The Journal of Chemical Physics

The protein escape process at the ribosomal exit tunnel has conserved mechanisms across the domains of life

Cite as: J. Chem. Phys. **158**, 015102 (2023); https://doi.org/10.1063/5.0129532
Submitted: 05 October 2022 • Accepted: 14 December 2022 • Accepted Manuscript Online: 15 December 2022 • Published Online: 04 January 2023

Phuong Thuy Bui and Trinh Xuan Hoang

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

The protein escape process at the ribosomal exit tunnel has conserved mechanisms across the domains of life

Cite as: J. Chem. Phys. 158, 015102 (2023); doi: 10.1063/5.0129532
Submitted: 5 October 2022 • Accepted: 14 December 2022 •
Published Online: 4 January 2023

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

AFFILIATIONS

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

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

ABSTRACT

The ribosomal exit tunnel is the primary structure affecting the release of nascent proteins at the ribosome. The ribosomal exit tunnels from different species have elements of conservation and differentiation in structural and physico-chemical properties. In this study, by simulating the elongation and escape processes of nascent proteins at the ribosomal exit tunnels of four different organisms, we show that the escape process has conserved mechanisms across the domains of life. Specifically, it is found that the escape process of proteins follows the diffusion mechanism given by a simple diffusion model, and the median escape time positively correlates with the number of hydrophobic residues and the net charge of a protein for all the exit tunnels considered. These properties hold for 12 distinct proteins considered in two slightly different and improved Gō-like models. It is also found that the differences in physico-chemical properties of the tunnels lead to quantitative differences in the protein escape times. In particular, the relatively strong hydrophobicity of *E. coli*'s tunnel and the unusually high number of negatively charged amino acids on the tunnel's surface of *H. marismortui* lead to substantially slower escapes of proteins at these tunnels than at those of *S. cerevisiae* and *H. sapiens*.

I. INTRODUCTION

The ribosomal exit tunnel is a narrow structure connecting the peptidyl transferase center (PTC), where polypeptide polymerization takes place during translation, to the surface of the ribosome. It is the first structure encountered by the nascent polypeptides and is the only passage for nascent proteins to be released from the ribosome. The ribosomal exit tunnel is believed to play important roles in translation regulation[1–3] as well as co-translational protein folding.[4–6] The tunnel's dimensions, 10–20 Å in width and 80–100 Å in length, allow it to accommodate up to ~40 amino acids[7] but limit the size of the folded peptide inside the tunnel.[8] In general, the protein and RNA composition of the ribosome can vary in different domains and different species, leading to different structural details of the exit tunnel. A comparison of the tunnel structures of a range of species has shown certain similarities and differences.[9] For example, it has been shown that the upper part of the tunnel, near the PTC, is relatively conserved across species. On the other hand, the lower part of the tunnel is substantially narrower in eukaryotes than in bacteria, which may have implications for antibiotic resistance.[10]

The post-translational escape of nascent proteins at the ribosomal exit tunnel is the final release of a protein from the ribosome when the protein's N-terminus is no longer attached to the PTC. This process is a necessary step for a nascent protein to empty the ribosomal exit tunnel for the next translation process and to complete its own folding to the native state. Only a very few studies have addressed this process until recently. In earlier works,[11,12] by coarse-grained simulations in the Gō-like models, we have shown that the escape process is assisted by the folding of the nascent

protein and is akin to the diffusion of a Brownian particle in a linear potential field. In more recent studies, by using the atomistic tunnel of *H. marismortui*, it was shown that the roughness of the exit tunnel can increase the difficulty of nascent proteins to escape[13] and that the escape time is modulated by energetic interactions of the protein with the exit tunnel, such as hydrophobic and electrostatic interactions.[14] Another study with the *E. coli*'s tunnel suggests that electrostatic interaction can extremely delay protein escape.[15]

The present study is aimed at extending our understanding of the protein escape process at the ribosomal exit tunnels of different species. In particular, we consider the exit tunnels from four organisms, namely *E. coli*, *H. marismortui*, *S. cerevisiae,* and *H. sapiens*, which are representatives from all three domains of life (bacteria, archaea, and eukarya). Their ribosome structures have been experimentally determined at high resolutions, allowing us to have atomic details for the tunnel models used in the simulations. The belief is that the differences in the structural and chemical details of the exit tunnels considered will help us have a more complete picture of the protein escape process at ribosomal exit tunnels.

We used the same simulation approach as in the previous study[14] to study the escape process, but with a larger set of proteins and improved models for the nascent proteins. The Gō-like models in the present study, namely the Gō-MJ and Gō-MJ-nn models, incorporate the well-known Miyazawa–Jernigan's contact energy matrix in the depths of the Lennard-Jones potentials for native and non-native contacts, and thereby, to a certain degree, take into account the effects of the amino acid sequences in the escape and folding of these proteins. The energy parameters in the Gō-like models are also rescaled such that the melting temperature in the model matches the experimental melting temperature of each protein.

We will show that while there are significant variations in the escape times among the exit tunnels of different organisms, the mechanisms governing the protein escape are remarkably similar at different exit tunnels, suggesting that they are conserved across the domains of life.

II. MODELS AND METHOD

A. Improved Gō-like models

Gō-like models have been widely used to study protein folding dynamics due to their simplicity and effectiveness.[16–18] They are a class of models that emphasize the importance of native interactions[19] and can be applied to any protein with a known native structure. In this work, we used two variants of improved Gō-like models to simulate nascent proteins: the first one incorporates variable strengths of the potentials for native contacts, and the second one also includes attractive potentials for non-native contacts. These models partially take into account the effects of the amino acid sequence through the use of the Miyazawa–Jernigan matrix for inter-residue contact energies,[20] in a similar manner to other Gō-like models used in the literature.[21–24]

1. Gō-MJ model

The Gō-MJ model is modified from the one of Clementi *et al.*[25] by adding a variation in the strengths of the potentials for native contacts. Considering only the C_α atoms, the potential energy of a protein in a given conformation is given by

$$V_{\text{Gō–MJ}} = \sum_{\text{bonds}} K_b(r_{i,i+1} - r^*_{i,i+1})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta^*)^2 + \sum_{\text{dihedrals}} \sum_{n=1,3} K_\phi^{(n)}[1 - \cos(n(\phi - \phi^*))] + \sum_{j>i+3}^{\text{native}} \epsilon_{ij}^{\text{NC}} \left[5\left(\frac{r^*_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r^*_{ij}}{r_{ij}}\right)^{10} \right] + \sum_{j>i+3}^{\text{non-native}} \epsilon \left(\frac{\sigma}{r_{ij}}\right)^{12}, \tag{1}$$

where the terms on the right side correspond to the potentials on the bond lengths, bond angles, dihedral angles, native contacts, and non-native contacts, respectively, as described in detail elsewhere.[13,25] The native contacts are determined from an all-atom consideration[26] of the protein structure in the Protein Data Bank (PDB) and the atomic van der Waals radii.[27] r_{ij} is the distance between residue i and residue j, the * symbol denotes the native state's value, σ is an effective diameter of amino acids, and ϵ is an energy parameter. The value of $\epsilon_{ij}^{\text{NC}}$, which sets the potential depth for a native contact, is calculated as

$$\epsilon_{ij}^{\text{NC}} = \frac{n_{ij} e_{\text{HB}} + e_{\text{MJ}}(s_i, s_j)}{u} \epsilon, \tag{2}$$

where n_{ij} is the number of hydrogen bonds between residue i and residue j in the native state, $e_{\text{HB}} = 1.5$ kcal/mol is a hydrogen bond's energy, $e_{\text{MJ}}(s_i, s_j)$ is the inter-residue contact energy for the pair of amino acids of the types s_i and s_j given by the Miyazawa–Jernigan matrix[20] with the energy converted to kcal/mol and given in the absolute value, and u is a normalizing factor such that the average energy of all the native contacts is ϵ. Other parameters in the model are $\sigma = 5$ Å, $K_b = 100\,\epsilon$ Å^{-2}, $K_\theta = 20\,\epsilon$ (rad)$^{-2}$, $K_\phi^{(1)} = \epsilon$, $K_\phi^{(3)} = 0.5\epsilon$.

2. Gō-MJ-nn model

In the Gō-MJ-nn model, the last term in Eq. (1) is replaced by

$$\sum_{j>i+3}^{\text{non-native}} \epsilon_{ij}^{\text{NN}} \left[5\left(\frac{\sigma_1}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_1}{r_{ij}}\right)^{10} \right], \tag{3}$$

which provides attraction to the non-native contacts. The potential depth for a non-native contact is calculated as

$$\epsilon_{ij}^{\text{NN}} = f \frac{e_{\text{MJ}}(s_i, s_j)}{u} \epsilon, \tag{4}$$

where f is a factor that sets the relative strengths of non-native contacts. In the present study, we used $\sigma_1 = 5.5$ Å and $f = 0.4$.

In both the Gō-MJ and Gō-MJ-nn models, ϵ is the single parameter that sets the energy scale of the whole protein. Because temperature effects are important for the dynamics of proteins, especially for their diffusion in the ribosomal tunnel, it is important to have the correct energy scale for each protein. Following previous work,[14] we determined ϵ individually for each protein by fitting the melting temperature in the model to the experimental melting temperature, T_m. The melting temperature in the model is defined by $T_{\max}$, the temperature of the specific heat's maximum of a protein obtained by simulations. The parameter ϵ is calculated as $\epsilon = \frac{(273+T_m)}{503.2195 \times T_{\max}}$ (kcal/mol), where $T_{\max}$ is given in units of ϵ/k_B and

T_m is given in °C. The values of $T_{\max}$ and ϵ in both the Gō-MJ and Gō-MJ-nn models for a list of 12 proteins considered are given in Table S1.

B. Tunnel model

Models of the exit tunnels are constructed based on the PDB structures of the large ribosomal subunits of the organisms. The structures of the PDB IDs 7k00,[28] 1jj2,[29] 5gak,[30] and 4ug0[31] were considered for the ribosomes of *E. coli*, *H. marismortui*, *S. cerevisiae,* and *H. sapiens*, respectively. The model considers all the heavy atoms for ribosomal RNA but only C_α's for ribosomal proteins. To reduce computational time, we kept only atoms within a cylinder of radius R centered around an approximate chosen tunnel axis for the tunnel model. The value of R must be sufficiently large to enclose the atoms of the tunnel's wall. We have chosen $R = 30$ Å for the ribosome tunnels of *H. marismortui*, *S. cerevisiae,* and *H. sapiens* and $R = 45$ Å for those of *E. coli*. The model also ignores the motion of the ribosome; thus, all the tunnel atoms are kept fixed during the simulations.

For interactions of the tunnel with nascent proteins, the model used in this study is the T3 model described in Ref. 14, which contains three types of interactions: excluded volume, hydrophobic, and electrostatic. Details of the interaction potentials are given in Ref. 14. In short, the exclude volume interaction provides a short-range repulsion between the tunnel's atoms and the nascent chain's residues. The hydrophobic interaction gives rise to an attraction between hydrophobic residues (Ile, Leu, Phe, Met, Val, Pro, and Trp) of a nascent protein and those of the same type in ribosomal proteins via a 10–12 Lennard-Jones potential. The depth of this potential is constant for all pairs of hydrophobic residues and is equal to $\epsilon_{\text{hydr}} = 1.2$ kcal/mol. The electrostatic interaction is given by a screened Coulomb potential from the Debye–Hückel theory with Debye's screening length $\lambda_D = 10$ Å. The electrostatic interaction is considered between all charged residues of a nascent protein and all charged centers of rRNA and ribosomal proteins. In rRNA, each phosphorus atom is assigned with the charge $q = -1e$. In nascent and ribosomal proteins, lysine and arginine are given with the charge $q = +1e$, whereas aspartic acid and glutamic acid are given with $q = -1e$. The charges of amino acids are assumed to be concentrated on the C_α atoms.

C. Simulation method

A molecular dynamics (MD) method based on the Langevin equation of motion is used to simulate the motions of nascent chains. Details of the method are given in Ref. 11. We adopt a reduced unit system such that the mass unit is the average mass m of amino acids, the length unit is the effective diameter σ of amino acids, and the energy unit is kcal/mol. The friction coefficient of amino acids used in the simulations is $\zeta = 1\sqrt{m\sigma^{-2}(\text{kcal/mol})}$. Given that $m = 120$ g/mol and $\sigma = 5$ Å, simulation time is measured in the units of $\tau = \sqrt{m\sigma^2/(\text{kcal/mol})} \approx 3$ ps. This value of the time unit, suitable for the low-friction regime,[32] results in a much shorter timescale of the simulation folding times than the real folding times. It has been shown that the correct timescale can be reached by simulations by increasing ζ to its realistic value and using the high-friction estimate, $\tau_H = 3$ ns, of the time unit.[14,32,33]

For an isolated protein, the temperature of the specific heat's maximum $T_{\max}$ is determined from the temperature dependence of the specific heat by using replica-exchange molecular dynamics (REMD) simulations[34] and the weighted histogram analysis method.[35,36] For studying the protein escape at a ribosome tunnel, both the translation process and the escape process are simulated. In the translation process, a nascent chain is elongated at the position of the PTC at a constant rate corresponding to a growth time t_g per residue. t_g must be chosen sufficiently large such that the escape properties are converged given that the growth times in cells are orders of magnitude larger than in simulations. We used $t_g = 400\tau$ for most proteins and $t_g = 2000\tau$ for proteins that are kinetically trapped at the tunnel. The escape time is measured from the moment of complete elongation (the C-terminal residue is released from the PTC) until the nascent protein has fully escaped the tunnel. All simulations of the translation and escape processes of proteins are carried out at room temperature, $T = 300$ K. Typically, the escape time distribution and the escape probability are calculated from 1000 independent trajectories for each protein.

D. Diffusion model

The diffusion model[12] considers the protein escape process as the diffusion of a Brownian particle in a one-dimensional potential field $U(x)$, with x being the position of the particle. Such a process is governed by the Smoluchowski equation. Given the linear form $U(x) = -kx$ of the external potential, where k is a constant force acting on the particle, the distribution of the escape time can be obtained from an exact solution of the Smoluchowski equation and is given by[12]

$$g(t) = \frac{L}{\sqrt{4\pi D t^3}} \exp\left[-\frac{(L - D\beta k t)^2}{4Dt}\right], \tag{5}$$

where L is the diffusion distance equal to the tunnel length, D is the diffusion constant assumed to be position independent, $\beta = (k_B T)^{-1}$ is the inverse temperature, where k_B is the Boltzmann constant. Interestingly, the escape time distribution in Eq. (5) can fit the data from various simulations of protein escape in the Gō-like model.[12–14] It has been shown that the free energy of a protein at the ribosome tunnel is approximately linear along an escape coordinate,[11,14] which justifies the linear form of $U(x)$ in the diffusion model.

The distribution in Eq. (5) gives the mean value $\mu_t = L/(D\beta k)$ and the standard deviation $\sigma_t = \frac{\sqrt{2L}}{D(\beta k)^{3/2}}$ for the escape time.[12] Note that both μ_t and σ_t diverge when $k = 0$, for which $g(t)$ becomes a heavy-tailed Lévy distribution.

III. RESULTS

A. Differences in physico-chemical properties of nascent proteins and the ribosomal exit tunnels

This study considered 12 small globular proteins with known melting temperatures T_m. They consist of the B1 domain of protein G (1pga),[37] the Rop protein (1rop),[38] the SH3 domain (1shg),[39] the Z domain of Staphylococcal protein A (2spz),[40] the Cro repressor (1orc),[41] chymotrypsin inhibitor 2 (2ci2),[42] antifreeze protein (1msi),[43] cold-shock protein (1csp),[44] ubiquitin (1ubq),[45]

histidine-containing phosphocarrier protein HPr (1poh),[46] hyperthermophilic archaeal DNA-binding protein Sso10b2 (1udv),[47] and barnase (1a2p)[48] with the PDB IDs of their native structures enclosed in the parentheses. The references associated with the proteins correspond to the experimental studies in which T_m has been reported (see Table S1 for the values of T_m and other properties of the proteins). For convenience, we will call the proteins by their PDB IDs. These proteins have lengths between 56 and 108 amino acids and distinct native structures with two all-α, two all-β, and eight α/β proteins. Their T_m values range from 43.8 °C (for 1csp) to 157.5°C (for 1udv). Our analyses show that the fraction of hydrophobic residues in their amino acid sequences varies from ~21% to ~47%, the fraction of positively charged amino acids varies between ~6% and ~18%, and the fraction of negatively charged amino acids ranges from ~6% to ~19.6%. The protein net charges are from $-6e$ to $+3e$. These properties indicate that the proteins considered have a wide range of specificities, leading to diverse interactions with the exit tunnel.

All proteins in our considerations are small-sized and cannot be compared to any protein length distribution in proteomes (Fig. S1). However, it can be expected that the chain length does not impact the escape time, as shown in one of our previous studies.[12] The structural class, on the other hand, can have a minor effect on the protein escape, with α-proteins escaping somewhat more slowly than β-proteins.[12] We have checked that the structural class composition of our set of proteins is not too different from that of Richardson's Top2018 high-quality protein structures[49] with a strong dominance of α/β-proteins (see Table S2). Even though our protein set is quite small with only 12 proteins, its protein sequences have similar ranges of hydrophobicity and fractions of positively and negatively charged amino acids as found in various proteomes, whose sequences are taken from the UniProt database[50] (Figs. S2 and S3). For example, the range of fraction of hydrophobic amino acids in our 12 proteins is shared by 97% of the protein population in the human proteome, whereas the corresponding numbers for the ranges of fraction of positively and negatively charged amino acids are 92% and 88% [Figs. S2(a)–S2(c)]. In the proteomes of *S. cerevisiae* [Figs. S2(d)–S2(f)] and *E. coli* [Figs. S3(a)–S3(c)], these percentages are also very high, ranging from 84% to 98%. The proteins in *H. marismortui*'s proteome [Figs. S3(d)–S3(f)] tend to have a lower fraction of positively charged amino acids and a higher fraction of negatively charged amino acids than in other organisms, resulting in only 78% and 68% of the proteome sharing the ranges of these two fractions, respectively, with the 12 proteins considered. These statistics suggest that the chosen proteins to a good extent reflect the variabilities of hydrophobic and charge compositions of the proteins in the organisms considered, though they do slightly worse for *H. marismortui*. They can be considered representative of typical globular proteins in terms of hydrophobic and charge fractions in the amino acid sequences.

The ribosomal exit tunnels of the four organisms considered have notable differences and similarities. The differences in the shape of these tunnels can be visualized through the graphs representing their effective diameter d along the tunnel axis x shown in Fig. 1(a). For each position x, d is calculated as $d = 2\sqrt{(S/\pi)}$, where S is the tunnel's cross-sectional area accessible by a probe sphere of radius 3 Å. Although the effective diameter does not reflect all information about the shape of a tunnel, it already shows that the detailed shapes are different for different species. The diameter of *H. marismortui*'s tunnel appears to be the most uniform, while the other tunnels show stronger variations of d. The tunnel for *E. coli* is somewhat wider than the other tunnels.[9] The diameter profiles in Fig. 1(a) also show some similarities, such as that the average widths of the tunnels are more or less the same, the tunnels become wider near the exit, and the position at which the tunnel is narrowest appears to be about half-way from the opening of the tunnel for all tunnels. The d profile of *S. cerevisiae* looks the most similar to that of *H. sapiens*.

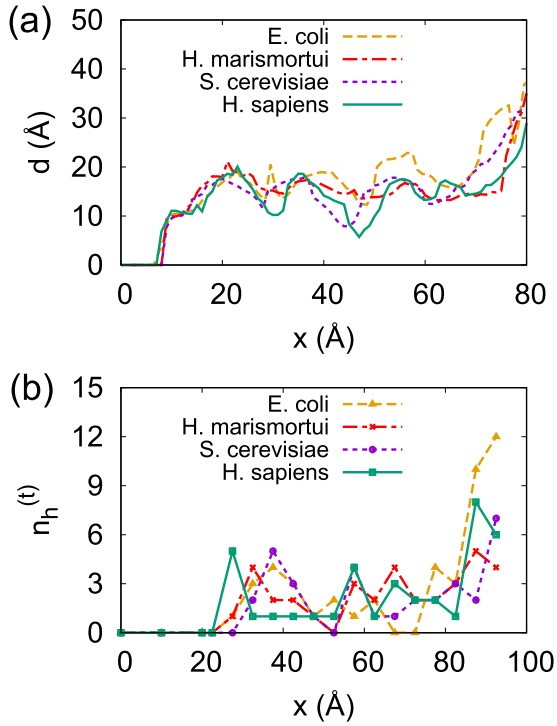

FIG. 1. (a) Dependence of the effective diameter, d, on the coordinate x along an approximate tunnel axis for the ribosomal exit tunnels of four organisms considered. The lines shown are for *E. coli* (dashed), *H. marismortui* (dashed-dotted), *S. cerevisiae* (dotted), and *H. sapiens* (solid), as indicated. The effective diameter d is calculated as $d = 2\sqrt{(S/\pi)}$, where S is the tunnel cross-sectional area accessible for a probe sphere with radius 3 Å. (b) Distribution of hydrophobic residues at the tunnel surface along the tunnel axis for the considered species, as indicated. For a given data point at position x, $n_h^{(t)}$ is the number of surface's hydrophobic residues within 5 Å from x along the tunnel axis.

We have inspected the tunnel surfaces to get information about the hydrophobic and charged amino-acid residues exposed on the surface from ribosomal proteins. The numbers of these residues for each tunnel are listed in Table I. The distribution of hydrophobic residues along the tunnel axis is shown in Fig. 1(b). It is found that *E. coli* has the highest number of hydrophobic residues on the tunnel surface, about 30% higher than the other organisms. The hydrophobic residues are the most abundant near the tunnel exit for *E. coli*, *S. cerevisiae*, and *H. sapiens* [Fig. 1(b)]. It is interesting to note that almost all the charges of amino acids on the tunnel surface are positive charges for *E. coli*, *S. cerevisiae*, and *H. sapiens*

TABLE I. Hydrophobic and charged properties of the ribosomal exit tunnels' surfaces in the organisms. For each organism, the listed properties are the number of hydrophobic residues ($N_h^{(t)}$), the numbers of positively ($N_+^{(t)}$) and negatively ($N_-^{(t)}$) charges of ribosomal amino acid residues that are found at the tunnel's surface. Note that these properties do not refer to the ribosomal RNA.

Organism	$N_h^{(t)}$	$N_+^{(t)}$	$N_-^{(t)}$
E. coli	46	30	4
H. marismortui	35	26	19
S. cerevisiae	33	32	4
H. sapiens	37	26	1

(Table I), suggesting that the charged amino acids play an important role in the function of the exit tunnel. Note that the ribosomal RNA is negatively charged, so only positively charged amino acids can significantly change the electrostatic potential inside the tunnel. An exception is found for *H. marismortui*, for which the number of negatively charged amino acids on the tunnel surface is much higher than in the other organisms, even though it is still significantly smaller than the number of positively charged ones (19 vs 26). The distinction in the electrostatics of *H. marismortui*'s tunnel may be related to the fact that this species can survive in extreme environmental conditions, such as at high temperatures, with high salt concentrations, or at high or low pH.

B. Conservation of the diffusion mechanism of the escape process

We have carried out simulations of the nascent chain's growth and the escape processes of all proteins considered in the Gō-MJ model at the four ribosomal tunnels and in the Gō-MJ-nn model at the human ribosomal tunnel only. In most cases, the protein can escape easily at the exit tunnel, but for several proteins at some of the tunnels, kinetic trapping can delay the escape. A kinetic trap is found in a simulation if the protein gets stuck in some state at the tunnel, leading to a much longer escape time, more than ten times longer than in an average trajectory. Kinetic trapping can be due to the roughness in the shape of the exit tunnel as well as the interactions between nascent proteins and the tunnel wall.[13,14] Interestingly, it has been shown that the probability of trapping a protein decreases with the growth time per residue t_g, and can become negligibly small at realistic translation rates.[14] In our study, we have simulated the easily escaped proteins with $t_g = 400\tau$, whereas those with kinetic trapping with the increased $t_g = 2000\tau$. The latter value of t_g reduces the trapping probability to below 5% and makes the statistics reliable. We have checked that a further increase in t_g produces very little changes in the escape time distribution and the median escape time.

Figure 2 shows that the escape probability, P_{escape}, increases sigmoidally with time and asymptotically approaches the value of 1 for all proteins in both the Gō-MJ and Gō-MJ-nn models at the human ribosomal exit tunnel. The proteins also escape efficiently at all other ribosomal exit tunnels. The proteins that are more likely to get kinetically trapped are 1pga, 1rop, 1orc, and 1udv, with 1udv being the slowest escaper. The median escape time, t_{esc}, the time at which $P_{\text{escape}} = 0.5$, varies among the proteins from a few hundred to a few thousand τ (see Table III).

We inspected the diffusion mechanism of the escape process by examining the escape time distributions of the proteins. The model mechanism is that of the diffusion model described in Sec. II B, which corresponds to the diffusion of a one-dimensional Brownian particle in a linear potential field. Interestingly, for all proteins

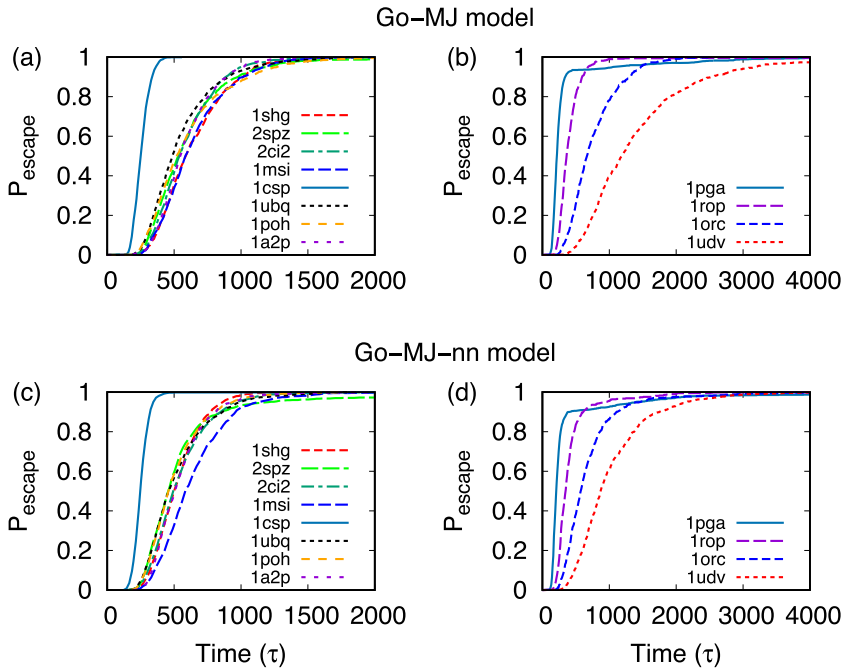

FIG. 2. The escape probability, P_{escape}, as a function of time at $T = 300$ K for proteins in the Gō-MJ [(a) and (b)] and the Gō-MJ-nn [(c) and (d)] model at the human ribosomal exit tunnel. The eight proteins in panels (a) and (c) were simulated with $t_g = 400\tau$, whereas the four proteins in panels (b) and (d) were simulated with $t_g = 2000\tau$. An increased value of t_g was used because the latter proteins have higher probabilities of kinetic trapping.

and all the tunnels considered, the escape time distribution follows relatively well that of the diffusion model. For example, Fig. 3 shows that the histogram of the escape times of the 2ci2 protein obtained by the simulations can be fitted to the distribution function in Eq. (5) for all the exit tunnels considered. Thus, the diffusion mechanism is conserved among the proteins and among the species, although the individual distributions can be different from each other. From the fits to the diffusion model, we can get the values of the parameters D and k, which can be considered as an effective diffusion constant of a protein at a tunnel and an effective mean force acting on the protein along the escape coordinate, respectively. These are highly collective quantities that reflect the complex dynamics of nascent proteins at the exit tunnels. The values of D and k are listed in Table II for all the proteins in each tunnel. They strongly vary with the protein and with the tunnel (Fig. S4). D is in the range from 0.4 to 1.2 $\text{Å}^2\tau^{-1}$. With $\tau = 3$ ps, the obtained values of D are of the order of 10^{-8} m^2 s^{-1}, i.e., about two orders of magnitude larger than diffusion constants of isolated proteins in water ($\sim 10^{-10}$ m^2 s^{-1}).[51] Note that D depends on the friction coefficient ζ, and the value of ζ used in the simulations is 100 times smaller than that of amino acids in water.[14] It is expected that, at realistic friction, D is smaller but of the same order of magnitude as that of isolated proteins. The force k varies more strongly than D. It is interesting that the obtained values of k are in the range of a sub-piconewton to a few tens of piconewtons, which is within the scale of molecular forces in proteins.[52]

C. Conservation of the effects of hydrophobic and electrostatic interactions on the protein escape time

To evaluate the effects of the tunnel's interactions on the protein escape, we investigated the dependence of the median escape time, t_{esc}, on the number of hydrophobic residues, N_h, and the net charge, Q, of a protein (Table III). Figure 4 shows that t_{esc} positively correlates with both N_h and Q for all the species considered, indicating that both hydrophobic and electrostatic interactions modulate the escape time. The Pearson's correlation coefficient R varies from about 0.42 to about 0.84 for different tunnels and protein models. The p-values for these correlations, calculated using the one-tailed Student's t-test, are given in the panels of Fig. 4. Except the one case shown in Fig. 4(b) for the *H. marismortui*'s tunnel, which shows a weak correlation between t_{esc} and N_h ($R = 0.418$ and $p = 0.088$), all other correlations have medium to high R-values and are statistically significant ($p < 0.05$). The correlation of t_{esc} with Q is higher than the correlation with N_h for all the tunnels except for *E. coli*. Thus, the effect of electrostatic interaction on the protein escape tends to be stronger than that of the hydrophobic interaction, though this still depends on the tunnel.

It is expected that the C-terminal segment of a protein is most relevant to its escape process.[15] We have checked that by calculating N_h and Q only for the C-terminal 50 residues, as shown in Fig. S9, the correlation of t_{esc} with N_h becomes very poor or almost disappears for all the tunnels, while the correlation with Q remains statistically

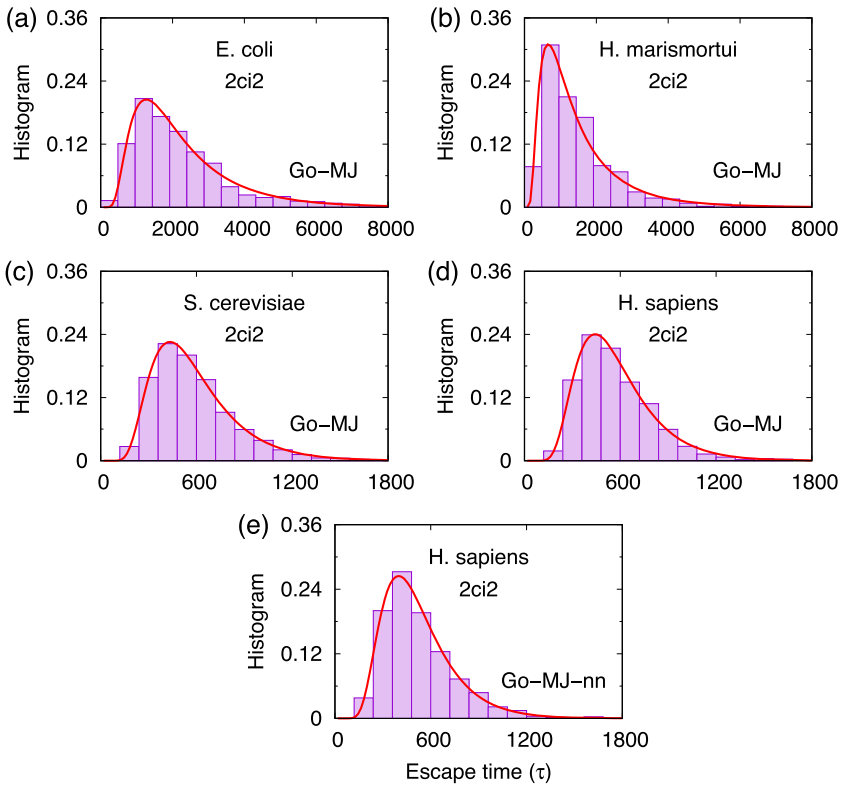

FIG. 3. Distributions of the escape time for the 2ci2 protein in the Gō-MJ model at the ribosomal tunnels of *E. coli* (a), *H. marismortui* (b), *S. cerevisiae* (c), and *H. sapiens* (d), and in the Gō-MJ-nn model at the *H. sapiens* ribosomal tunnel (e). The normalized histograms obtained by simulation (boxes) are fitted to the diffusion model (solid line) by using Eq. (5). The simulations were carried out at the temperature $T = 300$ K and with the growth time per residue $t_g = 400\tau$.

TABLE II. Diffusional properties of the protein escape process at ribosomal exit tunnels. The proteins are identified by their PDB ID (first column) and the chain length, N. For each protein, the properties given are the diffusion constant D and the pulling force k of the diffusion model, whose values, in units of $Å^2\ \tau^{-1}$ and pN, respectively, are obtained by fitting the histograms of escape times from simulations to the diffusion model (see text). The names of the organisms and the model for nascent proteins considered are given on top of the D and k columns.

		E. coli		*H. marismortui*		*S. cerevisiae*		*H. sapiens*		*H. sapiens*	
		Gō-MJ		Gō-MJ		Gō-MJ		Gō-MJ		Gō-MJ-nn	
Protein	N	D ($Å^2\ \tau^{-1}$)	k (pN)	D ($Å^2\ \tau^{-1}$)	k (pN)	D ($Å^2\ \tau^{-1}$)	k (pN)	D ($Å^2\tau^{-1}$)	k (pN)	D ($Å^2\tau^{-1}$)	k (pN)
1pga	56	1.174	12.010	0.817	14.288	0.779	21.121	1.103	12.880	1.180	12.507
1rop	56	1.093	5.135	1.030	1.822	0.771	14.495	0.891	9.028	0.911	9.484
1shg	57	0.510	6.543	0.497	4.059	0.703	11.182	0.761	6.171	0.704	8.241
2spz	58	0.504	12.466	0.405	11.927	0.580	16.731	0.935	5.674	0.947	6.875
1orc	64	0.797	5.881	0.338	1.905	0.754	9.484	0.877	4.597	1.008	4.763
2ci2	65	0.487	2.650	1.065	1.822	0.909	5.591	0.802	6.461	0.873	6.585
1msi	66	0.419	6.171	0.559	3.686	0.555	14.039	0.734	6.461	0.754	6.336
1csp	67	0.586	16.690	0.625	12.714	0.789	14.867	0.723	16.483	0.714	17.849
1ubq	76	0.459	10.146	0.536	5.342	0.570	14.122	0.948	6.129	1.108	5.591
1poh	85	0.712	8.532	0.571	9.235	0.962	10.271	1.148	4.597	0.954	6.543
1udv	88	0.379	2.112	0.516	0.396	0.578	5.011	0.641	3.437	0.672	4.555
1a2p	108	0.710	4.141	0.780	2.650	1.064	3.396	0.610	8.490	0.671	8.738

TABLE III. Median escape times of proteins at the ribosomal exit tunnels of different organisms. The proteins are listed with the number of hydrophobic residues, N_h, net charge, Q, and median escape times, t_{esc}, given in units of τ obtained by simulations. The names of the organism and the protein model used in the simulations are given on top of each t_{esc} column.

Protein	N_h	Q (e)	*E. coli* Gō-MJ t_{esc} (τ)	*H. marismortui* Gō-MJ t_{esc} (τ)	*S. cerevisiae* Gō-MJ t_{esc} (τ)	*H. sapiens* Gō-MJ t_{esc} (τ)	*H. sapiens* Gō-MJ-nn t_{esc} (τ)
1pga	12	−4	220.0	248.0	177.1	204.3	195.8
1rop	15	−4	495.6	1262.5	269.7	351.6	330.0
1shg	20	+1	819.0	1273.0	361.4	578.5	478.6
2spz	16	−2	463.6	595.9	295.7	523.2	436.5
1orc	18	+3	602.3	2587.9	391.0	649.5	564.5
2ci2	29	−1	1906.0	1175.0	525.5	523.1	476.7
1msi	31	0	1050.9	1261.7	365.2	572.4	566.5
1csp	22	−6	294.4	359.8	244.1	240.8	226.8
1ubq	26	0	599.0	941.4	356.5	477.1	437.9
1poh	26	−2	463.7	532.9	287.1	507.6	436.0
1udv	32	+3	2828.3	5843.2	930.0	1185.8	859.8
1a2p	29	+2	907.2	1202.0	727.3	536.1	485.1

significant and even slightly improves for some tunnels compared to the cases without the C-terminal cut-off. This result indicates that the impact of electrostatic interaction on the protein escape time is dominating over hydrophobic interaction for residues near the protein C-terminus. The impact of hydrophobic interaction on the protein escape seems to appear with a longer protein segment (the longest protein in our study is barnase with 108 residues), as indicated by the correlations in Fig. 4. However, in longer proteins, it is unlikely that residues too distant from the C-terminus can influence the escape time.

Following the previous work,[14] we tested the dependence of t_{esc} on a linear function of both N_h and Q. The form of the function chosen is $(1-s)N_h + sQ$ where $s \in (0,1)$ is a tunable parameter. We find that this function yields a better correlation with t_{esc} than both N_h and Q alone at some intermediate value of s (Fig. S6). Figure 5 plots the dependence of t_{esc} on the function $(1-s)N_h + sQ$ for the optimal value of s, i.e., the value that maximizes the correlation coefficient R, for all the tunnels and the protein models considered. The values of R in these plots range from 0.668 to 0.885, and all the p-values are <0.01. The best correlations are found for the tunnels of

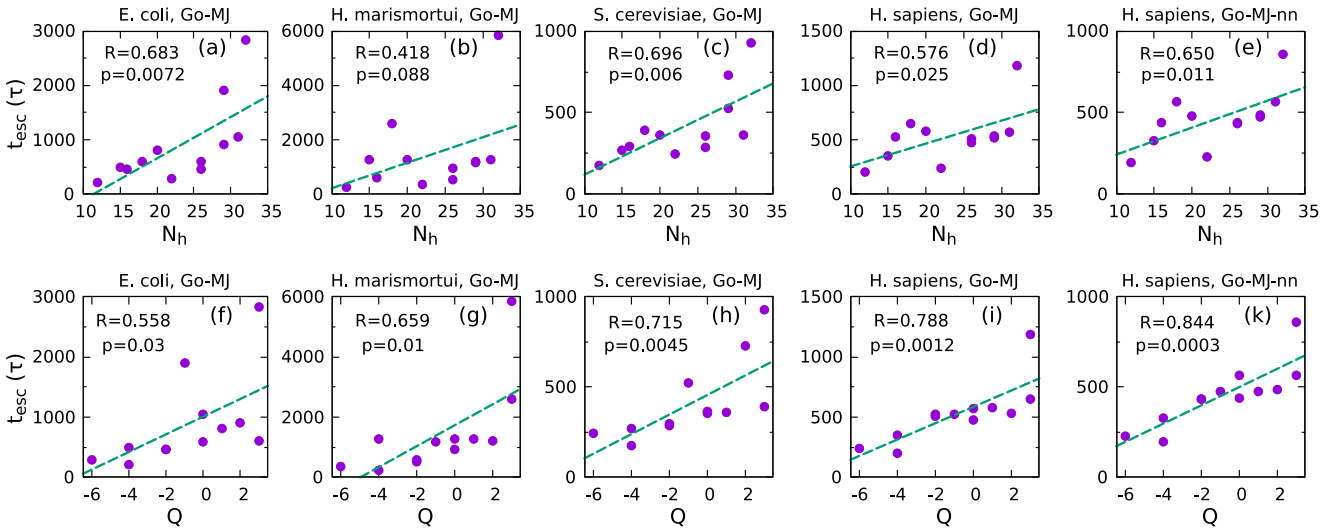

FIG. 4. Dependence of the median escape time, t_{esc}, on the number of hydrophobic residues, N_h, (a)–(e), and the total charge, Q, (f)–(k), of nascent proteins at the ribosomal exit tunnels of different species. The names of the species and the protein model are given on top of each panel. The dashed line represents a linear fit. The Pearson's correlation coefficient R and the corresponding p-value calculated using the one-tailed Student's t-test are given in each figure.

S. cerevisiae and *H. sapiens*, with R exceeding 0.8. For the *H. sapiens* tunnel, the Gō-MJ-nn model yields a better correlation than the Gō-MJ model.

The above results show that the effects of hydrophobic and electrostatic interactions on the protein escape time at different exit tunnels are qualitatively similar. Generally, increasing N_h and Q leads to an increased escape time, though the quantitative effects depend on the protein and the tunnel.

The common mechanism of these effects is that attraction between the protein and the tunnel slows down the protein escape while repulsion speeds it up.[14] The slowest escape is found for protein 1udv at *H. marismortui*'s tunnel with $t_{esc} \approx 5800\tau$ (Fig. S5) due to a strong electrostatic attraction between the protein and the tunnel. The 1udv protein has the highest net charge among the proteins ($Q = +3e$) (Table III), while the *H. marismortui*'s tunnel has the highest number of negatively charged amino acid residues on the tunnel's surface ($N_-^{(t)} = 19$) (Table I). We have checked that switching off the electrostatic interaction of 1udv drastically reduces its median escape time by ten times to about 600τ. An example for a strong effect of hydrophobic interaction is of the protein 2ci2 at the *E. coli*'s tunnel with $t_{esc} \approx 1900\tau$. The 2ci2 is among the proteins with the highest numbers of hydrophobic residues ($N_h = 29$) (Table III), and *E. coli* has the highest number of hydrophobic residues on the tunnel's surface ($N_h^{(t)} = 46$) (Table I).

D. Effects of non-native interactions on the escape process

The effects of non-native interactions on the protein escape time can be seen by comparing the results of the Gō-MJ and the Gō-MJ-nn models at the human ribosomal exit tunnel. It is shown that the two models yield qualitatively and quantitatively similar results in the escape time distribution (Fig. 3) as well as in the dependence of t_{esc} on N_h and Q (Figs. 4 and 5). A more careful examination shows that non-native interactions reduce escape time by 4%–28% depending on the protein (Table III). This reduction effect is consistent with a result for homopolymer models, which shows that self-attractive polymers escape faster than self-repulsive polymers for polymer lengths larger than about 60 residues (Fig. S7). Non-native interactions that are governed by attractive potentials energetically drive the protein escape because the chain can form more non-native contacts if it is found outside the tunnel. Even though the native contacts can be more competitive in energy compared to the non-native contacts, as assumed in the Gō-MJ-nn model, they are fewer in numbers and some of the native contacts cannot be formed once the protein has not been fully escaped. This effect can be seen for protein 2ci2 by looking at the distributions of native and non-native contacts and the radius of gyration of the protein conformations at the moment of complete translation at the human ribosomal tunnel (Fig. S8). These distributions show that, on average, the Gō-MJ-nn model yields more compact conformations with smaller numbers of native contacts and larger numbers of non-native contacts than the Gō-MJ model.

IV. DISCUSSION

It is suggested that the driving forces for the protein escape come from different sources, including (i) an enthalpic preference associated with the folding of a nascent protein near the tunnel,[11,53] (ii) an entropy gain of a chain emerging from the tunnel,[54] and (iii) the stochastic motion of a partially folded chain leading to a kind of diffusion process.[11,12] The mechanical forces of the first two types, in ribosome stalling and ribosome-bound systems, have been quantified experimentally[53,54] and also computationally,[54,55] giving values from several to about 12 pN. Apart from these sources, the force governing the protein escape may also come from interactions of

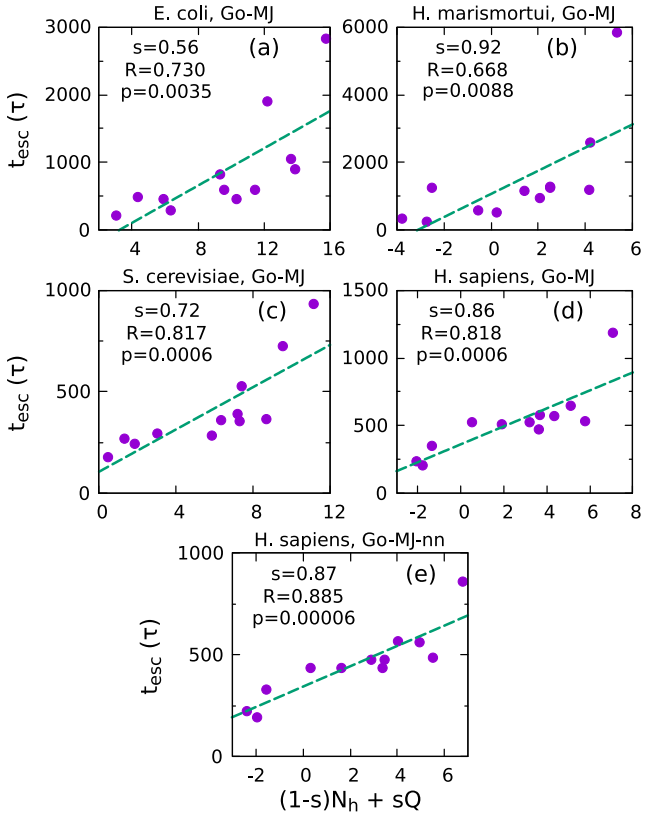

FIG. 5. Dependence of the median escape time, t_{esc}, on the function $(1-s)N_h + sQ$ of the number of hydrophobic residues, N_h, and the total charge, Q, of nascent proteins at the ribosomal exit tunnels of *E. coli* (a), *H. marismortui* (b), *S. cerevisiae* (c), and *H. sapiens* (d) and (e). The proteins are considered in the Go-MJ model (a)–(d) and in the Go-MJ-nn model (e). The names of the species and the protein model are given on top of each panel. The dependence is shown for the value of s that maximizes the correlation coefficient R. The values of s, R, and p are given in each panel. The dashed line shows a linear fit of the data.

nascent chains with the ribosomal tunnel. Electrostatic repulsion and attraction, as well as hydrophobic attraction, can speed up or slow down the escape process.[14,15]

The mean force from all the above sources acting on a nascent protein at the ribosome may be well represented by the force k in the diffusion model, which can be obtained by non-equilibrium methods via the escape time distribution. Interestingly, the magnitude of k found in the present study for different proteins and ribosomes (Table II) is similar to the mechanical forces reported in other studies.[53–55] For a tunnel that has energetic interactions with nascent proteins, the force acting on a protein by the tunnel may contribute to or counter balance the other forces depending on whether it is attractive or repulsive. If the attraction to the tunnel is sufficiently strong, the protein may have very long escape times and may not follow the diffusion model. The slowest escaping protein in our consideration is 1udv in *H. marismortui*'s model, having a very small force $k = 0.396$ pN. The escape time distribution of this protein has a long tail but still follows the diffusion model (Fig. S5). Note that the diffusion model predicts that for $k = 0$, both the mean escape time and the dispersion diverge (see Methods). It is possible that some proteins can have $k \leq 0$, resulting in infinite escape times.

In a recent study,[15] Nissley *et al.* have reported very long ejection times of some proteins at the *E. coli*'s tunnel, including the ones with PDB IDs 2jo6, 1u0b, and 4dcm, using similar coarse-grained simulations. We have checked that the fractions of hydrophobic amino acids as well as the fractions of positively and negatively charged amino acids in these three proteins are within the ranges given by the 12 proteins considered. Furthermore, 2jo6 and 4u0b have negative net charges for the whole chain (for the C-terminal 50 residues, the net charge is zero for 2jo6 and $-1e$ for 1u0b), suggesting that they are not slow escapers due to overall electrostatic repulsion with the tunnel. Indeed, our simulations of 2jo6 and 1u0b at the *E. coli*'s tunnel, using an averaged value of ϵ from the 12 proteins studied, show that they escape efficiently with the median escape time $t_{esc} \approx 666\tau$ for 2jo6 and $t_{esc} \approx 1270\tau$ for 1u0b (Fig. S10), i.e., within the same range of escape times as other proteins. It is also found that their escape time distributions are consistent with the conserved mechanism given by the diffusion model (Fig. S10). Including these two proteins in the initial set of 12 proteins, however, slightly deteriorates the correlations of t_{esc} with N_h and Q (Fig. S11). It would be interesting to check what specific detail causes the extreme delay of the escape process in Nissley *et al.*'s approach. We did not simulate 4dcm because the PDB structure of this protein is not contiguous containing missing residues. This protein, however, is expected to escape very slowly because it has a positive net charge of $Q = +8e$ in the C-terminal 50 residues, compared to $+5e$ in 1udv, the slowest escaping protein in the 12 proteins considered.

The previous work[14] has estimated the timescale of the protein escape times to be of the order of 0.1–1 ms by rescaling the simulation times to the values at realistic friction and simultaneously using the high-friction value of the time unit, $\tau_H = 3$ ns.[32,33] According to this estimation, the longest median escape times of the proteins in the present study are of the order of 10 ms, which is still shorter than the times needed by the ribosome to translate one codon. Given that the 12 proteins considered can be representative for most proteins in the proteomes in terms of hydrophobic and charge fractions, as discussed in Sec. III A, this result suggests that typical proteins escape efficiently at the ribosome tunnel and do not delay the ribosome's new translation cycle.

V. CONCLUSION

The ribosomal exit tunnel has many structural and chemical elements that could affect the post-translational escape of nascent proteins. These elements include the irregular shape of the tunnel, the exposed hydrophobic side-chains of the ribosomal proteins, and the charged amino acids on the tunnel's surface. The exit tunnels from different organisms from different domains of life, as considered in our study, show significant differences in structural and physico-chemical properties, despite certain similarities. The present study shows that despite all these differences in the exit tunnels, the protein escape process has conserved mechanisms across the domains of life. First, it is shown that the escape process follows the simple diffusion mechanism described by the diffusion model. This property holds true for 12 proteins with distinct native structures and diverse physico-chemical properties within two different

protein models that are given with and without non-native interactions. Second, the median escape time, t_{esc}, positively correlates with both the number of hydrophobic residues, N_h, and the net charge, Q, of proteins with a sufficient statistical significance in most cases. This property underlines the simple mechanism that attraction between the protein and the tunnel slows down the protein escape while repulsion speeds it up. The effects of hydrophobic and electrostatic interactions on the escape time are additive to each other, as indicated by improved correlations when considering the dependence of t_{esc} on a linear function of N_h and Q. These results reinforce our understanding of the protein escape process as one that is simple and predictable. The results also suggest that the impact of electrostatic interaction on the escape time is stronger than that of hydrophobic interaction and becomes dominant toward the C-terminal residues of nascent proteins. It is expected that proteins with a high positive net charge in the C-terminal segment may have unusually long escape times.

Our study also shows significant variations among the organisms when considering the quantitative effects of the exit tunnels on the escape process. The exit tunnels of *E. coli* and *H. marismortui* generally yield longer protein escape times than those of *S. cerevisiae* and *H. sapiens*. These observations are related to the facts that *E. coli* has a ~30% higher number of hydrophobic residues exposed inside the exit tunnel than the other organisms, and *H. marismortui* has a number of negatively charged amino acids on the tunnel's surface that is substantially larger than the other organisms (19 vs 1–4). From an evolutionary perspective, it could be that the exit tunnels of *S. cerevisiae* and *H. sapiens* have evolved to deal with a larger number of proteins in their genomes, suppressing the escape time. The argument for this hypothesis is that a too-slow escape of a nascent protein from the exit tunnel could hamper the ribosome's productivity; thus, it is beneficial to have an exit tunnel that allows all proteins in the genome to escape efficiently.

SUPPLEMENTARY MATERIAL

See the supplementary material for the list of 12 proteins considered with selected properties, for the structural class compositions of Richardson's Top2018 dataset and our protein set, for the distributions of protein length in various proteomes, for the histograms of hydrophobicity and charges in protein sequences of various proteomes, for the dependences of the diffusion constant D and the force k from the diffusion model on the chain length, N, of proteins, for the escape properties of protein 1udv, for the dependence of the correlation coefficient R between the median escape time and the function $(1-s)N_h + sQ$ on the parameter s for the tunnels and protein models considered, for the escape properties of self-repulsive and self-attractive homopolymers, for an analysis of the effects of non-native interactions on the escaping conformations, for the correlations of t_{esc} with N_h and Q calculated for the C-terminal 50 residues, for the histograms of escape times and the escape probabilities of 2jo6 and 1u0b proteins, and for the correlations of t_{esc} with N_h and Q by adding 2jo6 and 1u0b to the initial protein set.

ACKNOWLEDGMENTS

This research is funded by the Graduate University of Science and Technology under Grant No. GUST.STS.DT2020-VL03 for the postdoctoral fellowship of Phuong Thuy Bui. We also acknowledge the support of the Institute of Physics and the Center for Informatics and Computing at VAST for the use of their computer clusters.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Phuong Thuy Bui: Data curation (lead); Formal analysis (equal); Funding acquisition (lead); Investigation (lead); Visualization (lead); Writing – original draft (lead). **Trinh Xuan Hoang**: Conceptualization (lead); Formal analysis (equal); Funding acquisition (supporting); Investigation (supporting); Methodology (lead); Project administration (lead); Resources (equal); Software (lead); Supervision (lead); Validation (equal); Visualization (supporting); Writing – review & editing (lead).

DATA AVAILABILITY

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

REFERENCES

[1] H. Nakatogawa and K. Ito, Cell **108**, 629 (2002).
[2] K. Ito and S. Chiba, *Regulatory Nascent Polypeptides* (Springer, 2014).
[3] D. N. Wilson, S. Arenz, and R. Beckmann, Curr. Opin. Struct. Biol. **37**, 123 (2016).
[4] L. D. Cabrita, C. M. Dobson, and J. Christodoulou, Curr. Opin. Struct. Biol. **20**, 33 (2010).
[5] D. V. Fedyukina and S. Cavagnero, Annu. Rev. Biophys. **40**, 337 (2011).
[6] M. Thommen, W. Holtkamp, and M. V. Rodnina, Curr. Opin. Struct. Biol. **42**, 83 (2017).
[7] N. R. Voss, M. Gerstein, T. A. Steitz, and P. B. Moore, J. Mol. Biol. **360**, 893 (2006).
[8] A. Kosolapov and C. Deutsch, Nat. Struct. Mol. Biol. **16**, 405 (2009).
[9] K. Dao Duc, S. S. Batra, N. Bhattacharya, J. H. Cate, and Y. S. Song, Nucleic Acids Res. **47**, 4198 (2019).
[10] K. Kannan and A. S. Mankin, Ann. N. Y. Acad. Sci. **1241**, 33 (2011).
[11] P. T. Bui and T. X. Hoang, J. Chem. Phys. **144**, 095102 (2016).
[12] P. T. Bui and T. X. Hoang, J. Chem. Phys. **149**, 045102 (2018).
[13] P. T. Bui and T. X. Hoang, J. Chem. Phys. **153**, 045105 (2020).
[14] P. T. Bui and T. X. Hoang, Biophys. J. **120**, 4798–4808 (2021).
[15] D. A. Nissley, Q. V. Vu, F. Trovato, N. Ahmed, Y. Jiang, M. S. Li, and E. P. O'Brien, J. Am. Chem. Soc. **142**, 6103 (2020).
[16] D. Baker, Nature **405**, 39 (2000).
[17] R. D. Hills Jr. and C. L. Brooks III, Int. J. Mol. Sci. **10**, 889 (2009).
[18] S. Takada, Biophys. Physicobiol. **16**, 248 (2019).
[19] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pep. Protein Res. **7**, 445 (1975).
[20] S. Miyazawa and R. L. Jernigan, J. Mol. Biol. **256**, 623 (1996).
[21] J. Karanicolas and C. L. Brooks III, J. Mol. Biol. **334**, 309 (2003).
[22] R. B. Best, Y.-G. Chen, and G. Hummer, Structure **13**, 1755 (2005).
[23] T. Tanaka, N. Hori, and S. Takada, PLoS Comput. Biol. **11**, e1004356 (2015).
[24] Q. V. Vu, Y. Jiang, M. S. Li, and E. P. O'Brien, Chem. Sci. **12**, 11851 (2021).
[25] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[26] M. Cieplak and T. X. Hoang, Int. J. Mod. Phys. C **13**, 1231 (2002).

[27]J. Tsai, R. Taylor, C. Chothia, and M. Gerstein, J. Mol. Biol. **290**, 253 (1999).
[28]Z. L. Watson, F. R. Ward, R. Meheust, O. Ad, A. Schepartz, J. F. Banfield, and J. H. Cate, eLife **9**, e60482 (2020).
[29]T. M. Schmeing, K. S. Huang, S. A. Strobel, and T. A. Steitz, Nature **438**, 520 (2005).
[30]C. Schmidt, T. Becker, A. Heuer, K. Braunger, V. Shanmuganathan, M. Pech, O. Berninghausen, D. N. Wilson, and R. Beckmann, Nucleic Acids Res. **44**, 1944–1951 (2016).
[31]H. Khatter, A. G. Myasnikov, S. K. Natchiar, and B. P. Klaholz, Nature **520**, 640 (2015).
[32]T. Veitshans, D. Klimov, and D. Thirumalai, Folding Des. **2**, 1 (1997).
[33]D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997).
[34]Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[35]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[36]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[37]P. Alexander, J. Orban, and P. Bryan, Biochemistry **31**, 7243 (1992).
[38]P. F. Predki, V. Agrawal, A. T. Brünger, and L. Regan, Nat. Struct. Biol. **3**, 54 (1996).
[39]Y.-J. Chen, S.-C. Lin, S.-R. Tzeng, H. V. Patel, P.-C. Lyu, and J.-W. Cheng, Proteins: Struct., Funct., Bioinf. **26**, 465 (1996).
[40]A. Myrhammar, D. Rosik, and A. E. Karlström, Bioconjugate Chem. **31**, 622 (2020).
[41]G. I. Gitelson, Y. V. Griko, A. V. Kurochkin, V. V. Rogov, V. P. Kutyshenko, M. P. Kirpichnikov, and P. L. Privalov, FEBS Lett. **289**, 201 (1991).
[42]S. E. Jackson and A. R. Fersht, Biochemistry **30**, 10428 (1991).
[43]O. García-Arribas, R. Mateo, M. M. Tomczak, P. L. Davies, and M. G. Mateu, Protein Sci. **16**, 227–238 (2007).
[44]H. Welte, T. Zhou, X. Mihajlenko, O. Mayans, and M. Kovermann, Sci. Rep. **10**, 2640 (2020).
[45]D. Morimoto, E. Walinda, H. Fukada, Y.-S. Sou, S. Kageyama, M. Hoshino, T. Fujii, H. Tsuchiya, Y. Saeki, K. Arita *et al.*, Nat. Commun. **6**, 6116 (2015).
[46]R. Thapar, E. M. Nicholson, P. Rajagopal, E. B. Waygood, J. M. Scholtz, and R. E. Klevit, Biochemistry **35**, 11268–11277 (1996).
[47]K. Biyani, M. A. Kahsai, A. T. Clark, T. L. Armstrong, S. P. Edmondson, and J. W. Shriver, Biochemistry **44**, 14217 (2005).
[48]J. W. Bye, N. J. Baxter, A. M. Hounslow, R. J. Falconer, and M. P. Williamson, ACS Omega **1**, 669 (2016).
[49]C. J. Williams, D. C. Richardson, and J. S. Richardson, Protein Sci. **31**, 290 (2022).
[50]The UniProt Consortium, Nucleic Acids Res. **49**, D480 (2021).
[51]D. Brune and S. Kim, Proc. Natl. Acad. Sci. U. S. A. **90**, 3835 (1993).
[52]C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, and C. Bustamante, Science **334**, 1723 (2011).
[53]D. H. Goldman, C. M. Kaiser, A. Milin, M. Righini, I. Tinoco, Jr., and C. Bustamante, Science **348**, 457 (2015).
[54]B. Fritch, A. Kosolapov, P. Hudson, D. A. Nissley, H. L. Woodcock, C. Deutsch, and E. P. O'Brien, J. Am. Chem. Soc. **140**, 5077 (2018).
[55]S. E. Leininger, F. Trovato, D. A. Nissley, and E. P. O'Brien, Proc. Natl. Acad. Sci. U. S. A. **116**, 5523 (2019).