Role of cotranslational folding for β-sheet-enriched proteins: A perspective from molecular dynamics simulations

Peng Tao ◉ and Yi Xiao ◉*

School of Physics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

The formations of correct three-dimensional structures of proteins are essential to their functions. Cotranslational folding is vital for proteins to form correct structures *in vivo*. Although some experiments have shown that cotranslational folding can improve the efficiency of folding, its microscopic mechanism is not yet clear. Previously, we built a model of the ribosomal exit tunnel and investigated the cotranslational folding of a three-helix protein by using all-atom molecular dynamics simulations. Here we study the cotranslational folding of three β-sheet-enriched proteins using the same method. The results show that cotranslational folding can enhance the helical population in most cases and reduce non-native long-range contacts before emerging from the ribosomal exit tunnel. After exiting the tunnel, all proteins fall into local minimal states and the structural ensembles of cotranslational folding show more helical conformations than those of free folding. In particular, for one of the three proteins, the GTT WW domain, we find that one local minimum state of the cotranslational folding is the known folding intermediate, which is not found in free folding. This result suggests that the cotranslational folding may increase the folding efficiency by accelerating the sampling more than by avoiding the misfolded state, which is presently a mainstream viewpoint.

I. INTRODUCTION

Cotranslational folding (CTF) is the first step of protein folding *in vivo*, which refers to the process of translation and folding of the nascent peptide chain along a channel in the large subunit of the ribosome, that is, the ribosomal exit tunnel, from the N terminus to the C terminus. This kind of folding is different from the widely studied *in vitro* free folding (FF) due to the participation of the ribosome and the order of folding [1–4]. It is generally believed that the efficiency of *in vivo* folding is higher than that of *in vitro* folding [5–11], but how CTF as the initial step of *in vivo* folding improves the folding efficiency remains an open question. This question can be divided into two parts: (1) What are the structural features of the nascent peptide chain when it comes out of the ribosomal exit tunnel? (2) Are these structural features helpful for subsequent folding?

For the first question, since the length of the ribosomal exit tunnel is about 100 Å and the diameter is 10–20 Å [12,13], the nascent peptide chains move along a narrow and long pipeline during the translation process and their conformational changes must be affected by various factors, including spatial constraints, electrostatic interaction, and many more [14–17]. Both theoretical and simulation studies have shown that the ribosomal exit tunnel can promote the formation of helical structure [18–20], which is also consistent with the existing experimental data [21,22]; that is, the nascent peptide chain can only form helical conformation inside the ribosomal tunnel, and more complex structures can only be found in the vestibule of the tunnel. For example, the position of the formation of a β-sheet structure is about 25 amino acids from the peptidyl transferase center (PTC) of the ribosome [23]. Recently, several research groups have also found that larger structures can be formed in the vestibule, such as the entire zinc finger protein [24], spectrin [25], and the partial transmembrane segment of the potassium channel protein [26]. Since helical conformations prefer to form inside the ribosomal exit tunnel, does it mean that the ribosomes only affect the folding of helix-type proteins?

For the second question, since the nascent peptide chains are usually not completely folded when they come out of the ribosomal exit tunnel, they may form specific structures that do not exist in *in vitro* folding. For example, Holtkamp *et al.* found that the five-helix protein HemK was in a collapsed conformation just after exiting the ribosome, and then folded to a nearly native state based on this conformation [27]. It should be noted that this collapsed conformation, or intermediate state, does not exist in the folding process *in vitro* [27]. The same is true for the T4 lysozyme [28] and the HaloTag protein [29]. These results suggest that certain structures formed by CTF can guide the subsequent folding of the proteins. However, this may not hold for all proteins. For example, the folding pathways of Ig and SH3 in the ribosomal exit tunnel are not significantly different from those *in vitro* [30,31]. Does CTF only guide the sampling of specific proteins?

In order to answer the two questions above, we have studied the CTF processes of three proteins by using all-atom molecular dynamics simulations, two of which are all β (GTT and SH3, PDB (Protein Data Bank) ID: 2F21 [32,33] and 1YN8, respectively) and one is mixed α/β (CI2, PDB ID:
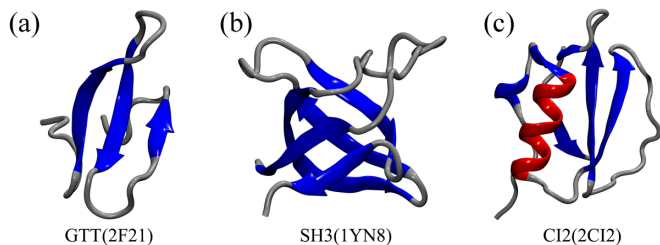
*yxiao@hust.edu.cn

FIG. 1. Cartoon representation of three proteins studied in this work. These figures were created by the VMD program [36].

2CI2 [34]). Their native structures are shown in Fig. 1. Considering that the entire simulation system will have more than 1 000 000 atoms after adding the water molecules into the large subunit of the ribosome, it is difficult to achieve all-atom simulations on the order of microseconds with current computational power, and so we still use our previous strategy to simplify the system; that is, we use a simple ribosomal exit tunnel model to replace the real ribosomal tunnel [35]. The construction and usage of this model will be described in detail in the Methods section.

II. METHODS

A. Construction and parameters of our simulation systems

All simulations in this paper were implemented using the AMBER package [37], and the protein force field used was FF14SB [38]. The initial structures of the simulations were created by the TLEAP program [39] and their dihedral angles of the main chains were set to 180°, resulting in fully extended peptides, where the N termini of the initial peptides were aligned to the positive direction of the x axis. Before starting energy optimization, each protein was wrapped in a rectangular TIP3P water box, and the minimum distances between GTT, SH3, and CI2 and the edges of the box were 12, 10, and 10 Å, respectively. Then three Cl^-, seven Na^+, and one Na^+ were added for GTT, SH3 and CI2, respectively, to neutralize the whole system. Next, two-stage energy optimization was carried out. In the first stage, all atoms of the peptide were constrained by a harmonic potential of 10 kcal/mol Å² (set NTR to 1 in PMEMD.CUDA), and only the energy of the solvent was optimized. In the second stage, the energy of the entire system was optimized. Similar to the previous study [40], each stage of optimization included 2000 steepest descent optimization steps plus 2000 conjugate gradient optimization steps. After the energy optimization was completed, the system was heated from 0 to 300 K in 100 ps, and then maintained at 300 K for another 100 ps. After that, a 200 ps equilibrium simulation was performed under the *NPT* ensemble. The time step of heating and equilibrium was 1 fs and the cutoff of nonbonded interaction was 10 Å.

Using the final structure obtained from the previous equilibrium process as the initial structure, three different types of production simulations were performed for each protein, namely, free folding simulation, and cotranslational folding simulation at the speed of residue/2 ns and residue/10 ns (denoted as FF, CTF-2, and CTF-10, respectively). For each type of simulation, eight to ten independent trajectories were run
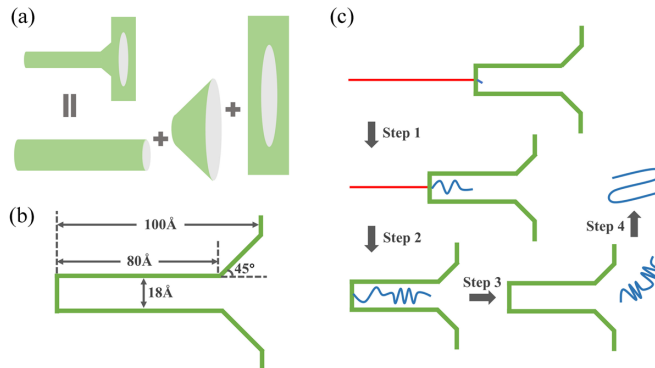

FIG. 2. (a) Schematic diagram of the model of the ribosomal exit tunnel. The green lines represent the rigid wall of the tunnel model. (b) Four geometric parameters of the tunnel model. (c) Flow chart of cotranslational folding simulation; residues with and without constraints are shown as red and blue lines, respectively.

by using different random number seeds under the *NPT* ensemble, generating >1.1 ms trajectories in total (summarized in Table S1 in the Supplemental Material [41]; all trajectories are shown in Supplemental Figs. S1–S24). The temperature was controlled by Langevin dynamics [42] at 300 K and the pressure was controlled by isotropic position scaling [43] at 1 bar. The time step and nonbonded interaction cutoff were set to 2 fs and 9 Å, respectively. The Particle Mesh Ewald (PME) method was applied to calculate the electrostatic interactions. The simulated structures were saved every 10 000 steps and all simulations were performed on the GPU nodes of our computational cluster. A more detailed description of the setting parameters can be found in previous studies [35,40,44].

B. Construction and usage of the ribosomal exit tunnel model

In order to simulate the CTF of nascent peptide chains, a rigid model of a ribosomal exit tunnel was constructed, which consisted of three parts, including a cylinder, a circular truncated cone, and an infinite plane [Fig. 2(a)] [35]. The geometric parameters of this model are shown in Fig. 2(b). It should be noted that the tunnel only has an effect on protein molecules, but has no effect on water molecules and ions. When an atom of the peptide reaches the tunnel wall, a completely elastic collision will occur; that is, the direction of the velocity of the atom is reversed and the magnitude is unchanged.

The simulation of CTF involves four steps. The first step is to place the tunnel along the x axis so that the peptide chain is coaxial with the tunnel [as shown in Fig. 2(c)]; then the bottom of the tunnel is moved to the minimum coordinate of the x axis of the first residue, namely, the minimum value of the x coordinate of all atoms of the residue. The second to last residue of the peptide is constrained by a harmonic potential of 10 kcal/mol Å² and only the first residue is allowed to move. In the second step, after a certain time, i.e., 2 or 10 ns, the bottom of the tunnel is moved to the minimum coordinate of the x axis of the next residue, and at the same time the constraint exerted on the next residue is removed, and so on, until the bottom of the tunnel moves to the minimum coordinate of the last residue on the x axis.

TABLE I. The helical tendency [48] and hydrophilic index [49] for all amino acids. The greater the helix tendency, the higher the probability of forming a helix, while the smaller the hydrophilic index, the more hydrophobic it is.

Amino acid	Helical tendency	Hydrophilic index
Ala (A)	1.42	1.8
Cys (C)	0.7	2.5
Asp (D)	1.01	−3.5
Glu (E)	1.51	−3.5
Phe (F)	1.13	2.8
Gly (G)	0.57	−0.4
His (H)	1.0	−3.2
Ile (I)	1.08	4.5
Lys (K)	1.14	−3.9
Leu (L)	1.21	3.8
Met (M)	1.45	1.9
Asn (N)	0.67	−3.5
Pro (P)	0.57	−1.6
Gln (Q)	1.11	−3.5
Arg (R)	0.98	−4.5
Ser (S)	0.77	−0.8
Thr(T)	0.83	−0.7
Val(V)	1.06	4.2
Trp (W)	1.08	−0.9
Tyr (Y)	0.69	−1.3

The third step is to fix the position of the tunnel so that the entire nascent peptide chain can move inside the tunnel. Since the left side of the tunnel is closed, the nascent peptide chain will come out spontaneously from the right side of the tunnel after a period of time. In the fourth step, when the nascent peptide chain comes out of the tunnel, that is, all atoms of the protein molecule are on the right side of the tunnel, the tunnel model will be removed and the nascent peptide is allowed to fold freely during the rest of the simulation. The code of the tunnel model is written in the FORTRAN language and embedded in the source code of PMEMD.CUDA [45,46].

In order to analyze the role of each component of the exit tunnel model, we also conducted a controlled study on it, that is, corelease folding (CRF) for GTT. Similar to CTF simulations, in corresponding CRF simulations, the residues were released one by one, but the original exit tunnel model was replaced by a flat wall (the last part of the exit tunnel model in Fig. 2).

C. Analysis of simulated trajectories

The calculation of the secondary structure uses the DSSP algorithm [47], which divides the secondary structure into seven classes. In order to simplify the analysis, we further transform these seven classes into three classes, among which Helix includes α helix, 3-10 helix, and π helix, Sheet includes parallel β sheet and antiparallel β sheet, and Turn includes Turn and Coil. The population value of a residue is the probability that the residue is in a helix or sheet in a trajectory. It should be noted that before the peptide chain exits from the pipeline, the constrained times for the residues are different, so when comparing with FF, the population
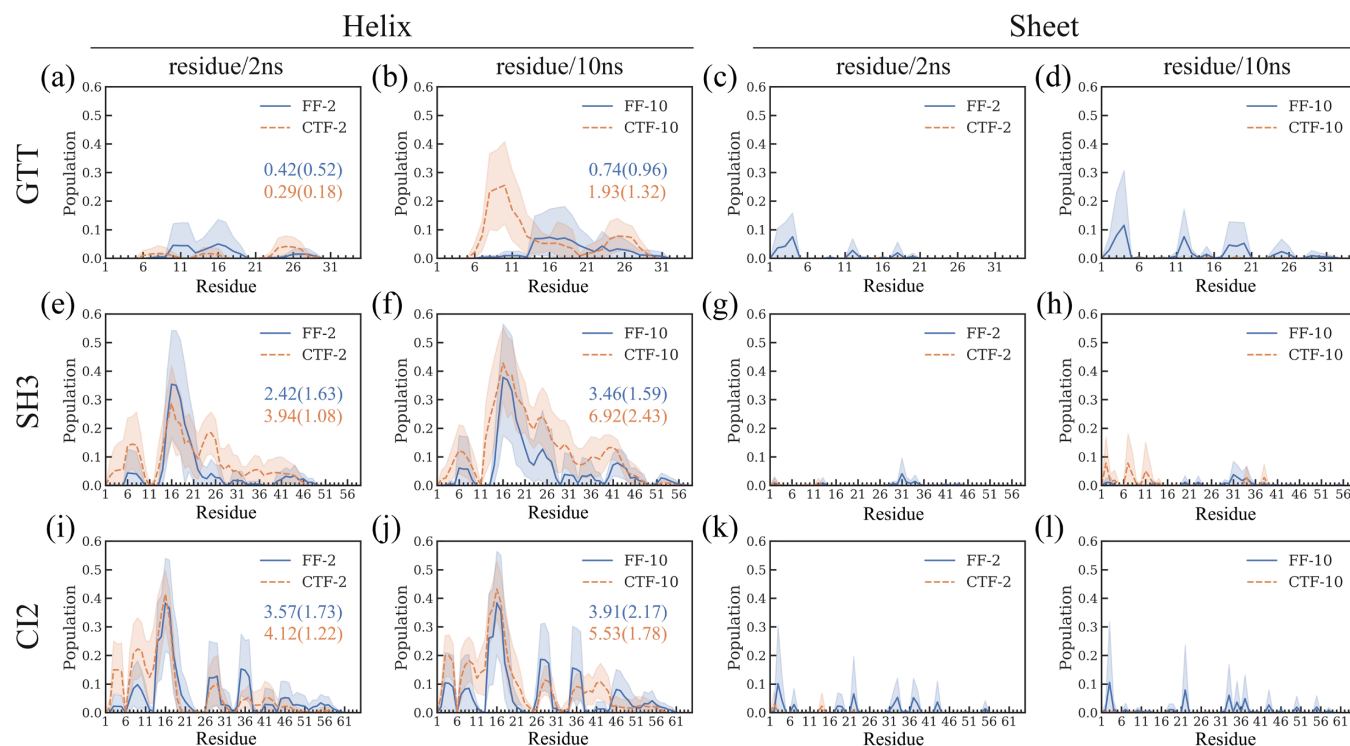

FIG. 3. The population where each residue is in a helical or β-sheet structure in cotranslational folding (orange dashed line) and free folding (blue solid line) for (a)–(d) GTT, (e)–(h) SH3, and (i)–(l) CI2 before exiting from the ribosomal tunnel. The corresponding shadows represent 95% confidence intervals of eight to ten independent trajectories. The numbers outside and inside the brackets represent the average and standard deviation of the cumulative probability of all residues, respectively. The coloring scheme for numbers is the same as for lines.
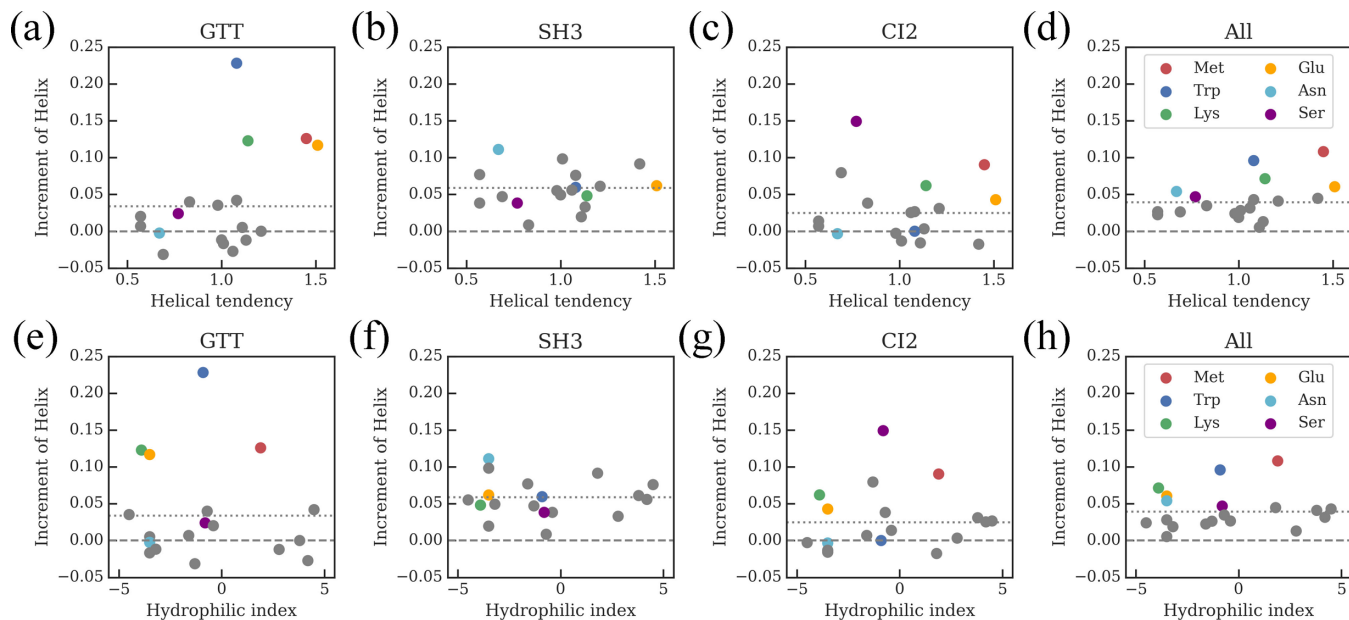
FIG. 4. At a translation speed of residue/10 ns, the relationship between the helix tendency (a)–(d) or the hydropathic index (e)–(h) of amino acids and the increment of the helical population. The increment is calculated by subtracting the helical population in cotranslational folding and free folding in Fig. 3. Each dot represents one kind of amino acid, and the dotted line represents the average value of the helical increment. The amino acids with the top six highest increments, namely, Met, Glu, Trp, Asn, Lys, and Ser are marked by red, orange, blue, cyan, green, and purple, respectively.

value of each residue is calculated by subtracting the part during the corresponding constrained time. Taking the CTF of GTT at the speed of residue/10 ns as an example, we take the trajectory of the first 350 ns (35 residues × 10 ns) to calculate the population of each residue. When comparing the population of the second residue, since the second residue is constrained for 10 ns in CTF, only the trajectory of the first 340 ns (350 ns–10 ns) trajectory in FF is used, and so on. The calculated result using this method in FF is marked as FF-10.
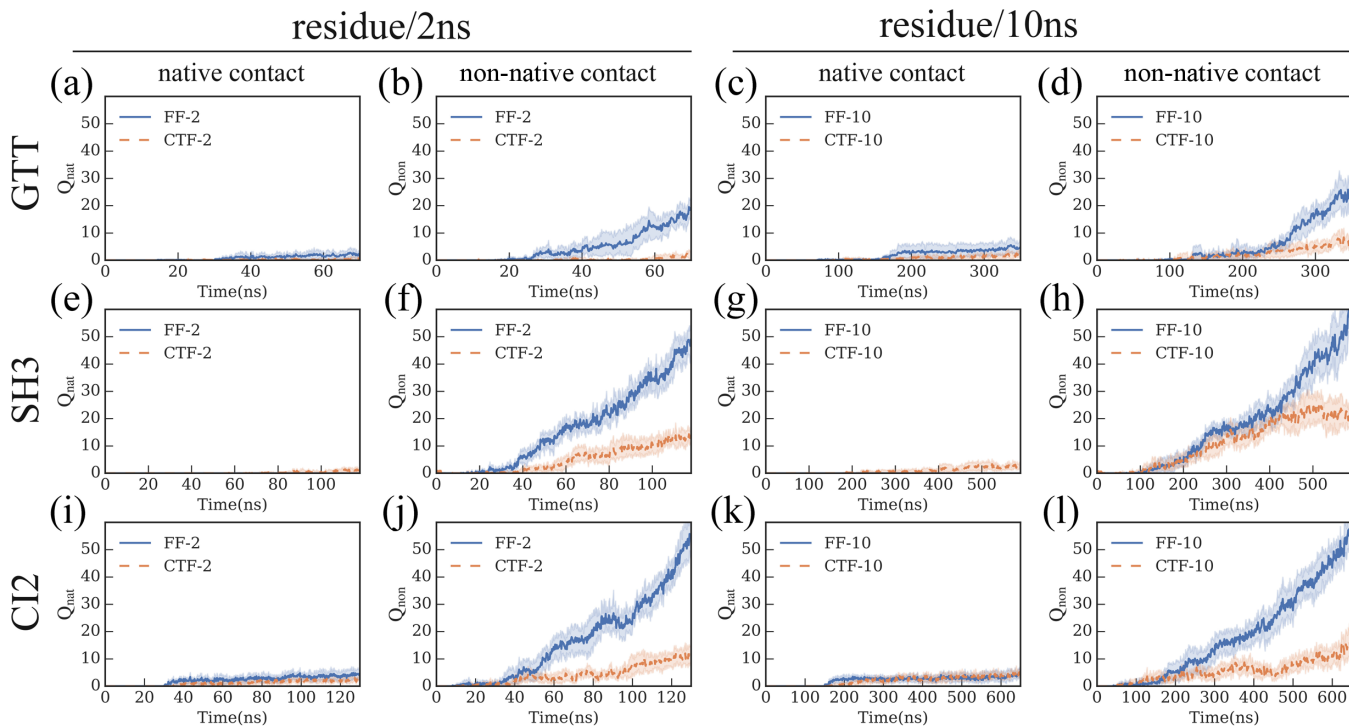

FIG. 5. The numbers of native contacts (Q_{nat}) and non-native contacts (Q_{non}) of (a)–(d) GTT, (e)–(h) SH3, and (i)–(l) CI2 change with time before exiting from the ribosomal exit tunnel. The meaning of the shadow is the same as in Fig. 3.
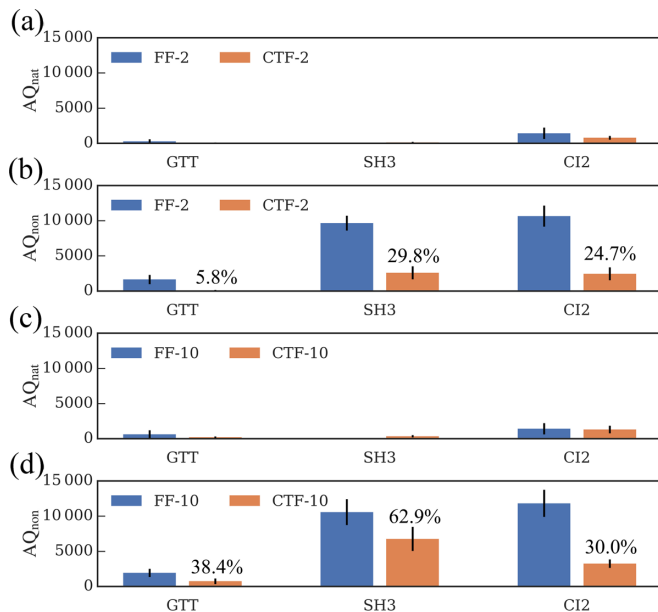
FIG. 6. The cumulative numbers of (a), (c) native contacts (AQ_{nat}) and (b), (d) non-native contacts (AQ_{non}) before exiting from the ribosomal exit tunnel. The labeled percentage represents the ratio of AQ_{non} in CTF to AQ_{non} in FF.

In this paper, residue-residue contact is defined based on the distance between C_α atoms between two residues. When two residues are separated by more than three residues in sequence, and the distance between their C_α atoms is less than 7 Å, the two residues are said to form a contact. Furthermore, if the formed contact also exists in the native structure, the contact is said to be a native contact; otherwise, it is called a non-native contact. Before exiting from the pipeline, the calculation of the number of contacts only considers the residues without constraints. Here we take the CTF of GTT at the speed of residue/10 ns as an example, when comparing the number of contacts, during the fifth 10 ns (40–50 ns), since only the first five residues are free to fold in CTF, thus only the first five residues will be considered in FF to calculate the number of contacts, and so on.

After the peptide chain exits from the ribosomal exit channel, we compare the structural ensembles under CTF and FF and analyze the secondary structures of the two ensembles, the number of contacts, the root-mean-square deviation (RMSD), and the radius of gyration (ROG). The helical tendency [48] and hydrophilic index [49] for all amino acids are listed in Table I. The calculations of RMSD and ROG exclude the loops at both ends (GTT: 7–29; SH3: 2–58; CI2: 5–64, only for C_α atoms). When clustering according to RMSD, each type of simulated trajectory is clustered into ten classes by using the hierarchical agglomerative approach (a bottom-up hierarchical clustering algorithm, denoted as HIERAGGLO in CPPTRAJ) [50]. The proportion of the first five classes and the RMSD values of the representative structures of these classes are summarized in Table S2 in the Supplemental Material [41]. All the above analyses do not consider the first-microsecond trajectory and are done with CPPTRAJ [51] and its parallel version CPPTRAJ.OMP [52].

III. RESULTS

A. The effect of cotranslational folding on the formation of helices depends on sequence and translation speed

First of all, we try to answer the first question: What conformation will the nascent peptide chain form in the process
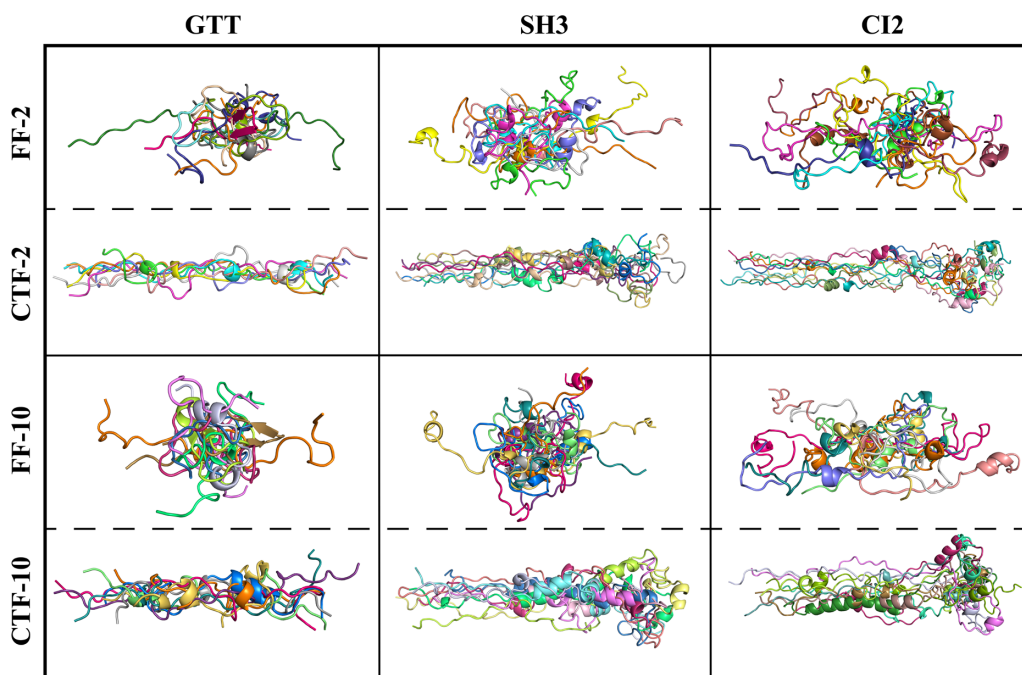

FIG. 7. The structural ensembles of the nascent peptides in the simulations at the specific time $N\Delta t$, where Δt is the translation speed and N is the residue number.
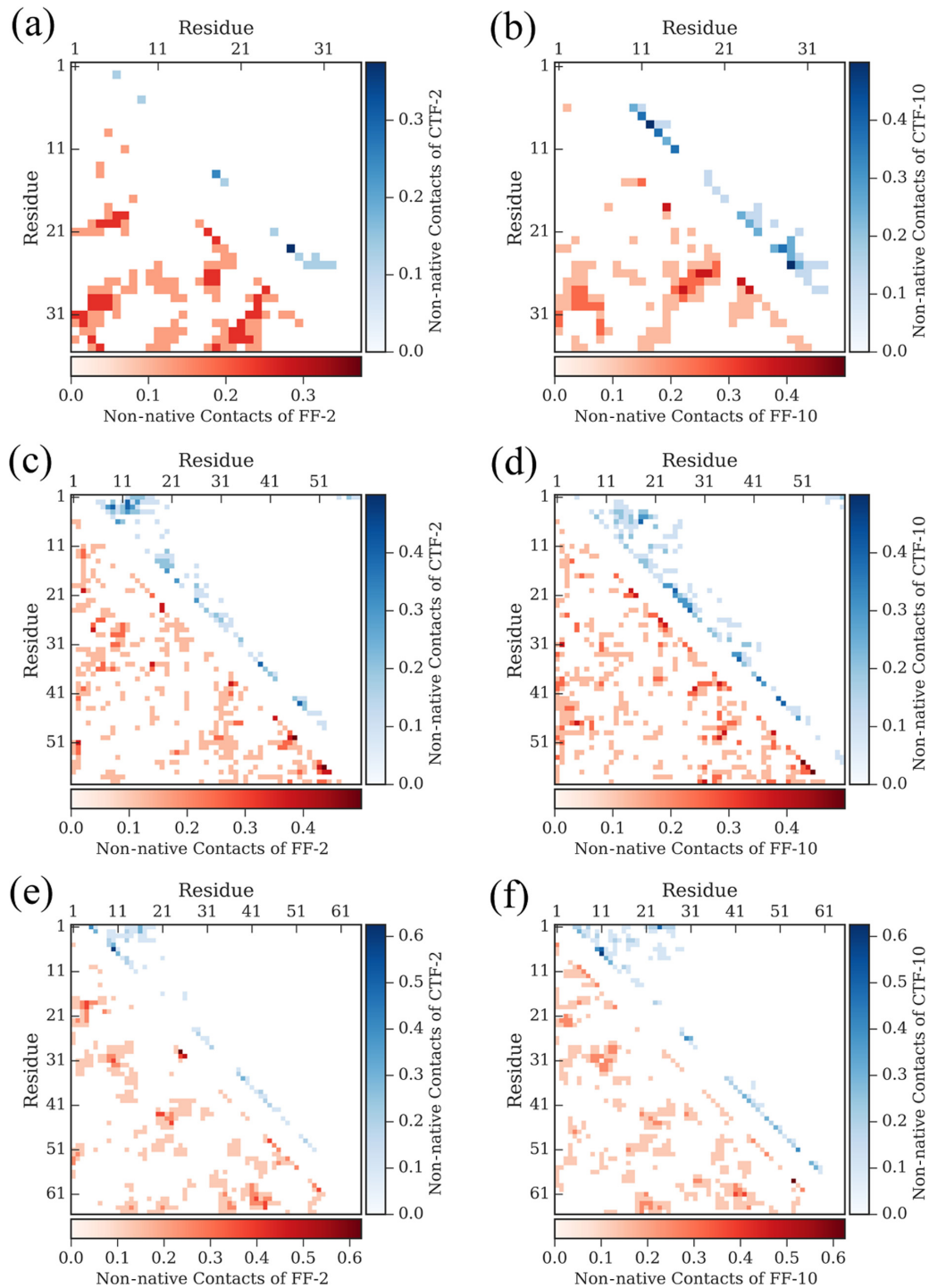
FIG. 8. The contact maps of the structural ensembles of (a), (b) GTT, (c), (d) SH3, and (e), (f) CI2 under free folding (blue, upper triangle) and cotranslational folding (red, lower triangle) when all the residues completely enter the tunnel (that is, the simulation time equals the translation speed multiplied by the number of residues). The darker the color, the higher the proportion of the contact in the ensemble.

of CTF? What are the differences between these conformations and those formed by *in vitro* folding? Starting from the secondary structure of the nascent chain, we analyzed the proportion of each residue forming a helix and β sheet before the nascent peptide chain completely entered into the tunnel, namely, during the simulation time t $\in$ [0, NΔt], where Δt is the translation speed and N is the residue number.

Figure 3 (left six subplots) shows the probability of each residue forming a helix at the translation speed of residue/2 ns and residue/10 ns. At the speed of residue/2 ns, except that the helical ratio of GTT in CTF is 31% lower than that of FF, the helical ratio of SH3 and CI2 in CTF is 64% and 15% higher than in FF, respectively. At the speed of residue/10 ns, the helical ratios of all three proteins in CTF are higher
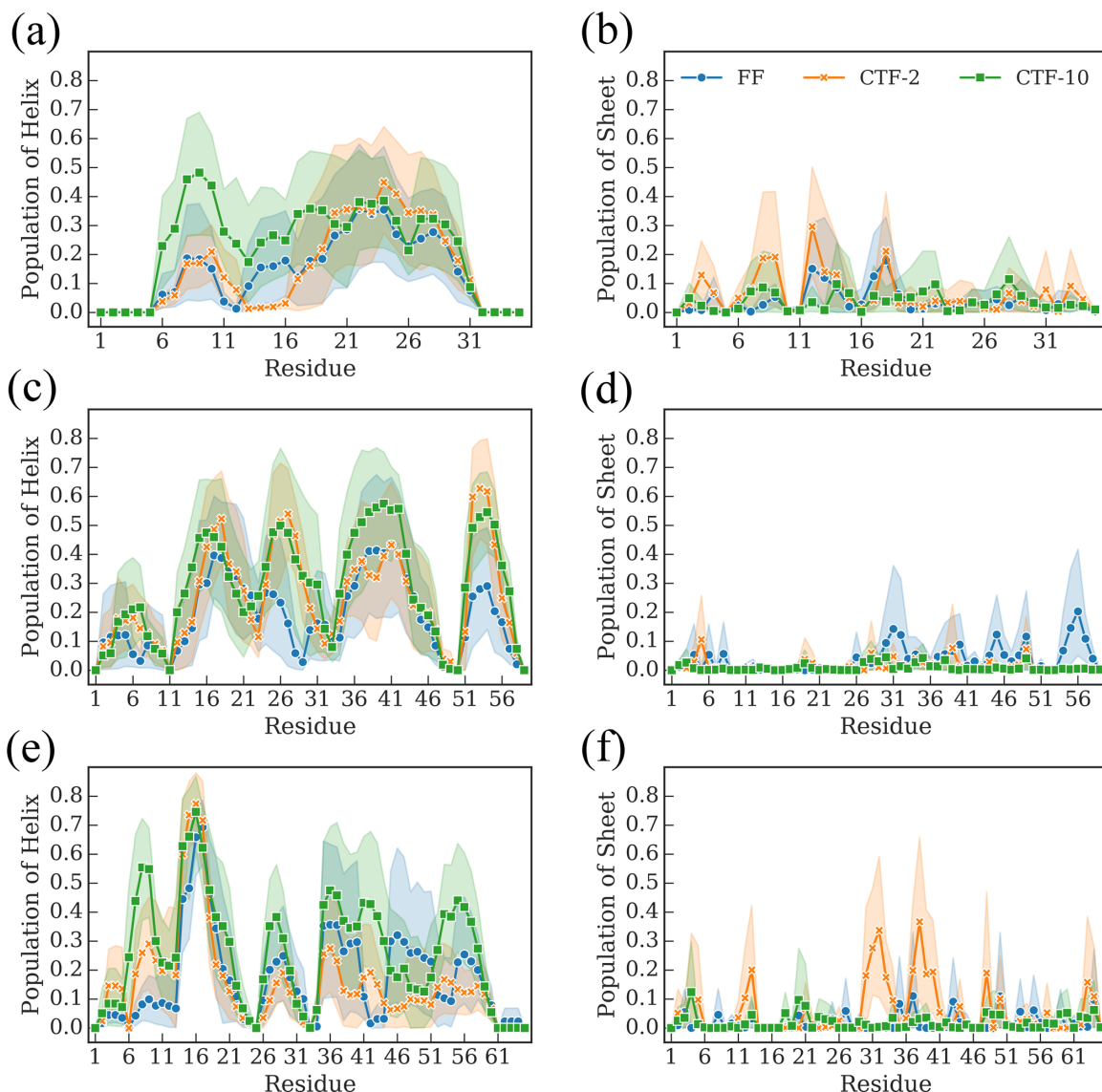
FIG. 9. After exiting from the ribosome tunnel, the population of Helix and Sheet of each residue of (a), (b) GTT, (c), (d) SH3, and (e), (f) CI2 under free folding (blue dots) and cotranslational folding at the speed of residue/2 ns (orange crosses) and residue/10 ns (green squares). The first-microsecond trajectory is not used in the analysis, and the subsequent analysis is also the same.

than those in FF, and the proportions of GTT, SH3, and CI2 increase by 161%, 100%, and 41%, respectively. This shows that the influence of the ribosomal channel on the formation of the helix depends on the system and the translation speed. It is worth mentioning that the increasing rates of the helix at the speed of residue/10 ns for all three proteins are higher than those of residue/2 ns. Unlike the helix, it is difficult for the β sheet to form, since the sampling is limited by the narrow and long space of the ribosomal exit channel, so the ratios of β sheets of these three proteins are relatively low at both translation speeds, and there is no significant difference compared with FF (right six subplots in Fig. 3). To investigate the effects of different parts of the tunnel model on the secondary structure, we conducted CRF simulations for GTT and found the behavior of the secondary structure under CRF is more similar to that of CTF (see Supplemental Fig. S25 [41]), which suggests that the change of secondary structure

is mainly due to the behavior of one-by-one translation rather than the spatial constraint of the cylinder.

From Fig. 3, it can also be found that the increases in the helical population of these three proteins during CTF are different (even for the same protein, each amino acid is different). In order to further analyze these differences, we calculated the difference in the helical population of each amino acid between CTF and FF. As shown in Fig. 4, at a translation speed of residue/10 ns, the difference is not significantly related to the hydrophilic index and the helix tendency of amino acids. However, some amino acids are more likely to adopt a helical conformation under CTF. In CTF-10, the amino acids with the top six highest increments are Met, Trp, Lys, Glu, Asn, and Ser, while in CTF-2 they are Ser, Tyr, Trp, Glu, Ile, and Thr (see Supplemental Fig. S26 [41]). In both cases, Trp, Gln, and Ser prefer to form a more helical structure during CTF.

TABLE II. After exiting from the ribosomal tunnel, the statistical results of the cumulative population of Helix (AP_h) and Sheet (AP_s), and Q_{nat} and Q_{non} of the three proteins in different simulations. The numbers outside and inside the brackets represent the average and standard deviation of corresponding quantities.

	GTT			SH3			CI2		
	FF	CTF-2	CTF-10	FF	CTF-2	CTF-10	FF	CTF-2	CTF-10
AP_h	4.86	5.25	7.94	10.93	14.81	17.07	10.98	10.13	16.91
	(4.20)	(3.13)	(3.19)	(4.60)	(4.92)	(5.92)	(4.94)	(3.80)	(6.31)
AP_s	1.38	2.23	1.33	2.20	0.85	0.52	1.11	3.70	1.26
	(1.38)	(1.79)	(2.00)	(1.57)	(0.72)	(0.62)	(0.96)	(1.34)	(0.93)
Q_{nat}	5.62	4.33	5.67	5.93	5.68	5.38	5.98	5.30	7.66
	(4.30)	(3.96)	(3.96)	(4.06)	(5.10)	(3.90)	(3.28)	(2.64)	(4.10)
Q_{non}	25.10	28.85	23.18	55.01	55.68	55.99	56.04	59.35	48.98
	(8.80)	(9.81)	(8.45)	(10.01)	(10.97)	(10.08)	(12.56)	(11.16)	(11.40)

B. Cotranslational folding can significantly reduce the number of non-native long-range contacts

In addition to the secondary structure, we also analyzed the formation of native and non-native contacts of the nascent chain before it completely entered the tunnel, and their numbers are recorded as Q_{nat} and Q_{non}, respectively. As shown in Fig. 5, for the two translation speeds, Q_{nat} of these three proteins is very small (less than 5), and there is no apparent difference compared with FF. This result is in line with our expectations because these three proteins are mainly β sheet, so native contacts are mainly long-range interactions. Surprisingly, for the two translation speeds, Q_{non} of these three proteins is significantly lower than that in FF, indicating that CTF can significantly reduce the non-native contacts. In
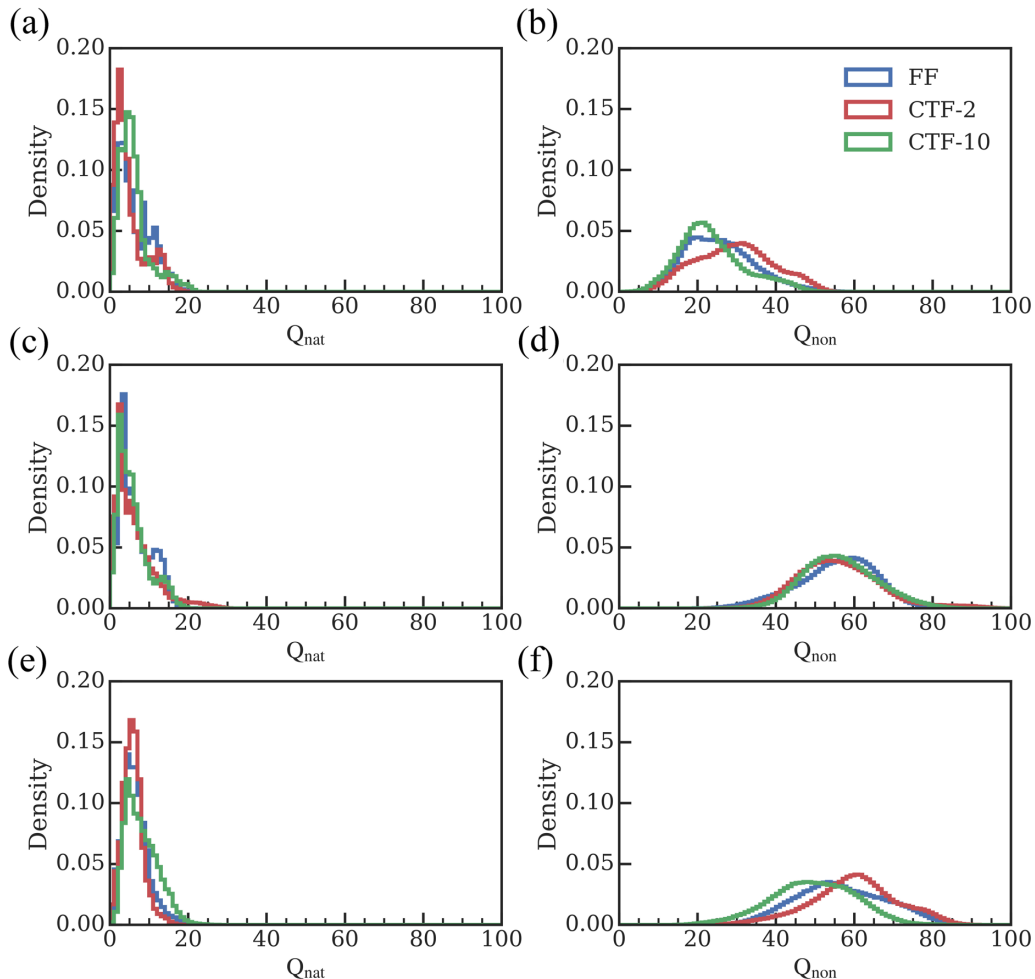

FIG. 10. After exiting the ribosome tunnel, the distribution of Q_{nat} and Q_{non} of (a), (b) GTT, (c), (d) SH3, and (e), (f) CI2 in free folding (blue) and cotranslational folding at the speeds of residual/2 ns (red) and residual/10 ns (green).
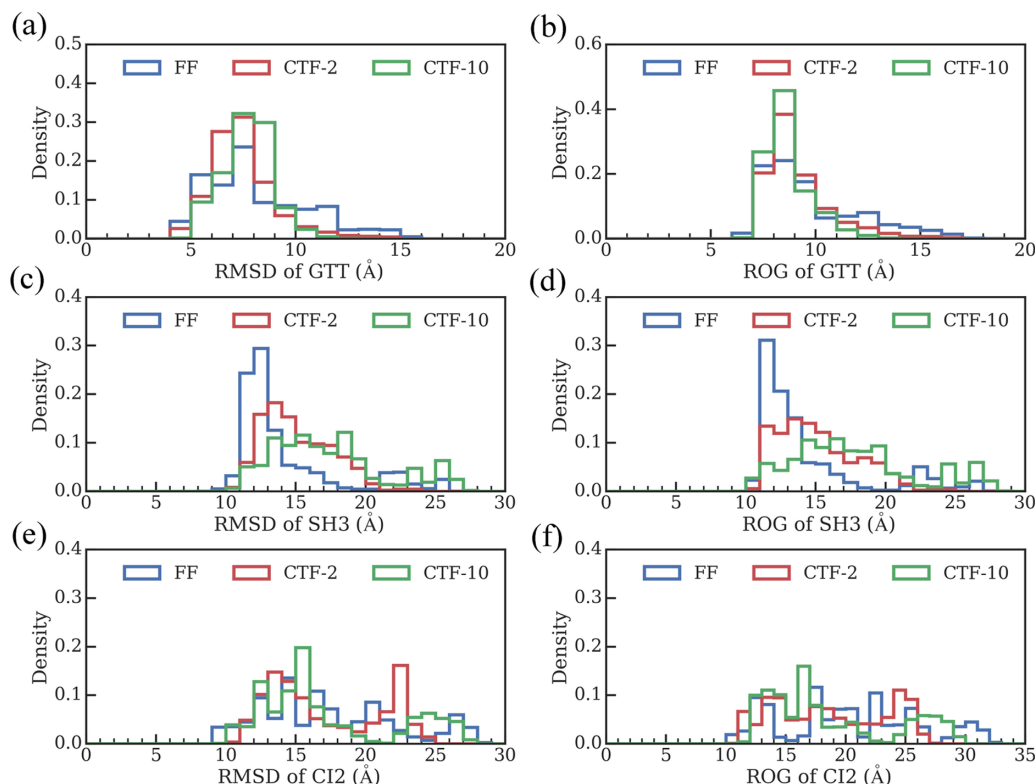
FIG. 11. After exiting from the ribosome exit tunnel, the distribution of root-mean-square deviation (RMSD) and radius of gyration (ROG) of (a), (b) GTT, (c), (d) SH3, and (e), (f) CI2 in free folding (blue) and cotranslational folding at the speeds of residual/2 ns (red) and residual/10 ns (green).

order to further quantify this reduction, we calculated the cumulative number of native and non-native contacts, recorded as AQ_{nat} and AQ_{non}, respectively (the area under the curve in Fig. 5). As shown in Fig. 6, at the speed of residue/2 ns, the AQ_{non} of GTT, SH3, and CI2 are decreased to 5.8%, 29.8%, and 24.7% of that in FF, respectively. Similarly, at the speed of residue/10 ns, the AQ_{non} of GTT, SH3, and CI2 are decreased to 38.4%, 62.9%, and 30.0% of that in FF, respectively, which are higher than the percentages in residue/2 ns. Furthermore, by comparing FF, CRF, and CTF simulations for GTT, we found that the behavior of contact in CRF is closer to that in FF (see Supplemental Fig. S27 [41]), indicating that the spatial constraint of the tunnel mainly reduces the non-native contacts, and thereby may prevent premature collapse to a local minimum state in subsequent folding.

Although we already know that CTF can reduce Q_{non}, maybe including long- or/and short-range contacts, it is unclear which part is reduced. Therefore, we extracted the structural ensemble of the peptide in the CTF simulation when it completely entered the tunnel and compared it with the structural ensemble in FF at the same time (Fig. 7), and analyzed the non-native contact map of these structural ensembles. As shown in Fig. 8, the contacts of the CTF ensemble are concentrated near the diagonal, which is mainly short-range contacts, while the contact of the FF ensemble includes both short-range contacts near the diagonal and a large number of long-range contacts. Therefore, mainly long-range non-native contacts are reduced by CTF.

C. Local minimum states under cotranslational folding have a higher helix propensity

Next, we turn our attention to the second question; that is to say, since CTF can make the nascent peptide chains have a higher helical propensity and fewer non-native long-range contacts, then how do these features help subsequent folding? Due to the limited simulation time, we did not observe any complete folding events, instead, most of the trajectories fell into the local minimum state. Therefore, we can only perform statistical analysis on the trajectories of subsequent simulations, such as whether the structural features caused by the previous CTF can be maintained in the subsequent folding. To simplify the analysis, all simulations do not include the first-microsecond trajectory. First, we counted the proportion of the secondary structure of each residue in the subsequent folding of the three proteins. As shown in Fig. 9, although the systems we simulated are three proteins dominated by a β sheet, the proportion of helix is much higher than that of a β sheet. In order to further quantitatively describe the difference between the helix (sheet) ratios in CTF and FF, we calculated the sum of the helix (sheet) ratios of all residues for each protein and denoted it as AP_h (AP_s). Table II gives the average values and standard variances of AP_h and AP_s in CTF and FF. At the speed of residue/2 ns, except that the helical ratio of CI2 in CTF is 7.7% lower than that of FF, the helical ratios of GTT and SH3 in CTF are 8.0% and 35.5% higher than in FF, respectively. At the speed of residue/10 ns, the helical ratios of all three proteins in CTF are higher than those in FF, and the

proportions of GTT, SH3, and CI2 increase by 63.4%, 56.2%, and 54.0%, respectively. It suggests that the proportion of the helix that CTF raises before exiting from the tunnel can be retained to a certain extent in the subsequent folding. Thus, the local minimum state under the CTF has a higher helix propensity, especially at the speed of residue/10 ns.

D. Cotranslational folding can help GTT enter the correct folding intermediate

Next, we calculate the distributions of Q_{nat} and Q_{non} of the three proteins in CTF and FF. From Fig. 10, we can intuitively see that these distributions are highly coincident. The average and standard variances of these distributions are given in Table II. For GTT and CI2, the order of the average Q_{nat} is CTF-10>FF>CTF-2, and Q_{non} is CTF-2>FF>CTF-10. For SH3, the rankings of the average values of Q_{nat} and Q_{non} are FF>CTF-2>CTF-10 and CTF-10>CTF-2>FF, respectively. These results show that the advantage of CTF in the nonnative contact established before exiting from the tunnel is not maintained in the subsequent folding.

In addition to the distributions of Q_{nat} and Q_{non}, we also calculated the distributions of root-mean-square deviation (RMSD) and radius of gyration (ROG), as shown in Fig. 11. Compared with the first two distributions (A,B), the latter four distributions (C–F) generally have more peaks. Furthermore, the peaks in FF and CTF are different, which indicates that the local minimum states in CTF and FF may be different. To verify this hypothesis, we clustered the subsequent trajectories according to the RMSD values and the results are shown in Fig. 12 and Table S2 [41]. Taking GTT as an example, we obtained ten classes (C1–C10, ranking by the size of the cluster) under three types of simulations, and then compared the five largest classes. As shown in Fig. 12, in FF, the first five classes do not tend to fold to the native state. In contrast, C1 in CTF-2 forms a misplaced hairpin, which may lower the barrier for the formation of the native hairpin (the first two β sheets), while C5 in CTF-10 is in the correct intermediate state, forming a native hairpin [the last two β sheets, Fig. 13(a)], which shows that CTF can help GTT enter the correct intermediate state.

IV. DISCUSSION

In this paper, we used all-atomic molecular dynamics simulation to study the CTF process of three proteins, GTT, SH3, and CI2, and found that the helical population of the nascent peptide chain can be enhanced by CTF in most cases and lower translation speed will produce more helix. This result is easy to interpret because the nascent peptide chain has a longer time to sample the local conformation at a lower speed. In addition, CTF can also significantly reduce the number of non-native long-range contacts. Some studies showed that, in this folding way, the nascent peptide chain can avoid falling into a misfolded state to improve folding efficiency [28,29]. This mechanism is reasonable and convincing. However, the nascent chain also rapidly collapses to local minimum states, most of which are misfolded states, in which Q_{non} is much larger than Q_{nat}. Therefore, the role of CTF is beyond preventing misfolding of nascent peptide chains. Moreover, although
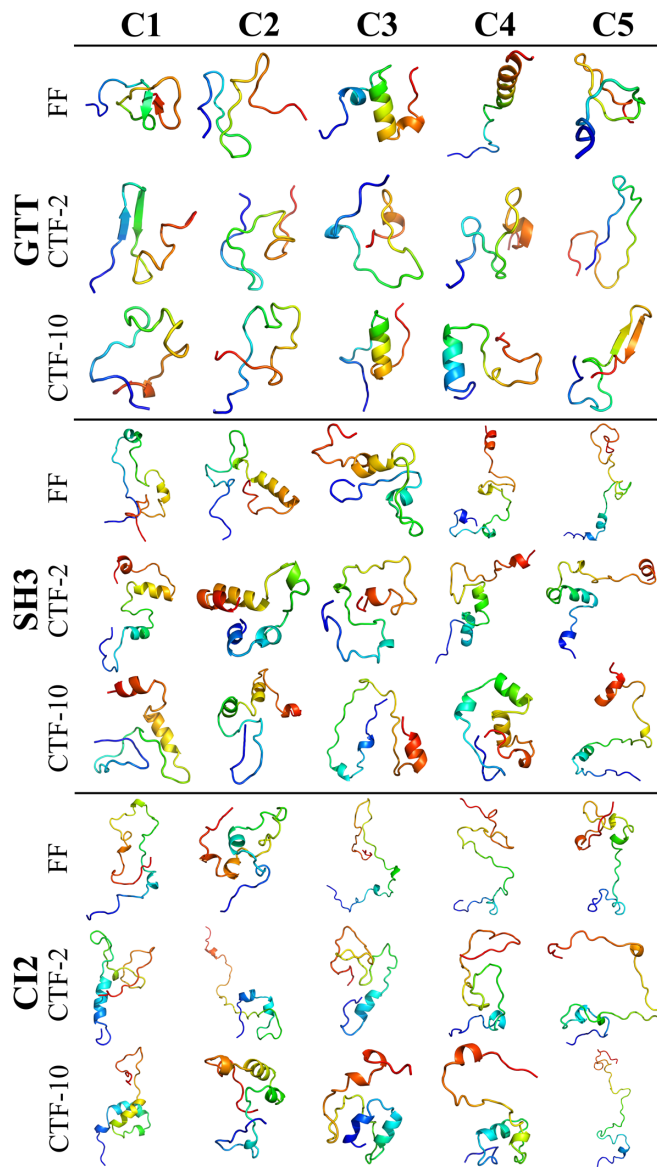

FIG. 12. After exiting from the ribosome exit tunnel, the representative structures of the five largest clusters of three proteins in three types of simulations. The RMSD values of the representative structures and the population of clusters are summarized in Table S2 [41]. This figure was prepared with PYMOL [53].

the distributions of Q_{nat} and Q_{non} are similar in CTF and FF, the local minimum states in CTF and FF are different. Through cluster analysis, we found that GTT is more likely to collapse to the correct intermediate state in CTF, which is not found in FF. Unfortunately, due to the limited simulation time, this phenomenon was not found for SH3 and CI2. However, for CI2, the probability of the formation of the native helix in CTF is significantly higher than that in FF, which may also be beneficial to the folding of CI2. As for SH3, we did not find that CTF has a significant effect on its subsequent folding, which is also consistent with previous research [30]. It should be noted that the failure to fold to the native state does not mean that the force field and simulation parameters used in this work are seriously flawed. On the one hand, the force field
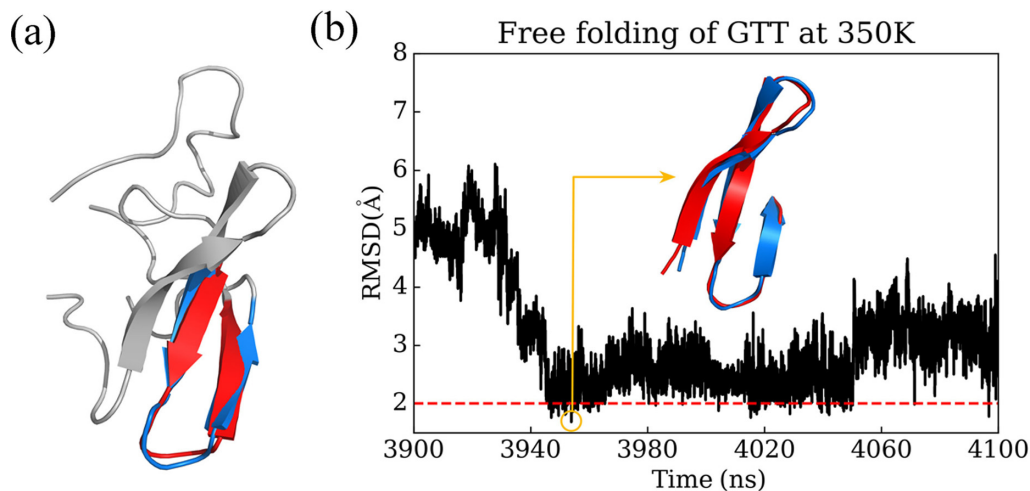
FIG. 13. (a) The fifth cluster C5 in CTF-10 has a native hairpin, which is the feature of the on-pathway folding intermediate. The RMSD of the alignment of the hairpin is 0.99 Å. (b) A short free folding simulation trajectory of GTT at 350 K, where GTT is successfully folded to its native state (defined by RMSD < 2 Å). The inset shows the structural alignment between the structure obtained by simulation (RMSD is 1.68 Å) and the native structure. Simulated and native structures are colored in red and blue, respectively.

and simulation parameters have been used to successfully fold five small proteins with different secondary structures, and the predicted folding times are in good agreement with the experiments [40]. On the other hand, for the GTT, there is an important reason for the failure to fold to the native structure; that is, the simulations were performed at room temperature (300 K), and most of the other works of FF simulations use a higher temperature to lower the barrier of folding or to speed up the folding speed [54]. To further illustrate that these parameters are reasonable, we also performed the FF simulations for GTT at 350 K, and finally successfully folded it to its native state [see an example in Fig. 13(b)]. As for SH3 and CI2, few reports have been made about using all-atom MD simulations to successfully fold them from extend initial structure, since their folding times at room temperature are more than 1 s [55,56].

Another useful result is that CTF can significantly increase the proportion of helices and maintain a high proportion of helices in subsequent folding. So what effect does a high proportion of helices have on folding? For proteins that are mainly α, a high proportion of helices may directly promote nascent peptide chains to form considerable native structures. For example, we recently found that CTF allows the formation of the entire N segment of HP35 (consists of two α helices) in the tunnel [35]. Then what about mainly β proteins? Cruzeiro proposed the vibrational excited states (VES) kinetic mechanism [57,58], who believed that CTF of a protein could be divided into three steps. The nascent peptide forms an all-helical structure in the tunnel at the first step, then the all-helical structure bends at a specific amino acid site, and the third step is that the helices transform into β sheets. According to this hypothesis, a higher ratio of helices means a better initial structure, so it will be helpful for subsequent folding. However, previous experiments showed that hydrophobic sequences were easy to form helices in the tunnel, while hydrophilic sequences were inclined to form stretch conformation [59]; that is, not all proteins were in a helical conformation when they exited from the tunnel, so

then what is the mechanism of CTF for these proteins? It is difficult to get an accurate answer to this question based on our current results; that is, the tendency of a residue to form a helix during CTF has no obvious relationship with its hydrophobicity, which is somehow inconsistent with the experiment. This inconsistency may be due to the fact that, to reduce the computational complexity, our exit tunnel model does not consider the potential of the real ribosome tunnel and its interaction with inside water molecules, but these factors may have an important impact on the dynamics of the nascent peptide [59,60]. In other words, a deeper understanding of the relationship between cotranslational folding and different sequence patterns requires more systematic and in-depth research.

As partially discussed above, the tunnel model and translation speed parameters used in this work have certain defects due to the current computing power. On the one hand, the tunnel model used in this work is rigid and regular, and the geometry of the real ribosomal tunnel is much more complicated, which may affect the CTF of the nascent peptide chain, such as the position where the fold begins [61]; not only that, but the interaction between the ribosome tunnel and the nascent peptide can regulate the folding of the nascent peptide and even the entire translation process [14]. For example, the restriction sites on the ribosomal tunnel can recognize certain sequences on the nascent peptide and cause translation pauses [62,63]. On the other hand, the actual translation speed in the cell is a few to tens of amino acids per second [64], which is about eight orders of magnitude slower than the translation speed used in this paper, which may make the sampling in the CTF more sufficient to achieve static balance sampling [65,66]. In addition, the translation speed in the cell is variable [67], and the change in translation speed may cause the nascent peptide to fold into different conformations [68,69], and the translation speed used in this paper is constant. In short, how these defects affect the folding of the nascent peptide, and whether these effects affect the conclusions of this paper need further study.

V. CONCLUSION

In summary, a general method, combining molecular dynamics simulations and a ribosomal tunnel model, is applied to study the cotranslational folding of three mainly β proteins. Before emerging from the ribosomal tunnel, more helical structures will be generated in cotranslation folding simulations, which will be maintained in subsequent folding. In addition, lots of long-range non-native contacts are decreased in cotranslational folding, but the distributions of non-native contacts in cotranslational folding and free folding are similar after the peptide exits the tunnel. Moreover, an on-pathway folding intermediate of GTT observed in the structure ensemble of cotranslational folding does not exist in free folding,

indicating that cotranslational folding can promote the formation of the correct folding intermediate by providing a suitable initial structure for subsequent folding. Overall, our study provides computational evidence of the positive role of cotranslational folding on folding efficiency.

[1] C. A. Waudby, C. M. Dobson, and J. Christodoulou, Trends Biochem. Sci. **44**, 914 (2019).

[2] C. M. Kaiser and K. X. Liu, J. Mol. Biol. **430**, 4580 (2018).

[3] F. Gloge, A. H. Becker, G. Kramer, and B. Bukau, Curr. Opin. Struct. Biol. **24**, 24 (2014).

[4] A. Gershenson and L. M. Gierasch, Curr. Opin. Struct. Biol. **21**, 32 (2011).

[5] M. Baiesi, E. Orlandini, F. Seno, and A. Trovato, Sci. Rep. **9**, 8426 (2019).

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

[7] M. V. Rodnina, Protein Sci. **25**, 1390 (2016).

[8] G. N. Jacobson and P. L. Clark, Curr. Opin. Struct. Biol. **38**, 102 (2016).

[9] G. Kramer, D. Boehringer, N. Ban, and B. Bukau, Nat. Struct. Mol. Biol. **16**, 589 (2009).

[10] S. Pechmann, F. Willmund, and J. Frydman, Mol. Cell **49**, 411 (2013).

[11] E. P. O'Brien, M. Vendruscolo, and C. M. Dobson, Nat. Commun. **5**, 2988 (2014).

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

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

[14] J. A. Farias-Rico, F. R. Selin, I. Myronidi, M. Frühauf, and G. Von Heijne, Proc. Natl. Acad. Sci. USA **115**, E9280 (2018).

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

[16] A. M. Knight, P. H. Culviner, N. Kurt-Yilmaz, T. Zou, S. B. Ozkan, and S. Cavagnero, ACS Chem. Biol. **8**, 1195 (2013).

[17] J. Lu, W. R. Kobertz, and C. Deutsch, J. Mol. Biol. **371**, 1378 (2007).

[18] H.-X. Zhou and K. A. Dill, Biochemistry **40**, 11289 (2001).

[19] C. Chen, E. Wang, P. Liu, and Y. Xiao, Phys. Rev. E **87**, 022701 (2013).

[20] G. Ziv, G. Haran, and D. Thirumalai, Proc. Natl. Acad. Sci. USA **102**, 18956 (2005).

[21] C. A. Woolhead, P. J. McCormick, and A. E. Johnson, Cell **116**, 725 (2004).

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

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

[24] O. B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Muller-Lucks, F. Trovato, J. D. Puglisi, E. P.

O'Brien, R. Beckmann, and G. von Heijne, Cell Rep. **12**, 1533 (2015).

[25] O. B. Nilsson, A. A. Nickson, J. J. Hollins, S. Wickles, A. Steward, R. Beckmann, G. von Heijne, and J. Clarke, Nat. Struct. Mol. Biol. **24**, 221 (2017).

[26] L. Tu, P. Khanna, and C. Deutsch, J. Mol. Biol. **426**, 185 (2014).

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

[28] C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, Jr., and C. Bustamante, Science **334**, 1723 (2011).

[29] A. J. Samelson, E. Bolin, S. M. Costello, A. K. Sharma, E. P. O'Brien, and S. Marqusee, Sci. Adv. **4**, eaas9098 (2018).

[30] E. J. Guinn, P. Tian, M. Shin, R. B. Best, and S. Marqusee, Proc. Natl. Acad. Sci. USA **115**, 12206 (2018).

[31] P. F. Tian, A. Steward, R. Kudva, T. Su, P. J. Shilling, A. A. Nickson, J. J. Hollins, R. Beckmann, G. von Heijne, J. Clarke, and R. B. Best, Proc. Natl. Acad. Sci. USA **115**, E11284 (2018).

[32] M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele, and J. W. Kelly, Proc. Natl. Acad. Sci. USA **103**, 10648 (2006).

[33] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. Guo, M. Gruebele, and D. E. Shaw, J. Mol. Biol. **405**, 43 (2011).

[34] C. A. McPhalen and M. N. James, Biochemistry **26**, 261 (1987).

[35] P. Tao, E. Wang, and Y. Xiao, Phys. Rev. E **101**, 052403 (2020).

[36] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).

[37] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, J. Comput. Chem. **26**, 1668 (2005).

[38] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11**, 3696 (2015).

[39] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, WIREs Comput Mol Sci **3**, 198 (2013).

[40] P. Tao and Y. Xiao, Phys. Rev. E **101**, 062417 (2020).

[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.105.024402 for tables with system details and simulations, plots of RMSD vs time, and secondary structure vs time for all simulations.

[42] B. P. Uberuaga, M. Anghel, and A. F. Voter, J. Chem. Phys. **120**, 6363 (2004).

[43] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[44] E. C. Wang, P. Tao, J. Wang, and Y. Xiao, Phys. Chem. Chem. Phys. **21**, 18219 (2019).

[45] A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **8**, 1542 (2012).

[46] R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **9**, 3878 (2013).

[47] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).

[48] P. Y. Chou and G. D. Fasman, Biochemistry **13**, 222 (1974).

[49] J. Kyte and R. F. Doolittle, J. Mol. Biol. **157**, 105 (1982).

[50] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham, J. Chem. Theory Comput. **3**, 2312 (2007).

[51] D. R. Roe and T. E. Cheatham III, J. Chem. Theory Comput. **9**, 3084 (2013).

[52] D. R. Roe and T. E. Cheatham III, J. Comput. Chem. **39**, 2110 (2018).

[53] W. L. DeLano, CCP4 Newsl. Protein Crystallogr. **40**, 82 (2002).

[54] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011).

[55] D. E. Otzen and A. R. Fersht, Biochemistry **37**, 8139 (1998).

[56] V. P. Grantcharova, D. S. Riddle, J. V. Santiago, and D. Baker, Nat. Struct. Biol. **5**, 714 (1998).

[57] L. Cruzeiro, Bio-Algorithms Med-Syst. **10**, 117 (2014).

[58] L. Cruzeiro, Z. Phys. Chem. **230**, 743 (2016).

[59] M. Bañó-Polo, C. Baeza-Delgado, S. Tamborero, A. Hazel, B. Grau, I. Nilsson, P. Whitley, J. C. Gumbart, G. von Heijne, and I. Mingarro, Nat. Commun. **9**, 5246 (2018).

[60] D. Suvlu, D. Thirumalai, and J. C. Rasaiah, J. Phys. Chem. B **125**, 817 (2021).

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

[62] D. N. Wilson and R. Beckmann, Curr. Opin. Struct. Biol. **21**, 274 (2011).

[63] G. W. Li, E. Oh, and J. S. Weissman, Nature (London) **484**, 538 (2012).

[64] M. Gouy and R. Grantham, FEBS Lett. **115**, 151 (1980).

[65] A. K. Sharma and E. P. O'Brien, Curr. Opin. Struct. Biol. **49**, 94 (2018).

[66] L. M. Alexander, D. H. Goldman, L. M. Wee, and C. Bustamante, Nat. Commun. **10**, 2709 (2019).

[67] S. Varenne, J. Buc, R. Lloubes, and C. Lazdunski, J. Mol. Biol. **180**, 549 (1984).

[68] F. Zhao, C. H. Yu, and Y. Liu, Nucleic Acids Res. **45**, 8484 (2017).

[69] F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M. V. Rodnina, and A. A. Komar, Mol. Cell **61**, 341 (2016).