Algorithms for Calculating Excluded Volume and Its Derivatives as a Function of Molecular Conformation and Their Use in Energy Minimization

Craig E. Kundrot,* Jay W. Ponder,† and Frederic M. Richards

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Received 2 August 1990; accepted 9 October 1990

A numerical method for calculating the volume of a macromolecule and its first and second derivatives as a function of atomic coordinates is presented. For N atoms, the method requires about $0.3\ N\ \ln(N)$ seconds of CPU time on a VAX-8800 to evaluate the volume and derivatives. As a test case, the method was used to evaluate a pressure-volume energy term in energy minimizations of the protein lysozyme at 1000 atm (1 atm = 1.013×10^5 Pa). R.m.s. gradients of 10^{-4} kcal/mol/Å were obtained at convergence. The calculated structures exhibited pressure-induced changes which were qualitatively similar to the changes observed in the 1000 atm structure determined by X-ray crystallography.

INTRODUCTION

Prior to the determination of the first three-dimensional protein structure, protein molecules were typically modeled as spheres or ellipsoids in calculations of hydrodynamic[1] or electrostatic properties.[2,3] With the advent of high resolution protein structures and advances in computational capabilities, methods using the overall shape of the protein and varying amounts of the detailed structure have been used in electrostatic,[4,5] packing,[6,7] and solvation energy[8–10] calculations. Iterative energy calculations such as energy minimization and molecular dynamics require efficient algorithms for the evaluation of such potentials and, in some applications, their derivatives.

This article presents a method for calculating the volume of a macromolecule and the first and second derivatives of the volume as a function of the atomic coordinates. Volume-dependent energy terms arise explicitly in studies involving high pressure, structural packing defects[7] and solvation energy.[10]

Although the volume of a small molecule tends to change slowly with conformation, the volumes of biological macromolecules such as proteins are very sensitive to conformation. Although proteins are dynamic structures in biological environments, they are well-packed structures on average[6] and are less compressible than ice.[11–13] The protein lysozyme contracts less than 1% and deforms only on the order of tenths of an Ångstrom unit at 1000 atmospheres (1 atm = 1.013×10^5 Pa) of hydrostatic pressure.[13] The crystallographic data in the lysozyme study are not sufficient to permit a detailed atomic interpretation of this pressure-induced deformation. In an effort to understand the mechanistic basis of this pressure-induced deformation, the TINKER force field[14] was modified by adding a pressure-volume energy term and the energy of the lysozyme structure minimized at 1 and 1000 atm.

Several types of surfaces and volumes are referred to in this work.[6] The accessible surface of the solute is the locus of points occupied by a probe sphere center as it is "rolled" across the solute. The excluded volume of a molecule is the volume enclosed by the accessible surface. The molecular surface is comprised of the contact surface and the reentrant surface. The contact surface is that part of the molecule's van der Waal surface, which is contacted by the surface of the probe sphere. The reentrant surface is the locus of points defined by that part of the probe sphere surface which faces the solute but does not contact solute atoms when the probe sphere contacts more than one solute atom. The molecular volume is the volume enclosed by the molecular surface, i.e., the volume inaccessible to any part of the probe sphere.

Previous workers have developed numerical[15] and analytical[16–19] methods for calculating molecular and excluded volumes. In this work, the excluded volume is computed using an adaptation of the analytical procedure of Connolly.[16] The first- and second-volume derivatives with respect to atomic positions are

*To whom all correspondence should be addressed. Present address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215.

†Present address: Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110.

CCC 0192-8651/91/030402-08$04.00

found by an efficient numerical approximation. Since the search for neighboring atoms is performed using a course cubic grid, the calculational time required for a molecule of N atoms is proportional to $N\ln N$. The energy minimization of the crystallographic structure of the lysozyme molecule at 1 and 1000 atm are presented as examples.

DERIVATION

The Excluded Volume

The pressure-volume energy term was defined using the excluded volume of the molecule for consistency with the derivation of the van der Waals equation of state. The algorithms of Connolly[16] were used to calculate the excluded volume to an accuracy of about 0.01%.

The First Derivatives

The first derivatives of the PV term with respect to atomic positions relate the change in atomic coordinates of solvent accessible atoms to the change in the excluded volume. One way to view these first derivatives is to view the solvent region as a continuum which exerts a force on each part of the accessible surface. Determining the resolved force acting upon each solvent accessible atom is then reduced to determining the accessible area of each atom and the resolved force acting on each patch of accessible surface.

The first step in obtaining the forces on an atom is to define the accessible area. The algorithm used in the program Access[20] was used to define the accessible area for each atom. The force acting on the accessible area was then calculated.

Because the geometry of several intersecting spheres is very complicated, Access solves a simpler problem (the intersection of circles in a plane) and uses numerical integration to combine the results. The atom in question, the "central" atom, is sectioned in planes along the z axis of the coordinate system used to describe the atomic coordinates. These z-sections are typically 0.1–0.2 Å apart. To cast the problem automatically in terms of the accessible surface, the radius of each atom is taken to be the van der Waals radius plus the probe sphere radius. A probe sphere radius of 1.4 Å was used to approximate the size of a water molecule. The surface of the central atom and its neighbors intersect each z section in a circle. The points where the neighbor's circles intersect the central atom's circle is calculated for each z section. The accessible area is calculated from the central atom arcs which are not occluded by neighboring atoms. In Access, the results of each z section are combined to find the accessible area of the atom.

The force acting on an atom's accessible surface can be similarly expressed as the combination of forces acting on each section. Consider one z section. The position of the z section and the spacing between sections defines ϕ_1 and ϕ_2; the z section intersects the central atom sphere at $\phi = (\phi_1 + \phi_2)/2$ (Fig. 1). Assume that the central atom circle is accessible to solvent over the range θ_1 to θ_2. The force acting on one point, B, on an atom whose center is at O is

$$\mathbf{dF} = P\,\mathbf{n}\,dS \tag{1}$$

where P is pressure, $\mathbf{n}$ is the normal vector $\mathbf{BO}/|\mathbf{BO}|$, and dS is the differential surface area. In spherical coordinates (Fig. 1)

$$\mathbf{n} = -\sin\phi\cos\theta\mathbf{i} - \sin\phi\sin\theta\mathbf{i} - \cos\phi\mathbf{k} \tag{2}$$

$$dS = a^2\sin\phi\,d\phi\,d\theta \tag{3}$$

where a is the radius of the sphere ($a = |\mathbf{BO}|$ = van der Waals radius + 1.4 Å) and $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the x, y, and z directions, respectively. Then the force acting on this accessible patch defined by θ_1, θ_2, ϕ_1, and ϕ_2 is

$$\mathbf{F} = \int_{\theta_1}^{\theta_2}\int_{\phi_1}^{\phi_2} -P[\sin\phi\cos\theta\mathbf{i} + \sin\phi\sin\theta\mathbf{j} + \cos\phi\mathbf{k}]a^2\sin\phi\,d\phi\,d\theta \tag{4}$$

$$\begin{aligned}\mathbf{F} = (-Pa^2/2)\,\{&(\sin\theta_2 - \sin\theta_1)[(\phi_2 - \phi_1) - (\sin2\phi_2 - \sin2\phi_1)/2]\mathbf{i} \\ &+ (\cos\theta_1 - \cos\theta_2)[(\phi_2 - \phi_1) - (\sin2\phi_2 - \sin2\phi_1)/2]\mathbf{j} \\ &+ (\theta_2 - \theta_1)(\sin^2\phi_2 - \sin^2\phi_1)\mathbf{k}\}\end{aligned} \tag{5}$$

The total force acting on the atom is obtained by adding the forces from each z section.

The Second Derivatives

The second derivatives of the PV energy relate the change in atomic coordinates of surface atoms to the change in the forces acting on the atoms. For example, if atom 1 defines part of the accessible surface of atom 0, then the pressure related force acting on atom 0 will change as atom 1 moves. For the force in the x direction, Fx, this can be written as

$$\frac{\partial Fx_0}{\partial x_1} = \frac{\partial}{\partial x_1}\left(\frac{\partial E}{\partial x_0}\right) = \frac{\partial^2 E}{\partial x_1 \partial x_0} \tag{6}$$

In terms of an accessible arc in a specific z section, the force is a function of the angles θ_1 and θ_2. The angles, in turn, are determined by the positions of the two respective neighboring atoms; atom 1 and

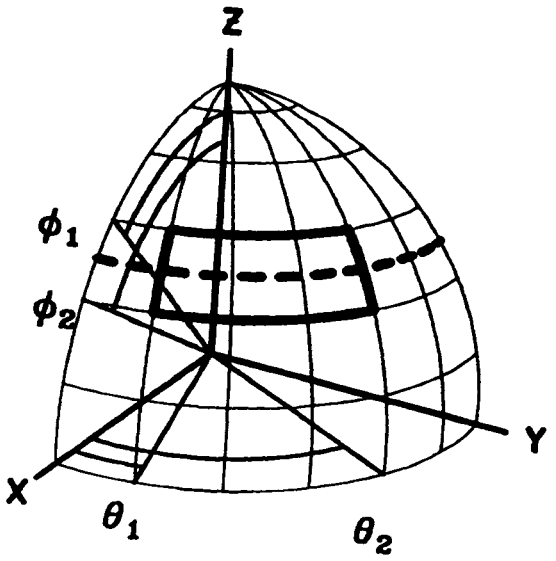

Figure 1. Arc definition. The arc segment shown (thick, solid lines) is defined in terms of an azimuthal angle, ϕ, and an equatorial angle, θ. The position of the azimuthal angle is determined by position of the z section (dashed line). $\phi = 0°$ and $\theta = 0°$ correspond to the z and x axes, respectively. The origin of this coordinate system is the atom center and the x, y, and z axes are parallel to the corresponding axes used in the protein coordinate data set.

atom 2. So to evaluate terms such as $\partial Fx_0/\partial x_1$, each accessible patch in a z section is analyzed separately. For a given patch, $\partial Fx_0/\partial x_1$ is evaluated using the equations below. The values of $\partial Fx_0/\partial x_1$ for each patch in the z section are summed to obtain the value in that z section, and the values of each z section are combined to obtain the final value of $\partial Fx_0/\partial x_1$. In short, the additive properties of the differential are used so that a much simpler and more tractable problem can be solved.

Consider one z section with three atoms, $i = 0$, 1, 2, with coordinates (x_i, y_i, z_i) and radii R_i (van der Waals radius + 1.4 Å). The central atom is $i = 0$ and atoms 1 and 2 define the arc angles θ_1 and θ_2, respectively. The force $\mathbf{F_0}$ is the pressure related force acting on atom 0 from the accessible patch defined by angles θ_1, θ_2, ϕ_1, and ϕ_2. We seek the derivatives of $\mathbf{F_0}$ with respect to the coordinates of atoms 0, 1, and 2. Note that the positions of the z sections are defined relative to (x_0, y_0, z_0), so that ϕ_1 and ϕ_2 are not a function of any atom coordinates.

We need to evaluate

$$\frac{\partial \mathbf{F_0}}{\partial x_i} = \frac{\partial \mathbf{F_0}}{\partial \theta_1}\frac{\partial \theta_1}{\partial x_i} + \frac{\partial \mathbf{F_0}}{\partial \theta_2}\frac{\partial \theta_2}{\partial x_i} \tag{7}$$

for $i = 0, 1, 2$ and analogous equations involving y_i and z_i. For most of this derivation, only the terms involving atom 1 will be written since the derivation involving atom 2 proceeds along the same lines. The above equations constitute level 1 of the chain rule applications.

Two types of terms are evaluated in level 2. The first is a straightforward differentiation of (5) to obtain the $\partial \mathbf{F_0}/\partial\theta$ type terms:

$$\begin{aligned}\partial \mathbf{F_0}/\partial\theta_1 = (\mathrm{Pa}^2/2)\,\{&\cos\theta_1[(\phi_2 - \phi_1) \\ &- (\sin 2\phi_2 - \sin 2\phi_1)/2]\mathbf{i} \\ &+ \sin\theta_1[(\phi_2 - \phi_1) \\ &- (\sin 2\phi_2 - \sin 2\phi_1)/2]\mathbf{j} \\ &+ (\sin^2\phi_2 - \sin^2\phi_1)\mathbf{k}\}\end{aligned} \tag{8}$$

For the second type of term, two new angles must be defined as illustrated in Figure 2

$$\beta_1 = \begin{cases}\tan^{-1}[(y_1 - y_0)/(x_1 - x_0)] & \text{quadrants I \& IV} \\ \tan^{-1}[(y_1 - y_0)/(x_1 - x_0)] + \pi & \text{quadrants II \& III}\end{cases} \tag{9}$$

$$\alpha_1 = \cos^{-1}(u_1) \tag{10}$$

where

$$u_1 = (r_0^2 + s_1^2 - r_1^2)/(2r_0s_1) \tag{11}$$

$$r_i = [R_i^2 - (z' - z_i)^2]^{1/2} \qquad (i = 0, 1) \tag{12}$$

$$s_1 = [(x_1 - x_0)^2 + (y_1 - y_0)^2]^{1/2} \tag{13}$$

and z' is the z coordinate of the z-section. Note that the arctangent takes on values between $-\pi/2$ and $\pi/2$. Since z' is defined by ϕ_1 and ϕ_2, $(z' - z_0)$ and r_0 are constants. Then

$$\theta_1 = \alpha_1 + \beta_1 \tag{14}$$

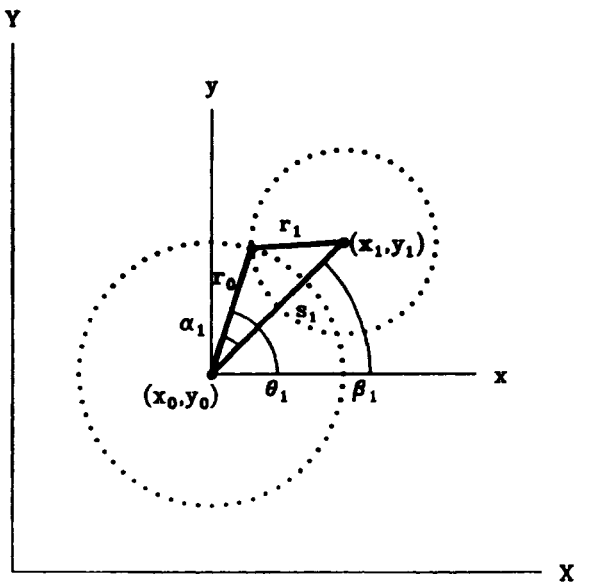

Figure 2. Geometry of a z-section. The X and Y axes are the axes of the protein atom coordinte system, the x and y axes are used to define θ_1, α_1, and β_1. The atoms 0 and 1 have coordinates (x_0, y_0) and (x_1, y_1), respectively, and the intersections of their atom spheres in this z section are represented by the dotted circles.

The particular point of intersection determines the sign of α. So

$$\frac{\partial \theta_1}{\partial x_i} = \frac{\partial \beta_1}{\partial x_i} + \frac{\partial \alpha_1}{\partial x_i} \tag{15}$$

for $i = 0, 1$ and analogous equations for y_i and z_i.

The $\partial\theta/\partial x$ type terms in level 2 are obtained differentiating (9) and (10) to produce the level 3 equations

$$\frac{\partial \beta_1}{\partial x_0} = (y_1 - y_0)/s_1^2 \tag{16}$$

$$\frac{\partial \beta_1}{\partial y_0} = -(x_1 - x_0)/s_1^2 \tag{17}$$

$$\frac{\partial \beta_1}{\partial z_0} = 0 \tag{18}$$

for $i = 0, 1$ and analogous equations for x_1, y_1, and z_1. Also

$$\frac{\partial \alpha_1}{\partial x_i} = [-1/(1 - u_1^2)^{1/2}] \frac{\partial u_1}{\partial x_i} \tag{19}$$

for $i = 0, 1$ and analogous equations for y_i and z_i.

The $\partial u/\partial x$ type terms of level 3 are obtained from (11) to produce the level 4 equations

$$\partial u_1/\partial x_i = (\partial u_1/\partial s_1)\,(\partial s_1/\partial x_i) \tag{20}$$

$$\partial u_1/\partial y_i = (\partial u_1/\partial s_1)\,(\partial s_1/\partial y_i) \tag{21}$$

$$\partial u_1/\partial z_i = (\partial u_1/\partial r_1)\,(\partial r_1/\partial z_i) \tag{22}$$

for $i = 0$ and 1. The derivatives on the right-hand side of (20)–(22) are obtained from (11)–(13) and are the level 5 equations

$$\partial s_1/\partial x_0 = -(x_1 - x_0)/s_1 \tag{23}$$

$$\partial s_1/\partial y_0 = -(y_1 - y_0)/s_1 \tag{24}$$

$$\partial u_1/\partial s_1 = (1/r_0) - (u_1/s_1) \tag{25}$$

$$\partial u_1/\partial r_1 = -r_1/(r_0 s_1) \tag{26}$$

$$\partial r_1/\partial z_0 = -(z' - z_1)/r_1 \tag{27}$$

and analogous equations for x_1, y_1, and z_1.

Equations (7)–(27) and analogous equations involving atom 2 were used to evaluate the second derivatives of the PV energy term.

CALCULATION OF ENERGY MINIMIZED STRUCTURE

Standard Potential Energy Functions

The basic TINKER potential energy functions[14] are variations of the usual set of functions found in molecular mechanics programs: bond stretching, bond angle bending, torsional angle rotation, van der Waals, and electrostatic terms (charge-charge, charge-dipole, and dipole-dipole). Actual values for the various parameters are taken directly from MM2[21] or from fits to small molecule thermodynamic and diffraction data, and to quantum mechanical calculations on protein subunits.[14] As described in the results, three different minimizations were performed, one including all long range interactions and two including only those van der Waals interactions within 8.0 Å and dipole-dipole interactions within 12.0 Å. A fifth-degree polynomial tapering function was applied at the cutoffs to keep the energy function continuous through the second partial derivatives with respect to atomic position (see also ref. 22). The tapering function was applied over the last 10, 25, and 50% of the van der Waals, dipole, and charge-charge interactions, respectively.

Pressure-Volume Potential Function

The pressure-volume function added to the total potential energy was

$$E_{PV} = 0.00001457 \times P \times V \tag{28}$$

where E_{PV} is in kcal/mol (1 kcal/mol = 4184 J/mol), P is measured in atm (1 atm = 1.013×10^5 Pa) and V is the excluded volume measured in Å^3 (1 Å = 10^{-10} m) using a 1.4 Å radius probe sphere. The pressure was fixed at 1000 atm during all calculations in which the E_{PV} term was used. The radii typically used in volume calculations represent the distance at which the van der Waals potential is at a minimum. To better approximate the "hard-sphere" boundary, the radii in this work used correspond to the distance at which the van der Waals potential produces a repulsive force of 1000 atm for atoms with 50% of their surface accessible to solvent (Table I). These radii were also used for the 1 atm structures. The work performed in going from the potential minimum to the radii in Table I is on the order of 0.2 kcal/mol, well below thermal energy (0.6 kcal/mol). Excluded volumes were computed with the algorithms used in the AMS and VAM programs of Connolly.[16] Derivatives of the excluded volume were computed using the formulas derived above. A numerical step size along the z axis of 0.01 Å or less was used in the derivative computations resulting in

Table I. van der Waals radii used in energy minimization.

Atom type	van der Waals radius (Å)
C	1.50
N	1.40
O	1.35
S	1.55
alcohol, amide, amine carboxyl H	0.80
other H	0.95

values accurate to within 1% or less. The last cycles of the minimization used a step size of 0.001 Å.

Energy Minimization

Since the observed X-ray pressure deformation and the computed pressure-volume work done are both very small, the ability to achieve complete convergence during energy minimization was crucial. Most optimization techniques currently applied to bipolymers provide an r.m.s. gradient per atom value of approximately 0.1 kcal/mole/Å (1 kcal/mole/Å = 6.948×10^{-11} N/molecule) even after many thousands of iterations. This level of performance was unacceptable for the current work because the residual force of 0.1 kcal/mole/Å obtained with most optimizers is of the same order of magnitude as the entire force exerted by the pressure-volume term. We therefore used a quadratically convergent Truncated Newton (TNCG) optimization procedure.[14] This method is a version of the classic Newton method for nonlinear optimization, but uses a preconditioned linear conjugate gradient procedure to partially solve the Newton equations at each cycle. Using the TNCG algorithm the calculation converged to an r.m.s. gradient of 0.0001 kcal/mole/Å in about 100 cycles or less. In the first part of the optimizations (i.e., r.m.s. gradient greater than 0.1 kcal/mole/Å) simple diagonal preconditioning was used. Symmetric successive over-relaxation (SSOR) or an Incomplete Cholesky preconditioner were then used to attain the final convergence.

APPLICATION

Test Case

Hen egg-white lysozyme served as the macromolecular test case for the volume-dependent terms because of our interest in examining the mechanistic basis of the deformations observed in the 1000 atm crystal structure. The crystal structure of lysozyme was determined using data extending to a nominal resolution of 2 Å.[13] The Hendrickson and Konnert restrained least-squares refinement program[24] was used to refine the 1 atm and 1000 atm crystal structures to an R factor ($R = \Sigma_{hkl} |F_o - F_c| / \Sigma_{hkl} |F_o|$) of 14.9%. The magnitude of the error in the crystallographic models was estimated by comparing the 1 atm model to a model derived from a "control" data set also collected at 1 atm but without the pressure cell. The control data was refined starting from the final 1 atm model. The r.m.s. difference was 0.047 Å between Cα positions and 0.063 Å for all atom positions.

Volumes for the structures are reported in Table II. To facilitate comparison of the crystallographic and calculated structures, the following atoms were deleted from the final structures prior to calculating the reported values: (1) residue side chains 61, 73, 97, 121, 125, and 128 which are not well-defined in the crystallographic model and (2) all hydrogen atoms since they are absent in the crystallographic model. To facilitate comparison to previously published work, the volumes were calculated using the commonly used set of radii[16,23] rather than the radii used in the energy minimization. The compressibility values calculated for lysozyme using the 1 atm and 1000 atm crystallographic structures are relatively insensitive to the probe sphere radius; 1.2 Å and 1.6 Å radii produce compressibility values within 1% of the value obtained with a 1.4 Å radius.

Computational Characteristics

The computational characteristics of the volume algorithms were as follows. The net pressure-related force acting on an entire molecule was used to evaluate the accuracy of the force calculation. The accuracy of the first derivative calculation is limited by the spacing between the z sections. The approximation made for each arc is that θ_1 and θ_2 are not a function of ϕ. As the interplane spacing is reduced, the approximation improves since θ_1 and θ_2 vary less over the ϕ range of an arc. The net pressure-related force acting on an entire molecule, which should be zero, was used to evaluate the accuracy of the force calculation. In practice, an interplane spacing of 0.06 Å appears to provide a good balance between accuracy and CPU time (Fig. 3), but the sensitivity of the truncated Newton method mentioned earlier required a spacing of 0.01 Å for all but the last cycles, where the spacing was 0.001 Å.

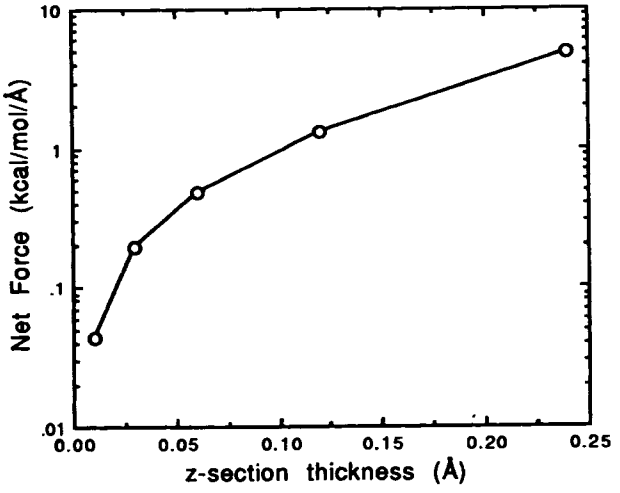

Figure 3. Trade-off between accuracy and spacing between z sections. The total pressure force acting on a lysozyme molecule is plotted against the spacing between z sections. A perfect calculation would produce a force of zero. The CPU time required to evaluate the pressure force is proportional to the number of z sections traversing the molecule.

Table II. Excluded volume values.[a]

	Excluded volume (Å³)			Pressure-induced change (%)			Minimization-induced change relative to 1 atm crystal structure (%)		
	Whole molecule	Domain I	Domain II	Whole molec.	Domain I	Domain II	Whole molec.	Domain I	Domain II
XRAY-1	23,892	15,705	9,482	−0.52	−0.52	−0.24			
XRAY-1000	23,767	15,624	9,818						
NOCUT-1	23,200	15,221	9,679	−0.90	−0.82	−0.54	−2.90	−3.08	−1.66
NOCUT-1000	22,991	15,096	9,627						
CUTOFF-1	23,147	15,351	9,613	−1.15	−0.83	−1.11	−3.12	−2.25	−2.33
CUTOFF-1000	22,880	15,224	9,506						
WATER-1	23,607	15,525	9,823	−0.80	−0.54	−0.41	−1.19	−1.15	−0.19
WATER-1000	23,419	15,441	9,783						

[a]An estimate of the experimental error in the crystallographic volumes is provided by a comparison of the XRAY-1 structure with a control 1 atm structure; the excluded volumes of these two molecules differed by 42 Å³ (0.18%), 48 Å³ (0.30%) and 33 Å³ (0.33%) for the whole molecule, domain I and domain II, respectively.

For N atoms, the CPU time required to evaluate the pressure-volume term and its derivatives are $O(N \ln N)$. The CPU time required as a function of the number of atoms is shown in Table III.

Calculated Structures

The 1-atm X-ray model was used as the starting point for each of a series of energy minimizations which generated three calculated 1 atm structures. Hydrogen atoms were added prior to the energy minimizations in idealized positions, forming hydrogen bonds where possible. The three optimizations and the resulting structures are: (1) CUTOFF-1, a single lysozyme molecule using a van der Waals cutoff of 8.0 Å and dipole-dipole cutoff of 12.0 Å, (2) NOCUT-1, a single lysozyme including all long-range interactions, and (3) WATER-1, lysozyme surrounded by 157 crystallographic waters conserved between the 1 and 1000 atm X-ray models and using the same 8.0/12.0 Å cutoffs as in (1). The calculations including water molecules were included because visual comparison of the structures using Insight version 2.3 (Biosym Technologies, Inc., San Diego, CA) suggested that explicit water molecules would act as "molecular doorstops." The water molecules filled voids on the protein surface and limited the motion of neighboring side chains.

Since the pressure related force on an accessible atom at 1 atm is at most 5×10^{-4} kcal/mol/Å per atom, the pressure term was not included in the 1-atm minimizations; the "1-atm" calculations were actually *in vacuo* calculations. The resulting structures will still be referred to as the 1 atm calculated structures. These 1-atm structures converged to a r.m.s. gradient value of 10^{-4} kcal/mol/Å or less.

In a second set of energy minimizations, the pressure-volume energy term was applied using 1000 atm of pressure to each of the 1-atm calculated structures. The resulting structures are referred to as CUTOFF-1000, NOCUT-1000, and WATER-1000. The two minimizations using cutoffs converged to r.m.s. gradient values below 10^{-4} kcal/mole/Å in 29 cycles (CUTOFF-1000) and 31 cycles (WATER-1000). The NOCUT-1000 structure did not converge to the same degree due to the combination of a dense Hessian matrix and very slow energy gradient evaluation. The NOCUT-1000 minimization was terminated after 43 cycles with an r.m.s. gradient value of 0.03 kcal/mole/Å and r.m.s. positional shifts of less than 0.0004 Å for the last five cycles. It is unlikely that further reduction of the gradient would lead to any significant structural change.

Table III. CPU Times.[a]

Molecule	Number of residues	Number of atoms	CPU time (s)		
			Volume	First Derivative	Second Derivative
Threonine	1	26	8.8	11.7	11.8
Enkephalin	5	75	22.2	31.4	28.4
Gramacidin A	16	276	108.1	145.3	133.6
Crambin	46	642	358.0	407.0	435.8
Lysozyme	129	1961	1300.8	1468.6	1442.5

[a]Timings were done on a DEC VAX 8820 for z sections of 0.0601 Å. For N atoms, the volume evaluations require about $0.09\ N \ln N$ s and the derivative evaluations require about $0.10\ N \ln N$ s. The time required is approximately proportional to the number of sections through the molecule.

Structural Comparison

The following comparisons can be thought of in terms of three transitions or operations: (1) the observed pressure-induced deformation, (2) the calculated pressure-induced deformation, and (3) the effect of energy minimization (comparing the 1 atm calculated structure to the 1 atm observed structure).

Table II shows that all three types of energy minimization produce a compression due to 1000 atm which is within a factor of about 2 of the observed crystal structures. The calculated compressibilities are too large for the whole molecule and the two domains. All calculations except the CUTOFF pair, agree with the crystallographic result that domain I is more compressible than domain II. However, the effect of energy minimization itself is several times larger than the observed or calculated pressure-induced deformations.

Though the pressure-volume energy term produces effects on the correct scale, the energy minimization procedure itself compresses the protein (Table II). Such changes prevent one from using the calculated structures to infer the mechanism of deformation in the observed structures.

DISCUSSION

Computational Cost and Performance

The new method described here, the calculation of the first and second derivatives of excluded volume with respect to atomic coordinates, is an efficient numerical method and whose computational time scales for N atoms is proportional to $N\log N$. The accuracy of the method is determined by the spacing between "slices" taken through the molecule.

Lysozyme Test Case

The details of the calculated deformation cannot be readily extrapolated to the crystal structures because of the large changes incurred by energy minimizing the 1-atm crystal structures. The excluded volume decrease observed upon energy minimization is reduced by the elimination of cutoffs for the van der Waals and electrostatic interactions and by the addition of explicit water molecules. This result confirms that the use of cut-offs for nonbonded potentials can have significant effects.[22]

There are at least two physical reasons why the energy minimized 1-atm structures are different than the observed crystal structure. Firstly, an energy minimized structure physically corresponds to a structure at 0 K. The volume of myoglobin at 80 K is 3% less than the volume at 300 K.[25] Extrapolating this result to 0 K produces a volume decrease of 4% compared to the 300 K. Unlike myoglobin, lysozyme does not possess large cavities and its volume contraction at low temperatures is probably smaller. So the contractions observed in these energy minimizations are of the magnitude expected for lysozyme at 0 K. Secondly, the force field used does not include a term for bulk solvent. The attractive van der Waals and electrostatic interactions between the protein and bulk solvent will tend to expand the protein. In short, there are physical reasons for the large (relative to the pressure-induced deformation) changes incurred by energy minimizing the protein crystal structure.

The lysozyme test case demonstrates the accuracy and feasibility of the volume derivatives. The volume and derivative calculations are accurate enough to permit energy minimizations to r.m.s. gradients of 10^{-4} kcal/mol/Å. The pressure-volume energy term produces pressure effects on the same scale as observed in the crystal structures.

Extensions

The current algorithm could be refined by defining an accessible arc more accurately. The approximation made for each arc is that θ_1 and θ_2 are not functions of ϕ. As the interplane spacing is reduced, the approximation improves since θ_1 and θ_2 vary less over the ϕ range of an arc. A computationally more efficient strategy would be to consider θ_1 and θ_2 as a function of ϕ. As a first approximation, θ can be regarded as a linear function of ϕ.

Fortran subroutines to evaluate the volume and the first and second derivatives are available as part of the TINKER molecular mechanics program package through J.W.P.

This work was supported by a grant from the Institute of General Medical Sciences to F.M.R. (GM22778). C.E.K. was supported as a predoctoral trainee on N.I.H. Training Grant GM07223 during part of this study. C.E.K. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research and this work was supported in part by the Jane Coffin Childs Memorial Fund for Medical Research. J.W.P. was supported as an N.I.H. postdoctoral fellow by an NSRA from the Institute of General Medical Sciences.

References

1. P.J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York, 1953, Chapter 14.
2. K. Linderstrøm-Lang, *Compt. Rendu Trav. Lab. Carlsberg*, **15,** 70 (1924).
3. C. Tanford and J.G. Kirkwood, *J. Am. Chem. Soc.*, **79,** 5333 (1957).
4. J. Warwicker and H.C. Watson, *J. Molec. Biol.*, **157,** 671 (1982).
5. R.J. Zauhar and R.S. Morgan, *J. Molec. Biol.*, **186,** 815 (1985).
6. F.M. Richards, *Annu. Rev. Biophys. Bioeng.*, **13,** 331 (1977).
7. A.A. Rashin, M. Iofin, and B. Honig, *Biochemistry*, **25,** 3619 (1986).

8. B.M. Pettitt and M. Karplus, *Chem. Phys. Lett.*, **121,** 194 (1985).
9. W.C. Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.*, (in press).
10. Y.K. Kang, K.D. Gibson, G. Nemethy, and H.A. Scherega, *J. Phy. Chem.*, **92,** 4735 (1988).
11. B. Gavish, E. Gratton, and C.J. Hardy, *Proc. Natl. Acad. Sci. USA*, **80,** 750 (1983).
12. K. Gekko and Y. Hasegawa, *Biochemistry*, **25,** 6563 (1986).
13. C.E. Kundrot and F.M. Richards, *J. Mol. Biol.*, **193,** 157 (1987).
14. J.W. Ponder and F.M. Richards, *J. Comp. Chem.*, **8,** 1016 (1987).
15. M.L. Connolly, *J. Appl. Cryst.*, **18,** 499 (1985).
16. M.L. Connolly, *J. Am. Chem. Soc.*, **107,** 1118 (1985).
17. T.J. Richmond, *J. Mol. Biol.*, **178,** 63 (1984).
18. K.D. Gibson and H.A. Scheraga, *Mol. Phys.*, **62,** 1247 (1987).
19. K.D. Gibson and H.A. Scheraga, *Mol. Phys.*, **64,** 641 (1988).
20. F.M. Richards, *Methods Enzymol.*, **115,** 440 (1985).
21. N.L. Allinger, *J. Am. Chem. Soc.*, **99,** 8127 (1977).
22. R.J. Loncharich and B.R. Brooks, *Proteins: Struc., Func. Genetics*, **6,** 32 (1989).
23. J.A. McCammon, P.G. Wolynes, and M. Karplus, *Biochemistry*, **18,** 927 (1977).
24. W.A. Hendrickson and J.K. Konnert, in *Biomolecular Structure, Function, Conformation and Evolution, Vol. 1*, R. Srinivasan, Ed., Pergamon Press, Oxford, 1981, pp. 43.
25. H. Frauenfelder, H. Hartmann, M. Karplus, I.D. Kuntz, Jr., J. Kuryan, F. Parak, G.A. Petsko, D. Ringe, R.F. Tilton, Jr., M.L. Connolly, and N. Max, *Biochemistry*, **26,** 254 (1987).