JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Trigger Factor Slows Co-translational Folding through Kinetic Trapping while Sterically Protecting the Nascent Chain from Aberrant Cytosolic Interactions

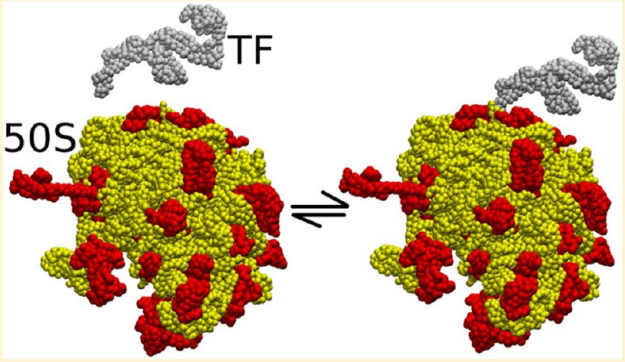
Edward P. O'Brien,[†] John Christodoulou,[‡] Michele Vendruscolo,[†] and Christopher M. Dobson*[,†]

[†]Department of Chemistry, University of Cambridge, Cambridge, U.K.

[‡]Institute of Structural and Molecular Biology, University College London and Birkbeck College, U.K.

Supporting Information

ABSTRACT: The *E. coli* chaperone trigger factor (TF) interacts directly with nascent polypeptide chains as they emerge from the ribosome exit tunnel. Small protein domains can fold under the cradle created by TF, but the co-translational folding of larger proteins is slowed down by its presence. Because of the great experimental challenges in achieving high spatial and time resolution, it is not yet known whether or not TF alters the folding properties of small proteins and if the reduced rate of folding of larger proteins is the result of kinetic or thermodynamic effects. We show, by molecular simulations employing a coarse-grained model of a series of ribosome nascent-chain complexes, that TF does not alter significantly the co-translational folding process of a small protein G domain but delays that of a large β-galactosidase domain as a result of kinetic trapping of its unfolded ensemble. We demonstrate that this trapping occurs through a combination of three distinct mechanisms: a decrease in the rate of structural rearrangements within the nascent chain, an increase in the effective exit tunnel length due to folding outside the cradle, and entanglement of the nascent chain with TF. We present evidence that this TF-induced trapping represents a trade-off between promoting co-translational folding and sterically shielding the nascent chain from aberrant cytosolic interactions that could lead to its aggregation or degradation.



INTRODUCTION

As the maintenance of a properly folded and fully functioning proteome is crucial for normal cellular function, cells have many mechanisms for maintaining protein homeostasis, including modulation of the rates of transcription, translation, and degradation of RNA and protein molecules. Indeed the regulation of translation is known to play a central role in protein homeostasis as the cellular concentration of a protein is only weakly correlated with its corresponding mRNA concentration,[1] and translation efficiency seems to be a better determinant of protein abundance in a cell than protein turnover.[2,3] During translation the protein nascent chain has the opportunity to fold concomitantly with its synthesis.[4] Such co-translational folding can in some cases reduce the probability of the formation of misfolded states,[5] which have a greater propensity to aggregate than do correctly folded proteins.[6] As it has been estimated that between 40% and 60% of bacterial proteins consist of two or more domains,[7] a sizable proportion of the proteins in bacteria have the potential to exhibit substantial degrees of folding during their biosynthesis. Thus, to understand protein folding *in vivo* and the maintenance of a healthy proteome in living cells, it is crucial to understand the nature of co-translational folding and the factors that influence it.

Molecular simulations can provide a detailed perspective on how co-translational folding can differ from folding in free solution. In particular, such studies have indicated that the ribosome can decrease the diversity of folding pathways and promote the formation of partially folded intermediate structures near the N-terminus of the nascent chain.[8−11] An interesting question that remains to be addressed fully concerns the effect of other cellular components that interact with the nascent chain, as these have the potential to alter substantially the co-translational folding properties of nascent polypeptide chains.

One of the first auxiliary factors to interact with nascent chains in *E. coli* is the chaperone trigger factor (TF, Figure 1a),[12] which consists of three regions referred to as the "head", "body", and "tail"; furthermore, two "arms" protrude from the body of TF. The tail of TF binds to the L23 ribosomal protein with a K_D as low as 50 nM,[13] and this binding orients the body and arms of TF toward the exit tunnel opening, through which

Received: March 8, 2012
Published: June 8, 2012

   dx.doi.org/10.1021/ja302305u | *J. Am. Chem. Soc.* 2012, 134, 10920−10932

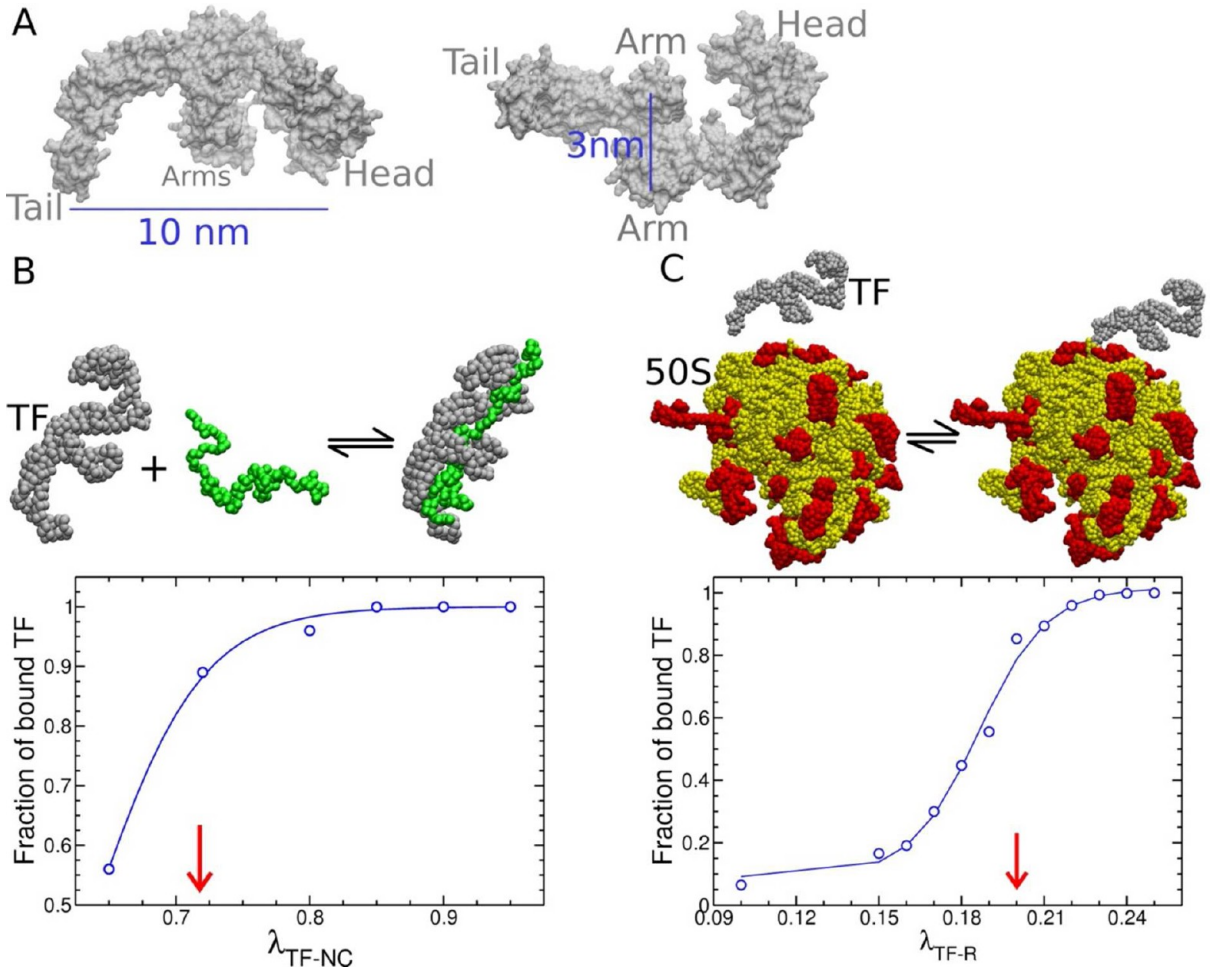


Figure 1. Affinity of TF for the nascent chain and ribosome was set by reproducing experimentally measured K_D values. (A) The TF structure consists of regions referred to as the "tail", "arms", and "head".[12] The head-to-tail distance is approximately 10 nm (left), while the distance between the arms is approximately 3 nm (right). These dimensions are sufficient to accommodate small folded domains.[14] (B) The energy scale for the interaction between TF and the nascent chain was set by reproducing the experimental K_D value of 1.7 μM TF binding to reduced α-lactalbumin, which is unable to fold in this form.[22] The fraction of TF bound to α-lactalbumin was calculated from simulations run at different $\lambda_{TF\text{-}NC}$ values; the parameter $\lambda_{TF\text{-}NC}$ uniformly scales the well-depths of the Lennard−Jones interactions between TF and this protein (see Methods). A $\lambda_{TF\text{-}NC}$ value of 0.72 (red arrow) was found to reproduce the experimentally measured fraction bound of 0.90 (which is calculated from K_D). (C) The energy scale between TF and its binding site on ribosomal L23 protein was set by reproducing the experimental K_D value of 2.1 μM TF binding to L23 on the ribosome.[27] The fraction of TF bound to the ribosome was calculated from simulations at different $\lambda_{TF\text{-}R}$ values, where $\lambda_{TF\text{-}R}$ corresponds to the Lennard−Jones well-depth between the tail of TF and the L23 ribosomal protein binding site (see Methods). A $\lambda_{TF\text{-}R}$ value of 0.20 was found to reproduce the fraction of bound TF predicted from the measured K_D value.

the nascent chain emerges.[14] With the concentration of ribosomes in the cytoplasm being around 20 μM (determined in *E. coli* cells that divide every 30 min[15]) and a concentration of TF of about 50 μM,[12] it is estimated that approximately 75% of all non-translating ribosomes, and up to 95% of all translating ribosomes are bound to TF. Therefore, to gain further insight into how chaperones can impact co-translational folding properties, a natural place to start is to explore the effects of interactions with TF.

TF has been found to perform a wide range of biological functions, including a peptidylprolyl isomerase activity that resides in the head domain,[11,15] although this function is not essential for its chaperone activity,[16] which originates primarily in its C-terminal domain.[17] TF increases the yield of biologically active proteins,[18,19] indicating that it plays an important role in promoting protein folding. In contrast to many other intracellular chaperones, however, TF does not use ATP to achieve this effect.[20] Therefore, the chaperone activity of TF may not result from the active refolding of proteins, but potentially from the indirect effect of suppressing competing reactions that reduce the folding yield, such as degradation[21] and aggregation.[18]

Several experimental reports have provided insight into the structural effects and overall impact of TF on nascent chain folding, which in turn have raised further fundamental questions. Domains containing as many as 150 residues have been shown to fold within the TF cradle, as indicated by cryo-EM,[14] cross-linking,[14] and proteolysis experiments.[21] The spatial and time-resolution of these techniques, however, is not yet sufficient to reveal the details of how the folding of a nascent chain is altered by TF. Another observation that remains to be explained in mechanistic terms is represented by time-dependent enzymatic activity measurements during the synthesis of β-galactosidase that indicate that the co-translational folding of this large protein is slowed down by the presence of TF both *in vitro* and *in vivo*;[19] this observation is

dx.doi.org/10.1021/ja302305u | J. Am. Chem. Soc. 2012, 134, 10920−10932

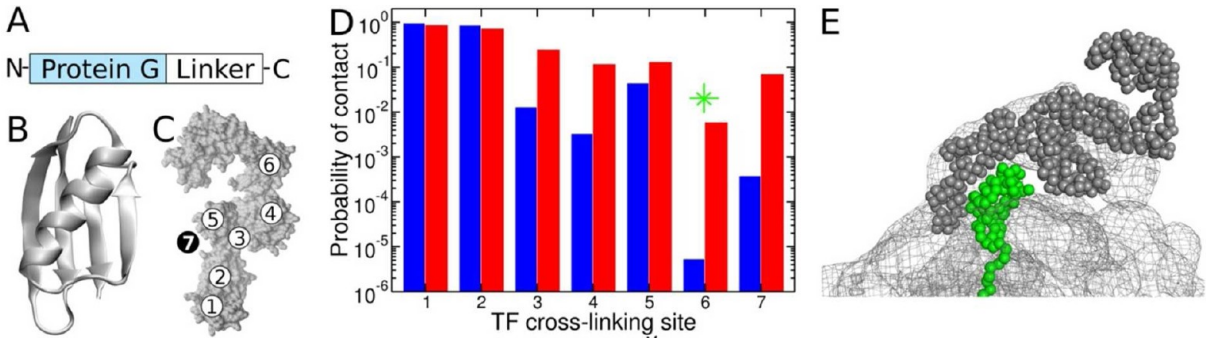


Figure 2. Comparison between simulations and cross-linking and cryo-EM data for the protein G RNC. (A) An unstructured polyglycine linker 35 residues in length was attached to the C-terminus of protein G, allowing this domain to fold, when fully synthesized with the linker, at the exit tunnel vestibule while still tethered to the P-site tRNA. The nascent chain is 91 residues in length and was simulated on an arrested ribosome. (B) Native structure of the protein G folding domain. (C) TF structure with cross-linking sites marked 1 through 7 and corresponding, respectively, to residues 61, 76, 404, 377, 322, 256, and 9 on TF;[14] site 7 is located on the exterior of TF. These sites, when labeled with the chemical compound BPA, can cross-link with amine groups on the nascent chain. (D) Comparison to experimental cross-linking results was performed by calculating the probability that an amine carrying side chain, or the N-terminus of the nascent chain, comes into contact during the simulations with one of the TF residues carrying a cross-linking agent. The probability of contact is shown for the folded (blue) and unfolded (red) populations in the simulations. Consistent with the experimental results, site 6, located in the head region of TF, is 1600 times more likely to come into contact, and presumably to cross-link, with a nascent chain amine group when the domain is unfolded as compared to when it is folded. (E) The positioning of TF and the folded protein G domain in the simulation is consistent with an electron density map (gray mesh) measured via cryo-EM.[14] The majority of TF and nascent chain residues are within the 1 σ surface shown. The cryo-EM data were measured on an *E. coli* ribosome containing a 104 residues long nascent chain consisting of an SH3 folding domain (62 residues) attached to a SecM sequence at its C-terminus. The TF-nascent chain conformation shown was chosen for illustration because it is close to the average center-of-mass position of TF in the simulations.

consistent with experiments that have shown TF can decrease the rate of refolding of a number of proteins when initiated from a chemically unfolded state.[18,22−24] It is not yet known if the slower folding of these larger proteins is caused by thermodynamic stabilization of the unfolded state or by kinetic effects that decrease the rate of structural transitions but do not significantly alter the equilibrium populations of the folded and unfolded states. Knowledge of the molecular and structural origins of these mechanisms would offer a better understanding of the role of chaperones in co-translational folding.

To address these open questions and to provide a molecular perspective on the effects of TF on co-translational folding, we have used a coarse-grained model of the ribosome nascent chain complex to simulate in the presence and absence of TF the biosynthesis of (i) a small (56 residues) protein G domain and (ii) the larger (216 residues) N-terminal domain of β-galactosidase. Only one out of the five domains of the latter was investigated as it was not possible to simulate the synthesis of the full-length protein on a reasonable time-scale using our approach. We find that protein G can fold under the cradle of TF when it is tethered to the ribosome by a polypeptide linker and that its folding kinetics, thermodynamics, and pathways are not significantly altered by the presence of TF. On the other hand, our simulations of the synthesis of the β-galactosidase domain reveal that its co-translational folding is under kinetic control. The molecular origin of this effect, in which the folding rate is reduced, arises primarily from a decrease in the rate of structural rearrangements within the domain. Additional contributions come from an increase in the effective exit tunnel length and entanglement of the domain with TF. Thus, we observe that TF slows the co-translational folding of this large domain through three different kinetic trapping mechanisms. Through an analysis of the structures obtained in the simulations, we find the accessibility of the nascent chain to other cytoplasmic proteins is greatly reduced by the presence of TF through steric occlusion, supporting the notion that the biological role of TF in the cell is to suppress nascent chain degradation and aggregation even if it is at the cost of reducing the co-translational folding rates. We discuss how these results may explain a series of different experimental observations on TF's behavior and TF's effect on heterologous protein expression levels.

RESULTS

Setting Physically Reasonable Energy Scales in the Simulation Model Yields Results Consistent with Experiments. Setting realistic intra- and interprotein interaction energies in the force field of co-translational folding in the presence of TF is crucial for simulating realistic behavior. If the interaction strength of TF with the nascent chain is set to be too weak or too strong, then excluded volume interactions or enthalpic interactions would, respectively, dominate between these two components. An effective way to set these energy scales is to utilize experimental native state stabilities to define the intraprotein interactions[25] and experimental dissociation constants (K_D) for interprotein interactions.[26] Therefore, the Lennard−Jones well-depths for the interactions between TF and nascent chains (Figure 1B) and TF and ribosomes (Figure 1C) were chosen such that they reproduce the experimentally measured K_D values[22,27] while still capturing the sequence effects of each residue in the nascent chain and the TF molecule (see Methods). For the protein G and N-terminal β-galactosidase domains studied here, we set intraprotein nonbonded parameters such that their native state stabilities equal that measured experimentally in bulk solution (see Methods). This approach ensures that realistic equilibrium populations of ribosome-bound TF and folded protein are obtained in this model.

To verify whether the resulting model is consistent with experimental cross-linking and cryo-EM results, we simulated a ribosome nascent chain (RNC) complex containing protein G tethered to the P-site tRNA by means of a 35-residue polyglycine linker (Figure 2A and B) and with TF bound to the ribosome, which resulted in a nascent chain of 91 residues.

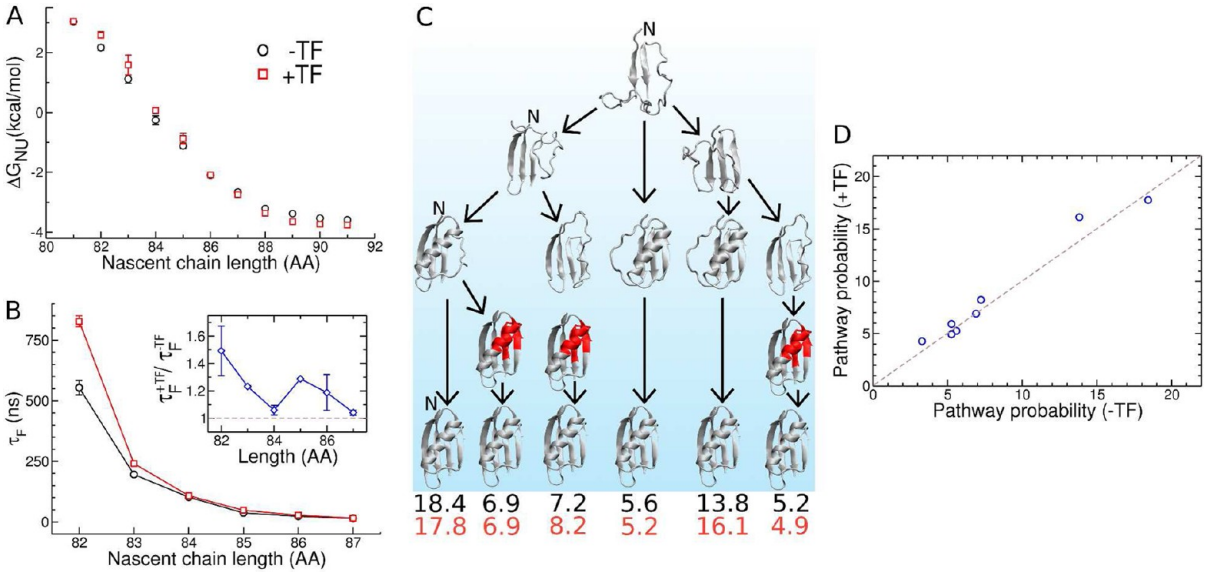


Figure 3. TF has minimal effect on the co-translational folding properties of protein G. (A) The stability of the native state, N, of protein G with respect to the unfolded state, U, as a function of nascent chain length is largely unaltered by the presence of TF (+TF, red squares). Values of ΔG_{NU} ($= -k_B T \log[P_N/P_U]$, where P_N and P_U correspond to the probability of the domain being folded or unfolded, respectively) in the absence of TF (−TF) are displayed as black circles. Error bars correspond to the standard error about the mean. (B) The mean folding time (τ_F) of protein G as a function of nascent chain length is only slightly increased by the presence of TF (red squares) below a chain length of 84 residues and displays no significant increase at longer nascent chain lengths. The ratio of τ_F in the presence (τ_F^{+TF}) and absence of TF (τ_F^{-TF}) is shown as an inset; at most there is a 50% increase in the time required for folding at 82 residues in length, with a much smaller difference at longer nascent chain lengths. (C) The six most probable folding (transition) pathways in both the presence and absence of TF with the percentage probability of taking each pathway given below (−TF, black; +TF, red). The accumulation of folded structure along the pathway is indicated by the appearance of elements of secondary structure (α-helices and β-strands). The bottom set of structures corresponds to the folded state. The red portions of a structure indicate incomplete docking and folding of the C-terminal β-hairpin onto the α-helix. The fully unfolded structure is not shown but was sampled during the simulations. (D) TF does not alter the distribution of the co-translational folding pathways of protein G. The percentage probability of taking a particular pathway in the presence of TF (+TF) versus taking that same pathway in the absence of TF (−TF) is shown for the 10 most probable pathways, which account for more than 50% of the routes of co-translational folding. They fall on or near the identity line (dashed) demonstrating that this distribution of pathways is not altered by the presence of TF.

For a similar length construct of SH3, consisting of a 62-residue folding domain and 42-residue linker, it was found that the domain could cross-link to residue 256 of TF (located on the head domain) when unfolded, but when folded no such cross-linking occurred (Figure 2C).[14] These results indicate that the folding status of the domain appears to determine where it is positioned under the TF cradle. Using the contact probability as an approximation for the probability of cross-linking, we calculated from the simulations the probability that any amine-containing residues in the protein G nascent chain would come into contact with TF residue 256. The results indicate that protein G is over 1600 times more likely to come into contact with TF residue 256 when unfolded compared to when it is folded (Figure 2D). Thus, the overall structural properties of the folded and unfolded ensembles under the TF cradle in this model are consistent with the available experimental data.

To validate the positioning of this folded domain and of TF with respect to the ribosome, we superimposed a cryo-EM electron density map obtained for the SH3-RNC complex at 88 K onto the average structure of the protein G-RNC system from a simulation at 88 K (Figure 2E). We observe that the majority of residues comprising TF and the nascent chain fit well within this density map, indicating that in these simulations the relative positions of TF, the nascent domain, and the ribosome are reasonable.

Trigger Factor Has Minimal Effects on the Cotranslational Folding Properties of a Small Protein Domain. Small protein domains have been observed to fold under the TF cradle,[14,21] but because of the limited resolution of the experimental probes used to date it is not known if TF alters their co-translational folding properties. We therefore simulated a series of arrested RNC complexes with different nascent protein G chain lengths in the presence and absence of TF. These protein G constructs contain an unstructured poly-glycine linker (Figure 2A) that varies between 25 and 35 residues in length, yielding nascent chain lengths ranging from 81 to 91 residues. We find the native state stability of protein G relative to that of its unfolded state changed by no more than 0.4 kcal/mol in the presence of TF (Figure 3A), indicating that the latter has little effect on the thermodynamics of co-translational folding. The co-translational folding kinetics of protein G are only moderately slowed by TF (Figure 3B); indeed, at nascent chain lengths greater than 84 residues there is very little (<20%) difference between the mean folding time in the presence or absence of TF. At shorter nascent chain lengths the mean folding time can increase by 50% because of the presence of TF; the folded state, however, is unstable at these lengths (Figure 3A) and consequently the number of co-translational folding events per unit time (i.e., the folding flux) is negligible.

Even though such global co-translational folding properties are largely unaltered by TF, it is still possible that this chaperone alters the probability distribution of microscopic folding routes, as these are exponentially sensitive to changes in barrier heights on the energy landscape. To test if this situation is likely to occur, we analyzed the transition paths[28] taken by

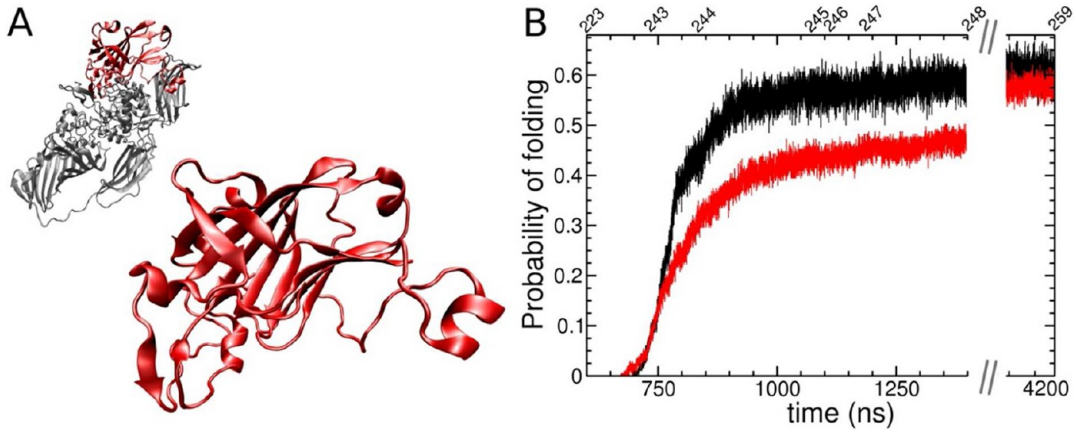


Figure 4. Trigger factor slows down the co-translational folding of the N-terminal β-galactosidase domain. (A) Full length β-galactosidase (upper left) is 1023 residues long and has been found experimentally to fold co-translationally in the absence of TF and to exhibit delayed co-translational folding in its presence.[19] Since simulations of the biosynthesis of full length β-galactosidase are not possible using the approach that we describe in this work because of its size, we simulated the biosynthesis of its N-terminal domain (red), which is 216 residues long and has a contact order[49] of 30 residues, the latter indicating that this domain makes long-range contacts in its folded state. (B) TF slows down the co-translational folding of the N-terminal domain. The probability of domain folding during continuous synthesis is shown as a function of the simulation time in the presence (red) and absence (black) of TF. The numbers 223 through 259 on the top axis indicate the nascent chain length, and their location indicates the time at which a new amino acid was added to the nascent chain. Between nascent chain lengths 223 and 243, new residues were added every 7.5 ns, and the intervening lengths are not indicated on this graph. Immediately prior to 1400 ns, the nascent chain is 247 residues in length and the N-terminal β-galactosidase domain is fully outside of the ribosome exit tunnel. We note that these Langevin dynamics simulations were run at a viscosity much lower than that of water, a situation that accelerates dynamical processes. A linear mapping between the calculated folding time of protein G in free in solution and that measured experimentally indicates that 90 ns of simulation time corresponds to approximately 60 ms in experiments carried out in aqueous solvent (see Methods).

protein G during folding after a temperature quench at a nascent chain length of 84 residues, where protein G is at its midpoint of stability (Figure 3A). The transition paths, which connect the unfolded and folded basins, were defined in terms of the extent and type of native tertiary structure that accumulates as a function of time (Figure 3C). The probabilities of each of the 10 most highly populated folding routes, which account for over 50% of the total number of routes taken by this protein, were then calculated (see Methods); these were found to be largely unchanged by TF (Figure 3D). Thus, we conclude that TF does not significantly alter the distribution of folding pathways of this small protein domain.

Trigger Factor Slows the Co-translational Folding of the N-Terminal Domain of β-Galactosidase through Kinetic Control. TF has been observed experimentally to slow down the co-translational folding process of β-galactosidase.[19] It is not known, however, if this effect is a result of thermodynamic control, i.e., a result of the stabilization of the unfolded state, or through kinetic control, i.e., as a result of a reduction in the rates of structural transitions without a change in their equilibrium populations. To probe the molecular origins of this TF activity, we have simulated the continuous biosynthesis of the N-terminal domain of β-galactosidase, the first domain of full-length β-galactosidase to emerge from the exit tunnel, in the presence and absence of TF. This N-terminal domain (Figure 4A) is 216 residues in length; nascent chain lengths up to 259 residues were therefore simulated to allow this domain to emerge fully from the exit tunnel.

We simulated a total of 410 independent synthesis trajectories of this RNC, half of which were carried out in the presence of TF and the other half in its absence. On the basis of the recent experimental observation that TF binds only to nascent chains that are 100 or more amino-acids in length,[29] the simulations were initiated from a nascent chain length of 143 residues. This length is too short for this domain to be fully outside the exit tunnel, and hence in this configuration it is unable to fold completely. Synthesis was then carried out in the simulations by adding amino acids to the C-terminus of the nascent chain at a series of time intervals (see Methods). We find by this method that the presence of TF slows down co-translational folding of the N-terminal domain (Figure 4B). At a nascent chain length of 247 residues, where the domain is completely outside the exit tunnel and hence sterically able to fold,[8] the fraction of trajectories in which the domain is completely folded is 0.58 in the absence of TF and 0.46 in its presence. Thus, we estimate that at this nascent chain length TF decreases the population of this domain in its folded state by 12%.

As mentioned above, such a decrease in the folded population could arise from a destabilization of the native state in the presence of TF (a thermodynamic effect) or from a folding reaction that is slowed in the presence of TF but leads to no change in the equilibrium population of the native state (a kinetic effect). Such kinetic effects have the potential to arise as a consequence of the out-of-equilibrium nature of translation. One way to test if a given reaction is under kinetic control is to observe it over much longer time scales and see if the final population of the species of interest changes. We applied this approach by extending the simulation time by a factor of 3, from 1400 to 4200 ns, which allows the nascent chain to elongate to 259 residues. We find that the folding probability, both in the presence and absence of TF, saturates at around 0.6 (Figure 4B), indicating that in both cases they have similar equilibrium folded populations. Thus, our simulations indicate that the slowdown in co-translational folding, which is consistent with experimental observation,[19] is due to kinetic control of this reaction.

The saturation of the folding probability at around 0.6 in the current study is likely to arise from an additional process that

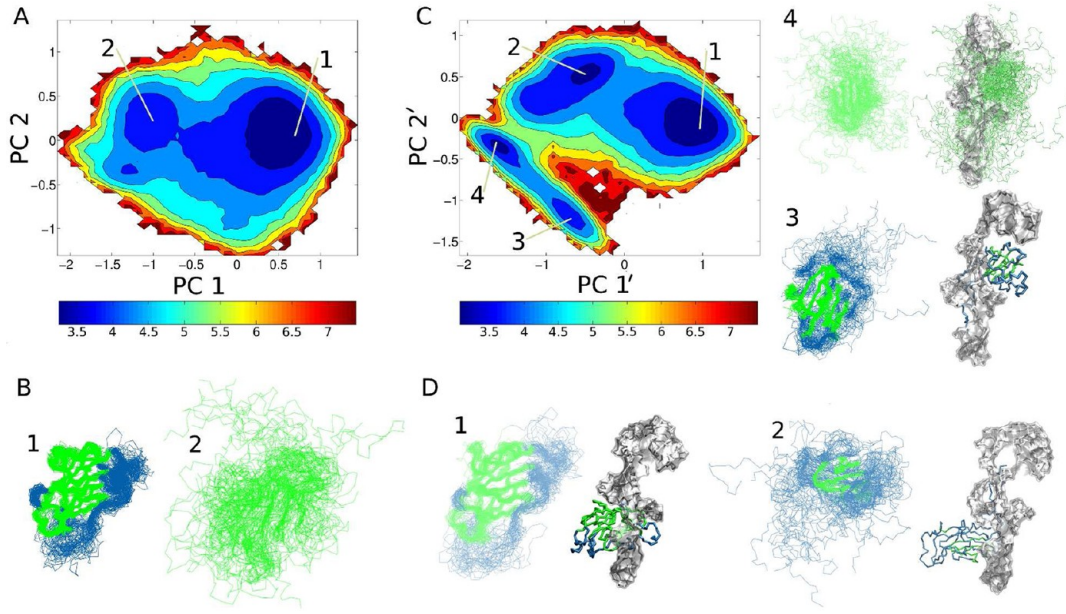


Figure 5. Principal component analysis reveals that additional unfolded states of the N-terminal domain of β-galactosidase are populated in the presence of TF. (A) The free energy surface, denoted F(PC 1,PC 2), at a nascent chain length of 247 residues was calculated as $-k_BT \log[P(\text{PC 1,PC 2})]$, where k_BT = 0.62 kcal/mol at 310 K and P(PC 1,PC2) is the probability of sampling a domain conformation that has a particular value of PC 1 and PC 2. Regions of isostability are shown in the same color. The energy scale, in units of kcal/mol, is shown beneath the figure. (B) Clustering of the structures in these two states (see Methods) reveals that basin 1 corresponds to the fully folded domain and basin 2 corresponds to an unfolded ensemble. (C) In the presence of TF the free energy surface projected onto principle components PC 1′ and PC 2′, which are different from PC 1 and PC 2 in panel A, exhibits four distinct basins labeled 1 through 4. (D) Structural clustering reveals basin 1 corresponds to the fully folded domain (left) that on average is located under the TF cradle (right); basin 2 corresponds to a partially folded C-terminal ensemble that is located in part under the cradle; basin 3 corresponds to a partially folded N-terminal ensemble that protrudes out of the cradle to the right of the TF body and above its arms (see also Supplementary Figure S1); basin 4 corresponds to a heterogeneous unfolded ensemble that samples throughout the cradle. In these clustered structures the domain is shown with and without the image of TF. In basin 2, the spatial proximity of the head and body of TF in the structure shown results in an apparent connection between these two segments in this surface area representation; there is, however, is no direct link between them.

occurs on much longer time scales, namely, topological frustration during the folding of this complex native structure. In this slow phase, partial unfolding must first occur to correct misfolded β-strand arrangements before the native state can be reached, leading to large energetic barriers to folding. Such topological frustration has been observed for other complicated folding architectures, and it is typical for such phenomena to give rise to slow phases in folding reactions.[30,31]

Structural Analysis of the Unfolded Conformations of β-Galactosidase Indicates the Mechanisms by Which Trigger Factor Slows down Co-translational Folding. Having identified kinetic effects as being central to the role of TF in the co-translational folding of β-galactosidase, it is important to determine the molecular mechanisms by which they arise. Because the present simulations (Figure 4B) reproduce the experimental observation that TF slows down the co-translational folding of β-galactosidase,[19] we analyzed the conformations sampled in the simulations to clarify the possible nature of this mechanism. To achieve this goal we first examined the nature of the very large and heterogeneous structural ensemble of conformations sampled during the simulations. We applied principal component analysis (PCA) to identify highly populated clusters of closely related structures (see Methods) from the last 116 ns of each synthesis simulation at a nascent chain length of 247 residues, where TF leads to the largest difference in the folding probability (Figure 4B). PCA is a mathematical procedure that projects the simulation time-series data onto a vector space that maximizes the variance in the data and separates out distinct subpopulations.

To visualize the data in this new vector space, we plotted the free energy surface as a function of principal components 1 and 2, i.e., we took the logarithm of the probability of finding a value (PC 1, PC 2) during the simulations (Figure 5A). We find that in the absence of TF there are two free energy minima, with one centered close to (0.5, 0.0), and the other close to (−1.0, 0.2). By clustering the simulation structures at these points (see Methods) we find that the first basin corresponds to the correctly folded ensemble of structures (Figure 5B), which has an average radius of gyration (R_g) 3% larger than that of the crystal structure ($R_{g,C}$) and an average Q value of 0.75 (the Q value is the fraction of native contacts and is typically less than 1.0 even in the native basin due to thermal fluctuations). The second free energy basin corresponds to heterogeneous unfolded conformations with an R_g that is 18% larger than $R_{g,C}$ and an average Q value of 0.45.

We applied the same analysis procedure to the conformations sampled in the last 116 ns of simulations in the presence of TF, again involving a nascent chain length of 247 residues. For this system we observed four basins of attraction in the two-dimensional free energy surface (Figure 5C), two more than in the absence of TF. These results indicate therefore that TF has partitioned the unfolded conformations into additional states that are separated by appreciable free energy barriers. Further analysis (Figure 5D) reveals that these basins, labeled 1 through 4 in Figure 5C, correspond to (1) a fully folded

dx.doi.org/10.1021/ja302305u | J. Am. Chem. Soc. 2012, 134, 10920−10932

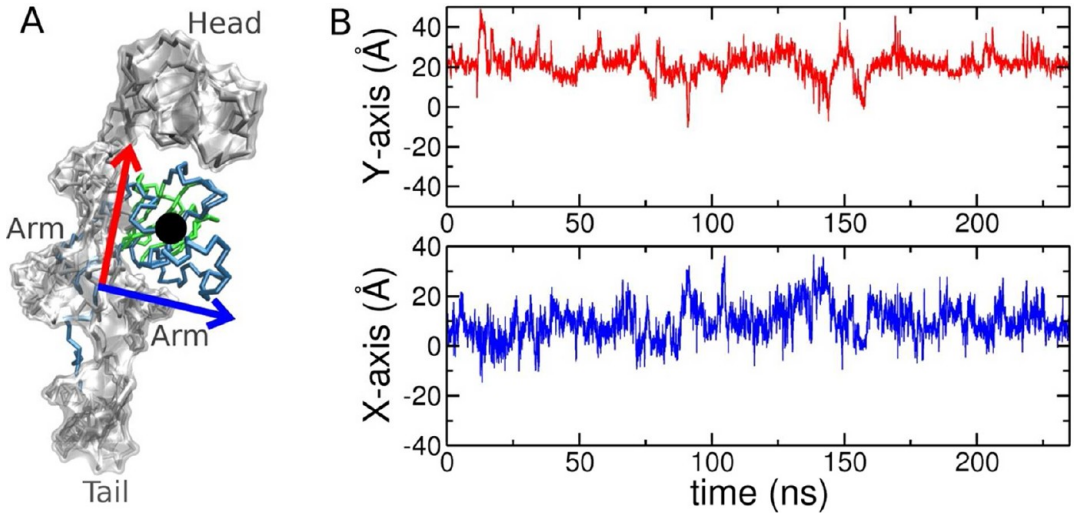


Figure 6. Trigger factor increases the effective length of the exit tunnel in a subset of biosynthesis trajectories of the N-terminal domain of β-galactosidase. (A) The coordinates of the center-of-mass of the folded portion of the N-terminal domain of β-galactosidase (black dot) were monitored in a local coordinate system whose origin is located between the arms of TF, and whose positive x-axis (blue arrow) and positive y-axis (red arrow), respectively, point to an arm and the head of TF. (B) A typical center-of-mass trajectory for those conformations populating basin 3 in Figure 5C. Note that the center-of-mass does not sample negative values along the y-axis, demonstrating that when folding begins outside the cradle, away from the exit vestibule, the folded nascent segment is unable to reposition itself back under the cradle near the exit vestibule. This situation increases the effective length of the exit tunnel by 30 ± 1 residues (see main text) as more unstructured nascent chain residues are needed to stretch from the exit vestibule to the location where tertiary folding occurs outside the cradle.

domain ($R_g/R_{g,C}$ = 1.05, Q = 0.74) whose center-of-mass is located, on average, under the TF cradle, (2) a partially folded C-terminal structure ($R_g/R_{g,C}$ = 1.21, Q = 0.48) whose center-of-mass protrudes from the TF cradle, (3) a partially folded N-terminal structure ($R_g/R_{g,C}$ = 1.20, Q = 0.48) whose center-of-mass protrudes out of the TF cradle to the right of its body and above its arms, and (4) to a heterogeneous unstructured ensemble ($R_g/R_{g,C}$ = 1.34, Q = 0.33) whose conformations sample widely throughout the TF cradle and therefore whose average center-of-mass location is not meaningful.

These structural ensembles suggest two mechanisms by which TF can slow down co-translational folding. For the structures in basin 3 (Figure 5D) we observe that folding of the domain outside the TF cradle leads to an unstructured segment of the nascent chain that must stretch back through the TF cradle to the exit tunnel opening (Supplementary Figure S1). This finding suggests that one mechanism by which TF could slow co-translational folding is to increase the effective exit tunnel length, which due to the small dimensions of the tunnel tends to inhibit tertiary structure folding. Another mechanism is suggested by the unfolded conformations in basins 2 and 4, which are for the most part located under the cradle formed by TF and yet are unable to fold during the first 1400 ns of the simulations (Figure 4B). These results suggest that the close proximity of TF and the nascent chain at the exit tunnel opening could potentially decrease the rates of structural rearrangements within the nascent chain, which are crucial in the biased stochastic search for the native state.[32] In what follows, we test these two hypotheses and find strong support for both of them.

Mechanism 1: Trigger Factor Slows Co-translational Folding Predominantly by Decreasing the Rate of Structural Rearrangements within a Nascent Chain. If TF decreases the rate of structural rearrangements within the unfolded state as the result of the high local density of these two components, then starting simulations in the presence and absence of TF from the same structural ensemble of unfolded conformations should alter the rate of formation of tertiary nascent chain contacts. To test this prediction, we took the final conformation from each of the 89 RNC trajectories that populate basins 2 and 4 (Figure 5C) at a nascent chain length of 247 residues and carried out further simulations for 150 ns in the presence and absence of TF (see Methods). We observed that residues 1−50, which are located at the N-terminus of the domain, were unfolded in 87% of these starting structures. Therefore, we monitored in the simulations the time that it takes to form a tertiary contact between residue 4 and residue 180, residues that are in contact with each other in the native structure.[33] We observed that it took three times longer to form such a contact in the presence of TF than in its absence. More specifically we found that the average time was 79 ns versus 26 ns in these low viscosity simulations (see Methods), which in the high viscosity found in experiments would correspond approximately to 53 ms versus 17 ms (see Discussion). Thus, our simulations indicate that TF significantly slows the rate at which tertiary contact formation occurs in the unfolded state. As 88% of the trajectories populate basins 2 and 4 within the unfolded ensemble at a chain length of 247 residues, we can conclude that this decreased rate of structural rearrangements is the predominant mechanism by which TF delays co-translational folding of this domain. The extent of this TF-induced delay is likely to depend on the nascent chain segment being monitored, with those involving longer range contacts likely to experience even greater delays in folding due to TF.

Mechanism 2: Domain Folding Outside the Cradle Formed by TF Slows Folding through an Increase in the Effective Length of the Exit Tunnel. Folding outside the TF cradle requires that an additional number of unstructured nascent chain residues stretch from the exit vestibule to the folded segment (Supplementary Figure S1), which would delay folding due to the additional time interval for their synthesis. However, for this potential mechanism to cause any significant

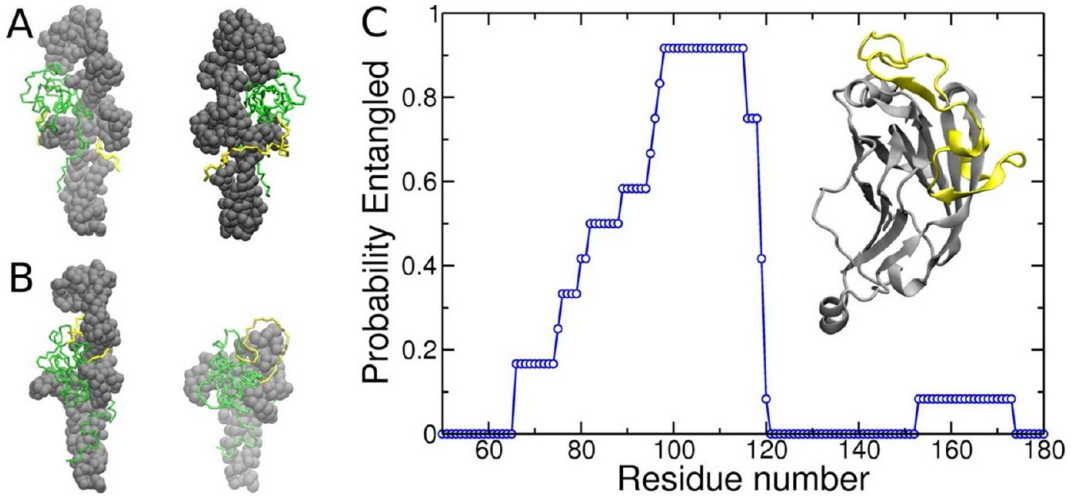


Figure 7. Entanglement between the N-terminal domain of β-galactosidase and trigger factor can occur during protein biosynthesis. (A) Illustration of a weakly entangled structure at a nascent chain length of 247 residues. These two structures show the same nascent chain−TF (gray) configuration from two different angles. The nascent chain is seen to form a tertiary structure outside the cradle (green), with the N-terminal portion of the nascent chain (yellow) wrapping around TF and back under the cradle on the other side. Few intranascent chain contacts are made between the segments separated by this entangled loop, making it relatively easy for thermal fluctuations to disentangle this configuration. (B) Illustration of a strongly entangled structure. The nascent chain is seen here to form a tertiary structure between the arms of TF, with a portion of the nascent chain (yellow) located closer to the C-terminus wrapped around TF between its body and head regions. This situation is more clearly seen in the right-hand structure where the head domain has been removed for visualization purposes. A large number of intranascent chain contacts are made between segments separated by this entangled loop, making it difficult to disentangle the structure. (C) Strong entanglement almost always involves residues between 65 and 120 in the nascent chain. This finding is illustrated in the plot of the probability that a residue along the nascent chain sequence is entangled with (i.e., wrapped around) TF given that strong entanglement has occurred. The residues that can strongly entangle correspond to a largely unstructured loop segment that is located on the surface of the folded N-terminal domain of β-galactosidase (colored in yellow in the structural inset).

delay in co-translational folding, nascent chain segments that begin to fold outside the cradle (Figure 5D) should not be capable of shifting their position back under the cradle. Such repositioning, if it occurred rapidly, would obviate the need for the additional unstructured residues. To test if this is the case, we monitored the center-of-mass location of the folded nascent chain portions outside the cradle in the trajectories that populate basin 3. The center-of-mass location was calculated in a local coordinate system whose origin is located between the arms of TF (Figure 1A), and whose positive x- and y-axes point to one arm and the head of TF, respectively (Figure 6A). In this coordinate system, the folded portion of the nascent chain will have repositioned itself if its center-of-mass switches to negative values along the y-axis. We find that 10 out of the 11 trajectories that initially populate basin 3 show no such effect and so remain in that configuration throughout the entire time of the simulation, with no repositioning of their folded segments. A typical time trace of the center-of-mass from one of these trajectories is shown in Figure 6B. Thus, TF also delays co-translational folding by increasing the effective exit tunnel length. However, basin 3 comprises only 12% of the unfolded population at this nascent chain length, so this mechanism represents a secondary contribution to the slowing of co-translational folding.

How much delay in complete domain folding is associated with a co-translational folding process that begins outside the cradle? To address this question we have analyzed the structures populating basin 3 and calculated the average number of residues in the nascent chain that stretch from the exit tunnel vestibule to the residue where tertiary folding begins outside of the cradle. We find that an additional 30 ± 1 nascent chain residues are needed to allow folding to occur outside the cradle. As a minimum of 24 residues is needed to stretch from the PTC to the exit vestibule where tertiary folding can occur,[9,34] this TF extension of the tunnel is therefore more than doubling its effective spatial length. *E. coli* translates proteins at rates of 10−20 residues per second,[35] indicating that this mechanism is likely to be able to delay folding *in vivo* by a time of 1.5 to 3 s.

Mechanism 3: Entanglement of the Domain with Trigger Factor Leads to Long-Lived Kinetic Traps. One of the limitations of a PCA analysis is that the results depend on the observables monitored; therefore it is possible that our clustering procedure may have missed other structural populations that could suggest additional contributions to the TF kinetic trapping mechanism. We hypothesized that a potential third delaying mechanism is entanglement of the nascent chain with the TF molecule. It is indeed known that at high enough concentrations polymers can intertwine and wrap around one another, dramatically slowing down the time scale at which their conformations are able to rearrange.[36] Entanglement becomes more likely as the polymers become longer or their persistence length decreases. By analogy, we reasoned that the high local concentration of TF and the unfolded β-galactosidase nascent chain may lead to such entanglement between these two components.

To test this hypothesis, we examined the final RNC-TF structure of each synthesis trajectory at a nascent chain length of 247 residues. We find that 11% of the structures (22 in total) exhibit entanglement, with at least one segment of the nascent chain wrapping completely around a portion of the TF structure and coming back into contact with another segment of the nascent chain (Figure 7). Of these entangled structures, 45% exhibit what we can classify as "weak" entanglement

(Figure 7A), and 55% exhibit "strong" entanglement (Figure 7B). Weakly entangled structures involve only the N-terminal segment of the nascent chain being wrapped around TF and then back under the cradle to form a small number of contacts (less than 5) with other segments of the nascent chain. Such structures are likely to be rapidly disentangled (i.e., be unwrapped from around TF) as a result of thermal fluctuations. On the other hand, strongly entangled structures involve more centrally located nascent chain segments that are wrapped around the TF structure and make a larger number of contacts with other nascent chain segments under the cradle. Such structures are likely to take much longer to disentangle and then to fold. To test these hypotheses, we simulated these weakly and strongly entangled structures for a time period of 150 ns (corresponding to approximately 100 ms at the higher viscosities found in experiment) and calculated the fraction of trajectories that disentangle during that time. We find that *all* of the weakly entangled structures disentangle at some point during the simulation, whereas *none* of the strongly entangled structures end up disentangling. Thus, the strongly entangled structures we have identified lead to very long-lived kinetic traps.

We further analyzed the structural properties of the strongly entangled structures and found that they have a high proportion of native-like interactions, with an average Q value of 0.63. To understand how this β-galactosidase domain can form substantial native structure while simultaneously forming an entangled structure with TF, we calculated the probability that a given residue along the nascent chain sequence is involved in strong entanglement. We find that all such cases, with one exception, involve residue numbers 65−120 along the sequence of β-galactosidase (Figure 7C). These residues, when highlighted on the native structure (Figure 7C, inset), can be seen to comprise a surface loop. Thus, this loop can wrap around TF while allowing a significant proportion of native-like structure to form. The favorable interactions between the nascent chain segments separated by this entangled loop create a significant free energy barrier to disentanglement, giving rise to the very slow total folding times for these structures. Thus, a third mechanism by which TF can slow co-translational folding arises from strong entanglement, which in the present study is observed in 6% of all synthesis trajectories. On the basis of polymer theory, we can generalize that such entanglement is increasingly likely as the size of the domain increases.[36]

Trigger Factor Sterically Occludes Other Cytosolic Components from Interacting with Nascent Chains, Thereby Providing Them with Protection from Degradation and Aggregation. It is often assumed that co-translational folding is biologically advantageous as it can promote folding and minimize misfolding of proteins with complex native topologies.[5] It is therefore important to ask what could be the biological benefit of a TF-induced slowing down of co-translational folding. One hypothesis is that TF inhibits aggregation of denaturated or partially folded nascent chains by shielding them from interactions with other proteins. This hypothesis is supported by the experimental observation that TF suppresses the aggregation of proteins when the folding process is initiated from a chemically denatured state.[18,23] Another hypothesis is that TF inhibits the untimely degradation of nascent chains, an idea supported by the protection afforded to some nascent chains from proteinase K digestion.[21] An implication of both of these hypotheses is that TF sterically occludes other cytosolic components from the vicinity of the nascent chain. To examine whether this steric occlusion mechanism occurs in our simulations, we calculated the average accessible surface area of β-galactosidase as it is synthesized by the ribosome. Instead of using a probe radius of 1.4 Å, which is commonly used to calculate the solvent-accessible surface of a protein, we used a probe radius of 10 Å. This larger probe radius better represents the steric factors involved when a large protease or an unfolded or partially folded protein molecule comes into proximity with the nascent chain (see Methods for details). We find that the presence of TF decreases the accessible surface area of the nascent chain to such cytosolic components by more than 50% (Figure 8), indicating that

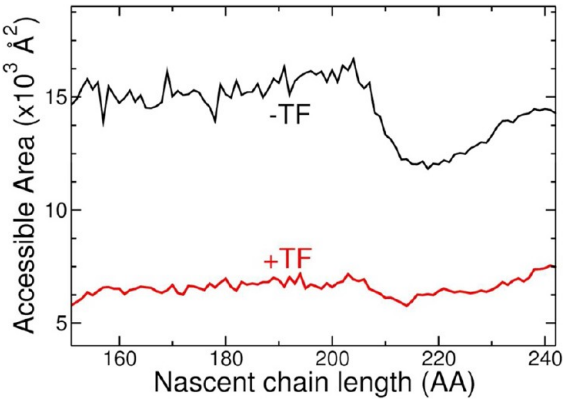


Figure 8. Trigger factor sequesters the nascent chain and reduces its interactions with other cytosolic components. The accessible surface area of the nascent chain of the N-terminal domain of β-galactosidase is shown in the presence (black) and absence (red) of TF at different lengths during its synthesis on the ribosome. TF reduces the accessible nascent chain surface area by more than 50%. A probe radius of 10 Å was used to compute the accessible area as this was considered to be a reasonable representation of the radius of a typical proteolytic enzyme that could potentially degrade the nascent chain and also the pervaded blob size of an unfolded protein that could potentially interact with the nascent chain (see Methods).

proteases and unfolded proteins in the cytoplasm of *E. coli* are much less likely to come into contact with the nascent chain when TF is present. Thus, our simulations support these hypotheses and indicate that TF suppresses aggregation and degradation of nascent chains at the molecular level through steric occlusion. We note also that at β-galactosidase nascent chain lengths between 200 and 220 residues there is a decrease in the total accessible area of the nascent chain (Figure 8) as a result of a collapse transition in the unfolded domain, a phenomenon we have observed previously for other proteins.[8]

■ DISCUSSION

By simulating the biosynthesis of protein G and the N-terminal domain of β-galactosidase in the presence and absence of TF we have provided molecular insights into the impact of this key chaperone on co-translational folding of protein domains of very different sizes. Our finding that protein G can be accommodated fully and also fold under the TF cradle is consistent with results from cross-linking and cryo-EM experiments of another small domain.[14] The use of molecular simulations in this study indicate that the co-translational folding properties of this small protein are not affected by the presence of TF. For the β-galactosidase domain, however, we observe that TF slows down co-translational folding as a result

of the generation of kinetic traps. These results are consistent with the experimental observation that TF slows the rate of co-translational folding of full-length β-galactosidase[19] but also provides an explanation for the molecular origin of this behavior. On the basis of our findings we suggest that such TF activity may be generic for large proteins, as other proteins have also been shown to exhibit decreased rates of refolding at certain TF concentration ranges.[18,22,37] Furthermore, the predominant mechanism of kinetic trapping that we observe results from a general decrease in the rates of nascent chain structural rearrangements. This mechanism appears likely to be insensitive to the particular details of a nascent chain, although its specific amino acid sequence may modulate its behavior to some extent. Far more important to this mechanism, we believe, is the size of the domain, as this controls the volume of the TF cradle the domain occupies. Larger domains will have less free space within the cradle, and thus should exhibit a greater decrease in their ability to undergo conformational rearrangements.

On the technical side, we note that the time scales of folding and translation in our simulations are several orders of magnitude smaller than those of the analogous events taking place in the cell because of the low-viscosity Langevin dynamics that we utilize as well as the coarse-grained representation of the RNC; the latter smoothes the energy landscape. This difference in time scales allows us to obtain statistically significant results by enhancing the rate of conformational interconversion; however, it does not affect the equilibrium properties calculated (e.g., the equilibrium constant) as they depend on the ratio of the time scales rather than their absolute magnitude. Similar reasoning applies to the continuous translation process that we have modeled; in this case, co-translational folding is an out-of-equilibrium phenomenon but is governed by the ratio of time scales within the system. The ratio of the measured protein G folding time in free solution[38] to the time scale of amino acid addition in *E. coli*[39] is 0.048 (i.e., 2.4 ms/50 ms). Therefore, to keep this same ratio in the coarse-grained simulations, we added an amino acid to the growing nascent chain approximately every 75 ns (= 3.6 ns/0.048) of simulation time, where 3.6 ns is the folding time calculated for the coarse-grained representation of protein G[8] (see Methods). For these reasons we expect our results to be relevant to the longer time scale of these processes in the cell.

By analyzing the simulations to determine the structures sampled in the unfolded ensemble we find that TF slows the co-translational folding of the β-galactosidase domain through kinetic trapping of the unfolded state. This effect arises from three distinct molecular mechanisms: (1) decreased rates of structural rearrangements in the nascent chain unfolded ensemble when TF is present, which in the present study occurs in 88% percent of the trajectories that remain unfolded (basins 2 and 4, Figure 5B); (2) folding of the domain outside the cradle formed by TF, which increases the effective tunnel length by 30 ± 1 residues and occurs in 12% of the trajectories that remain unfolded (basin 3); and (3) strong entanglement of the nascent chain with TF that inhibits the transition from the unfolded to the folded state and occurs in 6% of all the trajectories. Thus, a decrease in the rate of structural rearrangements is the primary mechanism of the delaying activity of TF, with lesser contributions from the other two mechanisms. We note that these mechanisms need not be mutually exclusive, and indeed their probabilities in the present study (88%, 12%, 6%) add up to more than 100% for this reason. For example, we observe mechanisms 2 and 3 to occur concomitantly in some of the trajectories. In addition, the probability of a given mechanism may differ from one system to another.

It has been suggested that TF delays protein folding by binding preferentially to hydrophobic segments in unfolded or intermediate states of the nascent chain,[40] thereby thermodynamically stabilizing the unfolded state. The extent of such destabilization may, however, not be very significant. If TF binds only to unfolded nascent chains, and the unbound TF concentration is unchanging, then according to a thermodynamic cycle the maximum that TF can destabilize the native state by ($\Delta\Delta G_{NU}$) is equal to the difference in the binding free energy of TF for a translating and non-translating ribosome ($\Delta\Delta G_{TF}$). Therefore, $\Delta\Delta G_{NU} = \Delta\Delta G_{TF} = -k_B T \log[K_{D,NT}/K_{D,AT}]$ where k_B is Boltzmann's constant, T is the temperature, $K_{D,NT}$ is the dissociation constant of TF from a non-translating ribosome, and $K_{D,AT}$ is the dissociation constant of TF from a translating ribosome. $K_{D,NT}$ has a value of 2.1 μM and $K_{D,AT}$ can be as small as 53 μM for some nascent chains.[12,26] Therefore, by this estimate, TF cannot destabilize the native state of a domain by more than 1.9 kcal/mol. This estimate is consistent with the observation that TF is only able to shift a marginally stable protein to a thermodynamically stable unfolded state,[22] while proteins with greater stability remain folded in the presence of TF. Most protein domains acquire native state stabilities of greater than 2 kcal/mol in magnitude under standard conditions.[38] Thus, this maximum destabilization by TF is not likely to contribute significantly to the delaying action of TF, a conclusion consistent with the predominance of the kinetic trapping mechanism that we observe in this study.

Strong entanglement of the nascent chain with TF leads to long-lived kinetic traps in the unfolded ensemble that have the potential of keeping the domain and the associated TF bound together even after the latter has dissociated from the ribosome. This conclusion provides a molecular mechanism for two different experimental observations. The first is the slow kinetic phase ($t_{1/2}$ = 35 s) that has been observed in FRET measurements of a RNC-TF system and was interpreted as corresponding to the unbinding of TF from the nascent chain after the former had dissociated from the ribosome.[40] We find that strong entanglement involves looping of the nascent chain around the body of TF. This loop can form anywhere along the TF structure, with a slightly greater preference for it to be located between the PPIase head domain and its arms (Figure 7). This finding suggests that the rate of disentanglement of such structures and consequently their rate of unbinding from TF would increase if the PPIase domain were deleted from the TF structure (Figure 7B). Consistent with this prediction, experimental results utilizing a PPIase truncated TF structure reveal a decrease in the mean time that TF remains bound to the nascent chain ($t_{1/2}$ = 23 s) after dissociation from the ribosome.[40] Because this strong entanglement mechanism tends to keep larger domains in an unfolded conformation for longer periods of time, its presence also suggests why eukaryotic proteins, which are larger on average than bacterial proteins, exhibit increased folding yields when expressed in *E. coli* cells that contain TF mutants in which the head domain has been truncated.[41] In this view the eukaryotic proteins are expected to be more likely to become entangled due to their larger size, and removal of the TF head domain allows disentanglement and the folding of such proteins to occur more quickly in *E. coli*.

Strong entanglement was found to involve a loop region on the surface of the folded β-galactosidase domain. Such entanglement allows the domain to acquire a significant degree of native structure while remaining entangled with TF. This finding suggests that if this loop was shortened through protein engineering, then its entanglement probability would decrease. More generally, we expect the weights given to these different kinetic trapping mechanisms will be context dependent. For example, the entanglement with TF is more likely for large folding domains as compared to small ones, or for proteins with shorter persistence lengths.[36] Moreover, slower structural rearrangements may well occur for larger nascent chains with greater hydrophobic content.

While protein entanglement has been observed in the form of knotted structures in individual protein molecules,[42] it has not to our knowledge been discussed in the biologically important context of molecular chaperones, which by their nature involve the interaction of two or more proteins. These results, however, indicate that entanglement could play an important and novel role in the biogenesis of proteins in cells. As a consequence of entanglement, portions of the nascent chain can come into contact with the "exterior" of TF, i.e., outside the cradle of TF. For example, in our simulations, residue 9 that is on the exterior of TF comes into contact with unfolded protein G (Figure 2D), suggesting that nascent chain cross-linking to this site could be possible. Indeed, cross-linking of the nascent chain to TF residue 9 has been observed experimentally,[14] although it was considered not to be very significant by the authors of the study because the corresponding band in the gel was not as intense as the bands associated with the other cross-linking sites on the interior cradle of TF. Such a decrease in intensity of this cross-linking site could, however, be attributed to the relatively small population of entangled structures (12%) observed in these simulations. Thus, these experimental cross-linking data are consistent with the possibility that entanglement can occur between a nascent chain and TF for a subpopulation of the proteins that are being synthesized.

Our results also shed light on the structural question of whether nascent chains of large domains accumulate under the TF cradle or stretch out from under it.[14] We have found that in the least structured unfolded population of the β-galactosidase domain the nascent chain segments do indeed sample configurations outside the cradle (see Figure 5D, and the conformations in basin 4). On average, however, most of the nascent chain segments are located under the TF cradle. Thus, it appears that large domains tend to accumulate under the cradle, although there are significant statistical variations from this average behavior. This statistical perspective emphasizes the point that while nascent chains may cross-link with different residues of TF at different nascent chain lengths,[14] this does not mean at the molecular level that the nascent chain takes a narrowly defined spatial route through the cradle as it is synthesized.

β-Galactosidase consists of five domains. Its enzymatic activity arises through the formation of a homotetramer in which there are significant protein−protein interfaces involving domains 3 and 5, with the active site forming upon a dimerization step that involves domain 3.[43] The concomitant appearance during translation of full-length β-galactosidase protein and active enzyme indicates that in the absence of TF co-translational folding is rate-limiting in the self-assembly process of the active tetramer.[19] Although in this study we have examined only the impact of TF on the first domain of β-galactosidase, our results have implications for these downstream processes. It is possible that each domain will tend to interact with a different TF molecule, while TF undergoes its cycles of binding and unbinding to the ribosome nascent chain complex.[40] Therefore, the mechanisms of kinetic trapping that we observe for the N-terminal domain can also affect the behavior of the other domains. In this case, the TF-induced delay in folding of each domain may be sufficient to switch the rate-limiting step to one involving post-translational folding or tetramer assembly. The entanglement phenomenon that we observe, for example, leads to long-lived states in which TF remains bound to the domain even though the latter may exhibit substantial elements of native structure. Therefore, these TF-bound molecules have the potential to inhibit tetramerization by blocking the energetically favorable interfaces that can appear upon complex formation.

In conclusion, this study illustrates the utility of coarse-grained simulations in gaining molecular insights into the *in vivo* effects of protein folding. By incorporating experimental data into such simulations, in this case K_D values, and checking for consistency with other experimental data, this approach can play an important role in identifying the molecular processes underlying experimental observations and suggest novel mechanisms by which chaperones such as TF can effect co-translational folding *in vivo*.

METHODS

We describe here the methods that have been used in the simulations and analysis. Further details on the force field and Monte Carlo simulations used to calculate $\lambda_{TF\text{-}R}$ and $\lambda_{TF\text{-}NC}$ can be found in the Supporting Information.

Simulation Details of Production Runs. In the production simulations the ribosome was treated as a rigid unit, while the nascent chain and TF molecule were not restrained. Treating the ribosome as a rigid body rather than considering its dynamics in detail has been shown to have a negligible effect on the calculated co-translational folding properties.[8,9] Holding the ribosome rigid, however, allows the RNC to be simulated for orders-of-magnitude longer in simulation time, allowing statistically significant results to be obtained.

CHARMM version c35b5[44] was used to simulate all of the systems detailed here. Langevin dynamics with a collision frequency of 0.05 ps^{-1} and 15 fs integration time-step were used in all simulations. Initial configurations of protein G and β-galactosidase RNC systems were generated as described previously.[8] TF was initially placed in its aligned configuration on the ribosome, and during the simulations it was not observed to undergo dissociation, consistent with the experimental observation of TF-bound half-lives of up to 50 s.[12] Replica exchange simulations were used as the basis of the calculation of thermodynamic properties of protein G on stalled ribosomes. Typically, 16 temperature windows were used with temperatures ranging from 250 to 380 K, and swap attempts were made every 700 integration time-steps between neighboring temperatures. A total of 130,000 replica exchanges were attempted, with the first 40,000 discarded to allow for equilibration. Acceptance ratios in the range of 0.17 and 0.87 were achieved, with each replica sampling the extremes of the temperature range several times.

A total of 205 independent continuous translation simulations of β-galactosidase were carried out at 310 K in the presence and absence of TF. The simulations were initiated from a nascent chain length of 143 residues. Each trajectory was equilibrated at this chain length for 75 ns (5×10^6 integration steps). No stable tertiary structure forms in this case at this chain length because with a native state contact order of 30 residues the nascent chain segment outside the tunnel cannot make a sufficient number of native contacts to form a stable structure. Continuous translation was then initiated with a new amino acid added to the C-terminus every 7.5 ns (5×10^5 integration steps) up to

a nascent chain length of 242 residues; the procedure for covalently attaching new residues in CHARMM has been described previously.[8] In free solution, we find that protein G folds with a mean time of 3.6 ns due to the low viscosity used in the simulations, while experimentally its has been observed to fold with a mean time of 2.4 ms; therefore, 1 ns of simulation time corresponds to 0.67 ms of experimental time in aqueous solution. Thus, adding a new amino acid every 7.5 ns in the simulations corresponds to an experimental translation time of 5 ms per residues. This translation time, which is faster than the average experimental time of between 20 and 50 ms in *E. coli*,[39] was used for two reasons. First, for a domain to fold on the ribosome, a minimum of 24−30 residues must be added to the C-terminus of the domain. Thus, up to and including a nascent chain length of 242 residues, the N-terminal domain of β-galactosidase, which is 216 residues long, will not completely fold, and therefore this fast translation rate of unfolded β-galactosidase segments is unlikely to affect the overall results. Second, simulating the elongation of a β-galactosidase segment by an additional 100 residues at realistic translation rates would have taken approximately 2 years per trajectory on the supercomputers that we use, an unrealistic time-scale over which to carry out this study.

At nascent chain lengths between 243 and 247 residues, where the domain is sterically allowed to fold fully, we used the estimated time-scale of amino acid addition based on the mRNA sequence of β-galactosidase in *E. coli*.[45] The codons in this range are GUA, ACA, GUU, UCU, and UUA, and have average amino acid addition times of 73, 178, 26, 55, and 157 ms in an *E. coli* that divides every 150 min at 310 K. Mapping of the time-scale between these experimental translation rates and that which is appropriate in these low viscosity simulations results in amino acid addition every 109, 267, 39, 82.5, and 235.5 ns of simulation time, respectively.

A total of 150 temperature quench simulations were carried out at each nascent chain length of the stalled protein G-RNCs by equilibrating the system at 360 K, which is above the midpoint of unfolding of protein G in free solution, for 10^6 integration time-steps, and then quenching the system to 310 K and recording its first-passage time to the native basin, the latter being defined as having a Q value of greater than 0.65.

Analysis. Thermodynamic properties were calculated from replica exchange simulations using the histogram-free formulation of the WHAM equations.[46] In the simulations of TF and α-lactalbumin a TF molecule was considered to be bound to reduced α-lactalbumin if there was more than one interprotein contact between these two components in a given simulation configuration, otherwise TF was considered unbound. The calculated fraction of TF bound to α-lactalbumin (Figure 1B) was found to be insensitive to the definition of a bound state up to a threshold of 30 interprotein contacts. In the rigid-body Monte Carlo simulations a TF molecule was considered to be bound to its ribosomal binding site in a simulation configuration if 80% (38 out of 48) or more of the crystal contacts between these two components (Supplementary Table S1) were present; this criterion was used to calculate the fraction of TF bound to the ribosome at 293 K (Figure 1C). A cross-linking site on TF was considered to be in contact with a cross-linking site on the nascent chain if any of the amine carrying residues in the protein G-RNC was within 15 Å of it. The amine-carrying residues, which are the only groups that can cross-link with the TF cross-linking agent BPA,[14] of protein G are residues 1, 4, 10, 13, 28, 31, and 50. The resulting time-series was used to calculate the probability of contact between cross-linking groups at 310 K (Figure 2D).

The program PyMol was used to align the protein G-RNC simulation structure with the EM map from the SH3-RNC-TF data. A surface contoured at 1σ was used in the visual representation (Figure 2E).

The stability of the native state with respect to the unfolded state was calculated as $\Delta G_{NU} = -k_B T \log[P_N/P_U]$, where P_N and P_U are, respectively, the probability of being in the native and unfolded ensembles. Conformations observed in the protein G and β-galactosidase simulations were classified as being in the native ensemble if more that 65% of their native contacts were formed (Q = 0.65), and were classified as being unfolded otherwise. This definition was then used to calculate P_N and P_U and utilized in Figures 3A,D and 4B. The mean-first-passage-time $<\tau_F>$ was calculated as $N^{-1}\Sigma_{i=1}^{N}\tau_{F,i}$, where N is the number of independent temperature-quench trajectories, and $\tau_{F,i}$ is the simulation time taken to obtain a Q value of 0.65. The transition pathways generated from the temperature quench simulations were analyzed as described previously.[8]

Principal components of the β-galactosidase structural ensembles were calculated in the following manner. The time-series data of Q_i, the fraction of native contacts per β-strand, were calculated from the last 116 ns of simulation conformations at a nascent chain length of 247 residues in the presence and absence of TF. This procedure resulted in a total of 394,625 time points from all of the independent synthesis trajectories under these two conditions. As identified in the crystal structure by the program STRIDE,[47] there are 24 β-strands in this domain structure; therefore each time point has 24 values. The time series were concatenated into two matrices, corresponding to data in the absence and presence of TF, each of which had dimensions of 394,625 × 24. PCA was applied separately to these matrices, and the eigenvectors corresponding to the two largest Eigenvalues were defined as PC 1 and PC 2 in the absence of TF, and PC 1′ and PC 2′ in the presence of TF. The time-series data were then projected onto these two principal components. The free energy surface along these two components was then calculated from these new data as $F(\text{PC 1, PC 2}) = -k_B T \log[P(\text{PC 1, PC 2})]$, where $P(\text{PC 1, PC 2})$ is the probability of finding a conformation at point in the range of PC 1 $\pm$ δ, PC 2 $\pm$ δ where δ = 0.1.

Representative structures of the conformations of the N-terminal domain of β-galactosidase in each free energy basin were determined by first identifying the centroids of those basins in the free energy surface and then determining those conformations that were within δ = 0.0075 of that centroid in the time-series data. Finally, the representative ensembles shown in Figures 5B,D were constructed by performing an rmsd least-squares alignment of each centroid's structures against the three β-strands with the largest average fraction of native contacts within that cluster. The rmsd alignment was performed using the rmsd Trajectory Tool in VMD version 1.8.7.[48]

In basin 3 of Figure 5C the center-of-mass of the folded portion of the nascent chain that resided outside the TF cradle (typically residues 1−186) was projected onto a local coordinate system (Figure 6). This local coordinate system was constructed at each analysis time point by defining its origin as the average of the Cartesian coordinates of TF residues 325 and 388, which are on the tip of each TF arm. Two vectors were calculated by subtracting the coordinates of TF residue 325 and 344 from this origin. These two vectors point, respectively, from the origin to one arm of TF and from the origin to the body of TF, i.e., away from the exit tunnel, but are not necessarily orthogonal, and so the cross-product was calculated. The resulting vector starts at the origin and points toward the head domain of TF and away from its tail, and these two vectors defined the local two-dimensional coordinate system. The center-of-mass time series was projected onto this new coordinate system, and the results that were obtained are shown in Figure 6.

The accessible surface area of the nascent chain was computed using the Correl module of the molecular modeling program CHARMM and a probe radius of 10 Å. This probe radius was chosen as we are interested in the accessibility of the nascent chain to other macromolecules, such as proteases and unstructured proteins, which could potentially aggregate with the nascent chain. The probe radius representing an unstructured protein is best equated with its thermal blob size. In polymer physics a thermal blob is the length scale above that excluded volume interactions dominate over thermal energy. The radius of a thermal blob in an athermal solvent is approximately $((2l_p^2)/(|dl|))$,[36] where l_p is the persistence length of the protein, and has a value of around 5 Å, and d = 5 Å is the diameter of the protein if it is represented as a tube centered along its backbone. In this case, the thermal blob has a radius of 10 Å.

Standard errors about the mean were calculated by breaking up the data into five or more data sets and computing the average for each of them.

■ ASSOCIATED CONTENT

Supporting Information

Simulation and analysis methods. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

cmd44@cam.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.P.O. thanks Guenter Kramer, Anja Hoffmann, Annemarie Becker, and Bernd Bukau for stimulating discussions about trigger factor and its possible functions. This work was supported by an NSF postdoctoral grant (E.P.O.), the BBSRC and the Wellcome Trust (M.V. and C.M.D.), and the EPSRC (E.P.O., M.V., and C.M.D.). This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

■ REFERENCES

(1) Taniguchi, Y.; Choi, P. J.; Li, G. W.; Chen, H. Y.; Babu, M.; Hearn, J.; Emili, A.; Xie, X. S. *Science* **2010**, *329*, 533.

(2) Maier, T.; Schmidt, A.; Guell, M.; Kuhner, S.; Gavin, A. C.; Aebersold, R.; Serrano, L. *Mol. Syst. Biol.* **2011**, *7*, 511.

(3) Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. *Nature* **2011**, *473*, 337.

(4) Nicola, A. V.; Chen, W.; Helenius, A. *Nat. Cell Biol.* **1999**, *1*, 341.

(5) Clark, P. L.; King, J. *J. Biol. Chem.* **2001**, *276*, 25411.

(6) Tartaglia, G. G.; Pawar, A. P.; Campioni, S.; Dobson, C. M.; Chiti, F.; Vendruscolo, M. *J. Mol. Biol.* **2008**, *380*, 425.

(7) Apic, G.; Gough, J.; Teichmann, S. A. *J. Mol. Biol.* **2001**, *310*, 311.

(8) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2011**, *133*, 513.

(9) O'Brien, E. P.; Hsu, S. T. D.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 16928.

(10) Elcock, A. H. *PLoS Comput. Biol.* **2006**, *2*, 824.

(11) Wang, P. Y.; Klimov, D. K. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 925.

(12) Hoffmann, A.; Bukau, B.; Kramer, G. *Biochim. Biophys. Acta, Mol. Cell Res.* **2010**, *1803*, 650.

(13) Raine, A.; Lovmar, M.; Wikberg, J.; Ehrenberg, M. *J. Biol. Chem.* **2006**, *281*, 28033.

(14) Merz, F.; Boehringer, D.; Schaffitzel, C.; Preissler, S.; Hoffmann, A.; Maier, T.; Rutkowska, A.; Lozza, J.; Ban, N.; Deuerling, E.; Bukau, B. *EMBO J.* **2008**, *27*, 1622.

(15) Dong, H.; Nilsson, L.; Kurland, C. G. *J. Mol. Biol.* **1996**, *260*, 649.

(16) Kramer, G.; Patzelt, H.; Rauch, T.; Kurz, T. A.; Vorderwulbecke, S.; Bukau, B.; Deuerling, E. *J. Biol. Chem.* **2004**, *279*, 14165.

(17) Merz, F.; Hoffmann, A.; Rutkowska, A.; Zachmann-Brand, B.; Bukau, B.; Deuerling, E. *J. Biol. Chem.* **2006**, *281*, 31963.

(18) Huang, G. C.; Li, Z. Y.; Zhou, J. M.; Fischer, G. *Protein Sci.* **2000**, *9*, 1254.

(19) Agashe, V. R.; Guha, S.; Chang, H. C.; Genevaux, P.; Hayer-Hartl, M.; Stemp, M.; Georgopoulos, C.; Hartl, F. U.; Barral, J. M. *Cell* **2004**, *117*, 199.

(20) Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. *Nature* **2011**, *475*, 324.

(21) Hoffmann, A.; Merz, F.; Rutkowska, A.; Zachmann-Brand, B.; Deuerling, E.; Bukau, B. *J. Biol. Chem.* **2006**, *281*, 6539.

(22) Maier, R.; Scholz, C.; Schmid, F. X. *J. Mol. Biol.* **2001**, *314*, 1181.

(23) Zhou, J. M.; Xie, J. B. *Biochemistry* **2008**, *47*, 348.

(24) Suno, R.; Taguchi, H.; Masui, R.; Odaka, M.; Yoshida, M. *J. Biol. Chem.* **2004**, *279*, 6380.

(25) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13403.

(26) Kim, Y. C.; Hummer, G. *J. Mol. Biol.* **2008**, *375*, 1416.

(27) Rutkowska, A.; Mayer, M. P.; Hoffmann, A.; Merz, F.; Zachmann-Brand, B.; Schaffitzel, C.; Ban, N.; Deuerling, E.; Bukau, B. *J. Biol. Chem.* **2008**, *283*, 4124.

(28) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.

(29) Oh, E.; Becker, A. H.; Sandikci, A.; Huber, D.; Chaba, R.; Gloge, F.; Nichols, R. J.; Typas, A.; Gross, C. A.; Kramer, G.; Weissman, J. S.; Bukau, B. *Cell* **2011**, *147*, 1295.

(30) Gosavi, S.; Chavez, L. L.; Jennings, P. A.; Onuchic, J. N. *J. Mol. Biol.* **2006**, *357*, 986.

(31) Capraro, D. T.; Roy, M.; Onuchic, J. N.; Jennings, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14844.

(32) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. *Annu. Rev. Biophys.* **2010**, *39*, 159.

(33) Juers, D. H.; Heightman, T. D.; Vasella, A.; McCarter, J. D.; Mackenzie, L.; Withers, S. G.; Matthews, B. W. *Biochemistry* **2001**, *40*, 14781.

(34) Steitz, T. A.; Seidelt, B.; Innis, C. A.; Wilson, D. N.; Gartmann, M.; Armache, J. P.; Villa, E.; Trabuco, L. G.; Becker, T.; Mielke, T.; Schulten, K.; Beckmann, R. *Science* **2009**, *326*, 1412.

(35) Young, R.; Bremer, H. *Biochem. J.* **1976**, *160*, 185.

(36) Rubenstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: New York, 2003.

(37) Ugrinov, K. G.; Clark, P. L. *Biophys. J.* **2010**, *98*, 1312.

(38) De Sancho, D.; Doshi, U.; Munoz, V. *J. Am. Chem. Soc.* **2009**, *131*, 2074.

(39) Liang, S. T.; Xu, Y. C.; Dennis, P.; Bremer, H. *J. Bacteriol.* **2000**, *182*, 3037.

(40) Kaiser, C. M.; Chang, H. C.; Agashe, V. R.; Lakshmipathy, S. K.; Etchells, S. A.; Hayer-Hartl, M.; Hartl, F. U.; Barral, J. M. *Nature* **2006**, *444*, 455.

(41) Gupta, R.; Lakshmipathy, S. K.; Chang, H. C.; Etchells, S. A.; Hartl, F. U. *FEBS Lett.* **2010**, *584*, 3620.

(42) Taylor, W. R. *Nature* **2000**, *406*, 916.

(43) Jacobson, R. H.; Zhang, X. J.; Dubose, R. F.; Matthews, B. W. *Nature* **1994**, *369*, 761.

(44) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

(45) Fluitt, A.; Pienaar, E.; Vijoen, H. *Comput. Biol. Chem.* **2007**, *31*, 335.

(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(47) Heinig, M.; Frishman, D. *Nucleic Acids Res.* **2004**, *32*, W500.

(48) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(49) Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. *Protein Sci.* **2003**, *12*, 2057.