

Molecular simulation of protein dynamics in nanopores. II. Diffusion

Leili Javidpour,¹ M. Reza Rahimi Tabar,^{1,2} and Muhammad Sahimi^{3,a)}

¹Department of Physics, Sharif University of Technology, Tehran 11155-9161, Iran

²Institute of Physics, Carl von Ossietzky University, Oldenburg D-26111, Germany

³Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, California 90089-1211, USA

(Received 28 November 2008; accepted 13 January 2009; published online 27 February 2009)

A novel combination of discontinuous molecular dynamics and the Langevin equation, together with an intermediate-resolution model of proteins, is used to carry out long (several microsecond) simulations in order to study transport of proteins in nanopores. We simulated single-domain proteins with the α -helical native structure. Both attractive and repulsive interaction potentials between the proteins and the pores' walls are considered. The diffusivity D of the proteins is computed not only under the bulk conditions but also as a function of their "length" (the number of the amino-acid groups), temperature T , pore size, and interaction potentials with the walls. Compared with the experimental data, the computed diffusivities under the bulk conditions are of the correct order of magnitude. The diffusivities both in the bulk and in the pores follow a power law in the length ℓ of the proteins and are larger in pores with repulsive walls. D^+/D^- , the ratio of the diffusivities in pores with attractive and repulsive walls, exhibits two local maxima in its dependence on the pore size h , which are attributed to the pore sizes and protein configurations that induce long-lasting *simultaneous* interactions with *both* walls of the pores. Far from the folding temperature T_f , D increases about linearly with T , but due to the thermal fluctuations and their effect on the proteins' structure near T_f , the dependence of D on T in this region is nonlinear. We propose a novel and general "phase diagram," consisting of four regions, that describes qualitatively the effect of h , T , and interaction potentials with the walls on the diffusivity D of a protein. © 2009 American Institute of Physics. [DOI: 10.1063/1.3080770]

I. INTRODUCTION

In Part I of this series¹ we studied folding and stability of proteins in slitlike nanopores using discontinuous molecular dynamics (DMD) simulations. We considered the effect of a variety of factors on protein folding and stability, including the pore and protein sizes, the nature of the interaction between proteins and the pore walls (repulsive as well as attractive), and the temperature. In the present paper we study diffusion of proteins in the same slitlike nanopores, investigate the effect on transport of proteins of the same factors that affected their folding and stability, and link the results with those presented in Part I.

While the three-dimensional (3D) structure of native proteins is controlled by their amino-acid sequence,²⁻⁴ the kinetics of their folding, as well as the rate of their transport, especially in confined media, depend on the local environment. Part I studied the folding and stability of proteins in a slit nanopore. However, although diffusion of proteins in dilute solutions under bulk condition is relatively well understood, the important phenomenon of protein transport in a confined medium is not. Such understanding is essential to important industrial applications that involve protein transport in small pores. For example, applications of biocatalysts⁵ and biosensors entail much better understanding of the confinement effect on the rate of protein transport.

Moreover, protein purification using nanoporous membranes is also attracting attention in the pharmaceutical industry.⁶ Rosenbloom *et al.*⁷ fabricated silicone-carbide (SiC) nanoporous membranes⁸ and showed them to allow diffusion of proteins up to 29 000 Da but not larger.

Aside from the aforementioned industrial applications, transport of proteins across nanoporous *biological* membranes is also of fundamental importance to life processes. If a protein molecule is large but the membranes' pores are small enough that the protein is not allowed to pass through as a single unit, one has the phenomenon of *translocation*—i.e., squeezing—of the protein through the pores.⁹ Translocation is important to gene therapy,¹⁰ drug delivery,¹¹ passage of proteins through the endoplasmic reticulum¹² and of RNA through the nucleus pore membrane, DNA plasmid transport from cell to cell through the walls,⁹ and transport of the polypeptide chain from inner mitochondrial and chloroplast membrane through its matrix.¹³

An important and experimentally accessible¹⁴ quantity in such studies is¹⁵ the *dwelt time*—the typical time that a protein, or any other biopolymer such as DNA or RNA, spends in the nanopore and its dependence on the molecular weight, the pore's length, and other parameters. In general, one has two dynamics: in *slow translocation* the protein remains equilibrated at almost all the times on *both sides* of the pore, whereas *fast translocation* allows the protein to pass through the pore. Several mechanisms may induce the translocation.¹³ For example, one is based on an external

^{a)}Electronic mail: moe@iran.usc.edu.

electric field applied across the membrane, which has been studied by *in vitro* experiments, as well as theoretically and by Monte Carlo simulations.^{16–19} Another mechanism uses a chemical-potential difference across the membrane, as there is usually no *in vivo* strong electric field in such membranes,²⁰ an example of which is experiments on chaperone-assisted translocation of proteins.^{9,13,21}

The protein transport in nanopores that we study in this paper is different from translocation in that, in our study there is no electric field or external potential gradient applied to the pore system. Thus, transport of the proteins takes place by diffusion only. The present study represents the first step toward a comprehensive study of transport of proteins in tight pores in the presence of an external field, which is currently under study by our group. We report on the simulation of protein diffusion in nanopores and study the effect of the pore size, temperature, proteins' length, and nature of the interaction between proteins and the pore walls on the diffusion process. Although, as pointed out in Part I, limited Monte Carlo and MD simulations of protein dynamics in various models of nanopores have been carried out before, to our knowledge MD simulation of diffusion of proteins in nanopores has not been undertaken before. To carry out the simulation, we propose a new method for taking into account the effect of the interaction between proteins and the solvent in the pore.

Diffusion of biopolymers in confined media has been studied extensively^{22–30} in the past. In particular, transport of such biopolymers such as RNA and DNA has been studied extensively.^{28–30} Simulation studies of this kind usually assume the biopolymers to be flexible chains without any specific atomistic structure or assume the biopolymers to have a rigid structure without considering their atomistic details. Such details do, however, affect the interaction of a biopolymer with a pore's walls. In particular, the atomistic details of proteins' structure are important in computing their diffusivity in confined media at temperatures near the folding temperature T_f , where the proteins' structure is neither completely rigid nor completely flexible. In fact, as described in Part I, at such temperatures the interaction of proteins' atoms with the pores' walls is very important, as it affects the folding dynamics, as well as the folding temperature itself.

Some simulation works on diffusion of proteins (or proteinlike polymers with a compact structure at low temperatures) in confined media have been reported,³¹ which did include some details of the protein structure, with the aim of realistically capturing the effect of the interactions with the walls on the protein's diffusivity. However, including the details of the atomistic structure of a protein is only half the "story." The other half is the need for including the effect of the solvent. Doing so realistically in any molecular simulations would entail very intensive computations, and even then only time scales on the order of nanoseconds may be realized, unless one uses massively parallel computations.³²

The rest of this paper is organized as follows. In the next section we describe briefly the protein and pore models that we utilize. Section III presents the details of the MD simulations, while in Sec. IV we describe a new method, based on a coupling between the DMD simulation and the Langevin

equation, that aims to take into account the effect of the solvent on the protein diffusion. The results are described and discussed in Sec. V.

II. THE MODELS

We first describe the protein model that we use in the study and the model of the nanopore that we utilize in the simulation.

A. The protein model

A model of *de novo*-designed α family of proteins³³ that consists of four types of amino acids in their 16-residue sequence is utilized in the simulation. The model is simplified further³⁴ to a sequence of hydrophobic (HP) (*H*) and polar (*P*) residues, $\{PPHPPHPPHPPHPPHPP\}$. Periodicity in the *H-P* sequence of the 16-residue peptide α_{1B} is used in order to make three other sequences with lengths $\ell=9, 23$, and 30 residues as $PP(HPPHPP)_n$, with $n=1, 2, 3$, and 4, corresponding to $\ell=9, 16, 23$, and 30. The simulations reported in Part I indicated that they all fold into an α -helix with $\ell-4$ hydrogen bonds (HBs). Because the four proteins have similar native structures, any differences in their diffusion dynamics should be solely due to the differences in their lengths.

The proteins are represented by the protein intermediate-resolution model^{35,36} (PRIME), with several changes that are described below (see also Part I). It is an off-lattice, unbiased, intermediate-resolution model in which every amino acid is represented by four united atom (UA) groups or beads. A nitrogen UA represents the amide N and hydrogen of an amino acid, a C_α UA represents the α -C and its H, and a C UA represents the carbonyl C and O. The fourth bead *R* represents the side chain, all of which are assumed to have the same diameter as CH_3 (alanine). The interpeptide bond is assumed to be in the trans configurations, all the backbone bond lengths and bond angles are fixed at their ideal values, and the distance between consecutive C_α UAs is fixed according to experimental data. Table I presents all the relevant parameters of the model. PRIME is capable of interacting both intra- and intermolecularly via HB and HP interaction potentials.

The protein model that we use is, in our opinion, much more realistic than what many of the previous investigators utilized. For example, they used a simplified model for the amino acids that was based on one or two UA beads. Moreover, the side chains of the amino acid residues were not explicitly considered. The model that we utilize represents the amino acids using four UA beads, the side chains are considered explicitly, and the molecular model also includes HB interactions, hence honoring the proteins' structure realistically.

We note that other proteins may have a more complex structure than what we model in the present paper. In particular, they may have β -strands, sheets, loops, and tight turns. Whether the results of the present study are applicable to more complex proteins remain to be seen, although we expect many of them to be quite general because, as we

TABLE I. Values of the potential parameters for the proteins used in the DMD simulations.

Beads	Diameter (Å)
N	3.300
C_α	3.700
C	4.000
R	4.408
Bonds	Length (Å)
$N_i-C_{\alpha,i}$	1.460
$C_{\alpha,i}-C_i$	1.510
C_i-N_{i+1}	1.330
$C_{\alpha,i}-R_i$	1.531
Pseudobonds	Length (Å)
N_i-C_i	2.45
$C_{\alpha,i}-N_{i+1}$	2.41
$C_i-C_{\alpha,i+1}$	2.45
$C_{\alpha,i}-C_{\alpha,i+1}$	3.80
N_i-R_i	2.44
C_i-R_i	2.49
Bonds	Bond angles (deg)
$\angle N_i-C_{\alpha,i}-C_i$	111.0
$\angle C_{\alpha,i}-C_i-N_{i+1}$	116.0
$\angle C_i-N_{i+1}-C_{\alpha,i+1}$	122.0
$\angle R_i-C_{\alpha,i}-N_i$	109.6
$\angle R_i-C_{\alpha,i}-C_i$	110.1

mentioned above, the 3D structure of native proteins is controlled mainly by their amino-acid sequence, which the model represents carefully and accurately.

B. The nanopore model

We use a slit nanopore, modeled as the space between two two-dimensional (2D) structureless walls in the xy plane between $z = \pm h/2$, where h is the pore's height (size). Periodic boundary conditions are used in the x and y directions (the xy planes are parallel to the walls). The pore size h is varied in order to study its effect on the results described below.

III. DISCONTINUOUS MOLECULAR DYNAMICS SIMULATOIN

The PRIME is designed for use with the DMD simulation,³⁷ a very fast alternative to the classical continuous MD method. The simplicity of the time integration in the DMD simulation makes it possible to simulate transport of proteins in nanopores on long time scales—on the order of many microseconds—at least two orders of magnitude longer than the previous simulations in this area (see also Part I).

All the forces are represented by either hard-sphere or square-well potentials. Four types of forces act on the beads: the excluded volume effect (hard-core repulsion) and attraction between bonded and pseudobonded beads, between pairs of the backbone beads during the HB formation, and between HP side chains. Nearest-neighbor beads along the chain backbone are covalently bonded, as are the C_α and R beads or UAs. Pseudobonds are inserted between next-nearest-neighbor beads along the backbone in order to keep its angles fixed, between neighboring pairs of C_α beads to maintain their distances close to the experimental data, and between side chains and the backbone N and C UAs to hold the side-chain beads fixed relative to the backbone. All of these keep the interpeptide group in the trans configuration, and all the model residues as L isomers, as required.

The potential between a pair ij of bonded beads, separated by a distance r_{ij} , is given by

$$U_{ij} = \begin{cases} \infty, & r_{ij} \leq l(1 - \delta), \\ \infty, & r_{ij} \geq l(1 + \delta), \\ 0, & l(1 - \delta) < r_{ij} < l(1 + \delta), \end{cases} \quad (1)$$

where l is the ideal bond length and $\delta = 0.02375$ is the tolerance in the bond's length.^{35,36} There are also the HP interactions between the side chains with the H residues in the sequence, when there are at least three intervening residues between them. Then, the interaction is given by

$$U_{\text{HP}} = \begin{cases} \infty, & r_{ij} \leq \sigma_{\text{HP}}, \\ -\epsilon_{\text{HP}}, & \sigma_{\text{HP}} < r_{ij} \leq 1.5\sigma_{\text{HP}}, \\ 0, & r_{ij} > 1.5\sigma_{\text{HP}}, \end{cases} \quad (2)$$

where σ_{HP} is the HP side chains' diameter.

The HB interaction may occur between the N and C beads with at least three intervening residues. However, each bead may not contribute to more than one HB at any time, with the range of the interaction being about 4.2 Å and strength of ϵ_{HB} . The shape of the HB potential is similar to that of the HP described above. The HBs are stable when the angles in N–H–O and C–O–H are almost 180°. The angles are controlled by a repulsive interaction between each of the N and C beads with the neighboring beads of the other one.³⁷ Thus, if a HB is formed between the beads N_i and C_j , a repulsive interaction between neighbor beads of N_i , namely, C_{i-1} and $C_{\alpha,i}$, with C_j is assumed. The same is used for the neighbor beads of C_j , namely, N_{j+1} and $C_{\alpha,j}$, with the N_i bead.

If one of the N or C bead is at one end of the protein, it has only one neighbor bead in the backbone instead of two, and, hence, controlling the HB angles will be limited, causing the HBs with one of their terminal constituents to be less restricted and, thus, more stable than the other HBs. This may cause formation of the next non- α -helical HBs in a part of the protein between the N and C beads and of semistable structures that influence the simulation results. Thus, we modify the PRIME and proceed as follows.

Assume that the N-terminal bead, N_1 , has a HB with C_j . For $i=1$, the bead C_{i-1} does not exist to have a repulsive interaction with C_j and help control the HB angles. Therefore, we use $C_{\alpha,1}$. Not only can we consider the repulsion

between this bead and C_j but also define an upper limit for their distance, so as to control the freedom of motion of N_1 and C_j that constitute the beads in the HB. The potential U_{kl} of such interactions is given by

$$U_{kl} = \begin{cases} \infty, & r_{kl} \leq \frac{1}{2}(\sigma_k + \sigma_l), \\ \epsilon_{\text{HB}}, & \frac{1}{2}(\sigma_k + \sigma_l) < r_{kl} \leq d_1, \\ 0, & d_1 < r_{kl} \leq d_2, \\ \infty, & r_{kl} > d_2. \end{cases} \quad (3)$$

Two H atoms have chemical bonds with the nitrogen in the protein's N terminal and are free to rotate around the $N_1-C_{\alpha 1}$ bond while satisfying the constraints on the angles between the chemical bonds of N_1 . Thus, if a HB is formed, one of the two H atoms lies in a plane formed by N, O, and C, such that the angles in N–H–O and C–O–H are as close to 180° as possible. Therefore, we force the maximum distance between $C_{\alpha 1}$ and C_j to be the same as d_2 , the maximum distance between $C_{\alpha i}$ and C_j in the usual HBs. This allows us to control the angles in a HB that contains N_1 . We use a similar approach when the C terminal, C_ℓ , has a HB with N_i . The temperature dependence of d_2 , obtained from separate simulations, is given by (T is dimensionless, in units of $\epsilon_{\text{HB}}/k_B T$)

$$d_2 \approx 5.53 - 0.019/T \quad \text{for } N_1 - C_{\alpha j}, \quad (4)$$

$$d_2 \approx 5.69 - 0.044/T \quad \text{for } C_\ell - C_{\alpha i}. \quad (5)$$

Two unbonded beads that have no HB and HP interactions also interact through a hard-core repulsion:

$$U_{\text{HC}} = \begin{cases} \infty, & r_{ij} \leq \frac{1}{2}(\sigma_i + \sigma_j), \\ 0, & r_{ij} > \frac{1}{2}(\sigma_i + \sigma_j). \end{cases} \quad (6)$$

The interactions between a pair of beads, separated along the chain by three or fewer bonds, are more accurately represented by the interaction between the atoms themselves, not the UAs. Consequently, we modify the PRIME to account for the interactions between pairs of beads separated by three or fewer bonds: the beads are allowed to overlap by up to 25% of their bead diameters, while for those separated by four bead diameters the allowed overlap is 15% of their bead diameters.

The interaction potential U_{PW} between the walls and the protein's beads is assumed to be

$$U_{\text{PW}} = \begin{cases} \infty, & z_X \leq -(h/2 - d_{3X}), \\ -\epsilon_{\text{PW}}, & -(h/2 - d_{3X}) < z_X \leq -(h/2 - d_{3X} - d_{4X}), \\ 0, & -(h/2 - d_{3X} - d_{4X}) < z_X < h/2 - d_{3X} - d_{4X}, \\ -\epsilon_{\text{PW}}, & h/2 - d_{3X} - d_{4X} \leq z_X < h/2 - d_{3X}, \\ \infty, & z_X \geq h/2 - d_{3X}, \end{cases} \quad (7)$$

where z_X is the z coordinate of the center of a bead X and the walls. ϵ_{PW} —the same for all the beads—is assumed to be $\frac{1}{8}\epsilon_{\text{HB}}$, so selected to represent realistically the competition between protein folding and its beads' interaction with the walls. To estimate d_{3X} and d_{4X} , we assumed the pore's walls

to be made of carbon. Then, the interaction and size parameters between the C atoms in the walls and the various beads were calculated using the Lorentz–Berthelot mixing rules, namely, $\sigma_{\text{CX}} = \frac{1}{2}(\sigma_{\text{C}} + \sigma_{\text{X}})$, and $\epsilon_{\text{CX}} = \sqrt{\epsilon_{\text{C}}\epsilon_{\text{X}}}$, where $X = \text{N}, C_{\alpha}, \text{C}, \text{and R}$. Then, using separate simulations, the interaction potential U_{CX} between different beads was estimated. The distances at which U_{CX} and its second derivative were zero were taken as d_{3X} and $d_{3X} + d_{4X}$. The results are $d_{3X} = 2.85, 3.02, 3.14, \text{ and } 3.31$ and $d_{4X} = 0.96, 1.01, 0.98, \text{ and } 1.12$ for $X = \text{N}, C_{\alpha}, \text{C}, \text{ and R}$, respectively. More details are given in Part I.

IV. COMPUTING THE DIFFUSIVITY: COUPLING THE LANGEVIN EQUATION TO THE DMD

In the previous applications of the PRIME,^{35,36} the effect of the solvent was only implicitly simulated. In fact, the effect of the solvent was modeled simply as the HP attraction between the HP side chains. However, this is appropriate and more important to the folding dynamics but not so much to the protein motion in a solution. Since we wish to study diffusion of proteins in a solution (in the bulk as well as in small pores), it is essential to simulate correctly the dynamics of protein motion in a solution by including the effect of the solvent. To include this effect, we have developed a novel method which we now describe.

In the PRIME, the protein motion in a solution is only due to the collisions caused by the Andersen thermostat, which is well suited for the DMD simulations. It is not clear, however, that this would be sufficient for generating the correct dynamics of protein diffusion in a solution. If we assume that the Andersen thermostat collisions are equal to those of the solvent with the protein's atoms, then, if the atoms are all exposed to the solvent, so that the protein does not have "buried" parts, i.e., parts that are not exposed directly to the solvent (as is nearly true in the case of a protein with an α -helix configuration), we may hope that the thermostat collisions produce the correct dynamics. However, the time interval between the thermostat collisions is, by itself, an important physical factor. In order to produce the correct physical behavior of the motion, this time interval must be selected carefully, so that the average time intervals between the collisions of the solvent molecules with the protein's atoms are correctly produced. However, once fixed, the thermostat's rate of collisions will be independent of the spatial configuration of a protein and its radius of gyration, R_g . This is not physical and, in general, yields unreliable diffusion coefficients for proteins from simulations that are carried out over a wide range of temperatures below and above the folding temperature T_f , where a protein changes widely its spatial configuration and the radius of gyration, R_g (see below). Thus, in this paper we develop a new method by coupling the Langevin equation to the DMD simulation of a large molecule, such as a protein, with a small computational cost which should yield completely physical results. While a combination of the Langevin equation and atomistic-scale interactions has been used in the past³⁸ in the translocation problem, this is the first time, to our knowledge, that the

Langevin equation is coupled to the DMD simulations for simulating diffusion of biomolecules in a nanopore.

In the new approach that we propose the thermostat collisions, which are used to keep the temperature of the protein equal to that of the solvent, do not move the center of mass (c.m.) of the protein. Instead, using the Langevin dynamics, independent collisions with the c.m. of the protein are made. With the PRIME alone (no explicit solvent), this was not necessary because, in the presence of the Andersen thermostat all the atoms and the c.m. are automatically at a given temperature T . Use of the Langevin equation not only makes the temperature of the protein's c.m. equal to that of the solvent but also, more importantly, allows the effect of the solvent's viscosity η and its molecular collisions with the proteins to be explicitly included in the simulation.

To simulate diffusion of proteins using a combination of the DMD and the Langevin equation, one must, in principle, apply the equation to all the atoms in the proteins. This, however, would require very long simulations, even when one uses the DMD. To circumvent this problem, we break down each protein into several smaller parts or beads and apply the Langevin equation to each part separately. We represent each part as a bead with its radius of gyration to be used as an approximation of its hydrodynamic radius in the Langevin equation. Thus, groups of four, or five, or six amino acids are represented as one coarsened bead in proteins of various lengths, implying that such beads consist of 16, 20, or 24 UAs. Recall that each amino acid in the PRIME is modeled by four UAs. Therefore, a protein of length 9 is assumed to consist of two beads of four and five amino acids; a protein of length 16 has three beads, two with five amino acids and one with six, and a protein of length 23 has four beads, one with five and three with six amino acids. The longest protein, with a length of 30, is assumed to have five beads, each with six amino acids.

Such a coarsened structure still leaves enough structural details of the proteins to affect their interactions with the solvent and, at the same time, a dynamic necklacelike model to trace the effect of the solvent on their motion. We refer to the model as dynamic necklacelike because (i) the size of the coarsened beads is not constant but is determined and updated during the simulations by calculating the radius of gyration of each bead periodically and (ii) because the model is not a necklace or a bead-spring model. The interactions between the atoms determine their positions which, in turn, determine the position as well as the internal arrangement of the coarsened beads. We emphasize that the coarsened beads are used only when we switch from the DMD simulations to the Langevin equation. During the DMD simulations the full atomistic model of the proteins, described in Sec. II, is used.

The coupling between the DMD simulation and the Langevin equation proceeds as follows. We first carry out DMD simulation for a time period Δt . Suppose that the speeds of a bead of protein at the beginning and end of the period Δt are, respectively, \mathbf{v}_b and \mathbf{v}_e . Since the solvent's viscosity affects the proteins' velocity in the pore but the time scale over which this effect is important is much different from Δt , we apply at the end of the time period Δt the Langevin equation to the bead's c.m. to correct its velocity

due to the presence of the solvent's molecules. To do so, we represent the coarsened bead as a particle with a mass m and an effective radius equal to its radius of gyration, R_g . Then, the force \mathbf{F} on its c.m. is given by

$$\mathbf{F} = m(\mathbf{v}_e - \mathbf{v}_b)/\Delta t. \quad (8)$$

The discretized Langevin equation is given by

$$\Delta \mathbf{v} = \mathbf{v}_n - \mathbf{v}_b = \mathbf{F}\Delta t/m - \xi \mathbf{v}_b \Delta t/m + \Delta \mathbf{F}(\Delta t), \quad (9)$$

where $\xi = 6\pi R_g \eta$, $\Delta \mathbf{F}$ is a Gaussian random force (with zero mean and variance $2k_B T \xi \Delta t/m^2$), and \mathbf{v}_n is the speed *after* applying the Langevin equation (acting as the \mathbf{v}_b for the next application of the Langevin equation). Thus,

$$d\mathbf{v} = \mathbf{v}_n - \mathbf{v}_e = -\xi \Delta t \mathbf{v}_b/m + \Delta \mathbf{F}(\Delta t), \quad (10)$$

which yields \mathbf{v}_n , the velocity of the bead's c.m., corrected for the solvent effect. The DMD simulation is then continued for another time period Δt using \mathbf{v}_n as the new \mathbf{v}_b , the Langevin equation is applied again to correct the bead's c.m. velocity at the end of the period, and so on. In the simulations, Δt was taken to be $0.02m/\xi$, which is a small enough time step for integrating the Langevin equation over a time scale m/ξ .

If we use in the Langevin equation the bulk values of all the parameters and, in particular, the solvent's viscosity, we obtain the bulk diffusivity, even if diffusion of the proteins is simulated in a pore. The reason is that a bead or sphere near a wall or between the two walls of a pore experiences higher fluid viscosity than it does under the bulk conditions. There have been many studies of the changes in the viscosity of a fluid confined in a pore or near a solid surface, which contains a bead or particle.^{39–46} In particular, Faxén³⁹ and Brenner⁴⁰ derived expressions for the perpendicular and parallel components of the viscosity of a fluid containing a sphere near a wall, with the underlying assumption being that the fluid follows the Navier–Stokes equations. Recently, it was shown^{42–44} through careful comparison with experimental data that the Brenner–Faxén relations may be used for computing the viscosity of the same fluid between two parallel walls. The effect of two walls may be superposed linearly.⁴³ In the present work, we used this approach to compute the viscosity of the solvent containing the coarsened beads between the pore's walls.

Before doing so, however, one must address an important question: are the Brenner–Faxén relations valid at nanometer scales? Experimentally, this was recently demonstrated to be the case for a fluid containing a sphere of size of 27 nm near a wall.⁴² In very tight pores, however—those that consist of four or fewer layers of a fluid molecules—the usual definition of the viscosity breaks down, and the validity of the Navier–Stokes equations which is based on representing the fluid as a continuum is lost.^{47–49} However, the deviations become smaller^{47–49} as the size of the pore increases from five to six fluid molecules' diameter to larger values. For a slit pore that consists of ten layers of fluid molecules, there is very good agreement between the Navier–Stokes viscosity and what is computed by full MD simulations.^{47–49} In the problem that we study water is the solvent with a molecular diameter of about 0.3 nm. Therefore, roughly speaking, for nanopore sizes of about 3 nm or

larger, the Navier–Stokes viscosity may be assumed to be the correct viscosity, so that the Brenner–Faxén relations may be used with some confidence. Thus, only for the nanopore of size 1.75 nm—the smallest that we simulate—which contains about five layers of water molecules, the approximation may be considered crude. At the very least, we expect the results to be correct qualitatively.

The way the Brenner–Faxén relations are used in the simulations is as follows. Suppose that a coarsened bead, representing a portion of the protein, is a sphere of hydrodynamic radius a , which we approximate by its radius of gyration and is at a distance z from one of the pore’s walls (its distance from the other wall is $h-z$, with h being the pore size). Suppose also that the perpendicular and parallel components of the viscosity of the solvent that contains such a sphere are, respectively, $\eta_{\perp}=\lambda_{\perp}\eta_0$ and $\eta_{\parallel}=\lambda_{\parallel}\eta_0$, with η_0 being the bulk value of the viscosity. As mentioned above, it has been shown that a linear superposition is a very good approximation for determining the effective viscosity of the fluid containing the beads between two walls,⁴² according to which

$$\lambda^{\text{II}} \cong \lambda_{w1}^{\text{I}} + \lambda_{w2}^{\text{I}} - 1 = \lambda^{\text{I}}(z) + \lambda^{\text{I}}(h-z) - 1. \quad (11)$$

Here, λ^{II} is the two-wall correction factor, while λ_{w1}^{I} and λ_{w2}^{I} are the single-wall correction factors. To determine the single-wall effects, in the region $z/a > 1.5$, we used the first-order terms of the Brenner–Faxén relations, which have been shown^{41,43,45,46} to be less than 1% in error in this area:

$$\lambda_{\parallel}^{-1} \cong 1 - \frac{9}{16}\left(\frac{a}{z}\right) + \mathcal{O}\left[\left(\frac{a}{z}\right)^3\right], \quad (12)$$

$$\lambda_{\perp}^{-1} \cong 1 - \frac{9}{8}\left(\frac{a}{z}\right) + \mathcal{O}\left[\left(\frac{a}{z}\right)^3\right]. \quad (13)$$

For $z/a \leq 1.5$, we used the full Faxén relation instead of its first-order approximation:

$$\lambda_{\parallel}^{-1} \cong 1 - \frac{9}{16}\left(\frac{a}{z}\right) + \frac{1}{8}\left(\frac{a}{z}\right)^3 - \frac{45}{256}\left(\frac{a}{z}\right)^4 - \frac{1}{16}\left(\frac{a}{z}\right)^5 + \mathcal{O}\left[\left(\frac{a}{z}\right)^6\right]. \quad (14)$$

For $1.25 \leq z/a \leq 1.5$ we used the first-order approximation of the Brenner relation given above. For $z/a < 1.25$, where $\lambda_{\perp} > 10$, we turned off the Langevin dynamics in the perpendicular direction, i.e., in this region no “kicks” by the Langevin equation act on the coarsened beads. One must also impose some maximum value on η_{\perp} because the time step, $\Delta t = 0.02m/\xi$, is inversely proportional to η_{\perp} . While Δt determines the frequency of the switching between the DMD simulations and calculating the corrections to the speed changes by Langevin equation, very large values of η_{\perp} (i.e., very frequent switches) are not desirable, as it would add very significantly to the total computation times.

To calculate the diffusion coefficient of the proteins, we compute the mean-square displacements (MSDs) $R^2(t)$ and utilize the Einstein relation $\langle R^2(t) \rangle = 2dDt$, where D is the

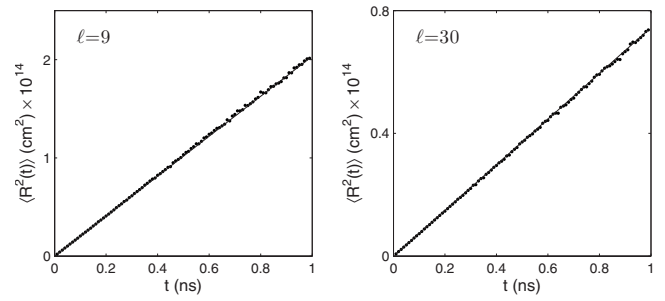


FIG. 1. MSDs $\langle R^2(t) \rangle$ vs time t for the smallest and largest proteins in the bulk. The solid straight lines represent fit of the results to Eq. (15) with $\beta \cong 1$.

diffusivity, d is the spatial dimension, and $\langle \cdot \rangle$ denotes an average over all the initial positions. In general,

$$\langle R^2(t) \rangle \sim t^{\beta}, \quad (15)$$

where $\beta=1$ corresponds to Fickian diffusion, whereas $\beta < 1$ represents anomalous diffusion. The MSDs are computed at each temperature T using all the data in the simulation time after the system reaches equilibrium.

We carried out simulations of the diffusion process in a pore system in which there was either an attractive or a repulsive potential, U^+ and U^- , respectively, between the proteins and the pore’s walls. Each case was studied over a range of temperature. Due to the large fluctuations of the proteins’ energy and structure near their folding temperature T_f , longer simulations were carried out near T_f . To reduce the central processing unit time, we used advanced algorithms in the computations, including link lists, neighbor lists, and false positioning which is special to the DMD. Temperature was held constant using the Andersen method.

V. RESULTS AND DISCUSSIONS

We studied diffusion of proteins of various lengths in nanopores of different sizes. In addition, we studied the effect of the nature of the interaction of the proteins with the pores’ walls—repulsive versus attractive interactions—on protein diffusion. A protein possesses its biological ability only at temperatures lower than its folding temperature T_f . Moreover, experimental works, including those on the transport of proteins, are usually carried out at room temperature. At high temperatures a protein is similar to a flexible polymer, the diffusion of which in confined media has been studied extensively. Therefore, except when we study the temperature dependence of the diffusivity, we focus mainly on protein diffusion at a low temperature, $T=0.09$. Note that, with³⁶ $\epsilon_{\text{HB}}=3300k_B$, the reduced temperature $T=0.09$ corresponds to 297 K. In what follows we present and discuss the results.

A. Diffusion in the bulk

Consider, first, protein diffusion under the bulk conditions. Figure 1 presents samples of time dependence of the MSD $\langle R^2(t) \rangle$. Over almost the entire simulated time interval, the MSDs follow Eq. (15) with $\beta \cong 1$, i.e., diffusion is Fickian.

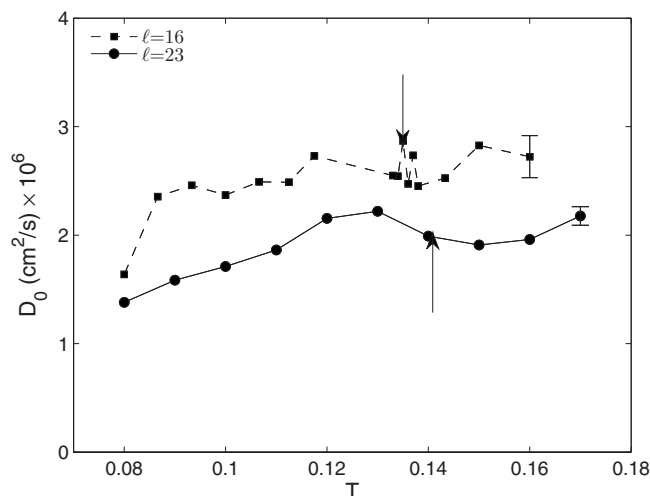


FIG. 2. Temperature dependence of the bulk diffusivity D_0 for two proteins. Arrows indicate the locations of the folding temperature T_f .

Figure 2 presents the temperature dependence of the bulk diffusivity D_0 for proteins of lengths $\ell=16$ and 23 . Far from the folding temperature T_f the diffusivity varies with T roughly linearly. This is expected, since at such temperatures the structure of a protein does not undergo fundamental changes. Near the folding temperature T_f , however, the structure does change and, therefore, nonlinear dependence on the temperature may be expected. We shall come back to this point shortly.

Figure 3 presents the dependence of the diffusivity D_0 on the protein's length ℓ at $T=0.09$. The results are represented very accurately by a power law,

$$D_0 \propto \ell^{-\nu}, \quad (16)$$

with $\nu \approx 0.84$. In contrast, the radius of gyration, R_g , of the same proteins under the bulk conditions varies perfectly linearly with the length ℓ . The reason is that, at low temperatures below their folding temperatures, the proteins take on the folded α -helix structure, which is like a cylindrical tube or a rod. Therefore, R_g should vary linearly with the protein's length ℓ , and Fig. 3 confirms this. At high enough temperatures, however, a protein is similar to a flexible polymer, and R_g no longer varies linearly with ℓ but as a power law with a Flory exponent of about 0.6.

How does the computed diffusivity compare with the experimental data? Konopa *et al.*²² reported on the diffusivity of green fluorescent protein (GFP), which is a 27 kDa

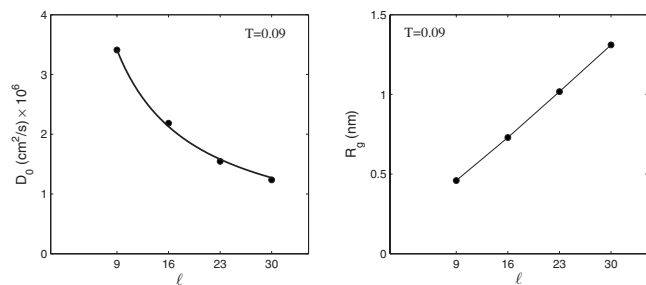


FIG. 3. Dependence of the bulk diffusivity D_0 and the radius of gyration, R_g , on the size ℓ of a protein.

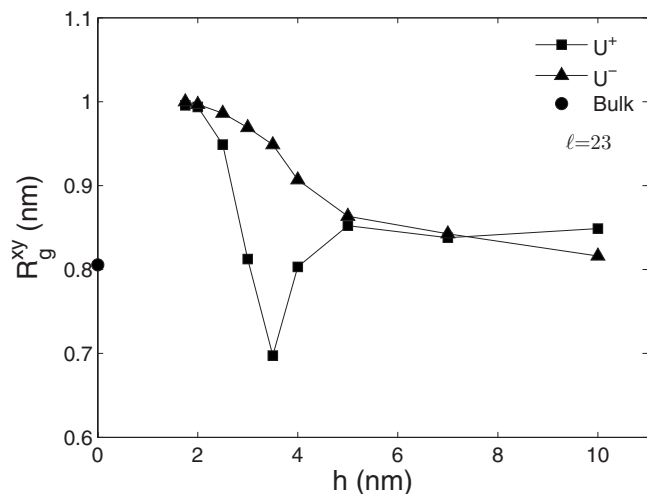


FIG. 4. Dependence of the 2D radius of gyration, R_g^{xy} , of a protein of length $\ell=23$ in the xy planes parallel to the pore's walls on the size h of the pores at $T=0.09$ with attractive (U^+) or repulsive (U^-) walls.

β -barrel protein with 238 amino acids.⁵⁰ The GFP is about 13 times heavier than the largest α -helix protein that we study, which is a 2.1 kDa protein with 30 amino acids. Assuming that the two proteins have nearly compact structures—a reasonable assumption—the sizes of the two proteins have a ratio $13^{1/3} \approx 2.3$. So, roughly speaking (roughly, since the shapes of the two proteins are different, and this affects their hydrodynamic radii), the protein that we study should have a diffusivity D about 2.3 times larger than that of the GFP in the bulk. The diffusivity of the GFP was reported²² to be $D \approx 8.7 \times 10^{-7}$ cm²/s, whereas for the protein that we study, $D \approx 1.2 \times 10^{-6}$ cm²/s, a factor of about 1.4 larger. Thus, the computed diffusivities are of the correct order of magnitude. See also Smith *et al.*⁵¹ for more data.

B. Diffusion in a nanopore

In addition to the pore size and the nature of the interaction between a protein and the pore's walls, an important factor in determining the diffusivity of a protein is its molecular size, represented roughly by its radius of gyration, R_g . The radius of gyration, R_g , of a protein in the bulk is relatively large at high temperatures, since it is in unfolded states with an essentially random-coil structure. However, as T is lowered, the protein begins to take on the compact structure of its native α -helical state and, thus, its R_g decreases. At low enough temperatures R_g attains its minimum value. The α -helix shape does not depend on the pore size h if h is relatively large and, as a result, the R_g of the folded protein at low temperatures in such pores is nearly the same as that in the bulk.

A better way of understanding the structure of a protein in a nanopore is perhaps through a study of its configuration in the xy planes parallel to the nanopores' walls. To do so, we computed R_g^{xy} , the 2D analog of R_g in the xy planes. Figure 4 presents R_g^{xy} of a protein of length $\ell=23$ at $T=0.09$ versus h . In the case of a pore with repulsive walls, R_g^{xy} is always larger than its corresponding value in the bulk because repulsion elongates the protein. In the case of a pore with attractive walls, however, there is an unexpected minimum in R_g^{xy} .

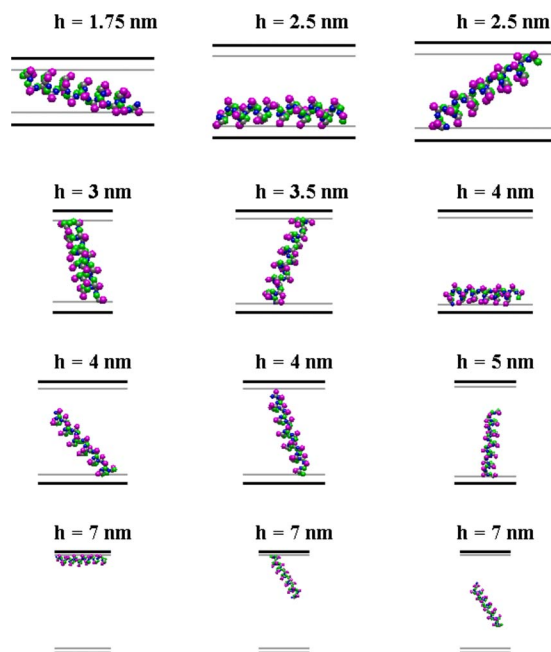


FIG. 5. (Color online) Configurations of a protein of size $\ell=23$ in various pores at $T=0.09$.

The minimum is related to a phenomenon that we call the proteins' "standing up," whereby both ends of a long enough folded protein interact with both walls, such that the protein is nearly perpendicular to the walls. This phenomenon, shown in Fig. 5, reduces R_g^{xy} . As the size of the pore increases, the phenomenon disappears and, therefore, R_g^{xy} increases again. The effect of this phenomenon on protein diffusion in small pores will be discussed shortly.

As discussed above, when a protein is (in a solution) at a temperature far below its folding temperature T_f , it is completely folded and its R_g is nearly equal to that in the bulk. Therefore, any change in the diffusivity of a protein in a nanopore at low enough T is strictly due to the interaction of the structure of a folded α -helix with the nanopore's walls. Our interest in the problem is also over such a temperature range. Note also that the folding temperature T_f for the bulk state and those for the pores are not the same. In fact, as shown in Part I, in a nanopore T_f also depends on whether the walls are attractive or repulsive. Therefore, due to the strong fluctuations of R_g and D in the region around T_f , the diffusivities in the pore and under the bulk conditions cannot be directly compared. The comparison is meaningful mainly at temperatures far from T_f .

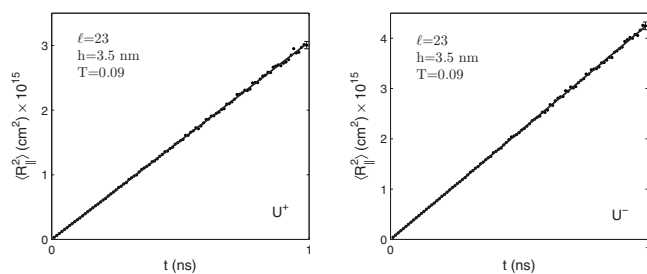


FIG. 6. MSDs $\langle R_{\parallel}^2(t) \rangle$, in the direction parallel to the walls, in pores with attractive (U^+) or repulsive (U^-) walls.

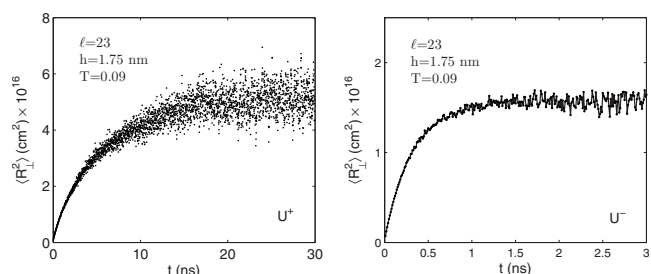


FIG. 7. Same as in Fig. 6 but for the MSDs $\langle R_{\perp}^2(t) \rangle$ perpendicular to the pore's walls.

Another aspect of diffusion of a protein in a nanopore which is fundamentally different from that in the bulk is that, in addition to the pore diffusivity being smaller than its bulk value D_0 , the diffusion process is anisotropic, with the MSD $\langle R_{\parallel}^2(t) \rangle$ in the xy planes parallel to the walls being much larger than $\langle R_{\perp}^2(t) \rangle$ in the direction perpendicular to the walls. Moreover, $\langle R_{\parallel}^2(t) \rangle$ grows with the time indefinitely and linearly, i.e., the diffusion process is Fickian. This is demonstrated clearly in Fig. 6. As one might expect, $\langle R_{\parallel}^2(t) \rangle$ in the pore with repulsive walls is larger than that in a pore with attractive walls. As the size of the pore decreases, the difference between the two MSDs also increases.

However, the MSD $\langle R_{\perp}^2(t) \rangle$ saturates after a finite time. We show in Fig. 7 the results at $T=0.09$ for a protein of length $\ell=23$ in two pores of size $h=1.75$ nm, one with attractive and one with repulsive interactions between the protein and the walls. Hence, the diffusivity in the direction perpendicular to the pores' walls is zero. It is clear that, in the pore with attractive walls, it takes a longer time for $\langle R_{\perp}^2(t) \rangle$ to saturate because there is more space for the proteins to diffuse in, but the fluctuations are also larger for the same reason. Thus, hereafter, by diffusivity we mean one associated with the proteins' motion in the planes parallel to the pores' walls.

Figure 8 presents the temperature dependence of the diffusivity of a protein of length $\ell=23$ in a pore of size $h=4$ nm with the two types of walls. These results should be

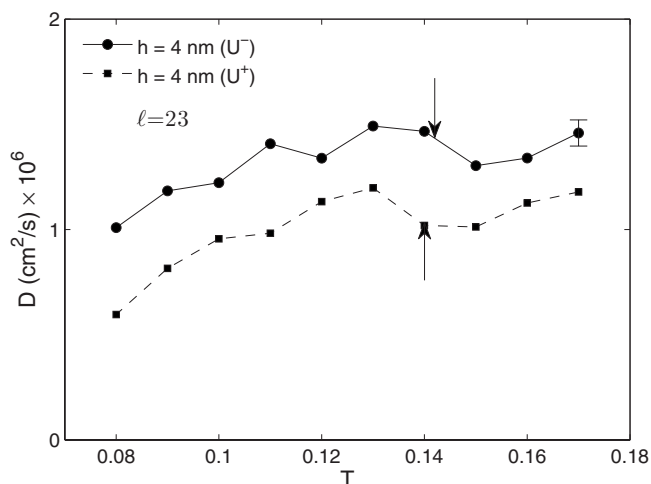


FIG. 8. Dependence of the diffusivity D of a protein of size $\ell=23$ on the temperature T in a pore of size h with attractive (U^+) or repulsive (U^-) walls.

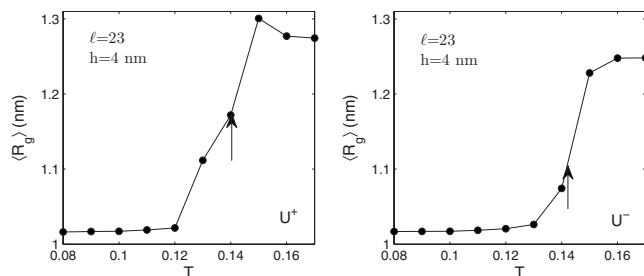


FIG. 9. Dependence of the average radius of gyration $\langle R_g \rangle$ on the temperature T in a pore of size h with attractive (U^+) or repulsive (U^-) walls. Arrows indicate the location of the folding temperature.

compared with those shown in Fig. 2 under the bulk conditions. The diffusivities in the pore with repulsive walls are larger than those with attractive walls, but the trends appear to be similar.

As discussed above, due to the strong thermal fluctuations near T_f , R_g varies strongly with T . Consequently, near T_f the effective diffusivity D no longer varies linearly with T . To see this better, we present in Fig. 9 the average radius of gyration $\langle R_g \rangle$ of the protein in the same nanopores of size $h=4$ nm. Far from the folding temperature T_f , R_g is constant. However, near T_f there are strong thermal fluctuations, as a result of which R_g changes rapidly.

Figure 10 presents the dependence of the diffusivity on the proteins' length ℓ in pores of size $h=4$ nm with attractive and repulsive walls at $T=0.09$. Similar to the bulk conditions, D depends on ℓ as a power law, as given by Eq. (16). The diffusivity in the pore with repulsive walls is larger than in the pore with attractive walls. This may be explained by observing that, in pores with repulsive walls the proteins are pushed toward the pore's center and, therefore, their diffusion is more efficient than that in the pore with attractive walls. We expect the same to be true if we impose an external potential gradient across the pore. For both pores, the exponent ν defined by Eq. (16) is close to 0.84 that we find for the bulk condition.

Figure 11 presents the dependence of the diffusivities on

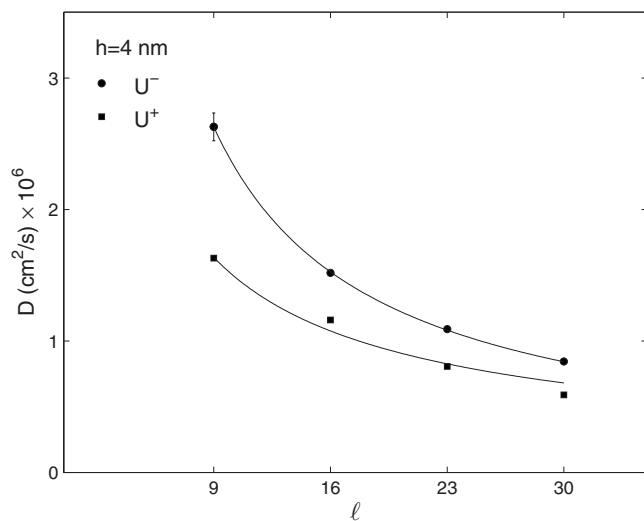


FIG. 10. Dependence of the diffusivity D on the protein's length ℓ at $T=0.09$ in a pore of size h with attractive (U^+) or repulsive (U^-) walls.

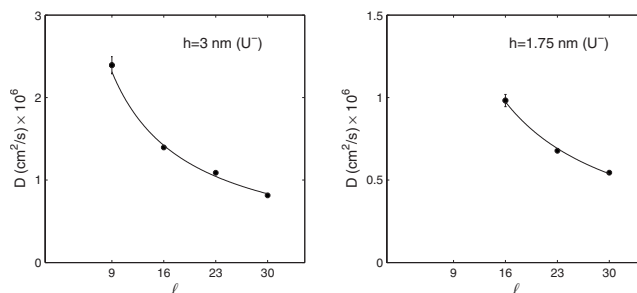


FIG. 11. Dependence of the diffusivity D on the protein's length ℓ at $T=0.09$ in two pores of size h with repulsive (U^-) walls.

the proteins' size ℓ in two smaller pores with sizes $h=3$ and 1.75 nm, both with repulsive walls. The qualitative features of the results are the same as those presented in Fig. 10 for pore size $h=4$ nm. Thus, at least for the range of pore sizes that we have studied, the scaling of the diffusivity in the pores with the proteins' size remains close to that in the bulk.

Figure 12 presents the ratio D/D_0 for a protein of size $\ell=23$ versus the pore size h at $T=0.09$. The results were computed using long simulations ($3 \mu\text{s}$). The diffusivities in the pores with repulsive walls are larger than those in pores with attractive walls because repulsion keeps the protein more or less near the pores' center, whereas with attractive walls the protein may spend considerable time near the walls and even attach itself to them. This is shown in Fig. 5. For the same reason, the increase in the value of D^+ with increasing pore size h is rather slow in pores with attractive walls.

However, a most interesting result is the dependence on the pore size h of D^+/D^- , the ratio of the diffusivities in pores with attractive and repulsive walls. Figure 13 presents the results. There seems to be no particular trend for this dependence. However, such "erratic" behavior may be explained by considering the configurations of the proteins in pores of various sizes.

Thus, consider Fig. 5. In the smallest pore, with size of 1.75 nm, the protein is highly confined, with h , the distance

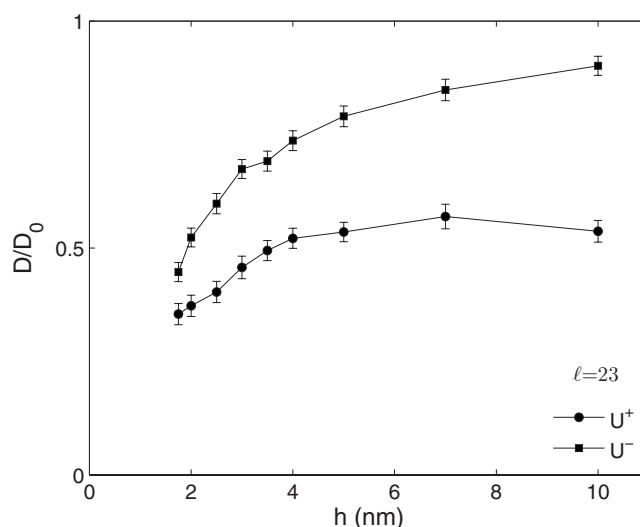


FIG. 12. Dependence on the pore size h of the diffusivity, normalized by its bulk value D_0 , at $T=0.09$ for a protein of length $\ell=23$. The pores have either attractive (U^+) or repulsive (U^-) walls.

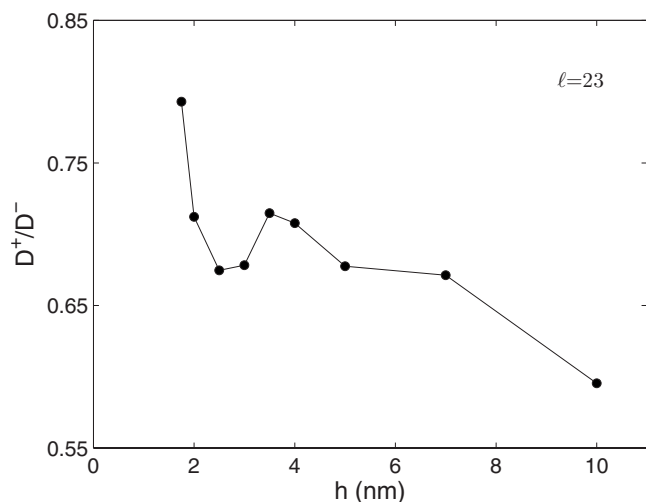


FIG. 13. Dependence on the pore size h of D^+/D^- , the ratio of the diffusivities in the pores with attractive (+) or repulsive (-) walls, for a protein of length $\ell=23$ at $T=0.09$.

between the walls, being just a little larger than the diameter ξ_d of the α -helix. The protein may be adsorbed on one wall only but often has contact with both walls through its ends. In the slightly larger pores, namely, with sizes $h=2$ and 2.5 nm, the protein interacts most strongly with only one of the walls, even if it is interacting with both walls, and the angle that its molecule makes with the walls (see Fig. 5) is rather small, as a result of which it will detach from one of the walls and adsorb completely on the other one. In the next two larger pores (3 and 3.5 nm), the protein may again attach to one or both walls, but the latter occurs more frequently. This is due to the fact that the angle between the protein and the walls is relatively large, making it difficult for the protein to separate completely from one wall to attach itself completely to the other wall.

In the nanopore of size of 4 nm, the distance h between the walls is nearly equal to R_{ee} , the end-to-end distance in the α -helix configuration. Here, if the protein is not adsorbed on one wall, it may have strong interaction with one wall or both. Finally in pores with a size $h > R_{ee}$ ($h=5$ nm and larger) the protein cannot have strong interactions with *both* walls *at the same time*. Another possibility is for the protein to be “free” of both walls (pore of size 7), but this is energetically unfavorable at low T and happens very rarely.

The two local maxima in Fig. 13 occur in pores of sizes $h=1.75$ and 3.5 nm. For this protein, $\xi_l \approx 3.5$ nm. Thus, we find that the first maximum occurs at a pore size nearly equal to the diameter ξ_d of the α -helix, while the second maximum occurs in a pore for which h is slightly less than the (end-to-end) length R_{ee} of the α -helix. Since in the two pores the protein interacts most often with both walls, its c.m. is close to the pores’ centers, hence more efficient diffusion. In effect, due to the opposite directions of the interaction forces between the two walls and the protein, the forces cancel each other, hence allowing the protein to diffuse efficiently. Therefore, we may argue that for α -helices with different sizes ℓ , the pore size at which the first maximum occurs should not vary much with ℓ , because the average diameter of the simulated α -helices is nearly the same. However, the h at which the second maximum occurs depends on ℓ .

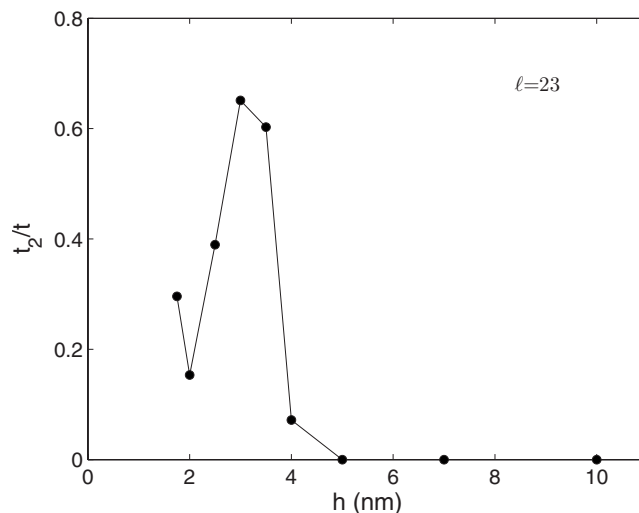


FIG. 14. Dependence on the pore size h of t_2/t , the ratio of the time t_2 that a protein spends interacting with both walls and the total time t spent in the pores, for a protein of length $\ell=23$ at $T=0.09$.

A more quantitative discussion of the two local maxima in the dependence of D^+/D^- on the pore size h is as follows. Suppose that t_2 is the time period over which a protein interacts with both walls, and t is the total time that the protein spends in a pore. By interacting with one wall we mean that at least one of the protein’s atoms is in the attractive range of that wall’s potential—the space between the distances d_{3X} and $d_{3X}+d_{4X}$ from the wall; see Sec. III. Thus, the protein interacts with both walls when, according to this definition, some of its atoms are interacting with one wall and some others with the other wall. Figure 14 presents the dependence of t_2/t on the pore size h . Consistent with the argument given above and Fig. 13, the two local maxima are at $h=1.75$ nm and $h=3$ nm, and for $h=3.5$ nm the value of t_2/t is large. Note that for pores of size $h > 5$ nm the time t_2 is insignificant simply because the pore size is large compared with the protein size. Note also that the minimum of R_g^{xy} in the pores with attractive pores, shown in Fig. 4, also occurs at $h=3.5$ nm. Therefore, the pore size dependences of all the quantities are consistent with one another.

Another quantitative way of studying the effect of confinement on protein diffusion is by computing the distribution of the distances Z_a of all the protein’s atoms from the two walls, as well as that of $Z_{c.m.}$, the distance of the protein’s c.m. from the walls, over the entire (simulation) time. Figure 15 presents the results for a protein of size $\ell=23$ at $T=0.09$ in pores of various sizes and repulsive walls. As the results indicate, the protein’s atoms and its c.m. are distributed mostly in the middle of the pore. This result, which is due to the lack of an energy scale for the interaction of a protein with the pore’s walls, also implies that the dependence of the two distributions on the temperature in a pore with repulsive walls cannot be significant. We shall come back to this point shortly. In the largest pore that we simulated, one with size $h=10$ nm, the two distributions appear to become mostly flat.

In pores with attractive walls, however, the results are quite different. Figure 16 presents the two distributions in the same pores but with attractive walls. Since the temperature is

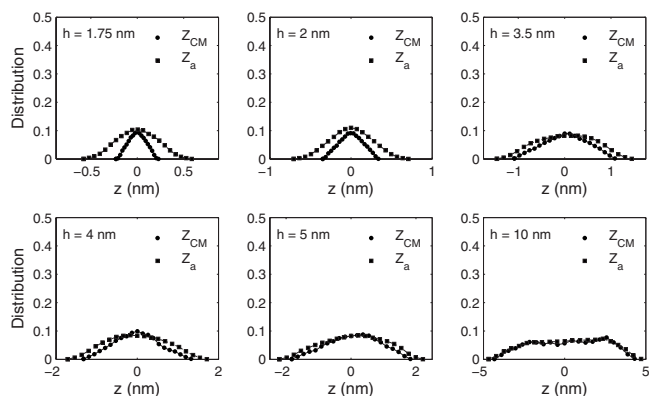


FIG. 15. Distributions of the distances Z from the pore walls of the atoms of a protein of length $\ell=23$ and those of the distances $Z_{c.m.}$ of its c.m. in pores of size h with repulsive walls at $T=0.09$.

low, the protein takes on a definite structure in which not all the atoms can interact directly with the walls, as they are at finite distances from them. The broadness of the distribution of $Z_{c.m.}$ correlates well with t_2/t , shown in Fig. 14. For larger t_2/t the distribution of $Z_{c.m.}$ is narrower and exhibits a pronounced peak in the middle of the pore, which is due to the fact that in such pores the protein's ends are attached to the walls most of the times; see Fig. 5. As the pore size increases though, the time scale t_2 becomes insignificant (see Fig. 14), and the peaks move toward the walls. The distributions become symmetrical, with one peak near each wall, only after very long times because, as described above, once a protein is very near a wall, it spends a long time there before diffusing toward the center and the opposite wall.

As mentioned above, Fig. 15 implies that temperature does not have a strong effect on the distributions of $Z_{c.m.}$ and Z_a if a pore's walls are repulsive. The same is not true when the walls are attractive. Shown in Fig. 17 is a comparison of the two distributions at two temperatures in pores of the same size, one with attractive and one with repulsive walls. It is clear that, whereas the two distributions are strongly affected as T increases when the walls are attractive, the same is not true when the walls are repulsive.

C. A phase diagram

We may summarize the effect on protein diffusion of the temperature T , pore size h , and interactions U^\pm with the

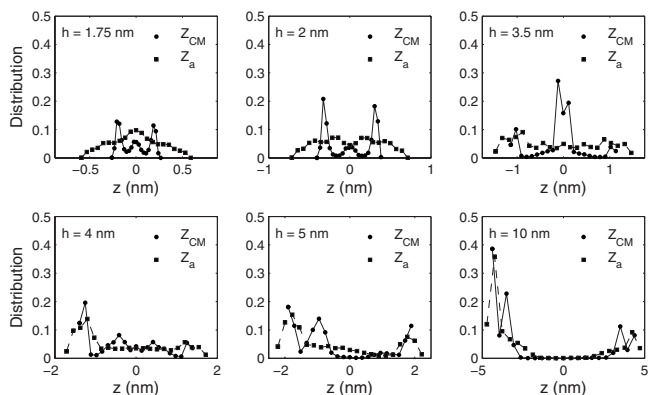


FIG. 16. Distributions of the distances Z from the pore walls of the atoms of a protein of length $\ell=23$ and those of the distances $Z_{c.m.}$ of its c.m. in pores of size h with attractive walls at $T=0.09$.

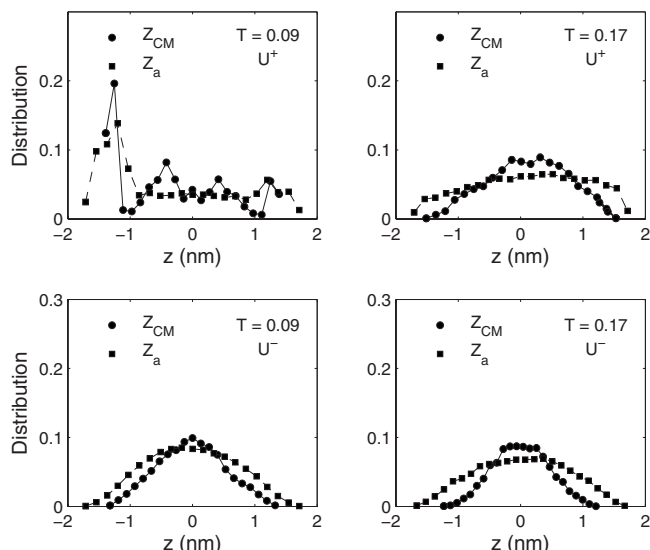


FIG. 17. Comparison of the distributions of the distances Z from the pore walls of the atoms of a protein of length $\ell=23$ and those of the distances $Z_{c.m.}$ of its c.m. in pores of size $h=4$ nm with attractive (U^+) or repulsive (U^-) walls at two temperatures T .

pore's walls in a "phase diagram." The phase diagram is shown in Fig. 18. We divide the phase space (T, h) into four regions for both U^+ and U^- and refer to the diffusivities in region number p as D_p^+ and D_p^- . Generally speaking, the above discussions indicate that we should *always* expect $D_p^+ < D_p^- < D_0$ in *all* the regions, which our simulations described above also confirmed. Moreover, we expect the difference $D_p^- - D_p^+$ to be larger in larger pores because in pores with attractive walls a protein is attached to the walls, whereas in the case of a pore with repulsive walls, it diffuses more or less freely in the area near the pore's center, and hence has larger diffusivity. In the discussions that follow, when we refer to the attractive interaction potential U^+ , we mean one that satisfies one condition: the interaction strength ε_{PW} is not so large (considering the energetic and entropic effects) that can prevent folding of a protein into its native α -helix structure.

Region 1 is bounded by the folding temperature T_f of a protein in a pore with repulsive walls. Because T is relatively high in region 1, in a pore with U^- the protein is unfolded and resembles a flexible chain. Thus, the theories for diffusion of flexible chains in pores are applicable. In the case of a pore with U^+ , however, the protein's atoms have significant interactions with the walls (if the pore is not too large). While a protein is not folded in this region, most of its atoms interact with the pore's walls. If the temperature is higher than a typical T at which the kinetic fluctuations are large enough to overcome all the possible interactions of protein atoms, then the average potential energy of the protein vanishes, and we expect U^+ to have a negligible effect on the protein structure and diffusion. Therefore, in region 1 and at relatively high T we must have $D_1^+ \approx D_1^- < D_0$. Moreover, in this region all the scaling relations derived previously for various properties of a flexible polymer chain in a confined medium should be valid for a protein because the fluctuations in the energy overcome all the internal and external

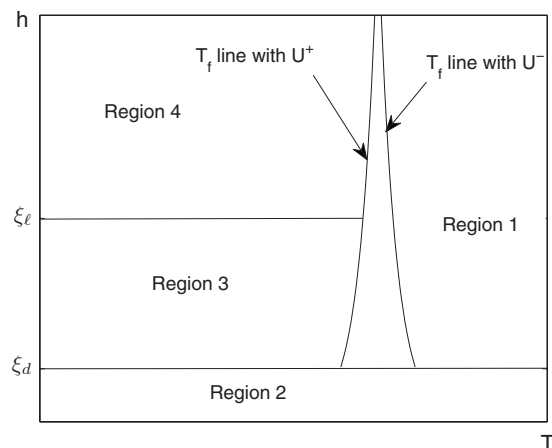


FIG. 18. The proposed phase diagram of the effect of pore size h and temperature T on a protein configuration and its diffusivity. ξ_d and ξ_l are, respectively, the diameter and length of the folded α -helix configuration.

interactions of the protein's atoms. Note that the curves that indicate the folding temperatures T_f cannot be continued down to the limit $h \rightarrow 0$ because in very tight pores the proteins cannot fold and T_f cannot be defined.

Most of region 2 is bounded by a pore size $h \approx \xi_d$, where ξ_d is the diameter of the α -helix. In this region, relative to the thermal fluctuations, the energetic interactions are more important. Moreover, a protein cannot attain its folded state due to the steric restrictions ($h < \xi_d$ in region 2). In the case of a pore with U^- the protein diffuses in the free space of the pore near its center and, if sterically allowed, may form some HBs in its structure. In the case of a pore with attractive walls, however, the protein's atoms interact with the walls. However, due to the small pore sizes in this region, we do not expect D_2^- to be significantly larger than D_2^+ .

Region 3 is bounded from below by the pore size, $h \approx \xi_d$, by the folding temperature line of proteins in a pore with attractive walls on the right and bounded from above by a pore size $h = h(\ell) \approx \xi_l$, where ξ_l is the length of the folded α -helix which varies linearly with ℓ at low T (see above). ξ_l is about the same as R_{ee} , the end-to-end distance of a protein in the α -helix configuration. For low h a protein is folded and highly confined in the pore. As the pore size increases, the protein interacts strongly with only one of the walls if the walls are attractive. In the case of a pore with repulsive interactions, the protein has freedom of diffusion between walls because the pore size is increasing. Near the upper boundary of this region, the protein is in the standing up configuration, described earlier.

However, two *anomalous* increases in the values of D^+ arise in region 3, where the pore size h is close to the border lines, comparable to either ξ_d or ξ_l . As discussed earlier, in the border areas in region 3, the proteins spend a significant time t_2 interacting strongly (in the sense defined earlier) with both walls *simultaneously*, which leads to the two anomalous peaks in the ratio D^+/D^- ; see Fig. 18.

Region 4 includes large pore sizes, $h > \xi_l$. For example, the simulations indicate that for an α -helix of size $\ell = 23$, region 4 includes pores with $R_g/h \approx 0.1$. In the case of a pore with attractive walls, a protein interacts with only one wall

(especially at lower T) but does not interact with both walls simultaneously due to the large size of the pore. It may, in fact, be adsorbed onto a wall, not diffusing freely in the space between the walls for a long time. Thus, a protein spends most of its time interacting strongly with the pore walls. The distribution of $Z_{c.m.}$ for an α -helix of length $\ell = 23$ in a pore size of $h = 10$ nm, shown in Fig. 10, confirms this. Hence, in region 4 we do not expect D_4^+ to increase significantly with increasing h , whereas D_4^- does increase with increasing h ; see Fig. 12.

Thus, in regions 3 and 4 D_p^- increases smoothly and simply with increasing pore size h and approaches the bulk value D_0 when h is large. D_p^+ , however, has a more complex behavior for an α -helix with two molecular length scales, i.e., the diameter ξ_d and length ξ_l of the cylinderlike shape of the α -helix. In fact, in these regions D^+ increases more slowly than D^- because, as discussed above, in these two regions a protein in a pore with attractive walls spends most of its time near the walls or is attached to them even in the larger pores.

VI. SUMMARY

DMD simulations were coupled to the Langevin equation, and together with the PRIME, an intermediate-resolution model of proteins, were used to study diffusion of proteins in nanopores. We considered pores with attractive or repulsive interaction potentials between their walls and the proteins. The diffusivity D of the proteins was computed as a function of the number of the amino-acid groups, temperature T , the pore size h , and the interaction potentials with the walls. We found that the diffusivity follows a power law in the length ℓ of the proteins and is larger in pores with repulsive walls. D^+/D^- , the ratio of the diffusivities in pores with attractive and repulsive walls, exhibits two local maxima in its dependence on the pore size h . We attributed them to the pore sizes and protein configurations that induce significant, long-lasting interactions with both walls of a pore. Indeed, we found the same type of h dependence for t_2/t , the ratio of the time t_2 , the time that the proteins spends interacting significantly with both pore's walls, and the total time t . Far from the folding temperature T_f , D increases linearly with T , but due to the thermal fluctuations and their effect on the proteins' structure near T_f , the dependence of D on T in this region is nonlinear. We propose a phase diagram, consisting of four regions, which describes qualitatively the effect of h , T , and interaction potentials with the walls on the diffusivity D .

The next step in this problem is to study transport of proteins in nanopores in the presence of an external potential gradient, such as a pressure gradient. Work in this direction is in progress.

ACKNOWLEDGMENTS

The work of L.J. was supported by the Iranian Nanotechnology Initiative.

¹L. Javidpour, M. R. Rahimi Tabar, and M. Sahimi, *J. Chem. Phys.* **128**, 115105 (2008).

²C. B. Anfinsen, *Science* **181**, 223 (1973).

- ³D. K. Klimov and D. Thirumalai, *Phys. Rev. Lett.* **76**, 4070 (1996).
- ⁴L. Mirny and E. Shakhnovich, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 361 (2001); H.-X. Zhou, *Acc. Chem. Res.* **37**, 123 (2004); J. Kubelka, J. Hofrichter, and W. A. Eaton, *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
- ⁵M. Dadvar, M. Sohrabi, and M. Sahimi, *Chem. Eng. Sci.* **56**, 1 (2000); M. Dadvar and M. Sahimi, *ibid.* **57**, 939 (2002); **58**, 4935 (2003).
- ⁶M.-E. Avramescu, Z. Borneman, and M. Wessling, *Biotechnol. Bioeng.* **84**, 564 (2003).
- ⁷A. J. Rosenbloom, D. M. Sipe, Y. Shishkin, Y. Ke, R. P. Devaty, and W. J. Choyke, *Biomed. Microdevices* **6**, 261 (2004).
- ⁸R. J. Ciora, B. Fayyaz, P. K. T. Liu, V. Suwanmethanon, R. Mallada, M. Sahimi, and T. T. Tsotsis, *Chem. Eng. Sci.* **59**, 4957 (2004); B. Elyassi, M. Sahimi, and T. T. Tsotsis, *J. Membr. Sci.* **288**, 290 (2007); **316**, 73 (2008).
- ⁹*Molecular Biology of the Cell*, edited by B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (Garland, New York, 2002); D. Görlich and I. W. Mattaj, *Science* **271**, 1513 (1996).
- ¹⁰E. Di Marzio and J. J. Kasianowicz, *J. Chem. Phys.* **119**, 6378 (2003) and references therein.
- ¹¹J. M. Tsutsui, F. Xie, and R. T. Porter, *Cardiovasc. Ultrasound* **2**, 23 (2004).
- ¹²W. Liebermeister, T. A. Rapoport, and R. Heinrich, *J. Mol. Biol.* **305**, 643 (2001).
- ¹³T. C. Elston, *Biophys. J.* **82**, 1239 (2002); T. Ambjörnsson and R. Metzler, *Phys. Biol.* **1**, 77 (2004); T. Ambjörnsson, M. A. Lomholt, and R. Metzler, *J. Phys.: Condens. Matter* **17**, S3945 (2005).
- ¹⁴J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13770 (1996); A. Meller, L. Nivon, E. Brandin, J. A. Golovchenko, and D. Branton, *ibid.* **97**, 1079 (2000); S. E. Henrickson, M. Misakian, B. Robertson, and J. J. Kasianowicz, *Phys. Rev. Lett.* **85**, 3057 (2000); A. Meller, L. Nivon, and D. Branton, *ibid.* **86**, 3435 (2001).
- ¹⁵W. Sung and P. J. Park, *Phys. Rev. Lett.* **77**, 783 (1996); M. Muthukumar, *ibid.* **86**, 3188 (2001); *J. Chem. Phys.* **118**, 5174 (2003).
- ¹⁶For a review see, for example, A. Meller, *J. Phys.: Condens. Matter* **15**, R581 (2003).
- ¹⁷D. K. Lubensky and D. R. Nelson, *Biophys. J.* **77**, 1824 (1999); Y. Kantor and M. Kardar, *Phys. Rev. E* **69**, 021806 (2004); R. Bundschuh and U. Gerland, *Phys. Rev. Lett.* **95**, 208104 (2005); S. Matysiak, A. Montesi, M. Pasquali, A. B. Kolomeisky, and C. Clementi, *ibid.* **96**, 118103 (2006).
- ¹⁸Y. Kafri, D. K. Lubensky, and D. R. Nelson, *Biophys. J.* **86**, 3373 (2004).
- ¹⁹R. H. Abdolvahab, F. Roshani, A. Nourmohammad, M. Sahimi, and M. R. Rahimi Tabar, *J. Chem. Phys.* **129**, 235102 (2008).
- ²⁰H. Salman, D. Zbaida, Y. Rabin, D. Chatenay, and M. Elbaum, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7247 (2001); R. Zandi, D. Reguera, J. Rudnick, and W. M. Gelbart, *ibid.* **100**, 8649 (2003); J. K. Wolterink, G. T. Barkema, and D. Panja, *Phys. Rev. Lett.* **96**, 208301 (2006).
- ²¹D. Klotsa, R. A. Romer, and M. S. Turner, *Biophys. J.* **89**, 2187 (2005).
- ²²M. C. Konopka, I. A. Shkel, S. Cayley, M. T. Record, and J. C. Weishaar, *J. Bacteriol.* **188**, 6115 (2006).
- ²³A. Partikian, B. Ölveczky, R. Swaminathan, Y. Li, and A. S. Verkman, *J. Cell Biol.* **140**, 821 (1998).
- ²⁴F. Dumas, N. Destainville, C. Millot, A. Lopez, D. Dean, and L. Salomé, *Biophys. J.* **84**, 356 (2003).
- ²⁵F. Amato, C. Cosentino, S. Pricl, M. Ferrone, M. Fermeglia, M. M. Cheng, R. Walczak, and M. Ferrari, *Biomed. Microdevices* **8**, 291 (2006).
- ²⁶J. Kärger, *Diffus. Fundam.* **2**, 78.1 (2005).
- ²⁷S. Condamin, V. Tejedor, R. Voituriez, O. Bénichou, and J. Klafter, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5675 (2008).
- ²⁸Y.-L. Chen, M. D. Graham, and J. J. de Pablo, *Phys. Rev. E* **70**, 060901 (2004).
- ²⁹P.-K. Lin, C.-C. Fu, Y.-L. Chen, Y.-R. Chen, P.-K. Wei, C. H. Kuan, and W. S. Fann, *Phys. Rev. E* **76**, 011806 (2007).
- ³⁰Y. Yang, T. W. Burkhardt, and G. Gompper, *Phys. Rev. E* **76**, 011804 (2007).
- ³¹M. C. Bujan-Núñez and E. Dickinson, *Mol. Phys.* **80**, 431 (1993).
- ³²G. Jayachandran, V. Vishal, and V. Pande, *J. Chem. Phys.* **124**, 164902 (2006).
- ³³L. Regan and W. F. DeGrado, *Science* **241**, 976 (1988).
- ³⁴Z. Guo and D. Thirumalai, *J. Mol. Biol.* **263**, 323 (1996).
- ³⁵S. Sun, *Protein Sci.* **2**, 762 (1993); S. Sun, P. D. Thomas, and K. A. Dill, *Protein Eng.* **8**, 769 (1995); S. Takada, Z. Luthey-Schulten, and P. G. Wolynes, *J. Chem. Phys.* **110**, 11616 (1999).
- ³⁶A. V. Smith and C. K. Hall, *Proteins* **44**, 376 (2001); *J. Mol. Biol.* **312**, 187 (2001); H. D. Nguyen, A. J. Marchut, and C. K. Hall, *Protein Sci.* **13**, 2909 (2004); A. J. Marchut and C. K. Hall, *Comput. Biol. Chem.* **30**, 215 (2006).
- ³⁷B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **31**, 459 (1959); D. C. Rapaport, *J. Phys. A* **11**, L213 (1978); *J. Chem. Phys.* **71**, 3299 (1979); A. Belleman, J. Orban, and D. V. Belle, *Mol. Phys.* **39**, 781 (1980); S. W. Smith, C. K. Hall, and B. D. Freeman, *J. Comput. Phys.* **134**, 16 (1997).
- ³⁸K. Luo, T. Ala-Nissila, S. C. Ying, and A. Bhattacharya, *Phys. Rev. Lett.* **99**, 148102 (2007); C. Forrey and M. Muthukumar, *J. Chem. Phys.* **127**, 015102 (2007).
- ³⁹H. Faxén, *Ark. Mat., Astron. Fys.* **17**, 1 (1923).
- ⁴⁰H. Brenner, *Chem. Eng. Sci.* **16**, 242 (1961).
- ⁴¹J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics* (Kluwer, Dordrecht, 1983).
- ⁴²P. Holmqvist, J. K. G. Dhont, and P. R. Lang, *J. Chem. Phys.* **126**, 044707 (2007).
- ⁴³B. Lin, J. Yu, and S. A. Rice, *Phys. Rev. E* **62**, 3909 (2000).
- ⁴⁴L. P. Faucheux and A. J. Libchaber, *Phys. Rev. E* **49**, 5158 (1994).
- ⁴⁵L. Lobry and N. Ostrowsky, *Phys. Rev. B* **53**, 12050 (1996).
- ⁴⁶J. S. Halow and G. B. Wills, *AIChE J.* **16**, 281 (1970).
- ⁴⁷J. J. Magda, M. V. Tirrell, and H. T. Davis, *J. Chem. Phys.* **83**, 1888 (1985); I. Bitsanis, S. A. Somers, H. T. Davis, and M. V. Tirrell, *ibid.* **93**, 3427 (1990); K. P. Travis and K. E. Gubbins, *ibid.* **112**, 1984 (2000); R. Qiao and N. R. Aluru, *ibid.* **118**, 4692 (2003).
- ⁴⁸K. P. Travis, B. D. Todd, and D. J. Evans, *Phys. Rev. E* **55**, 4288 (1997).
- ⁴⁹G. Karniadakis, A. Beskok, and N. Aluru, *Microflows and Nanoflows* (Springer, New York, 2005).
- ⁵⁰Complete information about the GFP (including images of its structure) is given in <http://www.rcsb.org/pdb/explore/explore.do?structureId=1EMA>.
- ⁵¹C. Smith, R. Kirk, T. West, M. Bratzel, M. Cohen, F. Martin, A. Boiarski, and A. A. Rampersaud, *Diabetes Technol. Ther.* **7**, 151 (2005).