Molecular simulation of protein dynamics in nanopores. I. Stability and folding

Leili Javidpour, M. Reza Tabar, and Muhammad Sahimi

Additional information on J. Chem. Phys.
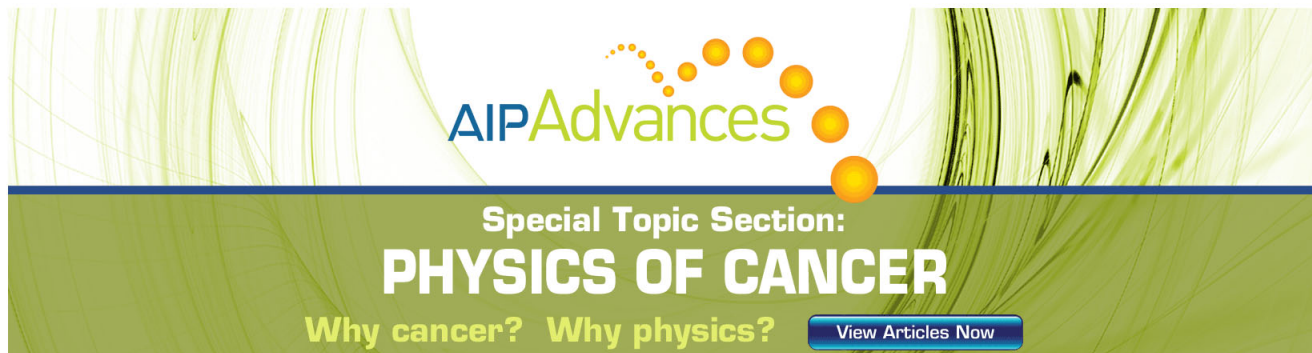
Molecular simulation of protein dynamics in nanopores. I. Stability and folding

Leili Javidpour,[1] M. Reza Rahimi Tabar,[1,2] and Muhammad Sahimi[3,a)]

[1]*Department of Physics, Sharif University of Technology, Tehran 11155-9161, Iran*
[2]*Institute of Physics, Carl von Ossietzky University, Oldenburg D-26111, Germany*
[3]*Mork Family Department of Chemical Engineering and Materials Science,*
University of Southern California, Los Angeles, California 90089-1211, USA

Discontinuous molecular dynamics simulations, together with the protein intermediate resolution model, an intermediate-resolution model of proteins, are used to carry out several microsecond-long simulations and study folding transition and stability of α-*de novo*–designed proteins in slit nanopores. Both attractive and repulsive interaction potentials between the proteins and the pore walls are considered. Near the folding temperature T_f and in the presence of the attractive potential, the proteins undergo a repeating sequence of folding/partially folding/unfolding transitions, with T_f decreasing with decreasing pore sizes. The unfolded states may even be completely adsorbed on the pore's walls with a negative potential energy. In such pores the energetic effects dominate the entropic effects. As a result, the *unfolded* state is stabilized, with a folding temperature T_f which is *lower* than its value in the bulk and that, compared with the bulk, the folding rate decreases. The opposite is true in the presence of a repulsive interaction potential between the proteins and the walls. Moreover, for short proteins in very tight pores with attractive walls, there exists an unfolded state with only one α-helical hydrogen bond and an energy nearly equal to that of the folded state. The proteins have, however, high entropies, implying that they cannot fold onto their native structure, whereas in the presence of repulsive walls the proteins do attain their native structure. There is a pronounced asymmetry between the two termini of the protein with respect to their interaction with the pore walls. The effect of a variety of factors, including the pore size and the proteins' length, as well as the temperature, is studied in detail. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2894299]

I. INTRODUCTION

A most important class of molecules in living cells consists of various types of proteins. Their importance to biological systems cannot be overstated:[1] they catalyze and regulate cells' activities when they act as enzymes. Muscles and other tissues are made of proteins, while as antibody proteins are a vital part of the immune system. As is well known, proteins with globular structure fold into compact configurations in which they are biologically active. An important issue is understanding the mechanisms by which proteins attain their folded structure, factors that contribute to the folding, and the environmental conditions that make the folding transition possible.[2–4] Such understanding is important not only as a scientific issue but also due to debilitating illnesses that afflict people, such as Alzheimer's and Parkinson's diseases that are believed to be the result of accumulation of toxic protein aggregates,[5–8] as well as the fact that industrial production of enzymes and therapeutic proteins based on the DNA recombinant method also involves protein folding.[9]

Although the three-dimensional (3D) structure of native proteins is controlled mostly by their amino acid sequence,[2–4] their transport properties and the kinetics of their folding depend on the local environment. But, whereas protein folding in dilute solutions under bulk conditions that are typically used in *in vitro* studies is relatively well understood, the more important problem of protein folding in a confined environment is not. For example, the environment inside a cell in which proteins fold is crowded, with the volume fraction of the crowding agents (such as RNA and ribosomes) may be in the 20%–30% range. Thus, even in the absence of interactions between proteins and other cellular molecules, the fact that a fraction of the space inside the cell is occupied implies that proteins' movement is limited, and their stability is affected.

Experiments[10,11] indicate that confinement can have a stabilizing effect on the proteins' native structure. Brinker *et al.* reported[11] that confinement of denatured proteins in the limited space of the cage model, first suggested by Anfinsen,[2] *accelerates* folding when compared with that in bulk solutions. Several groups have studied[12] many proteins with varying native-state architectures in cylindrical nanopores. These study indicated, however, that *in vivo* folding is *not* always spontaneous, rather, a subset of proteins may require molecular chaperones.

Protein (enzyme) immobilization using porous solid support, via adsorption, encapsulation, and covalent linking, has also been used for a long time.[13,14] Practical applications

a)Author to whom correspondence should be addressed. Electronic mail: moe@iran.usc.edu.

128, 115105-1

involving proteins, such as biocatalysis[15] and biosensors, also entail not only better understanding of the folding in confined media but also transport of proteins in such media. In the pharmaceutical industry, protein purification using nanoporous membranes is gaining attention.[16] Silicon-carbide (SiC) nanoporous membranes[17] allow[18] diffusion of proteins up to 29 000 Da, but exclude larger ones. Despite the fundamental and practical significance of transport of proteins in confined media, current understanding of the phenomena is limited.

The goal of this paper and its sequel is twofold: First, we use molecular dynamics (MD) simulations to study the folding and stability of an important class of proteins in slit nanopores, namely, α–*de novo*–designed proteins. We study the effect of the pore size and the nature of the interaction of the pore walls with the proteins on their folding and stability. There have already been several studies in which the entropic effect of confinement (an environment with purely repulsive walls) or crowding on protein folding and stability has been studied. A good review of the subject was given recently by Minton.[19] The main goal of such studies was to understand how the GroEL from *E. coli*—perhaps the best understood chaperone that are used in studies of cystosolic proteins—can help folding of the proteins. These studies have all reached the same conclusion, namely, that proteins are stabilized in such environments, unless the confining environment is too small, with a volume only slightly larger than that of the folded state. Computational studies of this issue include Monte Carlo[20] and (Refs. 5 and 21) simulations of proteins' behavior in confined environments. In particular, Lu *et al.*[5] and Cheung *et al.*[21] studied folding of proteins in *spherical* pores of different radii. The latter group studied the phenomenon as a function of the volume fraction of a crowding agent, modeled by a bed of hard spheres with repulsive interaction with the proteins. However, while a spherical pore may be a reasonable model for the cavity of GroEL-GroES complex, it is not so for the pores of membranes, biocatalysts, and sensors that are of prime interest to us. Instead, slit and cylindrical pores are more appropriate, particularly for the types of applications that we are interested in, namely, those in which the pore space consists of interconnected slit channels or cylindrical pores.

Compared with the case of a confined environment with repulsive walls, there have been very few studies in which the effect of attractive walls on protein folding and stability has been studied.[5,22] These studies utilized, however, cage-like structures as the model of the confined environment, which have little resemblance to the type of pore structure that we are interested in. Attractive walls are important to such applications as separation and purification of proteins by nanoporous membranes, as well as sensors and biocatalysts, in which adsorption of proteins on the pores' surface may occur.

Second, we will utilize, in Part II, a novel combination of the MD simulation and the Langevin equation to study protein transport in nanopores. To our knowledge, such a combination has never been proposed before, nor has there been any simulation of transport of proteins in nanopores

which is, in fact, of utmost importance to such practical applications as membrane purification of proteins, biocatalysis, and sensors.

In addition, the protein model that we use (see below) is, in our opinion, much more realistic than what the previous investigators[5,20–22] 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 (see below), the side chains are considered explicitly, and the model also includes hydrogen bond interactions, hence honoring the proteins' structure very realistically.

The rest of this paper is organized as follows. In the next section we describe the models of the proteins and nanopores that we utilize in our study. Section III describes the simulation method, while the results are presented and discussed in Sec. IV. The last section summarizes the paper.

II. THE MODELS

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

A. The protein model

We simulate *de novo*–designed α-family of proteins,[23] which consists of four types of amino acids in their 16-residue sequence, simplified further[24] to a sequence of hydrophobic (H) and polar (P) residues, {PPHPPHHPPHPPH-HPP}. We used the periodicity in the H-P sequence of the 16-residue peptide α_{1B}, in order to make three other sequences with lengths, $\ell = 9$, 23, and 30 residues as $\ell = 9$, with PP(HPPHHPP)$_n$, 2, 3, and 4, corresponding to protein lengths of 9, 16, 23, and 30. The four proteins have similar native structures, so that the differences in their behavior can be attributed to their lengths. The simulations (see below) indicated that they all fold onto an α-helix with $\ell - 4$ hydrogen bonds (HBs).

The proteins are modeled by the protein intermediate resolution model (PRIME),[25,26] an intermediate-resolution model which has been highly successful in reproducing several important and experimentally determined features of the proteins under bulk conditions. In order to use the PRIME in a nanopore, we have modified a few of its features which will be described below. 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 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 bonds' lengths and bond angles are fixed at their experimentally measured values, and the distance between consecutive C_α UA is also fixed according to experimental data. Table I presents all the relevant parameters of the model.

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

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

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, we mentioned above, the 3D structure of native proteins is controlled mainly by their amino acid sequences,[2–4] 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 flat 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 pore size h is varied in order to study its effect on the results described below.

III. MD SIMULATION

We have utilized discontinuous MD (DMD) simulation in order to study the dynamics of proteins in nanopores and, in particular, their folding and stability. Their transport in nanopores will be studied in Part II. In what follows we provide the details of the simulations.

A. DMD simulation

The PRIME has been designed for use with the DMD simulations. The DMD is an extremely fast alternative[27] to the classical continuous MD simulations. In particular, due to the simplicity of its time integration, the DMD simulation makes it possible to simulate the dynamics of proteins on long time scales—on the order of many microseconds, at least two orders of magnitude longer than the previous

simulations.[5,21,22] Four types of forces act on the beads: the excluded-volume effect (hard-core repulsion) and the attraction between the bonded and pseudobonded beads, between pairs of the backbone beads during the HB formation and between hydrophobic (HP) side chains. Nearest neighbor beads along the chain backbone are covalently bonded, as are the C_α and R beads or the UAs. The pseudobonds are between next nearest neighbor beads along the backbone 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 backbone N and C UAs to hold the side-chain beads fixed relative to the backbone. All of this 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 the bonded beads, separated by a distance r_{ij}, is given by

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

Here, l is the ideal bond length and $\delta=0.023\,75$ is the tolerance in the bond's length.[25,26] There are also 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

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

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, but 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 the strength ϵ_{HB}. The shape of the HB potential is similar to that of the HP potential 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. Thus, if a HB is formed between beads N_i and C_j, a repulsive interaction between the 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

TABLE II. Proteins' length ℓ, radius of gyration R_{gb} under bulk conditions, and the size of the nanopores used in the simulations. R_{gb} and the pore sizes are in nm.

ℓ	R_{gb}	h_1	h_2	h_3	h_4
9	0.459	1.50	1.89	4.40	7.55
16	0.729	1.75	3.0	7.0	12.0
23	1.02	2.44	4.18	9.75	16.7
30	1.31	3.14	5.38	12.6	21.5

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

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

Hereafter, all the energies are measured in units of ϵ_{HB}, unless specified otherwise.

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 at the same time 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 the maximum distance d_2 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),

$$
d_2 = 5.53 - 0.019T^{-1}, \quad \text{for } N_1 - C_{\alpha j},
\tag{4}
$$

$$
d_2 = 5.69 - 0.044T^{-1}, \quad \text{for } C_\ell - C_{\alpha i}.
\tag{5}
$$

There is hard-core repulsion between two unbonded beads that also have no HB and HP interactions,

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

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 develop a variant of 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 between a pore's walls and the protein beads is assumed to be

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

where z_X is z coordinate of the center of a bead X. ϵ_{PW}—assumed to be the same for all the beads—is taken to be $\epsilon_{HB}/8$, so chosen 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, $\sigma_{CX} = \frac{1}{2}(\sigma_C + \sigma_X)$ and $\epsilon_{CX} = \sqrt{\epsilon_C \epsilon_X}$, where $X = N$, C_α, C, and R. 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 (all in Å) are, $d_{3X} = 2.85$, 3.02, 3.14, and 3.31 and $d_{4X} = 0.96$, 1.01, 0.98, and 1.12, for $X = N$, C_α, C, and R, respectively.

The size of the nanopore is selected such that the ratio h/R_{gb} is the same for all the proteins of various lengths, where R_{gb} is the radius of gyration of the folded α-helix at low temperature, $T = 0.08$, in the bulk. Table II lists the relevant parameters. In addition to the cases listed in Table II, we also simulated the behavior of the proteins in the bulk. In the case of the shortest protein with length $\ell = 9$, we used a pore size larger than the diameter of cross section of the folded α-helices so that, sterically, it is possible for the protein to fold at low enough temperatures.

In all the cases but one, we simulated a pore in which there was an attractive potential U^+ between the proteins and the pore's walls. The sole exception was the case in which there was a purely repulsive interaction U^- between the pore's walls and the proteins in the smallest pore with $h = 1.75$ nm for a protein of length $\ell = 16$. Each case was studied over a wide range of temperature. To avoid trapping the system in a local energy minimum, we cooled it off smoothly from a high T, above the folding temperature T_f where the proteins are in the random coil state, to a low value. More

specifically, we cooled off the proteins of lengths of 9 and 16 from $T=0.16$, and from $T=0.17$ for those with lengths of 23 and 30.

The temperature range was divided, depending on ℓ, into 30–40 cases, nearly half of which were simulated for shorter times, while the others were simulated for longer times, in order to make sure that the shorter simulations were accurate. Due to the large fluctuations of the proteins' energy and structure near their folding temperature T_f, longer simulations were carried out there. For each case that we studied, the total simulated (real) time was about $5-7$ μs. To reduce the CPU times, we used advanced algorithms in the computations, such as link lists, neighbor lists, and false positioning which is special to the DMD. To maintain the system at a constant temperature, we used the Andersen thermostat.

Also computed was the density of states (DOS) of the proteins' potential energy at a given T, which allowed us to estimate the entropy at all the energy levels of the proteins and to compute the contributions of the different states to the free energy at any desired temperature. More details are given below.

B. The effect of the solvent: Coupling of the DMD to the Langevin dynamics

In the previous works on the effect of confinement on protein dynamics, the solvent molecules were not explicitly included in the simulations, since their viscosity η did not enter the computations. Their effect was represented by the attraction between the HP side chains. While this may be appropriate for studying the folding dynamics, it is not so for computing the proteins' diffusion coefficient. To explicitly include the solvent effect, we have developed a novel model based on a coupling between the DMD simulations and the Langevin dynamics. Since this is mostly relevant to the computation of the diffusion coefficients, we postpone its description to Part II of this series, although we also utilized it in some of the simulations described below.

C. Analyzing the results of the molecular simulations

To understand the results for each of the 21 cases studied in the bulk and the nanopores, we need an efficient and accurate method for analyzing the results of the DMD simulations. In particular, we should use a method for constructing the free energy and the potential of the mean force from the simulation results. Several methods are already available, such as the umbrella sampling,[28] multicanonical algorithm,[29] methods that are based on determining the free energy differences from nonequilibrium measurements,[30] and the weighted histogram analysis method (WHAM).[31] We used the WHAM which has been proven to be an efficient method[32] and can easily be implemented with the results of the DMD simulations. The WHAM was formulated for computing the DOS for a system at a fixed temperature. In another version of the WHAM, the simulations are carried out at different temperatures with a given Hamiltonian, and then the WHAM is used to compute the system's DOS.[33] It is this version of the WHAM that we used in the present study.

Thus, we used the WHAM to compute the DOS for the proteins, using the simulation results at all the temperatures. The DOS makes it possible to estimate the entropy at all the energy levels of the proteins and to compute the contributions of the different states to the free energy at any given temperature. Computing the entropy of the proteins enables us to calculate the thermodynamic quantities. For completeness, we provide a brief description of the WHAM and how it was implemented in our study.

Suppose that we divide the potential energy of the system into n bins, each with energy U_i, and that we have carried out simulation of m distinct cases at temperatures, T_k, $k=1,\ldots,m$. For each temperature we have a total number of N_k measurements of the energy, and for bin i of U, we have $n_{i,k}$ measurements. If τ_k is the correlation length between the consecutive steps of the simulations at T_k, we define $g_k=1+2\tau_k$. According to the WHAM, one estimates the DOS for U_i, with the minimum error, by

$$W(U_i) = \frac{\sum_{k=1}^m g_k^{-1} n_{i,k}}{\sum_{k=1}^m g_k^{-1} N_k \exp(\beta_k U_i - u_k)}, \tag{8}$$

where u_k is the free energy at T_k and $\beta_k=1/k_B T_k$, with k_B being Boltzmann's constant. If we define $P(U_i,\beta_k) \equiv W(U_i)\exp(\beta_k U_i)$, then, we obtain

$$\exp(u_k) = \sum_{i=1}^n P(U_i,\beta_k), \tag{9}$$

and by iterating the above two equations we estimate the DOS. Physically, $P(U_i,\beta_k)$ is the probability of finding energy U_i at temperature T_k, which is computed using the simulation results. Assuming the error of the $n_{i,k}$ measurements being of the order of $\sqrt{n_{i,k}}$, we obtain an estimate for the error $\delta P(U_i,\beta)$ of computing the probability $P(U_i,\beta)$ at (inverse) temperature $\beta=(k_B T)^{-1}$,

$$\delta P(U_i,\beta) = \left[\sum_{k=1}^m g_k^{-1} n_{i,k} \right]^{-1/2} P(U_i,\beta). \tag{10}$$

Having computed the DOS $W(U_i)$, we obtain the average potential energy at temperature T by

$$\langle U \rangle = \frac{\sum_{i=1}^n U_i W_i \exp(-\beta U_i)}{\sum_{i=1}^n W_i \exp(-\beta U_i)} = \frac{\sum_{i=1}^n U_i \exp(-\beta E_i)}{\sum_{i=1}^n \exp(-\beta E_i)}. \tag{11}$$

The second equality holds because the entropy of the energy level i is defined by, $S_i=k_B \ln(W_i)$, and we define the free energy of the ith level as $E_i=U_i-TS_i$. In a similar way, one computes the specific heat C_V at temperature T by $C_V=(k_B)^{-1}d\langle U \rangle/dT=(\langle U^2 \rangle - \langle U \rangle^2)/(k_B T)^2$, in which

$$\langle U^2 \rangle = \frac{\sum_{i=1}^n U_i^2 W_i \exp(-\beta U_i)}{\sum_{i=1}^n W_i \exp(-\beta U_i)}. \tag{12}$$

Equation (12) holds even if the energy in the bin i, $i=1,\ldots,n$ is in a 2D or 3D phase space, with one axis being
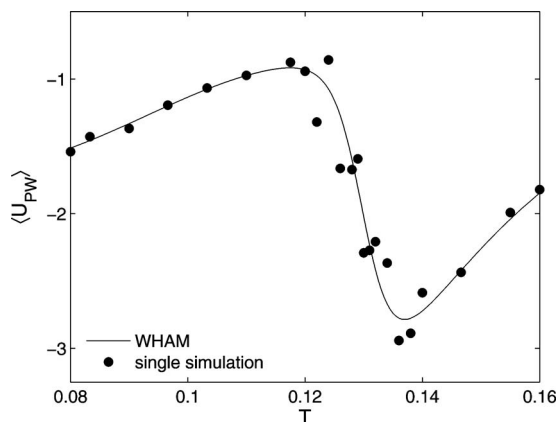
FIG. 1. Comparison of the computed temperature dependence of the average interaction potential $\langle U_{PW} \rangle$ between the protein and the pore walls using the WHAM and that obtained by a single simulation.

the energy and other two axes being other quantities, because in this case U_i has one value for each bin.

Thus, we may construct a 2D phase space, with one axis being the potential U and the other a quantity such as n_α, the number of α-helical HBs, or ϵ_{PW}. In this case all the bins, each with two indices indicating the two quantities, may be reindexed by only one, ranging from 1 to the total number of the variables. Then, all the above formulas hold, and we can compute the entropy and, hence, the free energy of the proteins, and also determine the statistical averages of the thermodynamic quantities at a given temperature. Similarly, we may construct a 3D phase space with, for example, the variables being n_α, ϵ_{PW}, and U and use the WHAM to determine the DOS in such a phase space.

The WHAM smoothens the results when using the data from all the simulations. Figure 1, for example, presents the average interaction potential $\langle U_{PW} \rangle$ between a protein and the walls of the pore, computed using the WHAM, with that calculated using a single simulation. Far from the folding temperature T_f the results are very similar. Near T_f, however, the data from a single simulation differ from those computed using the WHAM due to the large fluctuations in the potential energy. Figure 1 also provides evidence for the accuracy of the WHAM results.
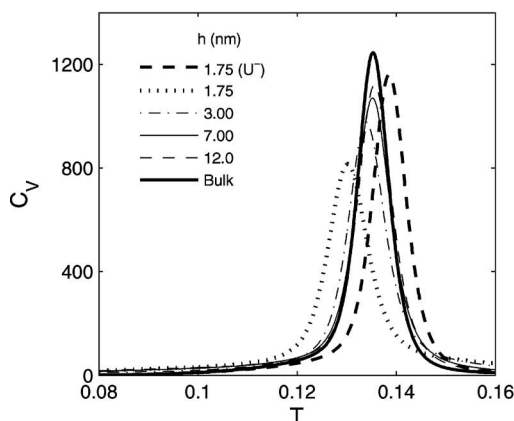
IV. RESULTS AND DISCUSSIONS

Due to the complexity of the dynamics of proteins in confined environments, there are many aspects of their behavior that one must study and, when meaningful, compare them with those under bulk conditions. In what follows we take up such issues and describe the results.

A. Folding temperature in nanopores

We first describe computation of the folding temperature T_f and the effect of confinement on T_f, which is of central importance to the rest of our discussions. The folding temperature T_f may be defined as the temperature at which (i) the specific heat C_V is maximum, (ii) the variance χ of the number n_α of α-helical HBs is maximum, (iii) $d\langle n_\alpha \rangle / dT$ is maximum, (iv) $\langle n_\alpha \rangle = 1/2$, or (v) the probabilities of being in the folded and unfolded states are equal or, equivalently, the free energy of the two states are equal, although the most commonly used definition is the first one. We estimated T_f for several pore sizes and proteins' lengths using the five methods together with the WHAM. All the methods yielded essentially the same estimates of T_f.

Figure 2, for example, displays the specific heat C_V and the variance χ of the number of the α-helical HBs, $\chi = \langle n_\alpha^2 \rangle - \langle n_\alpha \rangle^2$, of the protein of length $\ell = 16$ and their dependence on T, computed by the WHAM, where they are also compared with the results in nanopores of various sizes. The specific heat is in units of k_B. It has only one peak, indicating the two-state nature of its statistics. The locations of the peaks provide estimates of the folding temperature T_f. Similarly, the location of the maximum in χ provides an estimate of T_f.

The estimates of T_f for several pore sizes h, using the method (i), are shown in Fig. 3, which displays T_f versus $1/h$ (so that the intercept with the vertical axis represents the bulk). Regardless of how it is estimated, T_f of the proteins in nanopores with *attractive* walls *decreases*, relative to the bulk condition, and is smaller for smaller pores. This implies that there exists a temperature T with $T_{f,pore} < T < T_{f,bulk}$, at
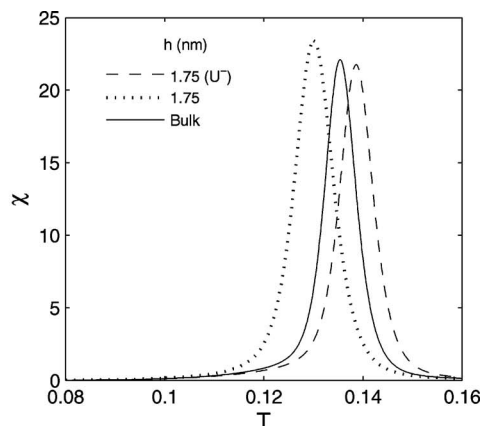

FIG. 2. The specific heat C_V and the variance χ of the number of α-helical HBs vs temperature for a protein of length $\ell = 16$ under bulk condition and in the pores.
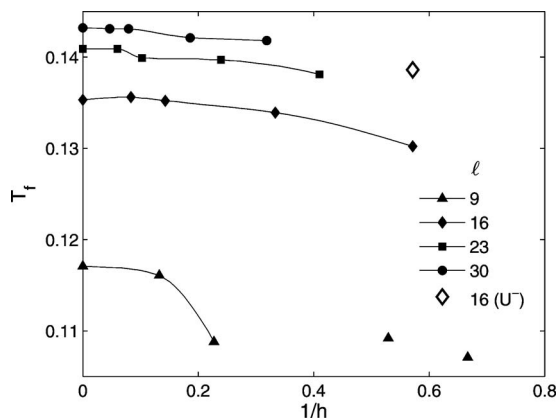
FIG. 3. Folding temperature T_f of proteins of length ℓ vs the pore size h. All the results are for attractive walls, except where noted by U^- for the repulsive ones.

FIG. 5. Average number $\langle n_\alpha \rangle$ of α-helical HBs vs T for a protein of length $\ell = 16$ under bulk condition and in the pores and computed using the WHAM. The circle indicates the location of the folding temperature.

which a protein is folded in the bulk, but not so in the nanopore. Therefore, one may say that the protein has been destabilized in a nanopore with attractive potential U^+. In a pore with purely repulsive walls, however, T_f *increases* relative to the bulk and, therefore, the protein is stabilized. In Sec. IV D we discuss this phenomenon in more detail.

B. α-helix folding in the bulk

Although the behavior of proteins in the bulk has been studied before, in this section we describe briefly some of the results under the bulk conditions, since they will be compared with those obtained in the nanopores. The simulated proteins in this work are single-domain structures and have a cooperative (two-state) folding dynamics. The DMD simulations indicate that the proteins that we consider (with lengths, $\ell = 9$, 16, 23, and 30) fold onto an α-helix configuration at low temperatures. To describe the results quantitatively, we use n_α, the number of α-helical HBs as the order parameter for the folding. Time dependence of n_α and the corresponding protein configurations are presented in Fig. 4.
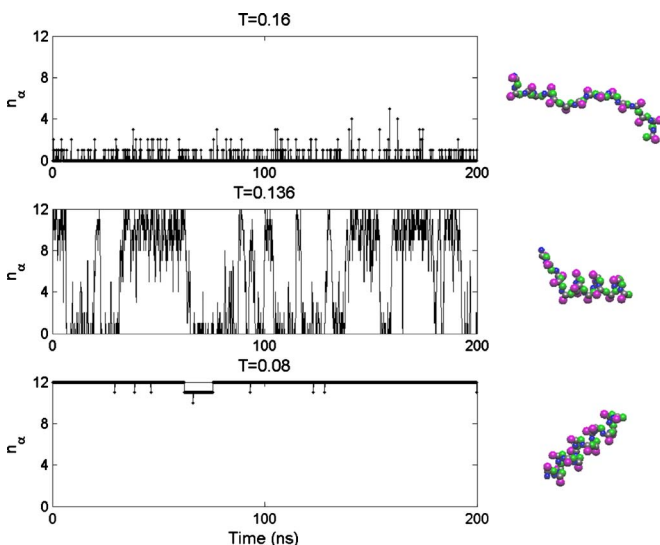

FIG. 4. (Color online) The instantaneous number n_α of the α-helical HBs for a protein of length $\ell = 16$ under bulk conditions. Side chains are shown in darker gray. All other beads are in the same brighter gray color.
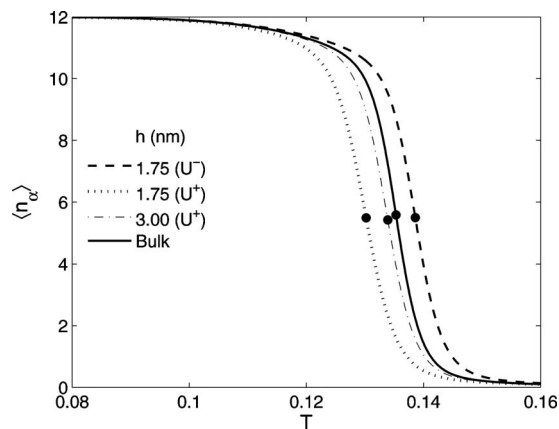
for three temperatures that are lower than, close to, and higher than the folding temperature, $T_f \simeq 0.135$, for the protein of length $\ell = 16$.

Proteins resemble a random coil at high enough temperatures and have no definite structure. Even if a HB is formed at such temperatures, due to the high kinetic energy, it has a short life. As the temperature is lowered, the α-helical HBs last longer and, hence, local α-helical structures are more likely to emerge, the formation of which is also aided by the fact that when a native α-helical HB is formed, it also helps other α-helical HBs to form in its surrounding. Close to T_f, the protein changes its structure regularly, from completely unfolded to partially and completely folded. Finally, at low enough temperatures the protein is always in the folded state with the nearly entire set of α-helical HBs formed.

Temperature dependence of the average $\langle n_\alpha \rangle$, computed by the WHAM, is shown in Fig. 5 for a protein of length $\ell = 16$, where it is also compared with those in a nanopore (see also below). It is clear that definite changes in the structure of the protein occur near T_f, which is clearly indicated by the changes in $\langle n_\alpha \rangle$.

Figure 6(a) presents temperature dependence of the average potential energy $\langle U \rangle = \langle U_{HB} + U_{HP} \rangle$, computed by the WHAM. All the energies are in units of ϵ_{HB}. As in the case of $\langle n_\alpha \rangle$, there is a sharp change in the average potential energy of the system at T_f.

Figure 7 depicts the free energy of the protein of length $\ell = 23$ and its relation with n_α and temperature in a 2D his-
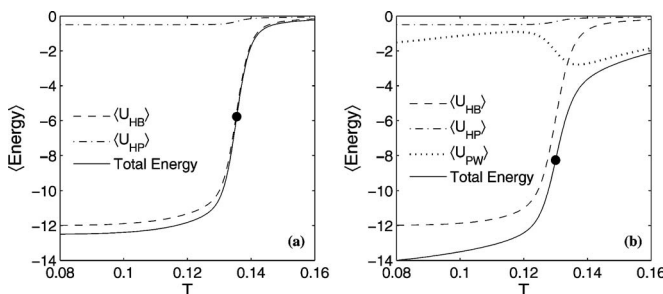

FIG. 6. Temperature dependence of the various contributions to the total energy for a protein of length $\ell = 16$ in (a) the bulk and (b) in a pore of size $h = 1.75$ nm. The circles indicate the location of the folding temperature.
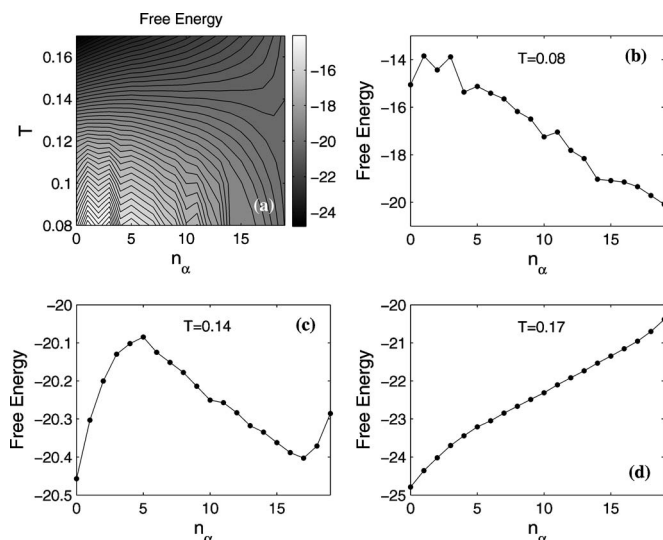
FIG. 7. Free energy of a protein of length $\ell = 23$ under bulk condition, as a function of the order parameter n_α and the temperature T.
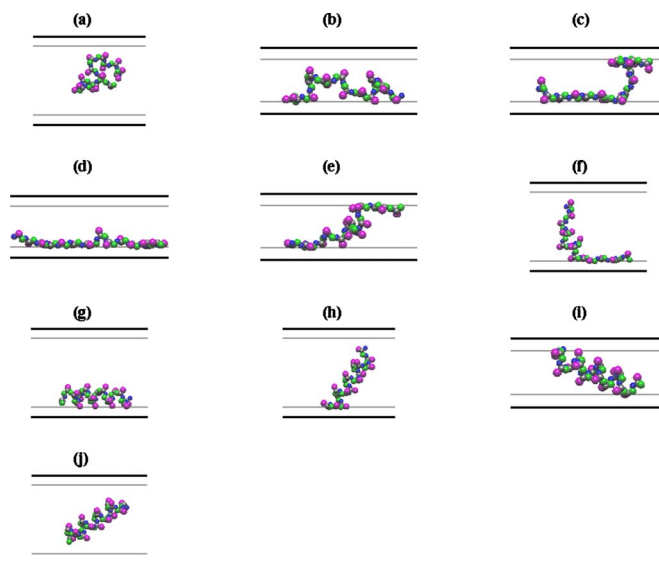

FIG. 8. (Color online) Protein configurations in a nanopore. In between the walls and the thin lines at a distance d_3, $U_{PW} = \infty$, beyond which the attractive potential U^+ acts for a distance d_4. (a) High-temperature, random coil structure. [(b)–(d)] Partial or complete adsorption on the walls. (e) Adsorption on the two walls at a temperature lower than that in (a)–(d). (f) Partially unfolded state near the folding temperature T_f. [(g)–(i)] Folded adsorbed states at $T < T_f$. (j) Folded state in the middle of the pore for a protein with large kinetic energy.

togram (computed using the WHAM). Two states of the protein, which have minimum free energy and are stable at low and high enough temperatures, are the folded and unfolded states. The potential energy of the folded state is at its minimum and, hence, is stable at low temperatures [Fig. 7(b)], but the unfolded state has high entropy and, thus, is stable at high temperatures [Fig. 7(d)], where the entropic part of the free energy dominates. Near the folding temperature T_f the free energies of the two states do not differ much, with an energy barrier existing between them, which is usually referred to as the transition-state ensemble [Fig. 7(c)] (see also below).

C. α-helix folding in a nanopore

At high temperatures a protein is similar to a random coil; see Fig. 8(a). However, the pore's walls restrict the number of its possible unfolded states at such temperatures. Due to the attractive potential between the nanopore's walls and the protein's atoms, portions of the protein may sometimes be adsorbed onto the walls. The protein may even be completely adsorbed on one wall. These are shown in Figs. 8(b)–8(d).

As the temperature is lowered, it becomes possible for more α-helical "turns" to appear, which attach the protein to the walls only laterally. In fact, in this case not all the protein's atoms that are in such turns may be in contact with the walls due to the definite structure of the protein; see Fig. 8(e). Near T_f the protein changes its state frequently, from completely folded to completely unfolded states. In contrast to the bulk conditions, however, a new phenomenon occurs near T_f, namely, attraction to the walls may give rise to an unfolding nucleus, because if the neighboring atoms of one part of the protein attach themselves to a wall, they will lose their native α-helical HBs. Such a configuration is presented in Fig. 8(f). Thus, it appears that the attractive interaction of the protein with the walls competes with folding to α-helix configuration. Hence, not only does the interaction

with the nanopore's walls disturbs folding but also folding to the definite α-helix structure disturbs the protein's interaction with the walls.

At still lower temperatures, the protein attains its native state and may be attached to one wall laterally or to one or both walls through its ends; see Figs. 8(g)–8(i). Moreover, if the protein has enough kinetic energy, it may overcome the attractive potential with the walls, as shown in Fig. 8(j). Thus, overall, the protein undergoes a repeating sequence of folding/partially folding/unfolding transitions, depending on the temperature and the nature of the interaction between the proteins and the pore's walls.

Figure 6(b) presents temperature dependence of the average potential energy and its various components for a protein of length $\ell = 16$, in the smallest pore simulated, namely, one with size $h = 1.75$ nm, and compares it with that of the same protein in the bulk [Fig. 6(a)]. As, expected, $\langle U \rangle$, $\langle U_{HP} \rangle$, and $\langle U_{HB} \rangle$ always decrease by decreasing T, but $\langle U_{PW} \rangle$ behaves differently. Near T_f U_{PW} exhibits an unexpected increase by decreasing T.

This behavior is shown in more detail in Fig. 9. Cooling the proteins at $T > T_f$ increases $|\langle U_{PW} \rangle|$, as well as the average number $\langle n_\alpha \rangle$ of the α-helical HBs which is, however, very small. Near the folding temperature $\langle n_\alpha \rangle$ is no longer negligible. In this range of T the proteins can only laterally attach themselves to the walls, hence decreasing $|\langle U_{PW} \rangle|$. Therefore, although the temperature is lower, but due to the internal interactions in the α-helix and the new structures that are formed in it, $|\langle U_{PW} \rangle|$ decreases. By lowering T further, nearly the entire α-helix is formed and $|\langle U_{PW} \rangle|$ increases again.

Using the WHAM, we also computed the DOS in the 3D phase space of $\langle n_\alpha \rangle$, U_{PW}, and the free energy which, as
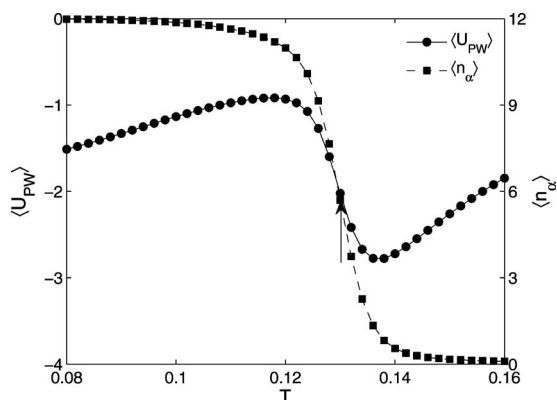
FIG. 9. The average interaction energy $\langle U_{PW} \rangle$ between a protein of length $\ell=16$ and the walls of a pore of size $h=1.75$ nm. The arrow indicates the location of the folding temperature.

FIG. 11. (Color online) Configurations of a protein of length $\ell=9$ in nanopores of sizes of 1.5 and 1.89 nm at $T=0.08$. The protein has not folded.

described above, was used to compute the free energy in the 2D phase space of $\langle n_\alpha \rangle$ and U_{PW} at any temperature; see Fig. 10. The free energy minima with high values of $\langle n_\alpha \rangle$ correspond to smaller values of $\langle U_{PW} \rangle$ and vice versa. Thus, Fig. 10 demonstrates clearly the competition between the quantities.

Folding of proteins through intermediates and metastability in protein folding have been studied widely under the bulk conditions, taking into account only the internal interactions in a protein.[34] Here, we present a case in a nanopore in which a state different from the true native state of the protein turns out to be at the minimum free energy at low temperatures. The effect is not related to the internal interactions of the protein, rather it is due to the energetic interactions of the protein with the pore walls and the entropic effect of confinement in very tight pores.
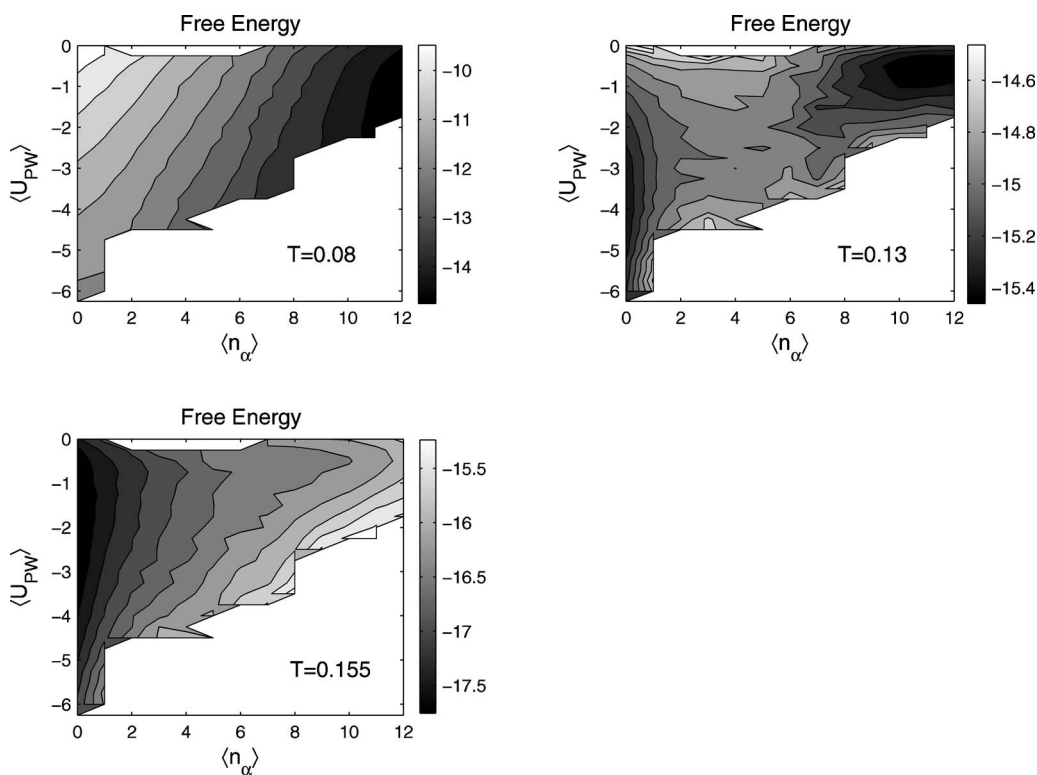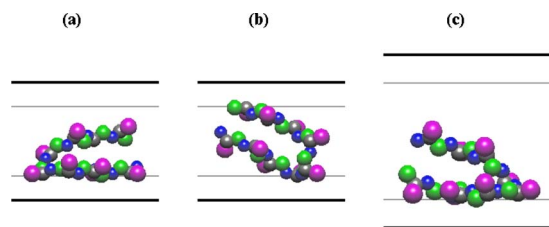
In two of the smallest pores that we have considered, namely, those with sizes $h=1.5$ and 1.89 nm, a small protein of length $\ell=9$ does not attain its true native state at low temperatures. Instead, it takes on a U shape with one or both of its end sides attached to the walls. This is shown in Fig. 11. The protein has four HBs, only one of which is α-helical (the native state has five α-helical HBs), and more of its atoms are close to the walls than those in the folded state. Although the potential energy of such unfolded states is nearly the same as those in the folded one, entropic effects which favor the unfolded states are also important. Upon further cooling at temperatures below the *apparent* folding temperature T_f, the protein becomes trapped in its U shape without having enough kinetic energy to overcome the energy barrier for attaining a folded state. Thus, such configurations do not represent truly folded states.

To understand better the nonfolding of the configurations in Fig. 11, we analyze the system more quantitatively. The potential $\langle U_{PW} \rangle$ at $T=0.08$ decreases from about $-0.9\varepsilon_{HB}$ in large pores to nearly $-1.9\varepsilon_{HB}$ in the two smallest pores that we consider. This implies that the decrease in $\langle U_{PW} \rangle$ compensates for the increase in $\langle U_{HB} \rangle$. For the system shown in
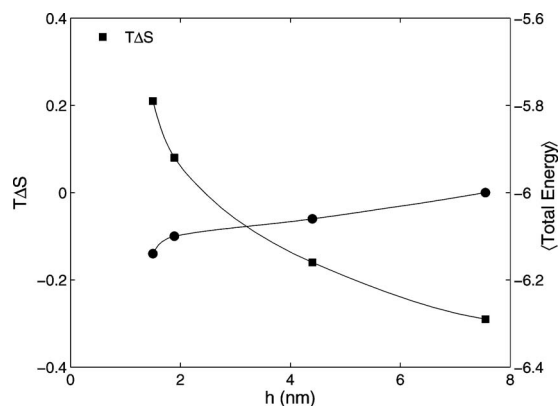

FIG. 10. Phase space for a protein of length $\ell=16$ in a pore of size $h=1.75$ nm.

FIG. 12. The total average energy $\langle E \rangle$ and entropy difference $\Delta S = S_U - S_f$ for all pores of size h and a protein of length $\ell = 9$. Entropy of the unfolded state is larger than that of the folded state in the two smallest nanopores. Moreover, the potential energy of the unfolded state in the two smallest nanopores is slightly smaller than that of the folded state in two largest nanopores.
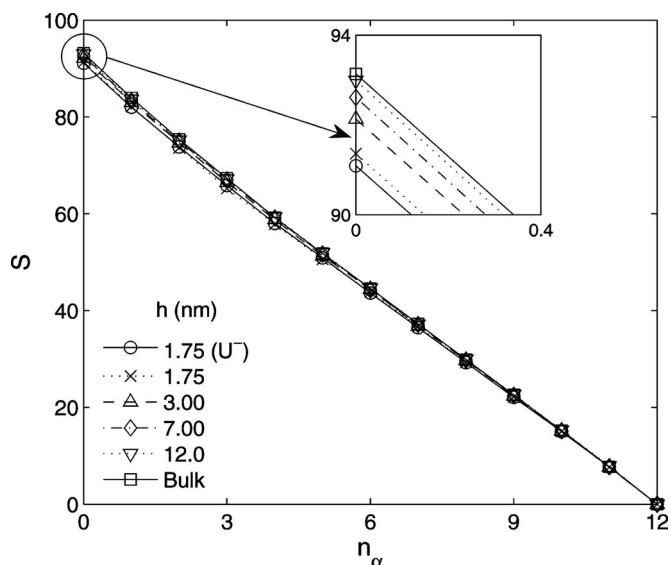

FIG. 13. Entropy of a protein of length $\ell = 16$ (in units of k_B, Boltzmann's constant) vs the order parameter n_α for various pore sizes. The entropy of the folded state ($n_\alpha = 12$) was set to zero for all the pore sizes and in the bulk. The inset shows the results for the unfolded state ($n_\alpha = 0$), indicating separation of the different cases shown.

Fig. 11, the total average energy, $\langle U \rangle = \langle U_{HB} + U_{HP} + U_{PW} \rangle$, of the unfolded U-shaped protein in the two smallest pores is even slightly smaller than $\langle U \rangle$ for the folded protein in larger pores. We also estimated the entropy of the unfolded U-shaped state and the folded state by summing over all the densities of state with $n_\alpha = 4$ and $U_{PW} \leqslant -1$ and $n_\alpha = 5$ and $U_{PW} \leqslant 0$. The differences between the entropies of the two states, i.e., $\Delta S = S_U - S_f$, were then computed as a function of the pore size h and are shown in Fig. 12 where, in order to make a direct comparison with the total energy, we present $T\Delta S$ ($T = 0.08$). In the small pores the entropy of the U-shaped unfolded state is larger than that of the folded state. However, while the U-shaped configuration can exist only near the walls, the folded state can exist without having any interaction with the walls. Hence, in the larger pores, the entropy of the folded state becomes larger because more free space is available for the protein between the walls.

Thus, the existence and stability of such unfolded states are directly due to the attractive interactions between the walls and a protein's atoms, which lower the energy of a non-native state and, therefore, help it compete strongly with the native state. The DMD simulations also indicate that the same protein in the same pore sizes, but with repulsive walls, can attain its native state at low temperatures.

D. Entropic effects

Using the WHAM, we estimated the DOS of the proteins as a function of n_α, the order parameter for the folding, in order to study the effect of the pore size and the type of its walls' interaction (attractive versus repulsive) on the proteins' entropies of the folded and unfolded states.

To compute the dependence of the entropy on n_α, we construct a 2D phase space with U and n_α being the main variables and compute the DOS in the 2D space using the WHAM (see above). Then, in order to estimate the DOS for a given n_α, we sum over all the DOSs with n_α held fixed at its specific value, but for different U. While most of the contributions to the potential energy are by the α-helical HBs, the values of U with a significant DOS and a fixed n_α is not spread out and, hence, the estimate is rather accurate. Then,

determining the DOS versus n_α is equivalent to computing the entropy versus this parameter.

Figure 13 presents the entropy of a protein of length $\ell = 16$ in the bulk and in pores of various sizes versus the folding order parameter n_α. For all the cases the entropy of the folded state ($n_\alpha = 12$) was set to zero, so that the difference between the entropies of the unfolded and folded states, $\Delta S = S_u - S_f$, shown in the inset of Fig. 13, becomes clearer. In contrast with the bulk, the entropy difference ΔS is smaller for the smaller pores. Since the entropy difference is the reason for the stability of unfolded states at high T, the implication is that in nanopores the unfolded states are less stable than the same proteins in the bulk. A folded state has a unique compact structure with a very low entropy, but the unfolded states have no definite structure and are more extended. Therefore, confining a protein in a nanopore decreases the entropy of its unfolded states much more than that of the folded ones. As a result, ΔS for proteins in a nanopore is smaller than the corresponding value in the bulk. Moreover, smaller pores result in smaller ΔS. This effect stabilizes proteins in a pore, hence increasing its T_f, as in the case of a protein of length $\ell = 16$ in a pore with a repulsive potential U^- and size $h = 1.75$ nm, shown in Fig. 13. In fact, for a pore of a given size, ΔS is smaller with repulsive walls than one with attractive walls, since the effect of confinement is stronger with repulsive walls.

To make such statements more quantitative, consider, $\delta S = \Delta S_p - \Delta S_b$, where the subscripts b and p denote the bulk and pore conditions, respectively. The DMD simulations indicated that for a pore of size h and the protein of size $\ell = 16$ one has,

$$-\delta S \propto h^{-1} \tag{13}$$

(the implied prefactor is about 3.2) which may be compared with

$$-\delta S \propto h^{-5/3}, \tag{14}$$

derived by de Gennes[35] for the entropy loss of a polymer chain confined to a tube of size $h < R_f$, where R_f is the Flory radius[35] of the chain, an estimate for the root mean square end-to-end distance of a flexible linear polymer, which is usually different from, but of the same order of magnitude as, R_g, the polymer's radius of gyration.

While in the DMD simulations the pore size h is comparable with or larger than the radius of gyration R_g of the protein at high T, we may expect an exponent smaller than the $5/3$ derived by de Gennes[35] for the chains. He argued that the entropy loss of a linear polymer confined in a tube of size $h < R_f$ should vary linearly with its length, based on which he derived the $5/3$ exponent which is, in fact, $1/\nu$, where ν is the exponent of 3D linear polymers, $R_f \sim N^\nu$, with N being the number of monomers in the polymer. Because in the simulations $h \gtrsim R_f$, we cannot *a priori* assume that the linear proportionality of the entropy loss with the proteins' length holds.

Thus, suppose that we have a fixed confining volume with dimension $h > R_f$ and imagine two cases: (i) a polymer (protein) with N monomers and (ii) two polymers that are far apart, each with length $N/2$. The total entropy loss in the second case is larger than that in the first case because in the second case we have configurations in which one polymer is far from the walls, while the other polymer is near the walls, hence reducing the entropy. In the first case, however, when the first half of the polymer is far from walls, so also is the other half. Therefore, the entropy loss of a polymer with length N is less than twice the entropy loss of a polymer of length $N/2$, implying that $\delta S \propto N^p$, where $p < 1$. Now, adopting de Gennes analysis, we expect the entropy loss of the polymer to be a function of h and $R_f = aN^{3/5}$ only (where a is the typical dimension of the monomers),

$$\delta S \cong \varphi_S(R_f/h), \tag{15}$$

where φ_S is a scaling function. But, since $\delta S \propto N^p$, if we assume a power-law relation for φ_S, we obtain,

$$\delta S \cong N^p(a/h)^{5p/3}, \tag{16}$$

implying a power of h which is smaller than $5/3$, consistent with Eq. (13). Note that the exponent of 1.0 is even smaller than $4/3$, the exact value of ν for 2D linear polymers.

The entropy differences can be used for obtaining accurate estimates of the folding temperature T_f. Consider, first, the bulk state. Recall that one way of estimating T_f is by defining it as the temperature at which the free energies of the folded and unfolded states are equal. Thus, one has $E_f - T_f S_f = E_{uf} - T_f S_{uf}$, so that $T_f = (E_{uf} - E_f)/(S_{uf} - S_f) = (E_{uf} - E_f)/(\Delta S_b + \delta S)$. Since under the bulk conditions δS vanishes, one has, $T_{fb} = (E_{uf} - E_f)/\Delta S_b$. But, the potential energy of the unfolded states is negligible and, therefore, we obtain

$$T_{fb} \simeq -\frac{E_f}{\Delta S_b}. \tag{17}$$

If the numerical values of the relevant quantities, computed by the DMD simulations, are used in Eq. (17), we obtain estimates of T_{fb} for proteins of length ℓ. In Fig. 14 compares
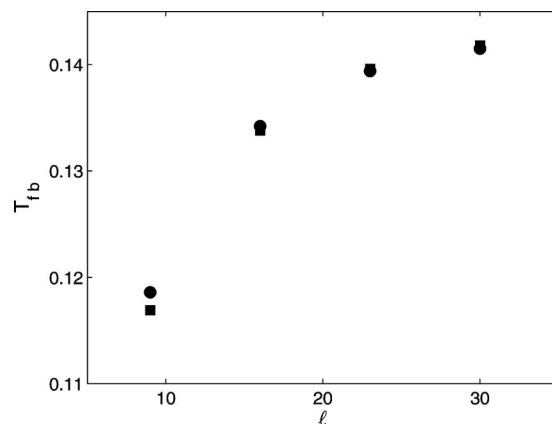

FIG. 14. Folding temperature T_{fb} in the bulk estimated by Eq. (17) (circles) and its comparison with those obtained directly from the WHAM (squares).

such estimates with those obtained directly by the free energy calculations (see above); the agreement is excellent.

Now, we use the same type of analysis to discuss what happens to the folding temperature T_f in a pore with repulsive walls—one in which only the effect of confinement is present, as a protein has no interaction energy with the walls other than the hard-core repulsion. Hence, we have $\delta S \ll \Delta S_b$, so that we can write $(T_f - T_{fb})/T_{fb} = 1/(1 + \delta S/\Delta S_b) - 1$ or

$$\frac{T_f - T_{fb}}{T_{fb}} \simeq -\delta S/\Delta S_b. \tag{18}$$

In the case of a pore with repulsive walls that we have studied (with size $h = 1.75$ nm and a protein of length $\ell = 16$), we have $-\delta S/\Delta S_b \simeq 0.022$, which is in complete agreement with the estimate obtained directly from the DMD simulations, $(T_f - T_{fb})/T_{fb} \simeq 0.025$.

E. Energetic effects

Many previous studies of the behavior of proteins in confined environments with repulsive walls[12,20,21] have stated that the entropy loss in such environments causes the instability of the unfolded states and, hence, the stability of proteins' folded states. In such environments, however, only entropic effects are important. We now describe a physical effect that stabilizes the *unfolded* states of proteins in nanopores with attractive walls, which we show competes with the entropy loss and wins over it.

As discussed above, in contrast with the folded states the unfolded states, due to their structure interact more strongly with the walls of a nanopore. To quantify this statement, we computed $\langle U_{PW} \rangle$ as a function of the order parameter n_α at the folding temperature T_f. The results for $\Delta \langle U_{PW} \rangle$ in nanopores of sizes $h = 1.75$, 3, 7, and 12 nm and the proteins of length $\ell = 16$ are -2.8, -2.2, -1.4, and -1.1, respectively. All the energies are in units of ϵ_{HB}. They indicate that, compared with the folded states with high n_α, $\langle U_{PW} \rangle$ is lower for the unfolded states with low n_α. While this part of the potential energy is present only in a nanopore, it is exactly what gives rise, relative to the bulk, to more stable unfolded states in the pore. In a pore with attractive walls, this effect is even
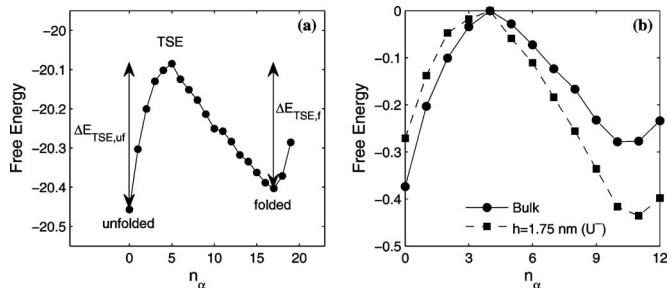
FIG. 15. (a) Schematic of the free energy of a protein, showing $\Delta E_{\text{TSE,f}}$ and $\Delta E_{\text{TSE,uf}}$. (b) Free energy vs order parameter n_α for a protein of length $\ell = 16$ in the bulk and in a pore with repulsive walls, at $T = 0.135$.

FIG. 16. Temperature dependence and comparison of the free energy changes ΔE (in units of $k_B T$) for the bulk state and for a pore with repulsive walls for a protein of length $\ell = 16$. (a) For the unfolded state and (b) for the folded state. Symbols indicate the folding temperature of each case, while TSE denotes the TSE. The pore size is $h = 1.75$ nm.

stronger than that of the entropy, such that it can even destabilize a folded state of a protein and reduce its T_f. If ε_{PW}, the strength of the interaction between the protein and the pore's walls, is reduced, this effect would weaken to the extent that the protein may be stabilized again in the nanopore.

Let us now define

$$\Delta \langle U_{\text{PW}} \rangle = (\langle U_{\text{PW}} \rangle_{\text{uf}} - \langle U_{\text{PW}} \rangle_{\text{f}})_{T=T_f}$$
$$= (\langle U_{\text{PW}} \rangle_{n_\alpha=0} - \langle U_{\text{PW}} \rangle_{n_\alpha=\ell-4})_{T=T_f},$$

as the difference between the potential energy of interaction with a nanopore's walls of the folded and unfolded states of a protein; it is a negative quantity. The DMD simulations for pores of size h and the protein of length $\ell = 16$ also indicate that

$$-\Delta \langle U_{\text{PW}} \rangle \propto h^{-0.5}. \tag{19}$$

F. Folding and unfolding rates

The rates of folding and unfolding of proteins are strongly dependent on the confinement of a nanopore, as well as the nature of the proteins' interaction with the walls. The rates are defined by $k_f = 1/\tau_f$ and $k_{uf} = 1/\tau_{uf}$, where τ_f and τ_{uf} are the average folding and unfolding times of the proteins. Suppose that the free energy difference between a folded state and the transition-state ensemble (TSE) is $\Delta E_{\text{TSE,f}}$ and similarly between an unfolded state and the TSE is $\Delta E_{\text{TSE,uf}}$. The folded state is the local minimum of E at a high value of the order parameter n_α; the TSE is local maximum for an intermediate value of n_α, whereas the unfolded state corresponds to the local minimum of E at a small value of n_α. These are shown schematically in Fig. 15(a).

According to Mirny and Shakhnovich,[2] the folding and unfolding rates of a protein may be estimated as $k_f \simeq c \exp(-\Delta F_{\text{TSE,uf}})$ and $k_{uf} \simeq c \exp(-\Delta E_{\text{TSE,f}})$, where c is a constant. (More precise estimators also exist, but we only wish to discuss qualitatively the effect of a pore size and the nature of the interaction with its walls on the folding and unfolding rates.) Therefore, we may use the WHAM to compute the free energy landscape at various T in order to estimate k_f and k_{uf} in units of the constant c.

Figure 15(b) presents the free energy of a protein of length $\ell = 16$ in the bulk and in a pore of size $h = 1.75$ nm with repulsive walls at $T = 0.135$. For both cases the free energy of the protein was set to 0 for $n_\alpha = 4$ which represents
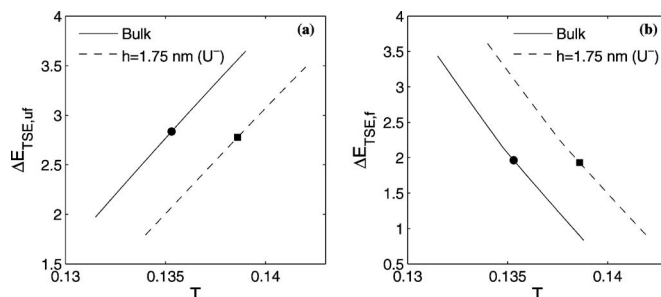
the TSE. It is clear that $\Delta E_{\text{TSE,f}}$ in the nanopore is larger than that in the bulk, whereas the opposite is true about $\Delta E_{\text{TSE,uf}}$.

To explain these results, we note that confinement affects the unfolded states much more strongly than folded ones because the structure in the latter is compact, whereas the unfolded configurations are more extended. The TSEs have some native α-helical HBs, as they are partially folded and, thus, they are not as extended as the unfolded states, but also not as compact as the folded states. We also note that, in a pore with purely repulsive walls, the interaction with the walls does not affect a protein's potential energy; only entropic effects are important. Moreover, most of the contribution to the free energy of a folded state is due to its low potential energy, but for the TSE and unfolded states most of the contributions are related to the entropy. Therefore, in a pore with purely repulsive walls, the free energy of a folded state is not affected much by confinement, but that of the TSE increases somewhat, whereas the free energy of the unfolded states increases the most due to their entropy changes in the nanopore.

Therefore, we expect the confinement in a nanopore to reduce the entropy of the TSEs, but not as much as that of the unfolded states. As a result, in a pore with purely repulsive walls, the entropy of the folded state is affected the least, the entropy of the TSEs is reduced somewhat, whereas that of the unfolded states is reduced the most. More quantitatively, we compare $\Delta S_{\text{TSE,f}} = S_{\text{TSE}} - S_f$ and $\Delta S_{\text{uf,f}} = S_{\text{uf}} - S_f$ in a pore of size $h = 1.75$ nm with repulsive walls with those under the bulk condition, for the protein length of 16 (see Fig. 13). In units of k_B, $\Delta S_{\text{TSE,f}}$ and $\Delta S_{\text{uf,f}}$ are, respectively, 59.2 and 93.1 for the bulk conditions and 58.0 and 91.1 for the pore with repulsive walls, hence demonstrating the effect of entropy on the TSE and the unfolded states in a pore with repulsive walls. Hence, as Fig. 15(b) indicates, in contrast with the bulk state, $\Delta E_{\text{TSE,f}}$ increases, whereas $\Delta E_{\text{TSE,uf}}$ decreases in a nanopore with purely repulsive walls.

Figure 16(a) presents the computed $\Delta E_{\text{TSE,uf}}$ (in units of $k_B T$) versus the temperature for a protein of length $\ell = 16$ in the bulk and in a pore with repulsive walls. Also shown are the corresponding T_f. Since one has $\Delta E_{\text{TSE,uf}}/k_B T \simeq \ln(c\tau_f) \simeq -\ln(k_f/c)$, τ_f—the folding time—at any T is smaller in a pore with repulsive walls or, equivalently, k_f—the folding
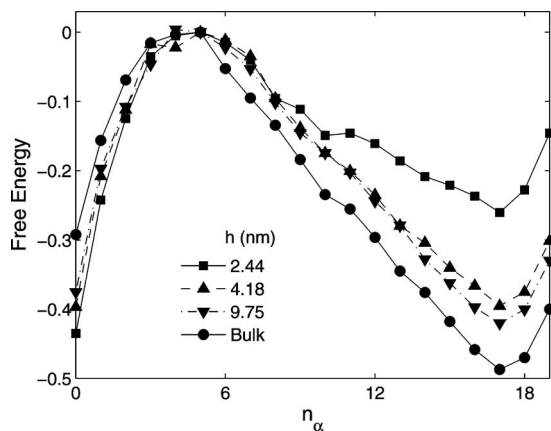
FIG. 17. Free energy vs the order parameter n_α for a pore of size h with attractive walls and in the bulk for a protein of length $\ell = 23$ at temperature $T = 0.138$.

rate–is larger than its bulk value. Figure 16(b) presents $\Delta E_{\text{TSE,f}}$ (in units of $k_B T$), where the opposite trends are presented.

In Fig. 17 we display the free energy of a protein of length $\ell = 23$ in the bulk and in pores of various sizes with attractive walls at temperature $T = 0.138$. For all the cases the free energy of the protein was set to 0 for $n_\alpha = 5$ which represents the TSE. The trends in Fig. 17 are opposite of those for the case of a pore with repulsive walls. That is, $\Delta E_{\text{TSE,f}}$ in a pore with attractive walls is *smaller* than that in the bulk, whereas $\Delta E_{\text{TSE,uf}}$ is greater.

To explain these results we recall that in a pore with attractive walls the interaction energy with the walls is much larger than the entropic effects. In such a pore the unfolded states interact much more strongly with the walls than the folded state which has a strictly compact configuration. Moreover, forming the α-helical structures reduces $|\langle U_{\text{PW}} \rangle|$. While the TSEs are partially folded states, in a pore with attractive walls the potential energy U_{PW} of a folded state is low; that of the TSEs is lower, while the potential energy of the unfolded states is the lowest. Therefore, because the energetic effects are stronger than the entropic effects, we expect $\Delta E_{\text{TSE,f}}$ to decrease but $\Delta E_{\text{TSE,uf}}$ to increases, trends that are seen in Fig. 20.

Figure 18 displays the temperature dependence of the free energy changes $\Delta E_{\text{TSE,uf}}$ and $\Delta E_{\text{TSE,f}}$ (in units of $k_B T$) for a protein of length $\ell = 23$ in the bulk and in pores of size h with attractive walls, where the corresponding folding temperatures are also shown. Given the relation between the free
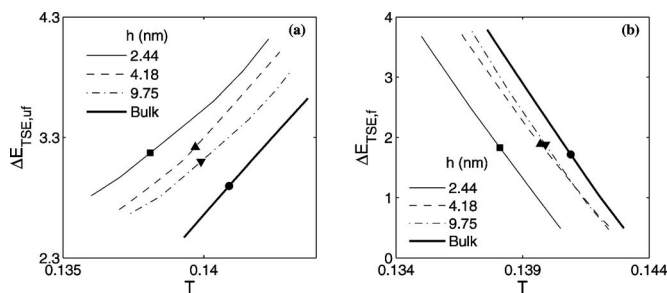

FIG. 18. Same as in Fig. 16, but for a pore of size h with attractive walls and a protein of length $\ell = 23$.
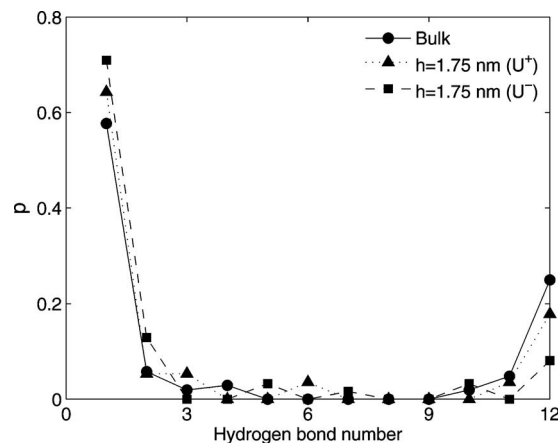

FIG. 19. The probability p of a HB being an unfolding nucleus for all the HBs (from the N-to the C-terminals) in the bulk and in a pore of size $h = 1.75$ nm with attractive (U^+) or repulsive (U^-) walls for a protein of length $\ell = 16$.

energy changes and the folding time τ_f, we see that in a pore with attractive walls at temperature T, τ_f is larger than the corresponding value in the bulk (or the folding rate k_f is smaller), while the reverse is true for the τ_{uf} of the unfolded states, and such increases or decreases are larger in smaller nanopores.

G. Role of proteins' termini and HBs in their unfolding

The importance of the N- and C-termini of proteins in their folding has recently been emphasized.[36–40] In addition, some have studied[39] whether it is the N-or C-terminal of a protein domain that folds or unfolds first, or which set of the residues may be more stable.[40] Here, we study this issue from interesting angles which, to our knowledge, has not been investigated before.

Suppose that at some point in time, when a protein is essentially in its folded state (with at least 90% of its α-helical HBs formed), it begins to unfold to reach $n_\alpha = 0$. The HB between C_i and N_{i+4} is numbered i, the number given to a residue is assigned from the N-terminal of a protein, while that of each HB is the lowest number of residues (the amino acids) that contains it. The direction of numbering is from the N-to the C-terminal. In the DMD simulations at any temperature the first HB from which the unfolding began was monitored. Such a HB acts as an unfolding nucleus.

In Fig. 19 we present the computed probability of a HB being an unfolding nucleus along a protein of length of 16 in the bulk, and in two nanopores of size of 1.75 nm, one with repulsive walls and another with attractive ones. In all the cases protein unfolding begins most probably from its ends. This may be due to finite-size effects.[41]

Note that, in all the cases, unfolding from an N-terminal is more probable than from a C-terminal. When an α-helical HB is formed, it also helps formation of α-helical HBs in its neighbors. Moreover, a protein helps formation of a HB in the direction of a N-terminal more than those in the C-terminal direction.[26] The effect that is indicated by Fig. 19 provides yet another evidence for the asymmetry between the two directions in an α-helix, which may be attributed to the steric effects. Interestingly, Fig. 19 indicates that, for a
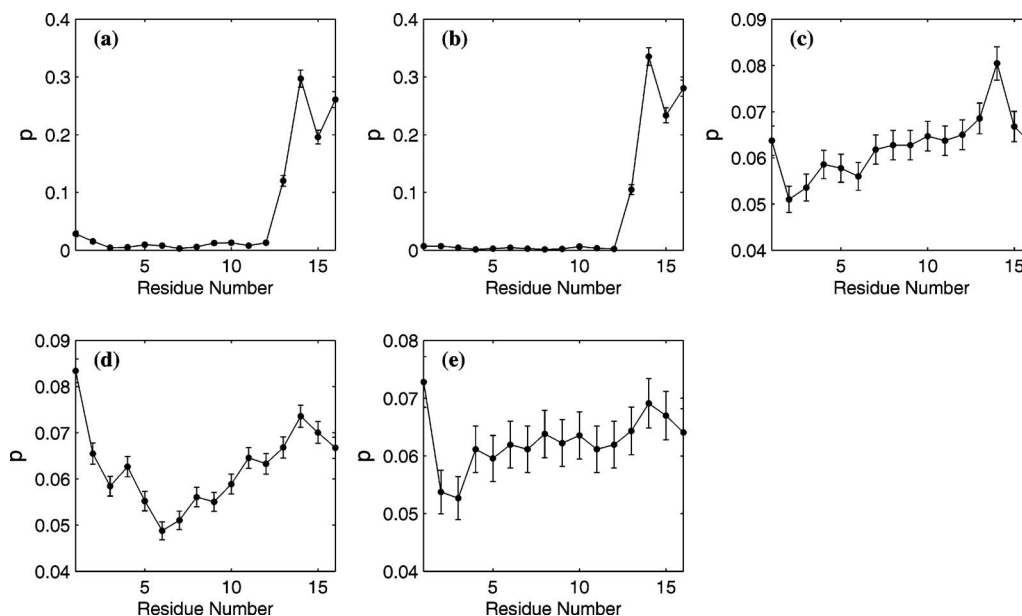
FIG. 20. The probability p of interaction with a pore's walls of the beads in each amino acid of a protein of length $\ell=16$, in contrast to other amino acids. The direction of the residue numbering is from the N-to the C-terminal. (a) $h=12$ nm and $T=0.08$; (b) $h=7$ nm and $T=0.08$; (c) $h=12$ nm and $T=0.135$; (d) $h=1.75$ nm and $T=0.13$; (e) $h=1.75$ nm and $T=0.16$.

protein confined in a nanopore, the probability of unfolding from a N-terminal becomes even larger, the probability of unfolding from the C-terminal decreases, whereas the probability of unfolding from the middle HBs is very low. In a similar manner, we may determine the HB in an unfolded protein which acts as a folding nucleus. Our simulations indicate that folding from the unfolded structure is less likely to begin from its termini.

There exists an asymmetry between the proteins' two termini. The DMD simulations indicated that at low temperatures, when a folded α-helix structure interacts with one wall of a pore and the interaction occurs only through one of its two termini, then, with overwhelming probability (more than 90% of the times), it is through its C-terminal which is adsorbed on the wall. This is demonstrated in Figs. 20(a) and 20(b), where we present the computed probability p of interaction with pore's walls of the beads in each amino acid of a protein of length $\ell=16$, in contrast with the other amino acids, at $T=0.08$ and in two different pores. The direction of the numbering of the residues is from the N-to the C-terminal. Moreover, we find that, at any temperature, binding of the proteins' residues to the pore's walls with attractive interactions is more likely to be by the resides that are closer to the C-terminal. This is demonstrated in Figs. 20(c)–20(e). This is, of course, related to the fact, described earlier, that in such pores unfolding from the C-terminal is more likely.

V. SUMMARY

Extensive DMD simulations were carried out in order to study the stability and folding of α–de novo–designed proteins in slit nanopores, which is of fundamental interest to protein separation and purification by nanoporous membranes, as well as to biocatalysis and nanoporous media that have been under study as sensors. To our knowledge, our study represents the most comprehensive one of its kind. The focus of the work was on nanopores with attractive walls, whereas the previous studies had considered the effect of repulsive walls. We considered the effect of a variety of factors on the stability and folding of proteins in nanopores, including the pore size, the nature of the interaction (repulsive versus attractive) of the proteins with the pore's walls, as well as the proteins' size.

Near the folding temperature T_f and in the presence of an attractive potential U^+ with the pore walls, the proteins experience a repeating sequence of folding/partially folding/unfolding transitions, with the folding temperature T_f, as well as the folding rate, decreasing with decreasing pore sizes. The unfolded states may even be completely adsorbed on the pore walls with a negative potential energy. In such pores the energetic effects dominate the entropic effects, as a result of which the unfolded state is stabilized, with a folding temperature T_f higher than its value in the bulk. The opposite trends are true in the presence of a repulsive interaction potential U^- between the proteins and the walls. For short proteins in very tight pores there exists an unfolded state with only one α-helical HB and an energy nearly equal to that of the folded state. The protein has, however, a high entropy, so that its free energy is lower than what is expected. Under such conditions, the proteins cannot fold onto their native structure. In the presence of repulsive walls, however, such proteins do attain their native structure, hence demonstrating once again the significant differences between pores with repulsive and attractive walls.

In Part II of this series we will focus on diffusion of proteins in the same types of pores and study the effect of a variety of factors on the rate of transport of proteins in nanopores which, to our knowledge, has not been studied before.

[1] C. Branden and J. Tooze, *Introduction to Protein Structure*, 2nd ed. (Garland, New York, 1998).

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

[3] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).

[4] L. 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).

[5] D. N. Lu, Z. Liu, and J. Z. Wu, Biophys. J. **90**, 3224 (2006).

[6] D. A. Kirschner, C. Abraham, and D. J. Selkoe, Proc. Natl. Acad. Sci. U.S.A. **83**, 503 (1986); E. H. Koo, P. T. J. Lansbury, and J. W. Kelly, *ibid.* **96**, 9989 (1999).

[7] K. A. Conway, S.-J. Lee, J.-C. Rochet, T. T. Ding, R. E. Williamson, and P. T. Lansbury, Jr., Proc. Natl. Acad. Sci. U.S.A. **97**, 571 (2000).

[8] D. G. Lynn and S. C. Meredith, J. Struct. Biol. **130**, 153 (2000).

[9] A. P. J. Midelberg, J Microbiol. Biotechnol. **6**, 225 (1996); A. D. Guise, S. M. West, and J. B. Chaudhuri, Mol. Biotechnol. **6**, 53 (1996); J. G. Thomas, A. Ayling, and F. Baneyx, Appl. Biochem. Biotechnol. **66**, 197 (1997).

[10] D. K. Eggers and J. S. Valentine, Protein Sci. **10**, 250 (2001).

[11] A. Brinker, G. Pfeifer, M. J. Kemer, D. J. Naylor, F.-U. Hartl, and M. Hayer-Hartl, Cell **107**, 223 (2001).

[12] E. E. Borrero and F. A. Escobedo, J. Chem. Phys. **125**, 164904 (2006); M. Hayer-Hartl and A. P. Minton, Biochemistry **45**, 13356 (2006); Y. C. Tang and H.-C. Chang, Cell **125**, 903 (2006); G. Ziv, G. Haran, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 18956 (2005); H. X. Zhou, J. Mol. Recognit. **17**, 368 (2004); H. X. Zhou and K. A. Dill, Biochemistry **40**, 11289 (2001); R. J. Ellis, Curr. Biol. **13**, R881 (2003); F. Tagaki, N. Koga, and S. Takada, Proc. Natl. Acad. Sci. U.S.A. **100**, 11367 (2003).

[13] G. F. Bickerstaff, *Immobilization of Enzymes and Cells* (Humana, Totowa, NJ, 1997); J. Livage, T. Coradin, and C. Roux, J. Phys.: Condens. Matter **13**, R673 (2001).

[14] C. Lei, Y. Shin, J. Liu, and J. Ackerman, J. Am. Chem. Soc. **124**, 11242 (2002).

[15] M. Dadvar and M. Sahimi, Chem. Eng. Sci. **56**, 1 (2000); **58**, 4935 (2003).

[16] M.-E. Avramescu, Z. Borneman, and M. Wessling, Biotechnol. Bioeng. **84**, 564 (2003).

[17] R. J. Ciora, B. Fayyaz, P. K. T. Liu, V. Suwanmethanond, 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).

[18] A. J. Rosenbloom, D. M. Sipe, Y. Shishkin, Y. Ke, R. P. Devaty, and W. J. Choyke, Biomed. Microdevices **6**, 261 (2004).

[19] A. P. Minton, J. Cell. Sci. **119**, 2863 (2006).

[20] G. Ping, J. Yuan, M. Vallieres, H. Dong, Z. Sun, Y. Wei, F. Y. Li, and S. H. Lin, J. Chem. Phys. **118**, 8042 (2003); H. C. Loebl and C. C. Matthai, Physica A **342**, 612 (2004).

[21] M. S. Cheung, D. Klimov, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 4753 (2005).

[22] J. Lendermann and R. Winter, Phys. Chem. Chem. Phys. **5**, 1440 (2003); C. H. Lee, J. Lang, C. W. Yen, P. C. Shih, T. S. Lin, and C. Y. Mou, J. Phys. Chem. B **109**, 12277 (2005); E. Rossinsky and S. Srebnik, Biopolymers **79**, 259 (2005); W. X. Xu, J. Wang, and W. Wang, Proteins **61**, 777 (2005); M. S. Cheung and D. Thirumalai, J. Mol. Biol. **357**, 632 (2006).

[23] L. Regan and W. F. DeGrado, Science **241**, 976 (1988).

[24] Z. Guo and D. Thirumalai, J. Mol. Biol. **263**, 323 (1996).

[25] S. 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).

[26] A. 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); Comput. Biol. Chem. **30**, 215 (2006).

[27] B. 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. Bellemans, 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).

[28] C. Bartels and M. Karplus, J. Comput. Chem. **18**, 1450 (1997); J. Phys. Chem. B **102**, 865 (1998).

[29] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992); U. H. E. Hansmann and Y. Okomoto, J. Comput. Chem. **14**, 1333 (1993).

[30] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); Phys. Rev. E **56**, 5018 (1997).

[31] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989); S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992); S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *ibid.* **16**, 1339 (1995).

[32] B. Roux, Comput. Phys. Commun. **91**, 275 (1995); S. Kumar, P. W. Payne, and M. Vasquez, J. Comput. Chem. **17**, 1269 (1996).

[33] Y. Zhou, M. Karplus, J. M. Wichert, and C. K. Hall, J. Chem. Phys. **107**, 10691 (1997); E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy, J. Phys. Chem. B **109**, 6722 (2005); J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, J. Chem. Theory Comput. **3**, 26 (2007).

[34] S. Schnabel, M. Bachmann, and W. Janke, Phys. Rev. Lett. **98**, 048103 (2007).

[35] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).

[36] N. Alexandrov, Protein Sci. **2**, 1989 (1993).

[37] M. Mallela, G. Krishna, and S. W. Wanglander, Proc. Natl. Acad. Sci. U.S.A. **102**, 1053 (2005).

[38] E. Jacob and R. Unger, Bioinformatics **23**, E225 (2007).

[39] J. Cellitti, R. Bernstein, and S. Marqusee, Protein Sci. **16**, 852 (2007).

[40] C. C. Chow, C. Chow, V. Raghunathan, T. J. Huppert, E. B. Kimball, and S. Cavagnero, Biochemistry **42**, 7090 (2003).

[41] U. H. E. Hansmanna and Y. Okamoto, J. Chem. Phys. **110**, 1267 (1999).