Protein escape at the ribosomal exit tunnel: Effect of the tunnel shape

Phuong Thuy Bui (iD), and Trinh Xuan Hoang (iD)

View Online   Export Citation   CrossMark

Protein escape at the ribosomal exit tunnel: Effect of the tunnel shape

View Online   Export Citation   CrossMark

Phuong Thuy Bui[1,2] (iD) **and Trinh Xuan Hoang**[3,4,a] (iD)

AFFILIATIONS

[1] Institute of Theoretical and Applied Research, Duy Tan University, Hanoi 100000, Vietnam
[2] Faculty of Pharmacy, Duy Tan University, Da Nang 550000, Vietnam
[3] Institute of Physics, Vietnam Academy of Science and Technology, 10 Dao Tan, Ba Dinh, Hanoi 11108, Vietnam
[4] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet,
  Cau Giay, Hanoi 11307, Vietnam

[a] Author to whom correspondence should be addressed: hoang@iop.vast.vn

ABSTRACT

We study the post-translational escape of nascent proteins at the ribosomal exit tunnel with the consideration of a real shape atomistic tunnel based on the Protein Data Bank structure of the large ribosome subunit of archeon *Haloarcula marismortui*. Molecular dynamics simulations employing the Go-like model for the proteins show that at intermediate and high temperatures, including a presumable physiological temperature, the protein escape process at the atomistic tunnel is quantitatively similar to that at a cylinder tunnel of length $L = 72$ Å and diameter $d = 16$ Å. At low temperatures, the atomistic tunnel, however, yields an increased probability of protein trapping inside the tunnel, while the cylinder tunnel does not cause the trapping. All-β proteins tend to escape faster than all-α proteins, but this difference is blurred on increasing the protein's chain length. A 29-residue zinc-finger domain is shown to be severely trapped inside the tunnel. Most of the single-domain proteins considered, however, can escape efficiently at the physiological temperature with the escape time distribution following the diffusion model proposed in our previous works. An extrapolation of the simulation data to a realistic value of the friction coefficient for amino acids indicates that the escape times of globular proteins are at the sub-millisecond scale. It is argued that this time scale is short enough for the smooth functioning of the ribosome by not allowing nascent proteins to jam the ribosome tunnel.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0008292

I. INTRODUCTION

After the determination of ribosome structures,[1,2] there has been increasing attention to understanding the role of the ribosomal exit tunnel in protein biosynthesis and in cotranslational protein folding (for recent reviews, see Refs. 3–6). Biochemical studies indicate that the tunnel plays an active role in the regulation of the protein translation process by blocking specific peptide sequences.[7,8] The mechanism of this blocking or translation arrest can be associated with certain ribosomal protein's motion, which alters the tunnel shape.[9] In cotranslational protein folding, the tunnel imposes a spatial confinement on the traversing nascent peptide while it is being synthesized by the ribosome. The narrow geometry of the tunnel was suggested to entropically promote the α-helix formation,[10] whereas it may sterically obstruct

the formation of the β-sheet.[11] Depending on the location within the tunnel, peptides can form simple α-helices[12] and small tertiary structure units.[13] The latter can be observed near the tunnel exit port where there is enough space to hold the structure. Simulations[14,15] and experiments[16–18] indicate that cotranslational folding starts inside the tunnel, with the structures ranging from a non-native compact conformation[16] and transient tertiary structures[15] to a small protein domain.[17] There are also considerations that the folding inside the ribosome tunnel is negligible, leading to a focus only on the folding of nascent chains as they emerge from the tunnel, as shown in studies with stalled ribosome-bound nascent chain experiments[19,20] and simulations.[21,22]

In recent works,[23,24] we suggested that the exit tunnel, as a passive conduit, has a significant impact on the early post-translational folding, i.e., shortly after the protein's C-terminus is released from

the peptidyl transferase center (PTC). This impact corresponds to a vectorial folding[23] induced by the tunnel and is associated with the escape process of a full length protein from the tunnel. In particular, the folding and escape of a nascent protein at the tunnel are concomitant with each other. Folding accelerates the escape process, whereas a gradual escape improves the protein foldability. Furthermore, we showed that the protein escape at the tunnel is governed by a diffusion mechanism and the escape time distribution can be captured by a simple model of a Brownian particle in a linear potential field.

The escape process also has an important meaning of itself. It should not be too quick because this would leave an escaped protein significantly unfolded outside the ribosome, which increases the chance of protein aggregation.[25] It cannot be also too slow because this would decrease the productivity of the ribosome. Interestingly, our previous study shows that the real length of the ribosome exit tunnel is close to a cross-over tunnel length[24] of 90 Å–110 Å for the diffusion of small globular proteins. For tunnels of lengths larger than this cross-over length, the diffusion is much slower. Thus, it was suggested that the ribosome tunnel length may have been selected by evolution to facilitate an appropriate escape time.

The previous works[23,24] on the protein escape process considered a highly simplified model of the ribosomal exit tunnel being a hollow cylinder with a repulsive wall. The real exit tunnel is highly porous for water-size molecules and effectively adopts an irregular shape[26] for polypeptides. The tunnel shape also depends on the type of organism.[27] The aim of our present study is to work with a realistic tunnel to test the validity of the previous findings and to investigate the effect of the tunnel shape on the protein escape process. For this purpose, we consider an atomistic model of the tunnel based on the resolved Protein Data Bank (PDB) structure of the large ribosome subunit of *Haloarcula marismortui*.[2] The atomistic tunnel incorporates all heavy atoms for the ribosomal RNA, but only the C_α's for the ribosomal proteins. The C_α-only representation is also used for nascent proteins within a standard Go-like model.[28–31] We have chosen the same coarse-grained level for ribosomal proteins for consistency in modeling with the chosen Go-like model. This choice, however, reduces the roughness of the tunnel's wall where amino acid side-chains are exposed. In order to accelerate the simulations, all the ribosomal atoms are kept fixed; thus, the tunnel acts solely as a passive channel for the escaping proteins in our consideration.

The main focus of our paper is on the effect of the tunnel shape on the escape process. In order to delineate this effect, we compared the escape of a nascent protein at the atomistic tunnel with that at an equivalent cylinder tunnel. The latter is described such that it produces a similar median escape time to that obtained with the atomistic tunnel. This comparison shows similarities and differences between the two tunnels. Remarkably, we find that the atomistic tunnel yields a non-zero probability of the trapping protein inside the tunnel, while the cylinder tunnel does not. Other issues being discussed are the dependences of the escape time on temperature, on protein's length and native state topology among small single-domain proteins, and on the friction coefficient for amino acids. An estimation of real escape time from the simulations is also presented. Interestingly, we find that the estimated escape time is relevant to the functionality of the ribosome.

II. MODELS AND METHODS

A. Protein and tunnel models

Nascent proteins are considered in a Go-like model[28–31] in which a protein is represented only by its C_α atoms. The intramolecular potential energy of a protein in a conformation is given by[31]

$$
\begin{aligned}
E = &\sum_{i=1}^{N-1} K_b (r_{i,i+1} - b)^2 + \sum_{i=2}^{N-1} K_\theta (\theta_i - \theta_i^*)^2 \\
&+ \sum_{n=1,3} \sum_{i=2}^{N-2} K_\phi^{(n)} [1 + \cos(n(\phi_i - \phi_i^*))] \\
&+ \sum_{i+3 < j} \epsilon \left[5 \left(\frac{r_{ij}^*}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij}^*}{r_{ij}} \right)^{10} \right] \Delta_{ij} \\
&+ \sum_{i+3 < j} \epsilon \left(\frac{\sigma}{r_{ij}} \right)^{12} (1 - \Delta_{ij}),
\end{aligned}
\tag{1}
$$

where N is the number of amino acid residues, r_{ij} is the distance between residues i and j, θ_i and ϕ_i are the bond and dihedral angles of the residue i, respectively, n is equal to either 1 or 3, the star superscript corresponds to the native state's value, and Δ is the native contact map with Δ_{ij} equal to 1 if there is a native contact between i and j and equal to 0 otherwise. Δ is defined based on an all-atom consideration of the protein PDB structure with a contact cut-off distance between two atoms equal to 1.5 times the sum of their atomic van der Waals (vdW) radii (the C3 map in Ref. 24). Energy is given in units of ϵ, which corresponds to the depth of the 10–12 Lennard-Jones potential given in the fourth term of Eq. (1). The parameters used in the model are b = 3.8 Å, σ = 5 Å, K_b = 100 ϵ Å^{-2}, K_θ = 20 ϵ (rad)$^{-2}$, $K_\phi^{(1)}$ = $-\epsilon$, and $K_\phi^{(3)}$ = -0.5ϵ.

To build up the atomistic model of the ribosomal exit tunnel, we used the crystal structure of *H. marismortui*'s large ribosomal subunit with the PDB code 1jj2.[2] Only a part of the subunit surrounding the tunnel was taken to model [Fig. 1(a)]. In particular, we excluded the ribosome's atoms that are further than 30 Å from a chosen axis x that originates from the PTC region and goes roughly through the middle of the tunnel. Furthermore, we kept only heavy atoms for the ribosomal RNA and only C_α atoms for the ribosomal proteins. The interaction potential between the C_α atoms of a nascent protein and the C_α atoms of a ribosomal protein is assumed to be repulsive and takes the form of $\epsilon(\sigma/r_{ij})^{12}$. The interaction potential between a nascent protein's C_α (a) and a ribosomal RNA heavy atom (b) is also repulsive and given by

$$
V(r) = \epsilon \left(\frac{R_a + R_b + R_+}{r_{ab}} \right)^{12},
\tag{2}
$$

where r_{ab} is the center-to-center distance between a and b, R_a = 2.5 Å is an effective radius of the amino acid, R_b is the atomic vdW radius of the ribosomal atom, and R_+ = 0.8 Å is an effective additive radius accounting for the fact that hydrogen atoms are not considered in the model of the tunnel. For example, if b is a carbon atom (with a vdW radius of 1.7 Å), then one gets $R_b + R_+$ = 2.5 Å, i.e., the same as R_a. We have checked that the above-mentioned value of R_+ yields an adequate escape behavior for the nascent proteins.[33]
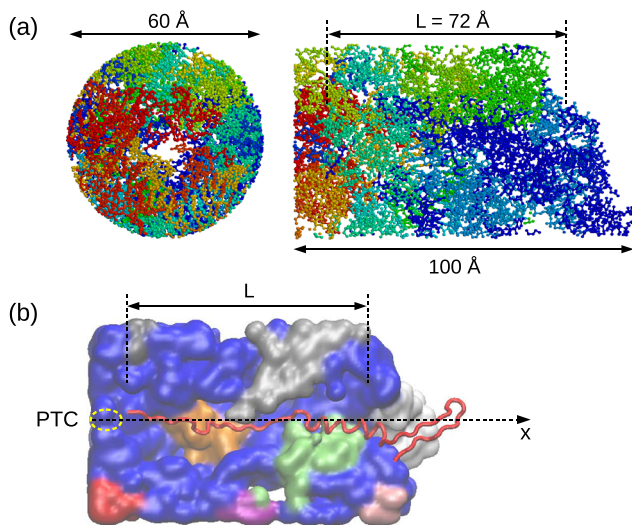
FIG. 1. (a) Two projected views of the ribosomal exit tunnel of *H. marismortui* represented by all heavy atoms within 30 Å away from a chosen axis that goes roughly through the middle of the tunnel. (b) A conformation of protein GB1 (red) obtained by simulation after the protein is grown from the PTC at the atomistic tunnel. The latter is shown from a cross-sectional plane that includes the tunnel axis. Assuming that *x* is the tunnel axis, the planes of the projected views are *y–z* and *x–y* in (a) and *x–z* in (b). The PTC region is identified as located between the bases A2486, U2620, and C2104 of the ribosomal RNA.[32] The tunnel length of *L* = 72 Å indicated by the arrows corresponds to an axial distance from the opening of the tunnel near the PTC to an inner edge of the tunnel exit port.

The simulations were carried out using the molecular dynamics method based on the Langevin equation of motion and a Verlet algorithm.[23] The amino acids are assumed to have an uniform mass, *m*. Temperature is given in units of ϵ/k_B, whereas time is measured in units of $\tau = \sqrt{m\sigma^2/\epsilon}$. We used the same value of the friction coefficient in the Langevin equation, $\zeta = 1m\tau^{-1}$, throughout this article, except in Subsection III C where the dependence on the friction coefficient is investigated. In each simulation, first, the polypeptide chain was grown in the tunnel from the PTC at a constant speed given by the growth time $t_g = 100\tau$ per amino acid. This growth time is slow enough to produce the converged properties of fully translated protein conformations in terms of the radius of gyration and the number of native contacts, i.e., the distributions of these quantities are similar to those obtained with a much larger growth time.[23] After the protein is completely translated, the simulation was run until it has fully escaped from the tunnel. The escape time was measured from the moment of complete translation. Because the exit port is irregular with a complex geometry, in order to capture the essential escape time, we have defined the tunnel region as the space within a cylinder of length *L* = 72 Å and radius 15 Å, centered about the tunnel axis *x*. This region starts from an opening of the tunnel near the PTC and ends at an inner edge of the exit port (Fig. 1), corresponding to positions from *x* = 10 Å to *x* = 82 Å. An amino acid residue is considered to have escaped from the tunnel if it is found outside the tunnel region. Typically, for each temperature, about

1000 independent growth and escape trajectories are simulated to obtain the statistics of the escape time.

B. Diffusion model

The escape of a fully translated protein at the ribosome tunnel is driven by (i) an enthalpic force associated with the folding of the protein outside the tunnel, (ii) an entropy gain as the chain emerges from the tunnel, and (iii) the stochastic motion of a partially unfolded chain. It has been shown that in the Go-like model, the free energy change of a protein along an escape coordinate is a monotonically decreasing function, which is approximately linear at intermediate and high temperatures.[23,33] Interestingly, all these effects can be effectively acquired in a diffusion model,[23,24] which describes the protein escape as the diffusion of a Brownian particle pulled by a constant force in one dimension. The particle diffusion in a potential field $U(x)$ is governed by the Smoluchowski equation[34]

$$\frac{\partial}{\partial t} p(x,t) = \frac{\partial}{\partial x} D\left(\beta \frac{\partial U(x)}{\partial x} + \frac{\partial}{\partial x}\right) p(x,t), \quad (3)$$

where $p(x,t)$ is a probability density of finding the particle at position *x* and at time *t* > 0, given that it was found at position *x* = 0 at time *t* = 0; *D* is the diffusion constant, assumed to be position independent; and $\beta = (k_B T)^{-1}$ is the inverse temperature with k_B the Boltzmann constant. The escape time is described as the first passage time of the particle reaching a distance *L* from an origin in the drift direction. Given an external potential of the linear form $U(x) = -kx$, where *x* is the coordinate of the particle and *k* is the force, the distribution of the escape time was obtained via an exact solution[35] and is given by[24]

$$g(t) = \frac{L}{\sqrt{4\pi D t^3}} \exp\left[-\frac{(L - D\beta k t)^2}{4Dt}\right]. \quad (4)$$

Using the distribution in Eq. (4), one obtains the mean escape time

$$\mu_t \equiv \langle t \rangle = \int_0^\infty t\, g(t)\, dt = \frac{L}{D\beta k}, \quad (5)$$

with $D\beta k$ as the mean diffusion speed, and the standard deviation

$$\sigma_t \equiv \sqrt{\langle t^2 \rangle - \langle t \rangle^2} = \frac{\sqrt{2L}}{D(\beta k)^{\frac{3}{2}}}. \quad (6)$$

Note that both μ_t and σ_t diverge when *k* = 0 for which $g(t)$ becomes a heavy-tailed Lévy distribution. It has been shown that *D* and βk depend on *L*, on the protein, and on other conditions such as the crowders' volume fraction outside the tunnel.[24]

III. RESULTS AND DISCUSSION

A. Effect of tunnel shape on escape process

To study the effect of the tunnel shape on the escape process, we consider the immunoglobulin binding (B1) domain of protein G (GB1) as a nascent protein. This protein has a length of *N* = 56 amino acids and was considered in our previous studies.[23,24] The

folding temperature of GB1 was found as $T_f = 1.004 \, \epsilon/k_B$.[24] Experimentally, the melting temperature of the wild-type GB1 at pH 5.5 has been reported to be 80.5 °C.[36] We will study the escape process at various temperatures but will focus on the simulation temperature $T = 0.85 \, \epsilon/k_B$, which after unit conversion corresponds to a physiologically relevant temperature of 26.3 °C.

In order to delineate the effects of tunnel shape on the escape process, we sought for an equivalent cylinder tunnel that yields a similar escape time to that obtained at the atomistic tunnel for the protein GB1. We found that among the cylinder tunnels of the same length as that of the atomistic tunnel, i.e., $L = 72$ Å, the one with diameter $d = 16$ Å satisfies quite well the last requirement over a wide range of temperatures. Figure 2(a) shows that the median escape times t_{med} for the atomistic tunnel and the cylinder tunnel are quite close to each other for various temperatures from 0.85 ϵ/k_B to 1.6 ϵ/k_B. Figure 2(b) also shows that the mean escape times μ_t for the two tunnels agree very well with each other at intermediate and high temperatures ($T > 0.85 \, \epsilon/k_B$). For $T \leq 0.85 \, \epsilon/k_B$, both t_{med} and μ_t for the atomistic tunnel are larger than for the cylinder tunnel, and the differences increase as the temperature decreases. These differences indicate that, at low temperatures, it is more difficult for proteins to escape from the atomistic tunnel than from the cylinder one. It appears that the physiological temperature 0.85 ϵ/k_B corresponds to a borderline behavior of the escape process in which the effect of tunnel shape starts to get in.

Figure 2 also shows the linear dependences of t_{med} and μ_t on T^{-1} in log–log scales (the dashed lines) as it would be found for a Brownian particle diffused in a potential field with a constant
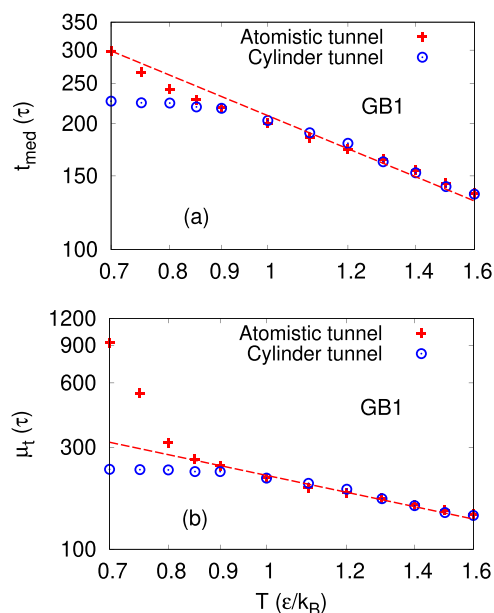

FIG. 2. Log–log plots of the temperature dependence of the median escape time t_{med} (a) and the mean escape time μ_t (b) for protein GB1 at the atomistic tunnel (crosses) and at a cylinder tunnel of length $L = 72$ Å and diameter $d = 16$ Å (circles). The straight line (dashed) has a slope of −1 and is fitted to the data points of the cylinder tunnel at high temperatures ($T \geq 1 \, \epsilon/k_B$).

βk. Our previous study[24] shows that this linear behavior is found for a homopolymer chain with self-repulsion and, thus, can be applied for intrinsically disordered proteins. For foldable proteins such as the GB1, this linear dependence can be observed only at high temperatures at which the proteins are unfolded during the escape.

Note that one can also have an equivalent cylinder tunnel of a length different from that of the atomistic tunnel. For example, we found that the cylinder tunnel of $L = 82$ Å and $d = 13.5$ Å also yields similar median and mean escape times to those of the atomistic tunnel (see Fig. S1 of the supplementary material) with an equally good agreement as in Fig. 2. For the most relevant comparison, we consider only the equivalent cylinder tunnel of $L = 72$ Å and $d = 16$ Å.

We now examine more carefully the escape processes of GB1 at the atomistic and cylinder tunnels at $T = 0.85 \, \epsilon/k_B$. Figure 3(a) shows that at the atomistic tunnel, the histogram of the escape time for this protein obtained from the simulations follows quite well the distribution function $g(t)$ given by the diffusion model in Eq. (4). Figure 3(b) shows that the probability of protein escape P_{escape} has a sigmoidal dependence on the time t, with P_{escape} reaching the value of 1 at $t \approx 900\tau$. This result means that the protein can effectively escape from the tunnel without significant delays compared to the median escape time, $t_{med} \approx 230\tau$. Figure 3(b) also shows the dependence of the probability $P_{C\text{-term-}\beta}$ of forming the C-terminal β-hairpin inside the tunnel on time. The time dependence of this probability is obtained by averaging over multiple escape trajectories. The C-terminal β-hairpin is said to be formed inside the tunnel if it forms at least half of its native contacts and when all of its residues (41–56) are located within the tunnel. We tracked this β-hairpin because a previous study showed that, at low temperatures, the GB1 protein can escape from a cylinder tunnel through two different pathways depending on whether the C-terminal β-hairpin is formed inside the tunnel or not.[23] Figure 3(b) shows that at the atomistic tunnel, only a small fraction of about 2% of the escape trajectories have this β-hairpin formed inside the tunnel. We have checked that the trajectories having this hairpin formed typically correspond to longer escape times than other trajectories.

Figure 3(c) shows the histogram of conformations observed during the escape process as functions of the number of native contacts N_c and the number of residues escaped from the tunnel N_{out} for protein GB1 at the atomistic tunnel. The histogram shows a high-density cloud of conformations having intermediate values of N_c and N_{out}, indicating that the protein folds during the escape. The blurring of the cloud, however, suggests that the protein adopts a wide range of conformations during the escape process. Given that the maximum N_c for GB1 is 120, the histogram shows that during the escape, the protein can form up to two-thirds of all of its native contacts. Note that conformations of $N_{out} = 0$, i.e., completely located within the tunnel, are also present in the histogram.

Figures 3(d)–3(f) show that the equivalent cylinder tunnel of diameter $d = 16$ Å produces not only a similar escape time distribution but also a similar dependence of P_{escape} on time and a similar histogram of escaping protein conformations to those obtained with the atomistic tunnel. Note, however, that there are differences. First, the escape time distribution at the atomistic tunnel is slightly more narrow than the one at the cylinder tunnel, while the median escape time at the cylinder is slightly smaller than at the atomistic tunnel (220τ vs 230τ). Second, for the cylinder tunnel, the
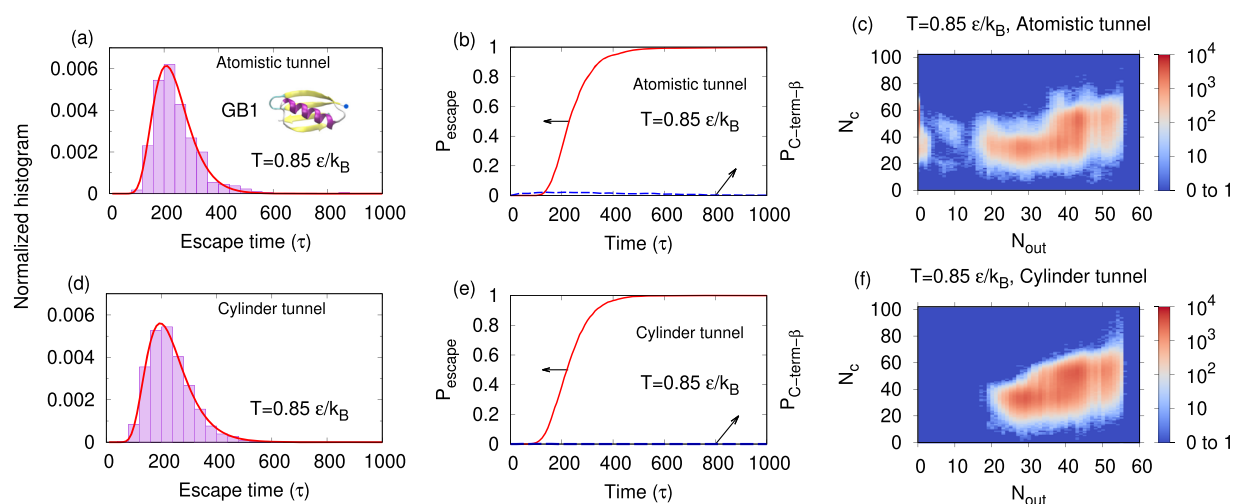
FIG. 3. Distributions of the escape time [(a) and (d)], time dependences of the escape probability P_{escape} (solid) and the probability of C-terminal β-hairpin formation inside the tunnel $P_{C\text{-}term\text{-}\beta}$ (dashed) [(b) and (e)], and histograms of conformations as functions of the number of residues escaped from the tunnel N_{out} and the number of native contacts N_c [(c) and (f)] for protein GB1 at the atomistic tunnel (upper panels) and at an equivalent cylinder tunnel (lower panels) of length L = 72 Å and diameter d = 16 Å at temperature T = 0.85 ϵ/k_B. The native conformation of GB1, shown in (a) as the inset, has 120 native contacts.

escape probability P_{escape} reaches 1 faster at the time about 500τ. Third, the histogram in N_c and N_{out} less dispersed in the case of the cylinder tunnel. These differences indicate that the protein escapes relatively more easily at the cylinder tunnel than at the atomistic tunnel.

Other differences at the two tunnels can be seen at the probability of forming the C-terminal β-hairpin and the histogram of conformation during the escape process. Figure 3(e) shows that for the cylinder tunnel, the probability $P_{C\text{-}term\text{-}\beta}$ is zero at all times,

indicating that the C-terminal β-hairpin does not form inside the cylinder tunnel. Figure 3(f) shows that the histogram of conformations during the escape process for the cylinder tunnel does not include conformations of small N_{out} (less than about 16). These results are different from those at the atomistic tunnel and indicate that the pathways at the atomistic tunnel are more diverse than at the cylinder tunnel.

The differences between the escape processes at the two tunnels magnify as the temperature is lowered. In Fig. 4, we show the same
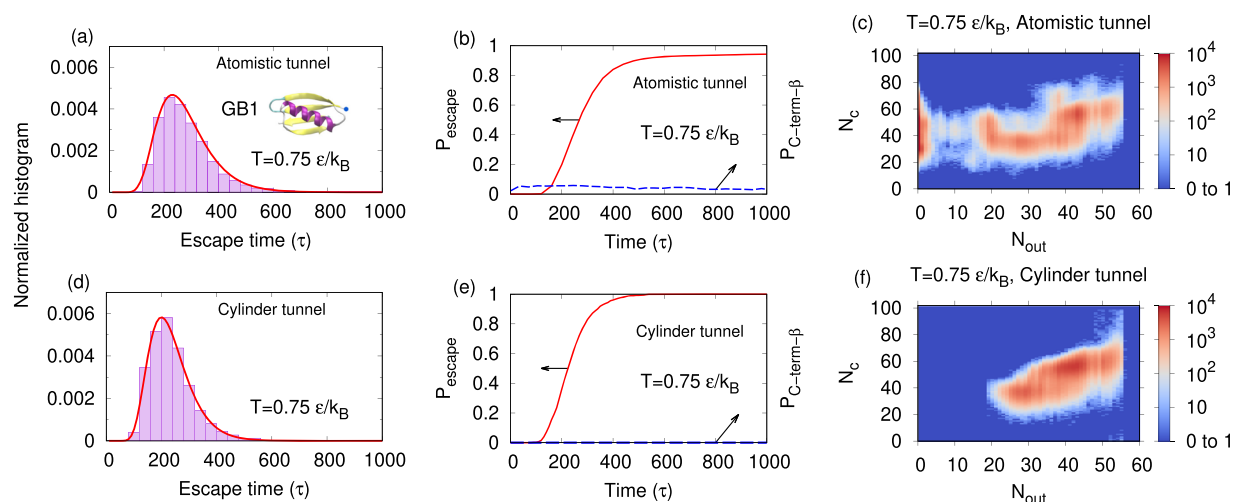

FIG. 4. Same as in Fig. 3, but for T = 0.75 ϵ/k_B. The panels show the distributions of the escape time [(a) and (d)], the time dependences of the escape probability P_{escape} (solid) and the probability of C-terminal β-hairpin formation inside the tunnel $P_{C\text{-}term\text{-}\beta}$ (dashed) [(b) and (e)], and the histograms of conformations as functions of the number of residues escaped from the tunnel N_{out} and the number of native contacts N_c [(c) and (f)] for protein GB1 at the atomistic tunnel (upper panels) and at an equivalent cylinder tunnel (lower panels) of length L = 72 Å and diameter d = 16 Å.
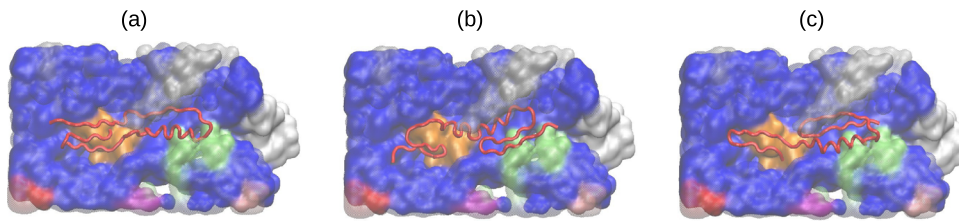
FIG. 5. Examples of the conformations of GB1 that are trapped inside the atomistic tunnel during the escape process at $T = 0.75\ \epsilon/k_B$. The number of native contacts of the conformations are $N_c = 27$ (a), $N_c = 34$ (b), and $N_c = 47$ (c), whereas all of them have $N_{out} = 0$, as indicated.

plots as in Fig. 3, but for $T = 0.75\ \epsilon/k_B$. Figures 4(a) and 4(d) show that the escape time distribution for the atomistic tunnel is significantly more broad than for the cylinder tunnel. Figures 4(b) and 4(e) show that, for the atomistic tunnel, the escape probability P_{escape} can reach only about 94% at $t = 1000\tau$, while for the cylinder tunnel, it can reach 100% at $t \approx 500\tau$. Figure 4(b) shows that at $T = 0.75\ \epsilon/k_B$, about 5% of the escape trajectories having the C-terminal β-hairpin formed inside the atomistic tunnel, while this fraction remains to be zero for the cylinder tunnel [Fig. 4(e)]. Figures 4(c) and 4(f) show that the histogram of escaping conformations for the atomistic tunnel is more complex than for the cylinder tunnel. There appears a significant number of conformations of low N_{out}, including those of $N_{out} = 0$, at the atomistic but the cylinder tunnel. We have checked that the trajectories that did not end with a successful escape after a long time compared to t_{med} are associated with conformations of $N_{out} = 0$. These conformations are identified as kinetic traps in the escape process.

Figure 5 shows several trapped conformations obtained at $T = 0.75\ \epsilon/k_B$ for GB1 at the atomistic tunnel. The number of native contacts N_c in these conformations is different, but all of them have the α-helix and at least one β-hairpin formed. The conformation shown in Fig. 5(c) also has a partial tertiary structure established by contacts between the α-helix and the N-terminal β-hairpin. These conformations did not appear at the cylinder tunnel, indicating that the irregular shape of the atomistic tunnel allows for and makes the formation of trapped conformations more easy inside the tunnel. Note that at the physiological temperature $T = 0.85\ \epsilon/k_B$, there were no kinetic traps. This can be understood as due to two reasons: the faster diffusion at this temperature helps the protein to avoid trapped conformations and the larger thermal fluctuations help the protein to get out from the traps. The trapping at an atomistic ribosome tunnel and the alleviation of trapping by increased thermal fluctuations have been also observed for the chymotrypsin inhibitor 2 (CI2) protein in an early simulation study by Elcock.[14] Our results here for GB1 are consistent with that previous work.

It is interesting now to compare the protein diffusion properties at the two tunnels using the diffusion model. Figure 6 shows the values of the diffusion constant D and the potential slope βk obtained by fitting the escape time histograms at various temperatures to the diffusion model, wherein it can be seen that for both the tunnels, D appears to increase linearly with temperature, whereas βk tends to adopt a constant value. The linear dependence of D on temperature agrees with that of an ideal Brownian particle. The atomistic tunnel, however, yields a lower D and a higher βk than the cylinder tunnel at intermediate and high temperatures ($T \geq 0.85\ \epsilon/k_B$), while the average diffusion speed given by $D\beta k$ is almost the same for the two tunnels. We have checked that at these temperatures, the escape time

distributions at the atomistic tunnel are slightly narrower than at the cylinder tunnel. At low temperatures ($T < 0.85\ \epsilon/k_B$), both D and βk for the atomistic tunnel deviate significantly from the average trends due to the impact of kinetic trapping. Also at low temperatures, the escape time distribution for the atomistic tunnel becomes broader than for the cylinder tunnel.

To better understand the differences between the two tunnels, we sought for a quantitative comparison between the shapes of the atomistic tunnel and the cylinder tunnel. To this extent, we have calculated the area S of the inner cross section of the atomistic tunnel as a function of the position x along the tunnel axis using a probe sphere. The effective diameter of the tunnel at a given
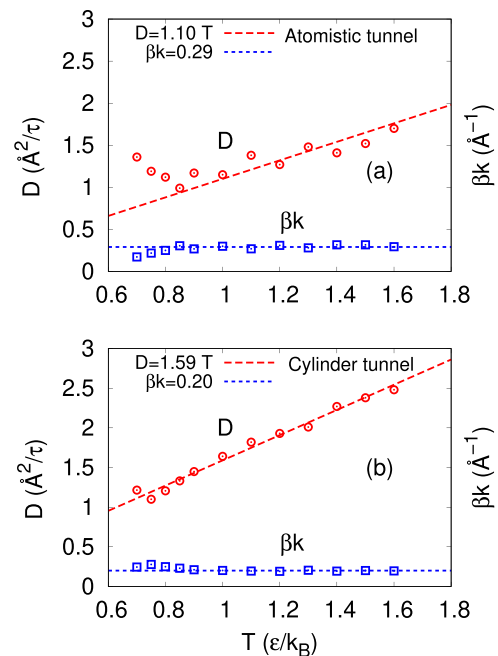

FIG. 6. Dependence of the diffusion constant D (circles) and the parameter βk (squares) on temperature for protein GB1 at the atomistic tunnel (a) and at the cylinder tunnel of $L = 72$ Å and $d = 16$ Å (b). The data are obtained by fitting the simulated escape time distribution to Eq. (4). Straight lines show fits of the simulation data for $T \geq 0.85\ \epsilon/k_B$ to a linear dependence on temperature in the case of D (dashed) and to a constant value in the case of βk (dotted). The functions of the fits are $D = 1.1T$ and $\beta k = 0.29$ Å^{-1} in (a) and $D = 1.59T$ and $\beta k = 0.2$ Å^{-1} in (b).

position then was calculated as $d = 2\sqrt{S/\pi}$. Figure 7(a) shows that the shape of the tunnel's cross section varies strongly with x. It is typically not circular and significantly deviates from that of the equivalent cylinder tunnel. The tunnel is also quite narrow near the PTC and becomes much wider at the exit port. Figure 7(b) shows that the effective diameter d of the atomistic tunnel varies, but not too strongly, between 15 Å and 20 Å, for the positions of x between 15 Å and 75 Å. Note that, for these positions, the diameter $d = 16$ Å of the equivalent cylinder tunnel lies within the variation range of the atomistic tunnel's diameter but is near the lower bound of this range. This can be understood as the irregular shape of the atomistic tunnel's cross sections makes it effectively smaller for nascent proteins.

The present model of the atomistic tunnel neglects the presence of amino acid side-chains. We have checked that by considering all the heavy atoms of ribosomal proteins in the tunnel model while keeping the C_α-only representation for GB1, the escape time distribution at $T = 0.85\ \epsilon/k_B$ changes only slightly with a small shift toward smaller values, but the escape probability reaches only about 95% at the time of 1000τ (see Fig. S2 of the supplementary material). We also found that due to the increased roughness of the tunnel surface, the all-heavy-atom tunnel requires a longer growth time of the protein, with $t_g = 400\tau$ per amino acid, to obtain the converged

properties of fully translated protein conformations. The results indicate that amino acid side-chains may have a relatively small but detrimental effect on the escape process. A proper consideration of this effect, however, would need models that include side-chain representations for both ribosomal and nascent proteins and allow the degrees of freedom of side-chain rotation. It can be expected that while the side-chain excluded volumes cause some obstruction to the escape process, the freedom of side-chain rotation can make the escape somewhat easier.

B. Dependence of escape time on protein

In order to study the dependence of the escape time on proteins, apart from GB1 we selected additional 16 single-domain proteins of lengths between 37 and 99 residues and belonging to different classes of all-α, all-β, and α/β proteins and carried out simulations for these proteins. We consider only the atomistic tunnel and the simulation temperature $T = 0.85\ \epsilon/k_B$. Figure 8 shows that the histograms of the escape times for the selected proteins are quite similar in the overall shape and the range of most probable values. The peak positions of the histograms vary but within the same order of magnitude. The distribution width is the most narrow for the all-β proteins. For most of the proteins, the histogram can be fitted quite well to the distribution function given by Eq. (4) of the diffusion model. For a few proteins, i.e., the ones with PDB codes 2spz and 1wt7, the agreement with the diffusion model is worse than for others with the appearance of a thick tail of large escape times [Figs. 8(m) and 8(n)]. For most of the proteins, we observed trajectories with trapped conformations of $N_{out} = 0$. The fraction of non-escaped trajectories at the largest time in the histograms shown in Fig. 8 (1200τ or 1800τ) is below 2% for 10 out of 16 proteins (see the caption of Fig. 8). At a larger time of 8000τ, this fraction falls below 2% for all proteins except 2spz for which this fraction remains at 9.7%. Thus, most proteins can escape efficiently. We have checked that the trapped conformations of 2spz typically have a two-helix bundle formed within the tunnel, resulting in an increased difficulty for its escape.

Figure 9(a) shows the median escape time t_{med} as a function of the chain length N for all 17 proteins considered including GB1. It can be seen that t_{med} is found in the range from 200τ to 500τ and does not seem to depend on N. However, the variation in the escape times among the proteins decreases with N. Figure 9(b) plots t_{med} against a topological parameter of the protein native state, the relative contact order (CO).[37] It shows a weak but visible trend that t_{med} decreases with CO. From the types of data points shown in Fig. 9(b), one can also see that the all-β proteins on average escape faster than the all-α proteins, whereas the α/β proteins can escape either as fast or as slow as the two other groups. From the fits of the simulated escape time distribution to the diffusion model, we obtained the values of D and βk for 17 proteins. Figures 9(c) and 9(d) show that both D and βk vary strongly among the proteins, but like for t_{med}, the variation decreases with N indicating a kind of convergence. It can be also noted that the strongest variation belongs to the α/β proteins, whereas the weakest belongs to the β proteins. It can be expected that proteins of length $N > 100$ have similar t_{med} and diffusion properties at the tunnel to the ones of length $70 < N < 100$.

It is interesting that all the proteins considered are able to escape from the atomistic tunnel successfully including the CHHC
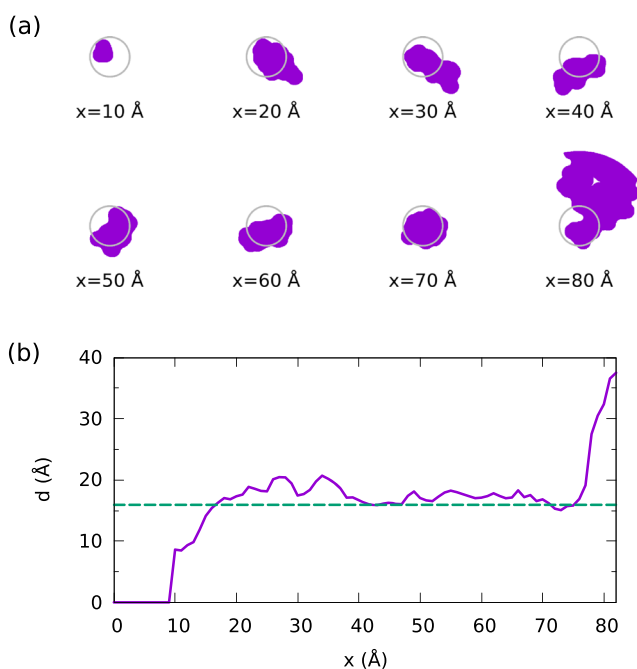

FIG. 7. (a) Inner cross sections (dark color) of the atomistic tunnel obtained by using a probe sphere of radius $R = 3$ Å at various positions x along the tunnel axis as indicated. For comparison, the cross section of an equivalent cylinder tunnel of diameter 16 Å is also shown (gray circle). (b) Dependence of the effective diameter d of the atomistic tunnel (solid) on x. For a given position x, d is calculated as $d = 2\sqrt{S/\pi}$, where S is the area of the tunnel's cross section. The horizontal dashed line indicates the constant diameter $d = 16$ Å of the equivalent cylinder tunnel.
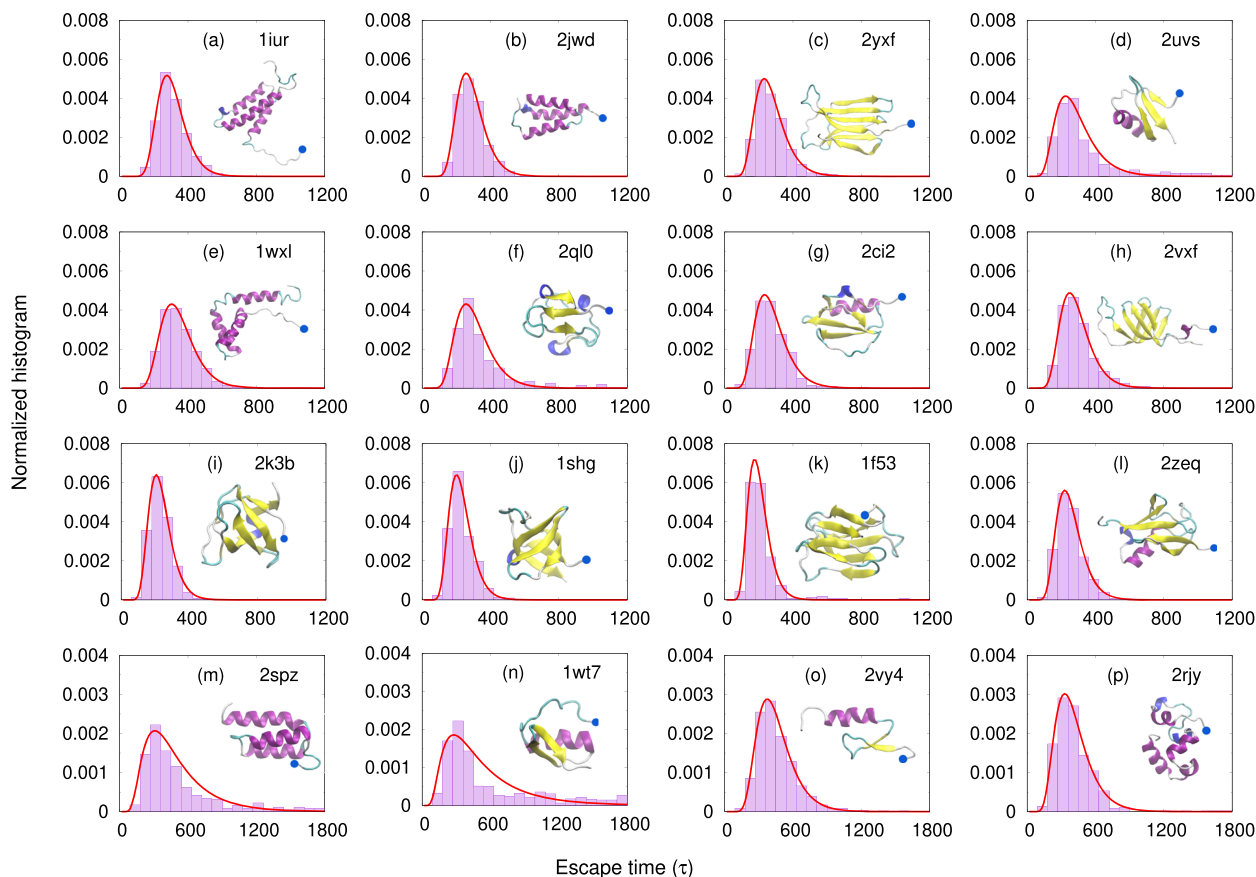
FIG. 8. Distribution of the escape time at the atomistic tunnel at $T = 0.85\ \epsilon/k_B$ for 16 small single-domain proteins (not including GB1) named by their PDB codes as 1iur (a), 2jwd (b), 2yxf (c), 2uvs (d), 1wxl (e), 2ql0 (f), 2ci2 (g), 2vxf (h), 2k3b (i), 1shg (j), 1f53 (k), 2zeq (l), 2spz (m), 1wt7 (n), 2vy4 (o), and 2rjy (p). The PDB code and the native state of each protein are shown inside the panels with the N-terminus indicated by a blue ball. In each panel, a normalized histogram of the escape times obtained from simulations (boxes) is fitted to the diffusion model (solid line). The fractions of non-escaped trajectories at the largest time in the histograms are 0% (1iur), 1% (2jwd), 0.3% (2yxf), 4.6% (2uvs), 0% (1wxl), 5.6% (2ql0), 0.6% (2ci2), 0.2% (2vxf), 1.2% (2k3b), 1.2% (1shg), 4% (1f53), 0.6% (2zeq), 10.9% (2spz), 13.7% (1wt7), 1% (2vy4), and 2.6% (2rjy).
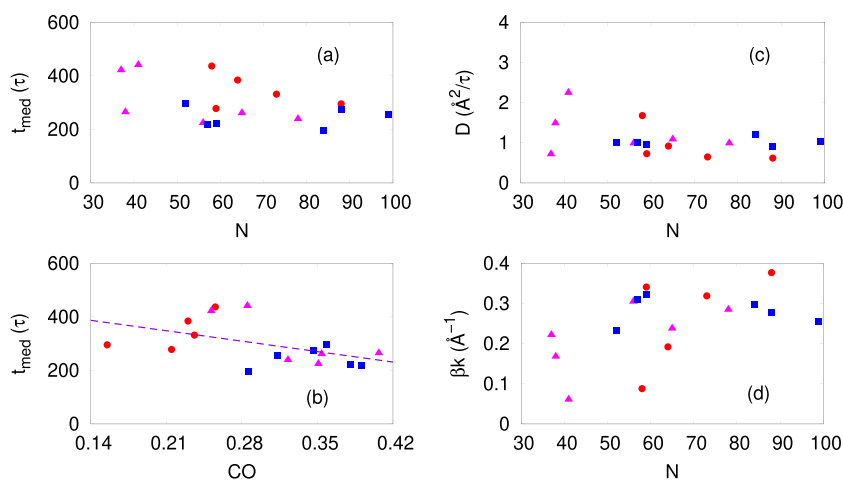

FIG. 9. [(a) and (b)] The median escape time t_{med} plotted against the protein length N (a) and the relative contact order CO (b) for GB1 and 16 proteins considered in Fig. 8. The values of t_{med} are obtained from simulations with the atomistic tunnel at $T = 0.85\ \epsilon/k_B$. The point type indicates the protein class, i.e., all-α (circles), all-β (squares), and α/β (triangles). [(c) and (d)] The diffusion constant D (c) and the potential parameter βk (d) plotted against the protein length N for 17 proteins considered in (a). D and βk are obtained by fitting the simulated escape time distribution to the diffusion model.
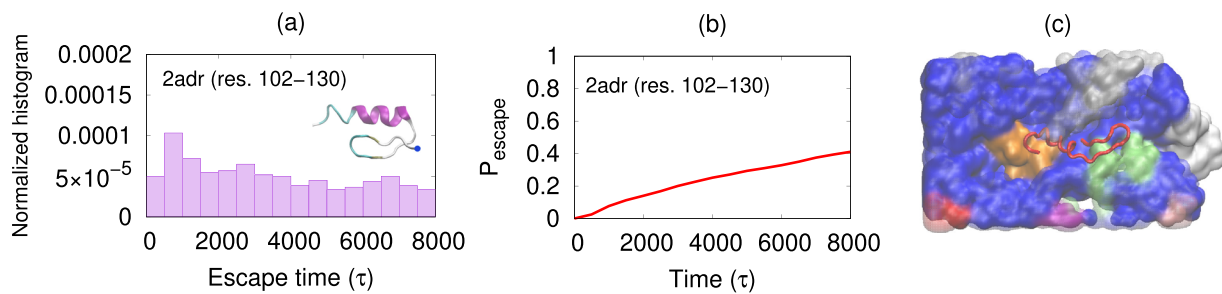
FIG. 10. (a) Histogram of the escape times at $T = 0.85\ \epsilon/k_B$ obtained from simulations at the atomistic tunnel for a 29-residue zinc-finger domain of ADR1 protein (pdb code: 2adr, res. 102–130) with the native structure of the domain shown in the inset. (b) Dependence of the escape probability, P_{escape}, on time for the system considered in (a). (c) A typical trapped conformation of the domain inside the tunnel.

zinc-finger domain (2vy4) that has the length of only 37 residues. In order to check whether a smaller zinc-finger domain can escape, we carried out the simulations for a 29-residue domain (res. 102–130) of the ADR1 protein with the pdb code 2adr. Figure 10 shows that this zinc-finger domain poorly escapes the tunnel with only 40% of the trajectories being able to escape after a long simulation time of 8000τ. A projected median escape time for this domain would be at least about 50 times larger than for GB1. Furthermore, the simulated escape time distribution as shown Fig. 10(a) cannot be fitted to the diffusion model. Figure 10(c) shows that a typical trapped conformation of this zinc-finger domain has the α-helix and the N-terminal β-hairpin formed and it is found deeply within the tunnel. The partial folding observed here is consistent with a recent experiment, which indicates that the 2adr zinc-finger domain can fold completely within the tunnel.[17] The difference between the zinc-finger domain of 2adr and that of 2vy4 is that the latter is eight residues longer and has a more developed β-hairpin with a coil-like flagging tail near the N-terminus. This flagging tail makes the β-hairpin formation inside the tunnel more difficult, allowing the N-terminus to reach out of the tunnel. It is suggested that the behavior of the 29-residue zinc-finger domain is similar to the diffusion with $k = 0$ in the diffusion model. A protein trapped entirely inside the tunnel would feel no free energy gradient and, therefore, has no indication on which direction to diffuse. It can be expected that the cross-over tunnel length[24] for the 29-residue zinc-finger domain is significantly shorter than the real length of the exit tunnel, and therefore, the protein is found in the slow diffusion regime.

C. Dependence of escape time on friction coefficient

This subsection is relevant only to the methodology used in the study, but it helps us to better interpret the previous results. All the simulations in Subsections III A and III B were done with the friction coefficient $\zeta = 1\ m\tau^{-1}$ for amino acids. This value of ζ may not be realistic for real proteins inside cells. Thus, we ask how the escape time would depend on ζ and whether one can extrapolate this dependence to obtain the real escape time. For this purpose, we carried out additional simulations for GB1 with $\zeta = 2, 4,$ and $8\ m\tau^{-1}$ at the atomistic tunnel with $T = 0.85\ \epsilon/k_B$. In these simulations, because the increased friction slows down the dynamics, the growth time per amino acid t_g was also increased to 200, 400, and 800τ, respectively,

for the given values of ζ. Figure 11(a) shows that the median escape time t_{med} is an almost perfect linear function of ζ. This linear dependence indicates that the simulation results are in the overdamped (large friction) regime. Figure 11(b) shows that the diffusion constant D of the escaping protein, obtained by fitting the simulated escape time distribution to the diffusion model, decreases with ζ like $D \sim \zeta^{-1}$. Together with the approximate linear dependence of D on temperature shown in Fig. 6, one finds a complete consistency with Einstein's relation $D = k_B T/\zeta^*$, where ζ^* is the friction coefficient of a Brownian particle. Thus, the protein at the tunnel behaves very much like a Brownian particle if one assumes that the quantity ζ^* is proportional to ζ and plays the role of an effective friction coefficient of the whole protein.
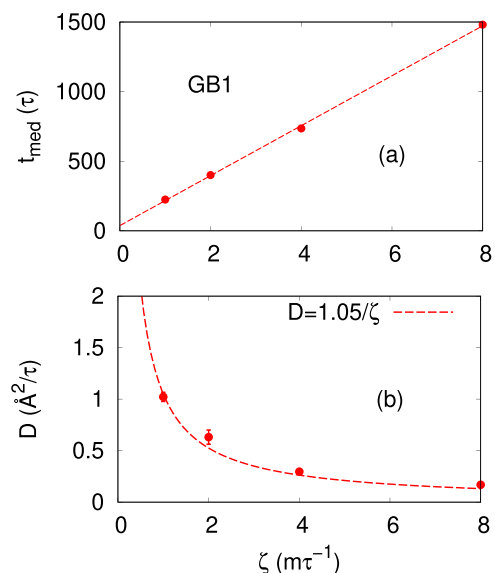

FIG. 11. Dependence of the median escape time t_{med} (a) and the diffusion constant D (b) on the friction coefficient ζ for protein GB1 at $T = 0.85\ \epsilon/k_B$. In (a), the dependence is fitted by a linear function, $t_{med} = 179.42\zeta + 37.62$ (dashed). In (b), the dependence is fitted by the function $D = 1.05/\zeta$ (dashed).

Given that $\sigma = 5$ Å, $m = 110$ g/mol, and $\epsilon \approx 0.7$ kcal/mol, the time unit in our simulation can be calculated as $\tau = \sqrt{m\sigma^2/\epsilon} \approx 3$ ps. The simulation's friction unit $m\tau^{-1} \approx 6 \times 10^{-11}$ g s^{-1}. The realistic friction coefficient of the amino acid in water can be obtained from the Stokes law, $\zeta_{water} = 6\pi\eta\sigma$, where $\eta = 0.01$ P is the viscosity of water at 25 °C. One obtains $\zeta_{water} \approx 9.4 \times 10^{-9}$ g s$^{-1} \approx 157m\tau^{-1}$. By extrapolating the linear dependence in Fig. 11(a), one finds that, at $\zeta = \zeta_{water}$, the median escape time for GB1 is $t_{med} \approx 3 \times 10^4 \tau \approx 90$ ns. This time appears to be too short for large scale motion like the protein escape.

Veitshans et al.[38] suggested that, at high friction, inertial terms in the Langevin equation are irrelevant, and the natural time unit is $\tau_H = \zeta\sigma^2/k_BT$. For water at room temperature, a direct calculation from the last formula gives $\tau_H \approx 0.6$ ns. With some scaling factor, Veitshans et al.[38] estimated that $\tau_H \approx 3$ ns. With these revised time units, the theoretical estimate of t_{med} is either 18 μs or 90 μs. The experimental refolding time of GB1 at neutral pH is about 1 ms.[39] We have checked that within the same Go-like model at $T = 0.85$ ϵ/k_B, the median refolding time is about 50% larger than t_{med}. Thus, the model prediction of the refolding time is smaller but within the same order of magnitude as the experimental value, given that some uncertainties are associated with the estimates. With the above estimates, and given the results of Sec. III B, it can be expected that the escape times of single-domain proteins are of the order of 0.1 ms, i.e., in the sub-millisecond scale.

IV. CONCLUSION

There are several remarks we would like to mention for the conclusion. First, the shape of the ribosomal exit tunnel appears to cause increased difficulty for nascent proteins to escape compared to a smooth cylinder tunnel. This difficulty is reflected by the appearance of kinetic traps in the escape pathways leading to lengthened escape times. We have shown that the trapped conformations are completely located inside the tunnel and usually have a significant development of tertiary structure. The formation of tertiary structure elements inside the tunnel correlates with the modulated shape of the ribosome tunnel, which has some narrow parts but also some wider parts that can hold a tertiary unit. In contrast, the equivalent cylinder tunnel of 13.5 Å diameter does not allow tertiary structure formation and yields no kinetic trapping. Second, thermal fluctuations are important for the escape of nascent proteins. We have shown that for GB1, a significant fraction of escape trajectories get trapped at $T = 0.75$ ϵ/k_B, but not at the physiological temperature $T = 0.85$ ϵ/k_B. Interestingly, at the latter temperature, almost all of the 17 single-domain proteins considered are able to escape efficiently, even the smallest one, the 37-residue 2vy4. Note that the trapping arises solely due to the folding of a protein within the tunnel, and thus, it depends on temperature. At high temperatures, folding is slow and diffusion is fast; therefore, a protein would have a low probability of getting trapped before escaping from the tunnel. At low temperatures, folding is fast while diffusion is slow, the trapping probability is increased. On the other hand, a protein would get out from a trap easier at a higher temperature.

Third, if a protein or peptide is too small, it cannot escape efficiently from the tunnel. The example of the 29-residue zinc-finger domain of 2adr shows that the protein is severely trapped inside the

tunnel with the median escape time about of two orders of magnitude larger than that of GB1. The trapped protein is not guided by a potential gradient toward the escape direction. This example reflects a relation between the protein size and a cross-over tunnel length for efficient diffusion, as predicted by our previous study with the cylinder tunnel.[24] Fourth, the escape time of single-domain proteins weakly depends on the native state topology and is almost independent of the protein size. Our model predicts that the protein escape time at the ribosome tunnel is of the order of 0.1 ms. The latter is much shorter than the time needed by the ribosome to translate one codon (tens of milliseconds), therefore not allowing nascent proteins to jam the ribosome tunnel.

One may ask to what extent hydrophobic and electrostatic interactions of a nascent protein with the ribosome exit tunnel can alter the above-obtained results. It is well-known that the tunnel's wall formed by the ribosomal RNA is negatively charged. We found that for the ribosome of H. marismortui, the tunnel's inner surface with $x < 82$ Å has only four hydrophobic side-chains that are clearly exposed within the tunnel, belonging to Phe61 of protein L4, Met130 of protein L22, and Met26 and Leu27 of protein L38, and about 10 exposed charged amino acid side-chains. These statistics indicate that the effect of hydrophobic interaction on the escape process can be considerably small, whereas the Coulomb interaction may have a strong effect on nascent chains. However, if the total charge of a nascent protein is neutral, the electrostatic forces on the protein's positive and negative charges may cancel out each other. Thus, it is reasonable to expect that the energetic interactions of nascent proteins with the tunnel can lead to specific changes in the escape behavior for individual proteins, but, on average, they give only higher-order corrections to what is obtained with the excluded volume interaction.

Finally, like for the cylinder tunnel, it is found that the escape time distribution at the atomistic tunnel for various proteins follows very well the one-dimensional diffusion model of a drifting Brownian particle. This consistent finding suggests that the protein escape at the ribosome tunnel may have been designed by nature to be simple, efficient, and predictable for the smooth functioning of the ribosome. This result also proves the usefulness of using simple stochastic models to understand the complex dynamics of biomolecules.

SUPPLEMENTARY MATERIAL

See the supplementary material for the dependences of the median escape time and the mean escape time on temperature for protein GB1 at the cylinder tunnel of length $L = 82$ Å and diameter $d = 13.5$ Å, and for the histogram of the escape time and the time dependences of the escape probability and the probability of the C-terminal β-hairpin formation inside the tunnel for protein GB1 at a tunnel model that considers all the heavy atoms of the ribosomal RNA and the ribosomal proteins.

ACKNOWLEDGMENTS

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] I. S. Gabashvili, R. K. Agrawal, C. M. T. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank, and P. Penczek, Cell **100**, 537 (2000).

[2] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, Science **289**, 905 (2000).

[3] D. V. Fedyukina and S. Cavagnero, Annu. Rev. Biophys. **40**, 337 (2011).

[4] F. Trovato and E. P. O'Brien, Annu. Rev. Biophys. **45**, 345 (2016).

[5] A. Javed, J. Christodoulou, L. D. Cabrita, and E. V. Orlova, Acta Crystallogr., Sect. D: Struct. Biol. **73**, 509 (2017).

[6] M. Thommen, W. Holtkamp, and M. V. Rodnina, Curr. Opin. Struct. Biol. **42**, 83 (2017).

[7] H. Nakatogawa and K. Ito, Cell **108**, 629 (2002).

[8] T. Tenson and M. Ehrenberg, Cell **108**, 591 (2002).

[9] R. Berisio, F. Schluenzen, J. Harms, A. Bashan, T. Auerbach, D. Baram, and A. Yonath, Nat. Struct. Mol. Biol. **10**, 366 (2003).

[10] G. Ziv, G. Haran, and D. Thirumalai, Proc. Natl. Acad. Sci. U. S. A. **102**, 18956 (2005).

[11] S. Kirmizialtin, V. Ganesan, and D. E. Makarov, J. Chem. Phys. **121**, 10268 (2004).

[12] J. Lu and C. Deutsch, Nat. Struct. Mol. Biol. **12**, 1123 (2005).

[13] A. Kosolapov and C. Deutsch, Nat. Struct. Mol. Biol. **16**, 405 (2009).

[14] A. H. Elcock, PLoS Comput. Biol. **2**, 0824 (2006).

[15] E. P. O'Brien, S.-T. D. Hsu, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, J. Am. Chem. Soc. **132**, 16928 (2010).

[16] W. Holtkamp, G. Kokic, M. Jäger, J. Mittelstaet, A. A. Komar, and M. V. Rodnina, Science **350**, 1104 (2015).

[17] O. B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Müller-Lucks, F. Trovato, J. D. Puglisi, E. P. O'Brien *et al.*, Cell Rep. **12**, 1533 (2015).

[18] R. Kudva, P. Tian, F. Pardo-Avila, M. Carroni, R. B. Best, H. D. Bernstein, and G. von Heijne, Elife **7**, e36326 (2018).

[19] L. D. Cabrita, S.-T. D. Hsu, H. Launay, C. M. Dobson, and J. Christodoulou, Proc. Natl. Acad. Sci. U. S. A. **106**, 22239 (2009).

[20] C. Eichmann, S. Preissler, R. Riek, and E. Deuerling, Proc. Natl. Acad. Sci. U. S. A. **107**, 9111 (2010).

[21] E. P. O'Brien, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, J. Am. Chem. Soc. **133**, 513 (2011).

[22] H. Krobath, E. I. Shakhnovich, and P. F. N. Faísca, J. Chem. Phys. **138**, 215101 (2013).

[23] P. T. Bui and T. X. Hoang, J. Chem. Phys. **144**, 095102 (2016).

[24] P. T. Bui and T. X. Hoang, J. Chem. Phys. **149**, 045102 (2018).

[25] C. M. Dobson, Nature **426**, 884 (2003).

[26] N. R. Voss, M. Gerstein, T. A. Steitz, and P. B. Moore, J. Mol. Biol. **360**, 893 (2006).

[27] K. Dao Duc, S. S. Batra, N. Bhattacharya, J. H. Cate, and Y. S. Song, Nucleic Acids Res. **47**, 4198 (2019).

[28] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).

[29] T. X. Hoang and M. Cieplak, J. Chem. Phys. **112**, 6851 (2000).

[30] T. X. Hoang and M. Cieplak, J. Chem. Phys. **113**, 8319 (2000).

[31] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).

[32] T. M. Schmeing, K. S. Huang, S. A. Strobel, and T. A. Steitz, Nature **438**, 520 (2005).

[33] P. T. Bui and T. X. Hoang, J. Phys.: Conf. Ser. **1506**, 012022 (2020).

[34] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, 1992).

[35] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes* (Chapman & Hall, 1965), pp. 219–223.

[36] R. Campos-Olivas, R. Aziz, G. L. Helms, J. N. S. Evans, and A. M. Gronenborn, FEBS Lett. **517**, 55 (2002).

[37] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).

[38] T. Veitshans, D. Klimov, and D. Thirumalai, Folding Des. **2**, 1 (1997).

[39] P. Alexander, J. Orban, and P. Bryan, Biochemistry **31**, 7243 (1992).