Microhydrodynamics Simulation of Protein Crystallization I. Static Calculations

J. T. W. M. Tissen, J. Drenth, H. J. C. Berendsen, and J. G. E. M. Fraaije
BIOSON Research Institute and Laboratory of Biophysical Chemistry University of Groningen, 9747 AG Groningen, The Netherlands

ABSTRACT A computer simulation method is proposed to study the effects of hydrodynamic interactions on protein crystallization. It is a combination of Stokesian dynamics and continuum hydrodynamics and is referred to as "microhydrodynamics." The method is checked against analytical expressions for Stokes drag and diffusion coefficients for unit spheres. For a number of protein molecules the diffusion coefficients have been calculated and compared with experimental values. It is shown that the method works well for stationary calculations. Using dynamical calculations interacting protein molecules will be simulated to study the events in the early stages of protein crystallization.

INTRODUCTION

In the process of determining the three-dimensional structure of protein molecules at the atomic resolution, protein crystallization is still believed to be the bottleneck. With the objective to develop physico-chemical models for nucleation, crystal growth, and related processes, (combinations of) different experimental and theoretical approaches are being used (Recently, the Dutch Protein Crystal Growth Interest Group was founded). In a series of papers we will treat one of these approaches, the computer simulation of hydrodynamic interactions, in more detail.

As protein crystallization is very sensitive to the conditions of the crystallizing environment it is very important to optimize pH, ionic strength, temperature, supersaturation, precipitant concentration, etc. One of the processes that affect crystallization is the freezing in of defects. Due to the relatively large free-energy effects (few tens of kT) that occur when a protein molecule (or building block) enters the crystalline phase the molecule has only very little time to adjust its orientation following the first contact with the crystal surface. Therefore, the probability of the occurrence of some kind of mismatch is relatively high. These defects are incorporated into the growing crystal, and this has serious implications for the quality as well as for the final size of the crystal. Incorporation of this type of defect (self-poisoning) can block the movement of steps or kinks on the crystal surface or can disrupt the intermolecular bonding networks, resulting in a cessation of crystal growth (see, e.g., Clydesdale et al. (1994)). In the search for the optimal crystallization conditions the ability of the protein molecules to reorient and to redissolve may be very important to minimize the number of defects, and the crystallization environment must be tuned so as to enhance this. In this respect steering effects must also be studied. These can be induced by long-range interactions. There are two types of these interactions: conservative, e.g., classical, electrostatic, van der Waals, etc.; and nonconservative or dissipative hydrodynamic interactions. Hydrodynamic interactions can play an important role in steering the incoming molecule, even more important than conservative electrostatic interactions (Brune and Kim, 1994).

To help analyze the data coming from the large number of crystallization experiments and improving understanding of the molecular processes taking place during the crystallization, computer simulations can be of great use. The value of a computer simulation is highly dependent on the choice of the model for the intermolecular interactions. For protein molecules in solution the electrostatic forces, van der Waals forces, external forces such as gravity, other systematic forces (if present), and also hydrodynamic interactions must be properly described. The hydrodynamic interactions are the only ones that are really long-ranged in the crystallizing environment. Their range is comparable to the size of the molecules, $\mathcal{O}$(nm). The electrostatic forces are much more short-ranged, $\mathcal{O}(\text{Å})$, as the salt content of the solution has a profound screening effect. The hydrodynamic interactions are distance dependent as well as shape dependent. It is therefore of great importance to use a model that properly takes these two dependencies into account. Here, the full mobility matrix is used to model the hydrodynamic interactions. Usage of simplified models could lead to neglect of physically important effects such as the earlier mentioned steering effect (Brune and Kim, 1994). In the simulation system we discriminate between the protein molecules and the solvent. A very important simplification is that we treat the protein molecules as being rigid and average the solvent effects. Thus, Newton's equations of motion are replaced by Langevin's equations of motion for the protein particles and by the Stokes equations for the solvent. This means that the total force on a protein molecule in solution is given by the sum of the hydrodynamic force, the systematic force, and the random force acting on it. The total system under consideration, protein molecules in a solvent, can be modeled by a method that is called "microhydrodynamics" (Kim and Karrila, 1991). By letting two, three, or more protein molecules interact, thereby sliding and rolling over each other, nucleation kinetics and thermodynamics can be studied. Furthermore, by applying the method to the attachment of building blocks to the crystal surface, crystal growth can be simulated. Thus, the microhydrodynamics technique can be used

to gain information on the physico-chemical aspects of the early stages of protein crystallization.

In this paper we describe the principle of the method and some first results for diffusion coefficients of individual protein molecules. In a following paper the dynamics of interacting molecules will be published. In the next section the simulation method will be described, followed by the results of static calculations for model systems and some proteins. In the final section the conclusions of the present study are drawn and it is explained how we will use dynamical simulations to study nucleation and crystal growth.

THEORY

In Fig. 1 the problem that is to be solved is schematically represented. Find the translational and rotational velocity of a protein molecule (b) resulting from the systematic force ($\mathbf{f}_s$) and the hydrodynamic force ($\mathbf{f}_h$) it experiences from interactions with another molecule (a) moving with velocity $\mathbf{v}$. The vectors are six-dimensional: 3 degrees of freedom for the translation and 3 degrees for the rotation. For the solvent a continuum hydrodynamics description is used: the Navier-Stokes equation. In case of a viscous, incompressible fluid this simplifies to the creeping flow equation

$$\nabla^2\mathbf{u} = \frac{1}{\eta}\nabla p \tag{1}$$

together with the continuity equation $\nabla \cdot \mathbf{u} = 0$, where $\mathbf{u}$ is the fluid velocity, η the fluid viscosity, and p the hydrostatic pressure.

The movements of the protein molecules are described by Langevin's equations of motion. For low Reynolds number the particle inertial force may be neglected as compared with the viscous force. This results in a balance between the systematic force and the hydrodynamic force. This leads to solving for

$$\frac{\partial \mathbf{r}}{\partial t} = -\frac{1}{\eta}\,\mathbf{M}\mathbf{f}_s + \boldsymbol{\mu} \tag{2}$$

Here, $\mathbf{M}$ is the mobility matrix and this is the inverse of $\mathbf{R}$, the resistance matrix. $\mathbf{r}$ is a six-dimensional generalized coordinate vector including rotations and $\boldsymbol{\mu}$ the stochastic velocity component.

To solve the mobility problem (given the forces, to find the resulting velocities) an expression for the mobility matrix $\mathbf{M}$ must be found. This can be achieved by applying the microhydrodynamics method.

In the next section the theory of the calculation method is briefly described. For a rigorous derivation of the algorithms the reader is referred to the excellent book by Kim and Karrila (1991). The simulation method is based on the so-called completed double-layer boundary integral equation method. This gives a complete description of the system (protein molecules in solution) and can be used to solve the mobility problem for the protein molecules for a large number of time steps, thereby yielding the trajectories of the proteins.

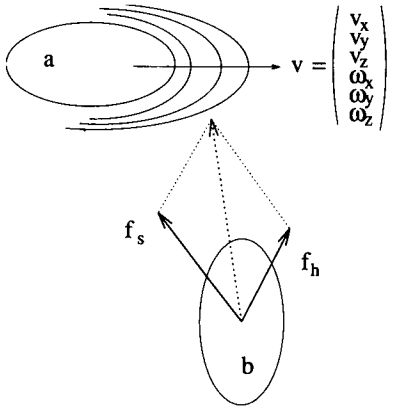

FIGURE 1 Particle b acted on by a six-dimensional systematic force $\mathbf{f}_s$ and a hydrodynamic force $\mathbf{f}_h$ caused by the flowfield induced by particle a, moving with a six-dimensional velocity $\mathbf{v}$.

Completed double-layer boundary integral equation method

The movements of the solvent molecules (Eq. 1) and the protein molecules (Eq. 2) are coupled by the stick boundary conditions through which it is assumed that the relative tangential velocity component of the fluid in contact with the rigid particle's surface is taken as 0. By satisfying the boundary conditions on discrete surface elements on the protein molecules a complete description of the system can be derived. This can be done by applying the boundary integral equation method (see, e.g., Weinbaum et al. (1990)).

The original, exact, solution is given by a sum of a single-layer potential and a double-layer potential (Ladyzhenskaya, 1969). These names come from the analogous problem of finding the potential caused by a charged particle as this can be done by mapping point charges (the single layer) on the surface as well as dipoles (the double layer). This procedure, however, can only solve the mobility problem via inversion of the resistance matrix (Youngren and Acrivos, 1975). Because this inversion procedure is very time consuming and has to be carried out each simulation time step, a method to directly solve the mobility problem was derived (Power and Miranda, 1987; Karrila and Kim, 1989; Karrila et al., 1989).

In this method the crucial step is to find the so-called Stokes double-layer density φ. By omitting the single layer and mapping only the Stokes double-layer (with some completion terms added) on the particle's surface (S), the double-layer density (φ) on each surface element can be determined by satisfying the stick boundary conditions on it. For the description of the molecular surfaces we use planar triangles (Juffer et al., 1991).

The final form of the completed double-layer boundary integral equation method for the mobility problems is

$$\varphi_j(\boldsymbol{\xi}) + (\mathcal{K}\varphi)_j(\boldsymbol{\xi}) + \varphi_j^n\langle\varphi^n, \varphi\rangle = -u_j^\infty + \left(f_i^r - \frac{1}{2}(\mathbf{f}^r \times \nabla)_i\right)\frac{\mathcal{G}_{ji}(\boldsymbol{\xi}, \mathbf{x}_c)}{8\pi\eta}, \quad \boldsymbol{\xi} \in S \tag{3}$$

Here, φ is the Stokes double-layer density, i, j denote the direction, $\boldsymbol{\xi}$ is a position at the center of a surface element, $\mathcal{K}$ is the double-layer operator, φ^n are null solutions, n corresponds to the null solution ($n = 1, \cdots, 6$), the first three null functions are proportional to the unit coordinate vectors, while the other three are (scaled) rotations about the coordinate axes, $\mathbf{u}^\infty$ is the velocity of the undisturbed fluid, $\mathbf{f}^t$ is the translational part of the force acting on the particle, $\mathbf{f}^r$ is the torque acting on the particle, $\mathcal{G}_{ji}$ is the singular solution, and $\mathbf{x}_c$ the position inside the particle on which the force/torque acts. The set of equations that is to be solved for φ can be written as

$$\mathbf{A}\varphi = \mathbf{b} \tag{4}$$

With knowledge of the molecular coordinates and orientation plus the description of the surface elements (left side) and with knowledge of the forces and torques acting on the particle (right side), Eq. 4 can be iteratively solved to give φ. From integration over the surface elements the velocity of the particle can then be calculated

$$\mathbf{v} = -\sum_{n=1}^{6}\int_S \varphi^n \cdot \varphi \, dS \tag{5}$$

Schematically, the simulation procedure is shown in Fig. 2.

First, a structure of the protein must be extracted from the Brookhaven Protein Data Bank or from previous (MD-) simulations (step 1). Then, the molecular surfaces must be discretized. Here, we use a triangulation procedure (step 2). From the molecular positions and orientations the left side of Eq. 4 can be constructed (step 4). Via summation over all atom pairs the intermolecular interaction energies can be calculated and from this the

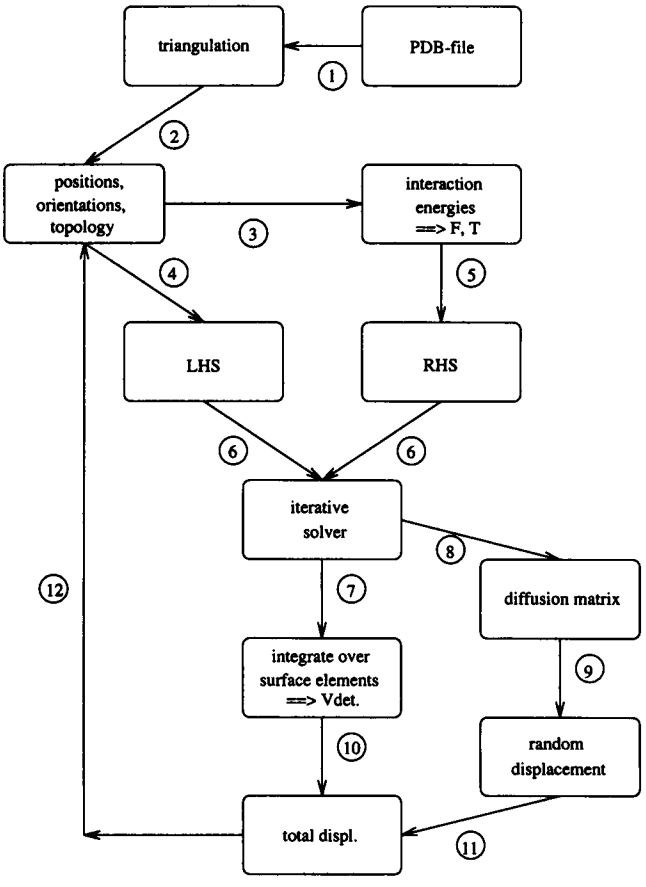

FIGURE 2 Flowchart of the simulation program.

forces and torques on the particles (step 3). These can be inserted in the right side of Eq. 4 (step 5). The set of equations is iteratively solved by a stabilized biconjugate gradient routine (da Cunha and Hopkins, 1993) (step 6). With the calculated double-layer density φ (step 7) the resulting velocities can be computed (step 10). After calculating the diffusion matrix (8) the random displacements can be computed (step 9) and added to the deterministic displacements (step 11). The positions and orientations of the particles can be updated (step 12) and this process is repeated for a large number of simulation time steps. Compared with the original CDL-BIEM program of Kim and co-workers we have made the following changes: 1) no preprocessor is needed to provide the program with an initial guess for the Stokes double-layer density, 2) the deflation terms of the eigenspace at +1 (see Kim and Karrila (1991)) have been omitted from the algorithm, 3) a different method to discretize the molecular surfaces is used, and 4) a stabilized biconjugate gradient routine (da Cunha and Hopkins, 1993) iteratively solves for the double-layer density.

TABLE 1 Calculated properties compared with analytical expressions (Eq. 6) for unit spheres

Expression	Number of surface elements			
	Exact	60	240	960
A [nm^2]	12.57	11.40	12.25	12.49
I	8.38	6.47	7.84	8.24
$\mathbf{u}/\mathbf{f}$ ($\times 6\pi\eta a$)	1.00	1.04	0.98	0.98
D^t [10^{-6} cm^2s^{-1}]	2.14	2.22	2.11	2.10
D^r [10^8 s^{-1}]	1.61	1.77	1.57	1.56

RESULTS

The properties that were compared with the exact solutions for unit spheres (radius 1 nm) are the surface area (A), moments of inertia normalized for unit radius and unit mass (I), Stokes drag ($\mathbf{f}$), and translational (D^t) and rotational diffusion coefficients (D^r):

$$A = 4\pi a^2 \qquad I = \int_S (x^2 + y^2)\, dS = \frac{8\pi}{3}$$
$$\mathbf{f} = 6\pi\eta a u \qquad D^t = \frac{kT}{6\pi\eta a} \qquad D^r = \frac{kT}{8\pi\eta a^3} \tag{6}$$

where a is the radius of the sphere, T is 293.15 K and η is 0.010019 poise (viscosity of pure water). Results are shown in Table 1 for spheres described by 60, 240, and 960 triangular surface elements. The effect of applying a tolerance smaller than 1.0 e − 3 in the iteration procedure is negligible, except for the 60 elements sphere. The calculation of the velocity vector $\mathbf{u}$ from an applied force is straightforward and gives better results for the better-triangulated spheres. The diffusion coefficients were calculated via the following (Brune and Kim, 1993):

$$D^t = \frac{kT}{3\eta}\mathrm{Tr}\,\mathbf{a} \qquad D^r = \frac{kT}{3\eta}\mathrm{Tr}\,\mathbf{c} \tag{7}$$

where $\mathbf{a}$ and $\mathbf{c}$ are the upper left and lower right parts of the mobility matrix, i.e., the translation-translation and rotation-rotation parts.

From the results shown in Table 1 it is concluded that for spheres it is sufficient to use 240 elements and a tolerance of 1.0 e − 3. More elements do not improve the results very much. The triangulation procedure as we employ it inevitably yields too-low moments of inertia. It is, of course, possible to adjust the triangulation such that the moments of inertia are exact for spheres, but for particles with complex shapes such as proteins this would be much more difficult. We therefore decided to ignore this effect.

For the infinite dilution limit we calculated the translational diffusion coefficients for some proteins at 20°C. The structures were taken from the Brookhaven Protein Data Bank. We have used 960 elements to describe the molecular surfaces. One of the triangulated proteins is shown in Fig. 3.

From the results in Table 2 it is clear that the static microhydrodynamics calculations reproduce the diffusion coefficients fairly well, although most of the calculated values are smaller than the experimental ones. There are different interplaying effects that may explain these differences.

1. A diffusing protein molecule carries a layer of water with it. The molecules as we model them do not have this water layer and thus have a volume that is too small. The calculated diffusion coefficient should therefore come out too high.

2. The presence of clefts in a protein enlarges the molecular surface area and hence decreases the diffusion. In reality these clefts are likely to be filled with immobilized

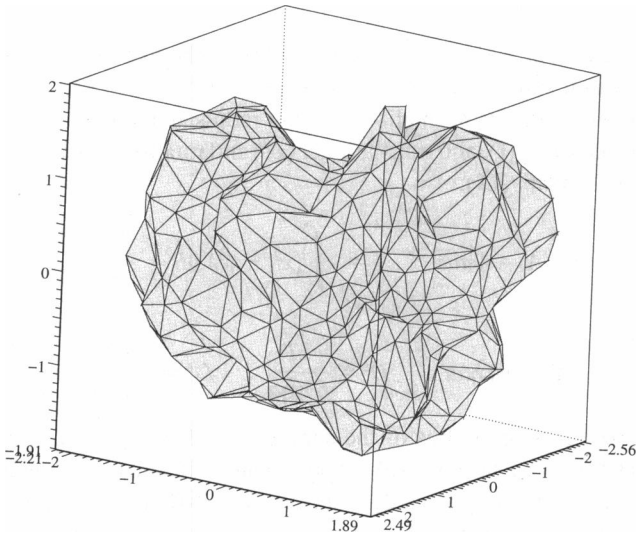

FIGURE 3 Ribonuclease molecule, triangulated with 960 surface elements. Axes denote coordinates in nanometers.

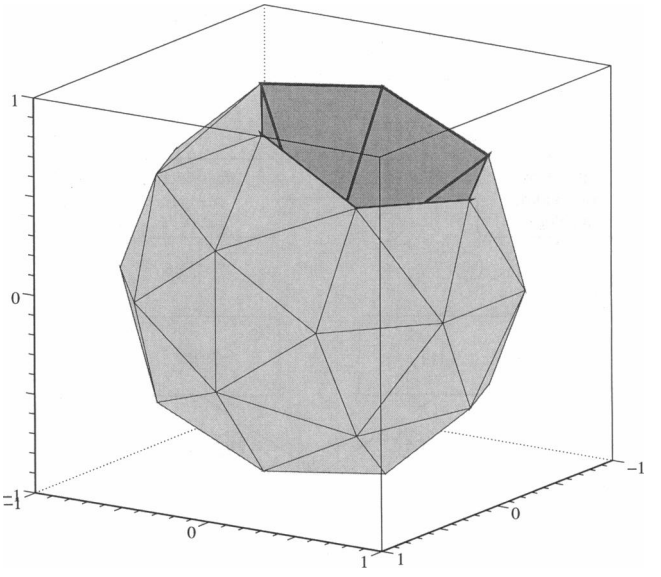

FIGURE 4 Unit sphere with a cleft (thick lines) in the upper part. Surface is described with 60 triangles.

TABLE 2 Calculated translational diffusion coefficients for proteins compared with experimental values (infinite dilution, 20°C)

Protein	D_{exp} ($\cdot 10^{-7}$ cm^{2}s^{-1})	D_{calc} ($\cdot 10^{-7}$ cm^{2}s^{-1})
Ribonuclease	11.9*	7.2
Myoglobin	10.4‡	6.5
Lysozyme	10.4*	9.1§
Chymotrypsinogen	9.5‡	6.4
Carboxypeptidase	8.7‡	6.0
Hemoglobin	6.9*	6.7
Lactate dehydrogenase	5.0‡	4.9
Aldolase	4.7‡	5.4

* (Cantor and Schimmel, 1980)

‡ (Tanford, 1967)

§ Brune and Kim (1993) calculated a value of $11.7 \cdot 10^{-7}$ cm^{2}s^{-1}

water and as a result the surface area of the real diffusing molecule (with the water) is smaller than in the simulation model. This surface area effect can readily be calculated for a sphere. Comparing the surface area and the diffusion coefficient for a unit sphere with a cleft (see Fig. 4) and a unit sphere without a cleft gives the following results. Without the cleft: area = 11.4 nm^2, $D = 2.2\ 10^{-6}$cm^{2}s^{-1}; with the cleft: area = 12.1 nm^2, $D = 2.0\ 10^{-6}$cm^{2}s^{-1}. Hence, if the protein molecule has some clefts filled with immobilized water our prediction for the diffusion coefficient would be too low.

3. Other features of the simulation method, such as the fact that we model the protein molecules as being rigid and that we neglect thermal fluctuations in the system, will also influence the outcome of the diffusion coefficient calculations. At this stage it is not quite possible to estimate the directions of these effects.

The calculations on stationary systems are very promising with respect to our plans to use the microhydrodynamics method to study protein crystallization. In effect, work is currently in progress to perform dynamical simulations of lysozyme molecules with systematic interactions plus hydrodynamic interactions; this work will be presented in a future paper.

CONCLUSIONS

From the results presented in the preceding section it is concluded that the microhydrodynamics method works well for static calculations. The method can handle arbitrary shapes and sizes with a full description of the hydrodynamic interactions.

Stimulated by these positive results we are now working on dynamic simulations from which we expect to get a better insight in the processes associated with protein crystallization. The equations for systems of more particles are analogous to those in the Theory section, and the resulting grand mobility matrix is $6N \times 6N$ for N particles. For the electrostatic part of the systematic force between the protein molecules the method of Juffer et al. (1991) can be used. Nucleation can be studied by simulating two (or more) protein molecules rolling and sliding over each other and eventually combining. With a simple description of the crystal surface the attachment of a building block to the crystal can be used to study crystal growth. For this purpose the crystal surface will have to be discretized, and this will pose an extra number of boundary conditions to the computational system.

In general, these simulations can yield information on the kinetics of the processes and on various energy terms, and may even reveal the energy barriers that the particles have to overcome before attaching. Also, the range and contributions of the different types of forces can be determined. With respect to steering effects and orienting effects this is very important. Considering the very limited time that a molecule has to reorient, the effect of an applied flow field (stirring, laminar flow, etc.) is very interesting. Because the set

of equations used in the computational scheme is linear, these external flow field terms can simply be added to the other forces. It may even be possible to check whether the absence of gravity enhances crystallization or not. The simulation results can then be compared with microgravity experiments. The effect of changes in the temperature can be studied via the random displacement terms, and a change in salt contents of the solvent affects the ionic strength and thus has an influence on the screening of the electrostatic interactions.

Using the microhydrodynamics technique the crystallizing environment can easily be changed, and the effects on protein crystallization can systematically be studied.

This work is sponsored by the Dutch Space Research Agency (SRON).

REFERENCES

Brune, D., and S. Kim. 1993. Predicting protein diffusion coefficients. *Proc. Natl. Acad. Sci. USA.* 90:3835–3839.

Brune, D., and S. Kim. 1994. Hydrodynamic steering effects in protein association. *Proc. Natl. Acad. Sci. USA.* 91:2930–2934.

Cantor, C., and P. Schimmel. 1980. Biophysical Chemistry. W. H. Freeman and Company, San Francisco.

Clydesdale, G., K. Roberts, and R. Docherty. 1994. Modelling the morphology of molecular crystals in the presence of disruptive tailor-made additives. *J. Cryst. Growth.* 135:331–340.

da Cunha, R., and T. Hopkins. 1993. PIM 1.0, Parallel Iterative Methods package for Systems of Linear Equations. Centro de Processamento de Dados, Universidade Federal do Rio Grande do Sul, Brasil and Computing Laboratory, University of Kent at Canterbury, United Kingdom.

Juffer, A. H., E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg, and H. J. C. Berendsen. 1991. The electric potential of a macromolecule in a solvent: a fundamental approach. *J. Comput. Phys.* 97:144–171.

Karrila, S., Y. Fuentes, and S. Kim. 1989. Parallel computational strategies for hydrodynamic interactions between rigid particles of arbitrary shape in a viscous fluid. *J. Rheology.* 33:913–947.

Karrila, S., and S. Kim. 1989. Integral equations of the second kind for stokes flow: direct solution for physical variables and removal of inherent accuracy limitations. *Chem. Eng. Commun.* 82:123–161.

Kim, S., and S. Karrila. 1991. Microhydrodynamics: Principles and Selected Applications. Butterworth-Heineman, Boston.

Ladyzhenskaya, O. A. 1969. The Mathematical Theory of Viscous Incompressible Flow. Gordon and Breach, New York.

Power, H., and G. Miranda. 1987. Second kind integral equation formulation of stokes flow past a particle of arbitrary shape. *SIAM J. Appl. Math.* 47:689–698.

Tanford, C. 1967. Physical Chemistry of Macromolecules. John Wiley, New York.

Weinbaum, S., P. Ganatos, and Z.-Y. Yan. 1990. Numerical multipole and boundary integral equation techniques in stokes flow. *Annu. Rev. Fluid Mech.* 22:275–316.

Youngren, G. K., and A. Acrivos. 1975. Stokes flow past a particle of arbitrary shape: a numerical method of solution. *J. Fluid Mech.* 69: 377–403.