*Proc. Natl. Acad. Sci. USA*
Vol. 90, pp. 3835–3839, May 1993
Biophysics

# Predicting protein diffusion coefficients

**(linear operator theory/Fredholm theorems)**

DOUGLAS BRUNE AND SANGTAE KIM

Department of Chemical Engineering, University of Wisconsin, Madison, WI 53705

*Communicated by R. Byron Bird, December 17, 1992*

**ABSTRACT** **Diffusion coefficients for proteins in water are predicted. The numerical method developed is general enough to be applied to a wide range of protein surface shapes, from rodlike to globular. Results are presented for lysozyme and tobacco mosaic virus, and they are compared with actual data and with predictions made by less general methods.**

The prediction of the diffusion coefficients of a protein in water is of both practical and theoretical importance. In practical terms, diffusion coefficients are required in the design of mass transfer equipment used in the biochemical and food-processing industries. In theoretical terms, diffusion coefficient predictions have been used to estimate the number of water molecules tightly bound to a protein in solution (1).

In this work, the protein is modeled as a rigid body immersed in a newtonian solvent. Hydrodynamic interaction with other proteins is neglected, leading to a prediction of an infinite dilution diffusion coefficient. This model was originally studied by Einstein, who considered a spherical rigid body representing a sugar molecule. Others have extended his approach by using shape factors (2).

Brenner (3, 4) generalized Einstein's work to particles of arbitrary shape, deriving the following relation for translational and rotational diffusion coefficients:

$$D^{t} = \frac{kT}{\mu}\frac{1}{3}\,\mathrm{tr}(\mathbf{a})$$

and

$$D^{r} = \frac{kT}{\mu}\frac{1}{3}\,\mathrm{tr}(\mathbf{c}).$$

In these equations, $\mu$ is solvent viscosity, $k$ is Boltzmann's constant, $T$ is the absolute temperature, and $\mathbf{a}$ and $\mathbf{c}$ are mobility tensors for the particle, which relate the steady-state translational velocity $\mathbf{U}$ and rotational velocity $\boldsymbol{\omega}$ of the particle to the hydrodynamic force $\mathbf{F}$ and torque $\mathbf{T}$ exerted on the particle by the solvent:

$$\begin{pmatrix}\mathbf{U}^{\infty}-\mathbf{U}\\ \boldsymbol{\Omega}^{\infty}-\boldsymbol{\omega}\end{pmatrix} = \mu^{-1}\begin{pmatrix}\mathbf{a} & \mathbf{b}^{t}\\ \mathbf{b} & \mathbf{c}\end{pmatrix}\begin{pmatrix}\mathbf{F}\\ \mathbf{T}\end{pmatrix},$$

in which $\mathbf{U}^{\infty}$ and $\boldsymbol{\Omega}^{\infty}$ represent the undisturbed velocity fields in the fluid. These mobility tensors are obtained by solving the steady-state Stokes equations for the particle. This can be done analytically for simple shapes such as spheres and ellipsoids, but in general it must be done numerically. Most numerical methods fail for shapes as complicated as a protein. In this paper we describe a boundary integral method of solving the steady-state Stokes equations that is not subject to the numerical instability of previous methods, and we apply it to the prediction of protein diffusion coefficients.

While the classical integral representation for Stokes flow (5) follows from an application of Green's identity to the fundamental solution (Green's function), the integral representation used here follows as a consequence of the Fredholm theorems on existence and uniqueness of solutions of Fredholm integral equations of the second kind (6).

## METHODS

This section describes the numerical methods used to determine the mobility tensors for the proteins. The primary method used is a variation of the method presented in ref. 7, called the completed double-layer boundary integral equation method. Readers who are interested in the mathematical foundations of this method may refer to the sections on the Fredholm–Riesz–Schauder theory in ref. 8.

Consider a particle moving with the rigid body motion $\mathbf{u} + \boldsymbol{\omega} \times (\mathbf{x} - \mathbf{x}_0)$ in reaction to an external force $\mathbf{F}$ and torque $\mathbf{T}$. The motion of the particle creates a disturbance velocity field in the solvent that can be expressed as

$$\mathbf{v}(\mathbf{x}) = \frac{1}{8\pi\mu}\sum_i [\mathbf{f}_i\cdot\mathcal{G}(\mathbf{x},\boldsymbol{\zeta}_i) + \mathbf{t}_i\times\mathbf{R}(\mathbf{x},\boldsymbol{\zeta}_i)] - \int_S 2\boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\Sigma}(\mathbf{x}_s,\boldsymbol{\xi})\cdot\hat{\mathbf{n}}(\boldsymbol{\xi})dS(\boldsymbol{\xi}).$$

The disturbance velocity is represented by the sum of point forces and point torques distributed inside the particle, plus a surface distribution of hydrodynamic dipoles of unknown density $\boldsymbol{\phi}(\boldsymbol{\xi})$ to be determined. Once $\boldsymbol{\phi}(\boldsymbol{\xi})$ is determined, the problem is completely solved.

It can be shown (7) that the governing equation for $\boldsymbol{\phi}$ is the boundary integral equation:

$$\boldsymbol{\phi}(\mathbf{x}_s) - \oint_S 2\boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\Sigma}(\mathbf{x}_s,\boldsymbol{\xi})\cdot\hat{\mathbf{n}}(\boldsymbol{\xi})dS(\boldsymbol{\xi}) + \sum_{\nu=1}^{6}\boldsymbol{\phi}^{\nu}\oint_S\boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\phi}^{\nu}dS(\boldsymbol{\xi}) = \frac{-1}{8\pi\mu}\sum_i [\mathbf{f}_i\cdot\mathcal{G}(\mathbf{x}_s,\boldsymbol{\zeta}_i) + \mathbf{t}_i\times\mathbf{R}(\mathbf{x}_s,\boldsymbol{\zeta}_i)]. \quad [1]$$

Given the total force and torque on the particle, this is an integral equation for an unknown vector density function $\boldsymbol{\phi}$, called the double-layer density. $\mathcal{G}$ and $\mathbf{R}$ are singularity solutions of the Stokes equations for a point force and torque acting on the solvent, respectively, $\boldsymbol{\Sigma}$ is the stress field in the solvent created by a point force, and $\hat{\mathbf{n}}$ is the surface normal. The $\boldsymbol{\phi}^{\nu}$ are known eigenfunctions of the integral operator. On the right-hand side of the equation, a number of force singularities $\mathbf{f}_i$ and torque singularities $\mathbf{t}_i$ located at $\boldsymbol{\zeta}_i$ appear.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

These are related to the hydrodynamic force and torque on the particle by

$$\sum \mathbf{f}_i = \mathbf{F}$$

$$\sum \mathbf{f}_i \times \boldsymbol{\zeta}_i + \sum \mathbf{t}_i = \mathbf{T}.$$

Once the double-layer density is obtained, the particle velocities can be computed by using the following relations:

$$\oint_S \boldsymbol{\phi}\cdot\boldsymbol{\phi}^\nu dS = u_\nu \quad \text{for} \quad \nu = 1\text{–}3$$

and

$$\oint_S \boldsymbol{\phi}\cdot\boldsymbol{\phi}^\nu dS = \omega_{\nu-3} \quad \text{for} \quad \nu = 4\text{–}6.$$

There are two advantages of using this method over previous integral equation methods for solving the Stokes equations. First, the governing equation is a Fredholm equation of the second kind for the unknown double-layer density. Previous methods (9, 10) result in a first-kind equation which is subject to numerical ill-conditioning when discretized finely. The second-kind method, on the other hand, has successfully tackled a range of complex hydrodynamic shapes, including those likely to be of interest in protein modeling (11). Second, ref. 7 showed that the spectral radius of the integral operator in this method is less than one, allowing fast iterative solutions of the resulting discretized equation—i.e., we solve the governing equations as

$$\boldsymbol{\phi}(\mathbf{x}_s) = +\oint_S 2\boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\Sigma}(\mathbf{x}_s, \boldsymbol{\xi})\cdot\hat{\mathbf{n}}(\boldsymbol{\xi})dS(\boldsymbol{\xi}) - \sum_{\nu=1}^{6} \boldsymbol{\phi}^\nu \oint_S \boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\phi}^\nu dS(\boldsymbol{\xi})$$

$$- \frac{1}{8\pi\mu} \sum_i [\mathbf{f}_i\cdot\mathcal{G}(\mathbf{x}_s, \boldsymbol{\zeta}_i) + \mathbf{t}_i \times \mathbf{R}(\mathbf{x}_s, \boldsymbol{\zeta}_i)].$$

Up to this point, the method is largely the same as described in ref. 7, but some changes were made to take advantage of an initial approximate solution.

The approximate solution was obtained by using a method introduced and described fully in ref. 12. It consists in locating a number of singular solutions of the Stokes equations inside the particle at chosen locations, then attempting to match the rigid body motion boundary conditions on the surface of the particle by varying the strengths of the singularities. If the velocities are given, a simple least-squares problem for the singularity strengths results. The total force and torque can be obtained by adding the individual singularity contributions. If the total force and torque are given, the singularity strengths are subjected to these constraints, and the velocities can be obtained.

To understand the impact of this singularity solution on the boundary integral method, consider the governing equations 1. The term involving the eigenfunctions $\boldsymbol{\phi}^\nu$ on the left-hand side is the result of using the double-layer density to represent the velocity at the surface of the protein. If the trial singularity solution is in fact an exact solution, the above equations decouple, because the surface velocity will be given exactly by the singularity solution:

$$\sum_{\nu=1}^{6} \boldsymbol{\phi}^\nu \oint_S \boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\phi}^\nu dS(\boldsymbol{\xi}) = \frac{-1}{8\pi\mu} \sum_i [\mathbf{f}_i\cdot\mathcal{G}(\mathbf{x}, \boldsymbol{\zeta}_i) + \mathbf{t}_i \times \mathbf{R}(\mathbf{x}, \boldsymbol{\zeta}_i)].$$

The remaining part of the equation,

$$\boldsymbol{\phi}(\mathbf{x}) - \oint_S 2\boldsymbol{\phi}(\boldsymbol{\xi})\cdot\boldsymbol{\Sigma}(\mathbf{x}_s, \boldsymbol{\xi})\cdot\hat{\mathbf{n}}(\boldsymbol{\xi})dS(\boldsymbol{\xi}) = 0,$$

is simply a statement that the double-layer density corresponding to an exact singularity solution is an eigenfunction of the integral operator with eigenvalue $-1$. Since these eigenfunctions are known (any linear combination of the rigid-body motions of the protein), the double-layer density corresponding to this exact singularity solution can be calculated. It is just the double-layer density corresponding to the rigid-body motion velocities obtained as part of the singularity solution.

Although the singularity solution obtained is rarely exact, the information it provides can still help in two ways: first, in reducing the size of the problem which the boundary integral method must solve, and second, in providing a useful initial guess for the double-layer density.

The problem size is reduced because of a reduced need for fine discretization. With an effective singularity solution, the double-layer density is a relatively smooth function of the coordinates, while with an ineffective range completer, large variations in double-layer density occur over short distances on the surface of the protein. These large variations must be resolved by making the spatial grid finer, increasing the number of equations which must be solved.

The singularity solution also provides a choice of two initial guesses for the double-layer density. The velocity at any surface point can be computed from the singularity solution either by evaluating each singularity's contribution at that point or by evaluating the rigid-body velocities at that point.

To implement these boundary integral equations numerically, an approximation of the equations must be constructed on a set of boundary elements that describe the surface of the protein. To implement the singularity method, it is necessary to choose singularity locations within the protein and locations on the surface to enforce the boundary conditions.

For this study, we focus on two widely different subjects, to illustrate the generality of our method. The first, tobacco mosaic virus, is long (3000 Å), thin (180 Å in diameter), and roughly cylindrical, composed of a strand of RNA surrounded by coat proteins. The second, lysozyme, is a small globular protein, roughly 50 Å in diameter.

For tobacco mosaic virus, it is not necessary to know the exact positions of each atom to be able to discretize the surface. The virus is large enough that the surface irregularities are of little hydrodynamic significance. The protein is modeled as a cylinder, and the cylinder is divided into ever smaller panels until the results are self-consistent. The singularities are placed on the centerline of the virus. Fig. 1 shows a sample discretization into a hexagonal prism.

For lysozyme, on the other hand, there is not such a simple method of discretization. The molecule is irregular enough that it is necessary to obtain the shape from the atomic coordinates. The atomic coordinates of the protein were obtained from the Brookhaven Protein Databank (6LYZ.BRK). A picture of the molecule is shown in Fig. 2*A*, using the standard van der Waals radii for each atom.

The surface defined by the atomic radii shows a large amount of small-scale irregularity. On the level of continuum hydrodynamics, small irregularities in the protein surface have only a small effect on results, but they significantly increase the computational time required to get the results. To produce a surface amenable to hydrodynamic calculations, small-scale irregularities have been smoothed by using the algorithms of Connolly (13). These algorithms perform the mathematical equivalent of rolling a probe sphere across the van der Waals surface of the molecule. At the places the probe sphere touches the molecule, a surface is defined. The surface resulting from a probe radius of 3 Å is shown in Fig. 2*B*.

To discretize the surface of the protein into boundary elements, the algorithms of Connolly (14) have also been used. These algorithms take the smoothed molecular surface

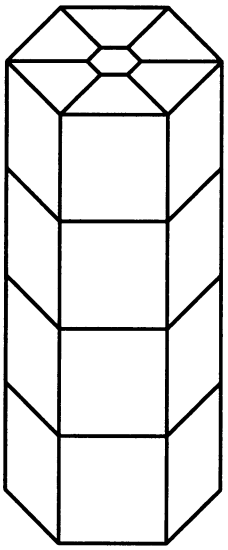

FIG. 1. Discretization of tobacco mosaic virus.

and divide it into a closed polygon composed of a number of triangles. The algorithms were designed for computer visualization of protein surfaces, and they produce a number of slender triangles which are unsuitable for use in a boundary element method. These have been combined with other triangles. The resulting surface is shown in Fig. 2*C*.

For the lysozyme molecule, singularities were located at the nitrogen atoms in the molecule, as long as the atoms were not within about 4 Å (about 2½ atomic radii) of the surface of the protein. This disperses the singularities throughout the protein, capturing the overall geometry, and avoids numerical difficulties associated with locating the singularities too near the surface.

The completed double-layer boundary integral equation method provides the rigid-body velocities of a protein given the forces acting on it. To predict the diffusion coefficients of the protein, its mobility tensors **a**, **b**, and **c** are needed. Referring to the defining relations for these tensors shows that solving for unit forces and torques in the coordinate directions will allow solution for successive columns of the tensors' cartesian representation.

As stated in ref. 7, the tensor **a** varies with the position in the particle taken as the origin of the coordinate system. In ref. 4, Brenner considers that the appropriate location to use for evaluation of the diffusion coefficient is the hydrodynamic center of mobility, the location at which the tensor **b** is symmetric. Wegener (15) also provides some proof of this assertion. When using the boundary element method of Kim and Karrila (7), however, it is convenient to choose the origin of the coordinate system so that null solutions of the equations are orthogonal to each other. The mobility tensors evaluated at this location must be translated to the hydrodynamic center of mobility.

It is possible to derive a relation giving the diffusion coefficients as a function of the origin of the coordinate system by starting with the following translation theorems for the mobility tensors of the particle (7):

$$\mathbf{a}^{(2)} = \mathbf{a}^{(1)} - \mathbf{r} \times \mathbf{c} \times \mathbf{r} - \mathbf{r} \times \mathbf{b}^{(1)} + \mathbf{b}'^{(1)} \times \mathbf{r}$$

$$\mathbf{b}^{(2)} = \mathbf{b}^{(1)} + \mathbf{c} \times \mathbf{r}$$

$$\mathbf{c}^{(2)} = \mathbf{c}^{(1)} = \mathbf{c}.$$

Here, **r** stands for the vector from the origin of the first coordinate system to the origin of the second. From these it is clear that the rotational diffusion coefficient does not

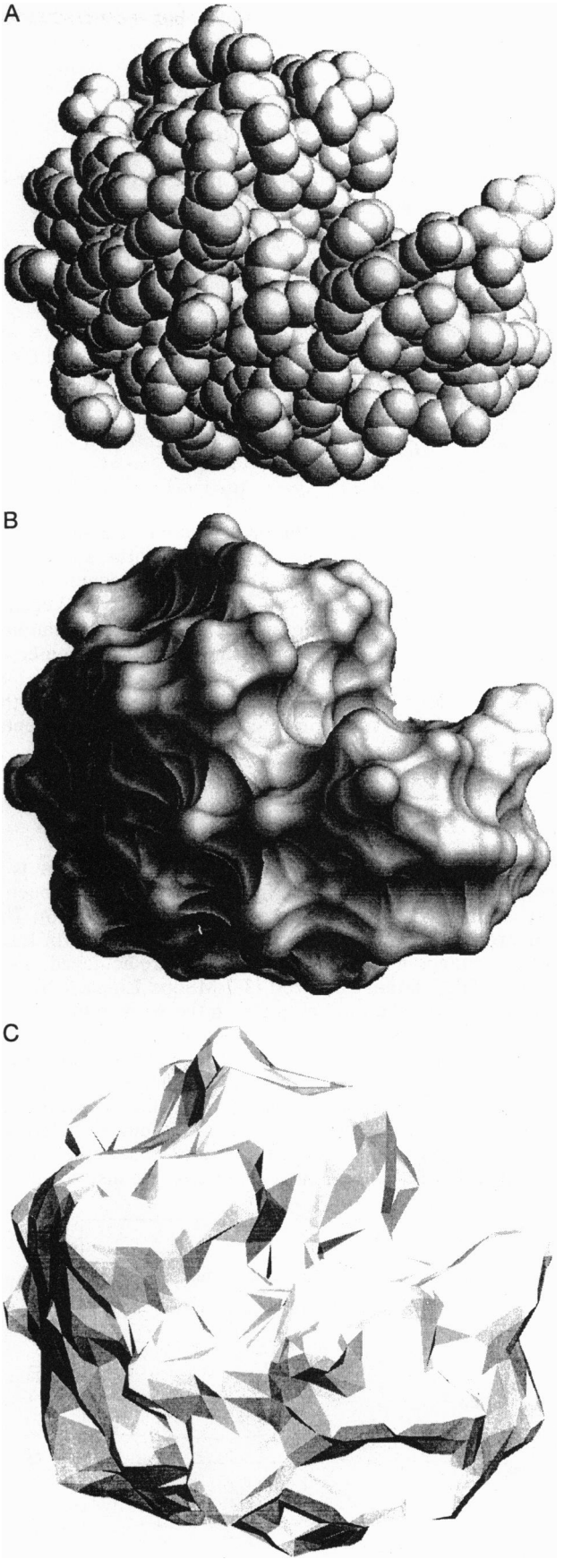


FIG. 2. Lysozyme. (*A*) van der Waals surface. (*B*) Smoothed surface. (*C*) Triangulated surface.

depend on the location of the origin, but the translational diffusion coefficient might. Taking the trace of the translation theorem for the **a** tensor, considering point 2 to be the center of mobility, yields

$$\mathrm{tr}(\mathbf{a}^{(cm)}) = \mathrm{tr}(\mathbf{a}^{(1)}) - \mathrm{tr}(\mathbf{r} \times \mathbf{c} \times \mathbf{r}) - \mathrm{tr}(\mathbf{r} \times \mathbf{b}^{(1)}) + \mathrm{tr}(\mathbf{b}^{t(1)} \times \mathbf{r}).$$

In this case, **r** is the vector from point 1 to the center of mobility and is given by a relation in ref. 7:

$$\mathbf{x}_{cm} - \mathbf{x}_1 = \frac{b_{23}^{(1)} - b_{32}^{(1)}}{(c_2 + c_3)}\mathbf{e}_1 + \frac{b_{31}^{(1)} - b_{13}^{(1)}}{(c_3 + c_1)}\mathbf{e}_2 + \frac{b_{12}^{(1)} - b_{21}^{(1)}}{(c_1 + c_2)}\mathbf{e}_3.$$

In this formula, $\mathbf{e}_i$ are the eigenvectors of **c**, $c_i$ are the eigenvalues, and $b_{ij}$ are components of **b**. Combining these two equations gives an expression for the translational diffusion coefficient at the center of mobility:

$$D_{cm}^{t} = D^{t} - \frac{kT}{3\mu}\left[\frac{(b_{23} - b_{32})^2}{(c_2 + c_3)} + \frac{(b_{31} - b_{13})^2}{(c_3 + c_1)} + \frac{(b_{12} - b_{21})^2}{(c_1 + c_2)}\right],$$

where the right-hand side can be evaluated with any convenient choice of origin. Note that the translational diffusion coefficient clearly reaches a minimum at the center of mobility, since the eigenvalues of $c$ are all positive. Wegener (22) also showed in a different way that the translational diffusion coefficient reaches a minimum at the center of mobility.

When the diffusion coefficients have been evaluated at the center of mobility, they can be compared with experimental values.

## NUMERICAL RESULTS AND DISCUSSION

The first protein chosen was lysozyme, because it is a relatively small well-characterized protein which has been a test subject for a number of predictive methods (1, 16). The convergence behavior for the diffusion coefficients of lysozyme is shown in Fig. 3. These results were achieved in 4.5 hr on a DECstation 5000/200 (3.7-Mflops Linpack benchmark) and will take only minutes on the workstations presently being introduced. In the case of lysozyme, we can compare the predicted diffusion coefficient to the experimentally determined value (1, 17) and to other predictions, as shown in Table 1. As a reference, we include the diffusion coefficient predicted by the method of Einstein, modeling the molecule as a simple sphere with a volume that of the surface we generated for our boundary method. We also include a

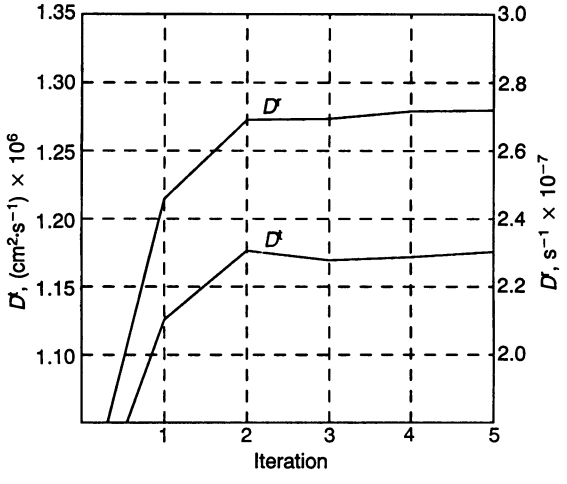


FIG. 3. Convergence of diffusion coefficients of lysozyme.

Table 1. Diffusion coefficients of lysozyme

| | Experimental | Predicted | | |
|---|---|---|---|---|
| | | Einstein | Ref. 1 | This work |
| $D^t$, $(cm^2 \cdot s^{-1}) \times 10^6$ | 1.11 ± 0.05 | 1.33 | 1.33 | 1.17 ± 0.04 |
| $D^r$, $s^{-1} \times 10^{-7}$ | 2.6 ± 0.7 | 3.8 | 3.55 | 2.8 ± 0.1 |

prediction by Venable and Pastor (1), using a method developed by de la Torre and Bloomfield (18). Ref. 1 results are given for the case of no assumed bound water molecules. As the number of assumed bound water molecules increases, the mobility of the molecule·water complex decreases, along with the predicted diffusion coefficient.

Results for tobacco mosaic virus are in similar good agreement with experimental values [computed translational diffusion coefficient of $(4.4 \pm 0.1) \times 10^{-8}$ $cm^2 \cdot s^{-1}$ vs. $(4.4 \pm 0.4) \times 10^{-8}$ $cm^2 \cdot s^{-1}$ experimental, and computed rotational diffusion coefficient of 335 ± 10 $s^{-1}$ vs. 333 ± 20 $s^{-1}$ experimental]. The roughly cylindrical shape was divided into an $n$-sided prism, and $n$ was increased. Consistent results occurred at about $n = 16$. The rotational diffusion coefficient given by Brenner's formulas can be considered an average value for rotations about any three orthogonal axes, and it well represents molecules such as lysozyme, where all axes are approximately equivalent. For tobacco mosaic virus, this is clearly not the case, and it was necessary to consider the individual components of the **c** tensor in the directions that diffusion was measured (orthogonal to the centerline) (17, 19).

In his prediction of the size of tobacco mosaic virus from its diffusion coefficient, Broersma (19, 20) has also achieved accurate results by using a singularity density distributed on the centerline of a cylinder.

For rotations about the centerline of the virus, we cannot compare experimental and predicted results, but we can compare with an exact solution. In ref. 21, a dimensionless torque of 270 is calculated for axial rotation of a cylinder of aspect ratio 10. We calculate a value of 280. Our estimate could be improved by finer discretization where the surface tractions are high (the edges of the cylinder), but for the purpose of calculating diffusion coefficients this refinement is not justified.

In summary, we have proposed a method for calculating the mobility tensors which determine the diffusion coefficients of a protein. This method is applicable to a wide range of proteins, as evidenced by the examples chosen. It needs information only about the surface shape of a protein. To facilitate use of this method, we have made our codes available for anonymous file transfer on the internet node doug.cae.wisc.edu, in the directory pub/microhydro.

This work was supported in part by grants from the Office of Naval Research and the National Science Foundation.

1. Venable, R. M. & Pastor, R. M. (1988) *Biopolymers* **82,** 1001–1014.
2. Tin, M. T. & Gusek, T. W. (1990) *Biotechnol. Bioeng.* **35,** 327–338.
3. Brenner, H. (1965) *J. Colloid Sci.* **20,** 104–122.
4. Brenner, H. (1967) *J. Colloid Interface Sci.* **23,** 407–436.
5. Lorentz, H. A. (1896) *Versl. Gewone Vergad. Afd. Natuurkd. K. Ned. Akad. Wet.* **5,** 168–174.
6. Power, H. & Miranda, G. (1987) *SIAM J. Appl. Math.* **47,** 689–698.
7. Kim, S. & Karrila, S. (1991) *Microhydrodynamics: Principles and Selected Applications* (Butterworth-Heinemann, Boston), pp. 117 and 355–493.
8. Friedman, A. (1982) *Foundations of Modern Analysis* (Dover, New York).

9. Youngren, G. K. & Acrivos, A. (1975) *J. Fluid Mech.* **69,** 377–403.
10. Wegener, W. A. (1986) *Biopolymers* **25,** 627–637.
11. Pakdel, P. & Kim, S. (1991) *J. Rheol.* **35,** 797–823.
12. Dabros, T. (1985) *J. Fluid Mech.* **156,** 1–21.
13. Connolly, M. L. (1987) *Visual Comput.* **3,** 72–81.
14. Connolly, M. L. (1985) *J. Appl. Cryst.* **18,** 499–505.
15. Wegener, W. A. (1985) *Macromolecules* **18,** 2522–2530.
16. Young, M. E., Carroad, P. A. & Bell, R. L. (1980) *Biotechnol. Bioeng.* **22,** 947–955.
17. Kuntz, I. D. & Kauzmann, W. (1974) *Adv. Protein Chem.* **28,** 236–345.
18. de la Torre, J. G. & Bloomfield, V. A. (1981) *Q. Rev. Biophys.* **14,** 81–139.
19. Broersma, S. (1960) *J. Chem. Phys.* **32,** 1626–1631.
20. Broersma, S. (1960) *J. Chem. Phys.* **32,** 1632–1635.
21. Chan, P. C., Leu, R. J. & Zargar, N. H. (1986) *Chem. Eng. Commun.* **49,** 145–163.
22. Wegener, W. A. (1981) *Biopolymers* **20,** 303–326.