
(30) Takano, T.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6371.

oxidized and semiquinone flavodoxin structures have been solved at 1.9 and 1.8 Å, respectively.[31] The method could be used to compare the monoclinic[32] and orthorhombic[33] crystal forms of yeast phenylalanyl transfer RNA. For proteins having several monomers in the asymmetric unit, such as superoxide dismutase,[34] the volumes of the different monomers could be compared.

Other Applications. The method could also be applied to the problem of protein compaction during refinement[27] and to following the volume variation during energy minimizations and molecular dynamics simulations. With modifications, the method could be used to measure empty spaces surrounded by atoms, such as internal cavities or packing defects.[4] Each internal void volume is surrounded by a separate piece of surface. For this case, the analytical partition method would be modified to consider the polyhedron defined by the centers of the atoms surrounding the cavity and to subtract from the volume of this polyhedron the volumes of the internal surface pieces of the cavity. The volumes of ligand-binding pockets on protein surfaces could be measured by developing a way to close off the mouths. The pocket volume would vary somewhat depending on how the capping was done. Interfacial void volumes, for example, those of hemoglobin subunit interfaces,[35] could be measured by defining a polyhedron from the centers of the atoms in the interface and subtracting from its volume the volumes of the surface pieces of these atoms.

Compatibility with Surface Areas and Graphics. The analytical partition method fits into a coherent scheme of methods and computer programs for calculating molecular surfaces, measuring their areas and volumes, and displaying them on both vector and raster computer graphics systems.[6,8] The importance of using compatible surface area and volume definitions has been emphasized.[36] The ability to display the analytical partition method graphically (Figures 7 and 8) not only communicates the method but also helps verify its correctness.

(31) Smith, W. W.; Burnett, R. M.; Darling, G. D.; Ludwig, M. L. *J. Mol. Biol.* **1977**, *117*, 195.
(32) Hingerty, B. E.; Brown, R. S.; Jack, A. *J. Mol. Biol.* **1978**, *124*, 523.
(33) Sussman, J. L.; Holbrook, S. R.; Warrant, R. W.; Church, G. M.; Kim, S.-H. *J. Mol. Biol.* **1978**, *123*, 607.
(34) Tainer, J. A.; Getzoff, E. D.; Beem, K. M.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1982**, *160*, 181.
(35) Greer, J.; Bush, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 303.
(36) Gates, R. E. *J. Mol. Biol.* **1979**, *127*, 345.

DNMR and Molecular Mechanics Studies of the Enantiomerization of Long-Chain (1,5)-Naphthalenophanes

Moon Ho Chang, Brian B. Masek,[1a] and Dennis A. Dougherty*[1b]

Contribution No. 7058 from the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125. *Received July 9, 1984*

Abstract: The enantiomerizations of several dioxa-(1,5)-naphthalenophanes (**1**: n = 14, 15, 16) have been studied by DNMR spectroscopy, and accurate activation parameters have been obtained. Large, negative entropies of activation are observed when n = 14 or 15. In sharp contrast, $\Delta S^{\ddagger}$ is very nearly zero when n = 16. Molecular mechanics calculations have been applied to these systems in an effort to obtain some insight into the underlying causes of this effect. The results suggest a model in which the polymethylene chains of all three compounds are quite unrestricted and conformationally flexible in the ground state. However, in the enantiomerization transition states when n = 14 or 15, the chain is quite restricted and this leads to the negative $\Delta S^{\ddagger}$. When n = 16, even in the enantiomerization transition state, the chain is relatively unrestricted and $\Delta S^{\ddagger}$ is small.

Macrocyclic organic molecules have been studied intensively in recent years. A major impetus for such work derives from the discovery of a large number of physiologically active substances that contain large rings, including the general classes of the ionophores[2] and the macrolide antibiotics.[3] Additionally, the fast-growing area of synthetic host–guest chemistry is dominated by macrocyclic compounds.[4] All such structures are of interest because they have quite specific binding capabilities, which implies they have well-defined, three-dimensional shapes. However, if one wished to a priori design a structure with a specific shape, a large carbocyclic ring would perhaps be the last starting point one would choose. Such structures are known to be extremely flexible, and to possess a large number of rapidly interconverting conformers.[5] It has been well documented that such "flexibility is the enemy",[6] when designing structures with specific binding properties. In nature, the ionophores and macrolides contain very specific substitution patterns along the chain,[2,3] the reproduction of which continues to be a focal point of modern synthetic organic chemistry. It seems certain that one role of such substituents is to diminish the conformational flexibility of the macrocycles, thereby leading to better defined topographies. In the area of synthetic host–guest chemistry, less subtle approaches to limiting ring flexibility have been adopted, primarily involving the incorporation of structurally rigid units (acetylenes, arenes, . . .) into the ring.

The present work was initiated with the goal of quantifying the effects of specific substituent patterns on macrocycle conformational dynamics. We believe that such information would be

(1) (a) NSF Predoctoral Fellow, 1981–1984. (b) Fellow of the Alfred P. Sloan Foundation, 1983–1985 Camille and Henry Dreyfus Teacher–Scholar, 1984–1989.
(2) Dobler, M. "Ionophores and Their Structures"; Wiley: New York, 1981.
(3) See, for example: Gale, E. F., et al. "The Molecular Basis of Antibiotic Action", 2nd ed.; Wiley: London, 1981.
(4) For a recent overview of the host–guest field, see: "Topics in Current Chemistry"; Vogtle, F., Ed.; Springer Verlag: Berlin, 1981 and 1982; Vol. 98 and 101.
(5) Dale J. *Top. Stereochem.* **1976**, *9*, 199–270. Anet, F. A. L.; Rawdah, T. N. *J. Am. Chem. Soc.* **1978**, *100*, 7166–7171, 7810–7814.
(6) Breslow, R. *Isr. J. Chem.* **1979**, *18*, 187–191.

0002-7863/85/1507-1124$01.50/0