Energetic dependencies dictate folding mechanism in a complex protein

Kaixian Liu[a,1], Xiuqi Chen[a,1], and Christian M. Kaiser[b,c,2]

[a]Cell, Molecular, Developmental Biology, and Biophysics Graduate Program, Johns Hopkins University, Baltimore, MD 21218; [b]Department of Biology, Johns Hopkins University, Baltimore, MD 21218; and [c]Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218

Edited by Michael T. Woodside, University of Alberta, Edmonton, AB, Canada, and accepted by Editorial Board Member Yale E. Goldman October 31, 2019 (received for review August 19, 2019)

Large proteins with multiple domains are thought to fold cotranslationally to minimize interdomain misfolding. Once folded, domains interact with each other through the formation of extensive interfaces that are important for protein stability and function. However, multidomain protein folding and the energetics of domain interactions remain poorly understood. In elongation factor G (EF-G), a highly conserved protein composed of 5 domains, the 2 N-terminal domains form a stably structured unit cotranslationally. Using single-molecule optical tweezers, we have defined the steps leading to fully folded EF-G. We find that the central domain III of EF-G is highly dynamic and does not fold upon emerging from the ribosome. Surprisingly, a large interface with the N-terminal domains does not contribute to the stability of domain III. Instead, it requires interactions with its folded C-terminal neighbors to be stably structured. Because of the directionality of protein synthesis, this energetic dependency of domain III on its C-terminal neighbors disrupts cotranslational folding and imposes a posttranslational mechanism on the folding of the C-terminal part of EF-G. As a consequence, unfolded domains accumulate during synthesis, leading to the extensive population of misfolded species that interfere with productive folding. Domain III flexibility enables large-scale conformational transitions that are part of the EF-G functional cycle during ribosome translocation. Our results suggest that energetic tuning of domain stabilities, which is likely crucial for EF-G function, complicates the folding of this large multidomain protein.

multidomain proteins | protein folding | protein translation | single-molecule optical tweezers | elongation factor G

Elongation factor G (EF-G) is a large protein composed of 5 domains (termed G, II, III, IV, and V) that catalyzes ribosome movement along the messenger RNA (mRNA) during translation (1). The structure of EF-G (2, 3) closely resembles that of elongation factor Tu in complex with transfer RNA (tRNA) and GTP (4), which delivers activated amino acids to the ribosome during translation. The overall shape of the C-terminal half of EF-G (domains III, IV, and V) is very similar to that of the ternary complex tRNA. As suggested by this "molecular mimicry," EF-G binds to the same site on the ribosome as ternary complexes. GTP binding in the N-terminal G-domain and hydrolysis after ribosome binding elicit large structural rearrangements within EF-G. Reorientation of the domains relative to each other is required for EF-G to facilitate ribosome translocation (5–7). Cross-linking the G-domain and domain V through engineered cysteines disrupts translocation activity and ribosome release, while GTPase activity is retained (8). This finding highlights the importance of conformational flexibility for EF-G function. In order to carry out its function, all 5 domains of EF-G must thus fold into their native structures and adopt the correct spatial arrangement relative to each other.

Large proteins with multiple domains, such as EF-G, constitute almost half of all proteins in extant proteomes (9). While some of their constituent domains are stable in isolation, multidomain protein folding and dynamics often are not well described as the sum of their parts in terms of folding (10) and stability (11). In the case of tandem repeat proteins, which exhibit minimal interactions between the native domains, high sequence identity among neighboring units increases their propensity to form off-pathway, misfolded structures (12, 13). Even though compact globular multidomain proteins are typically made up of dissimilar domains, their folding is severely hampered by interdomain misfolding (14–17). Assistance from molecular chaperones and cotranslational folding ensure efficient folding in the cell (18, 19). Because in vitro studies have revealed fast and robust folding of a set of isolated domains (10), it is generally assumed that domains fold cotranslationally as they emerge from the ribosome during synthesis. However, coupling of synthesis and folding can influence the conformational search of the nascent chain, as shown by theoretical and computational work (20) as well as experimental studies (21). Moreover, interactions with the ribosome are known to slow down folding (22). In addition, very limited information is available on how already folded domains affect the rates and stabilities of their unfolded neighbors (23, 24).

Once folded, domains of large globular proteins typically pack against each other, forming extensive interfaces (9, 11). Domain–domain interfaces tend to contain hydrophobic residues (9) and are overall similar to interfaces found in multisubunit complexes (25). For the latter, stability has been found to scale with buried surface area (26). Computational studies similarly suggest that domain–domain interactions provide stability that is proportional

Significance

Large proteins composed of multiple domains are abundant in all proteomes, but their folding and structural dynamics remain poorly understood. Using single-molecule force spectroscopy, we have defined how stabilizing interfaces among the domains of elongation factor G (EF-G) shape its folding pathway. Contrary to the expectation that multidomain proteins fold sequentially as they emerge from the ribosome, we find that folding cannot be completed until the full protein has been synthesized. This posttranslational folding mechanism results in a propensity for misfolding. It is dictated by an energetic coupling among domains that enables conformational flexibility crucial for EF-G function. EF-G thus provides an example of how distinct biological ends—robust folding and functionally important flexibility—come into conflict during protein biogenesis.

Author contributions: K.L. and C.M.K. designed research; K.L. and X.C. performed research; K.L. and X.C. contributed new reagents/analytic tools; K.L. and C.M.K. analyzed data; and K.L. and C.M.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.T.W. is a guest editor invited by the Editorial Board.

[1]K.L. and X.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: kaiser@jhu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1914366116/-/DCSupplemental.

First published November 27, 2019.

to the interface surface area (27, 28). Thermodynamic coupling is generally assumed to increase stability, but effects of tethering can offset stabilizing effects or even reverse them (11), making it difficult to predict the energetic effects resulting from interactions observed in protein structures.

Dissecting multidomain protein folding and domain coupling is challenging (29), and few mechanistic studies of nonrepeat proteins are available to date (14–17, 30). Single-molecule force spectroscopy is a powerful tool for dissecting complex folding pathways (31). In particular, optical tweezers have proven extremely useful in detecting and characterizing the transient states that constitute either productive intermediates or misfolded traps, both within (32–38) and among domains (15, 17, 39, 40). The approach affords precise control over the folding of large proteins, enabling selective unfolding of individual domains (17) and characterization of nonnative species that are prone to aggregation in ensemble measurements (41). As such, optical tweezers are an ideal tool for dissecting the folding and dynamics of large proteins that are difficult to study with other methodologies.

Here, we have used optical tweezers to define the folding and dynamics of the C-terminal EF-G domains (III, IV, and V). When unfolded, these domains form misfolded states that compete with otherwise rapid and efficient folding. The central domain III is the most dynamic part of the molecule. The N-terminal domains G and II form an extensive interface with domain III, but do not stabilize it. In contrast, interactions with the C-terminal domains IV and V drastically reduce its unfolding rate. Because IV and V are synthesized after III, these stabilizing interactions form posttranslationally only after synthesis of full-length EF-G is complete. Therefore, cotranslational folding, which facilitates efficient folding of the N-terminal part of the molecule, does not continue through the C-terminal domains. These results demonstrate how the folding mechanism of a large protein is determined by energetic dependencies among its constituent domains.

Results

Domain III Does Not Stably Fold When Emerging from the Ribosome. Previous work has established the importance of sequential, cotranslational folding for the 2 N-terminal domains of EF-G (16, 17). To determine whether cotranslational folding continues as protein synthesis progresses, we generated ribosome–nascent-chain complexes (RNCs) in which translation is stalled at codon 531 of EF-G (531_{RNC}), such that the first 3 domains (G, II, and III) have been extruded from the ribosome (Fig. 1*A*). After tethering individual 531_{RNC} molecules for mechanical manipulation with optical tweezers, we applied a continuously increasing force ("force ramp") to unfold the nascent chain. The resulting force–extension curves (FECs) exhibited clear unfolding transitions for the G-domain and domain II (Fig. 1*B*, red and yellow arrowheads), which resembled those for the 2 domains in EF-G (Fig. 1*C*). However, no transition was detected for domain III in 531_{RNC}, even though the domain shows a clear signature during unfolding of the isolated full-length protein (termed *EF-G* here; Fig. 1*C*, green arrowhead). These data indicate that only the 2 N-terminal domains (G and II) are natively folded in 531_{RNC}, whereas domain III is not stably structured.

The ribosome has been shown to destabilize proximal domains in nascent polypeptides (42–44). Therefore, it seemed possible that the apparent lack of domain III structure in the 531_{RNC} construct is the result of interactions with the ribosome. However, FECs obtained with an isolated protein comprised of the first 3 EF-G domains, termed *G–II–III* here, also lacked a clear unfolding signal for domain III (Fig. 1*D*). In contrast to 531_{RNC}, *G–II–III* exhibits folding and unfolding transitions in the force range below 5 pN (*SI Appendix*, Fig. S1), indicating that the isolated protein (*G–II–III*, Fig. 1*D*), but not the nascent chain (531_{RNC}, Fig. 1*B*) exhibits transient folding of domain III. "Force clamp" experiments, in which the molecules are held at a constant force of 3 pN, revealed domain III folding and unfolding transitions in both the *EF-G* and the *G–II–III* constructs (*SI Appendix*, Fig. S2). However, whereas domain III is largely folded in *EF-G*, unfolding only transiently, it is mostly unfolded in *G–II–III*. Therefore, it seems likely that contacts present in full-length *EF-G*, but not in *G–II–III*, stabilize domain III.

C-Terminal Neighbors Stabilize Domain III. The EF-G crystal structure reveals extensive interaction surfaces among individual domains that might be important for stability. Due to their tight coupling, the 2 N-terminal EF-G domains (G and II) have been referred to as a "superdomain" (6), burying 995 Å^2 of surface area [calculated with jsPISA (45) using the crystal structure of *Escherichia coli* EF-G (5)]. Energetic coupling with the G-domain is indeed required for domain II stability (17). The 2 C-terminal domains (IV and V) are coupled not only by noncovalent interactions (interface buried surface area, 332 Å^2). Domain IV is split into 2 parts (IVa and IVb; see bar domain in Fig. 1*A*) that bracket domain V, resulting in a discontinuous domain topology, a feature that increases cooperativity and stability (33). Therefore, the 2 N-terminal and the 2 C-terminal domains that flank domain III form 2 compact, stable structural units, which we refer to here as "G/II" and "IV/V."

The amount of surface area buried upon interaction of structural units has been proposed to be a measure of interaction energies, independent of the chemical nature of the interactions at the interface (26, 27, 46). Inspection of the EF-G structure reveals that domain III forms contacts with both the N-terminal G/II and the C-terminal IV/V units (Fig. 1*E*). The larger of the 2 interfaces is formed between III and G/II, with 684 Å^2 of buried surface area, compared to 490 Å^2 for III and IV/V (Fig. 1*E*). The domain interfaces likely stabilize domain III. Since G/II is insufficient to stabilize domain III (Fig. 1*D*), we expected that both G/II and IV/V are required for stable folding of domain III.

Surprisingly, an EF-G deletion construct termed *III–IV–V*, which lacks G/II, exhibits clear unfolding transitions for domain III (Fig. 2*A*, green arrowhead), suggesting that interactions with domains IV and V are sufficient to structurally stabilize domain III. When we unfolded fully structured *III–IV–V* molecules, we observed sequential transitions that fall into 4 clusters with distinct unfolding force and extension change ranges (Fig. 2*B*). The initial transitions in each FEC (green dots) occur in the range of 5 to 10 pN and line up reasonably well with a worm-like chain model (47, 48) for unfolding of domain III (green line), calculated based on the EF-G structure (*SI Appendix*, *Data Analysis* and Tables S1 and S2). Domain V (purple) unfolds next, most likely together with IVb, the C-terminal part of domain IV (*SI Appendix*, *Data Analysis*). The remaining part of domain IV (IVa) unfolds last. Unfolding of domain IVa (blue) occurs quickly after domain V has unfolded (Fig. 2*A*, *Inset*). In about one-third of all traces, the unfolding of domains IV and V is not resolved due to the time resolution of our measurement (1 ms) and occurs in one apparent step (gray). Taking into account the length changes of individual unfolding transitions (Fig. 2*C*), their order of occurrence in each force ramp, and the domain topology of *III–IV–V* (*SI Appendix*, *Data Analysis*), sequential domainwise unfolding of *III–IV–V* is therefore best described by the scheme shown in Fig. 2*C*. Unfolding of *EF-G* shows the same classes of transitions (*SI Appendix*, Fig. S3), and both *EF-G* and *III–IV–V* exhibit domain III hopping at a constant force of 3 pN (*SI Appendix*, Fig. S4). The C-terminal domains III, IV, and V therefore appear to adopt very similar structures in the *EF-G* and *III–IV–V* constructs and unfold through the same pathway in our experiments.

Native IV/V Is Required for Robust Domain III Folding. In natively structured molecules, domain III unfolds below 12 pN, whereas the other EF-G domains all unfold at higher forces. It is therefore possible to selectively denature domain III and study its

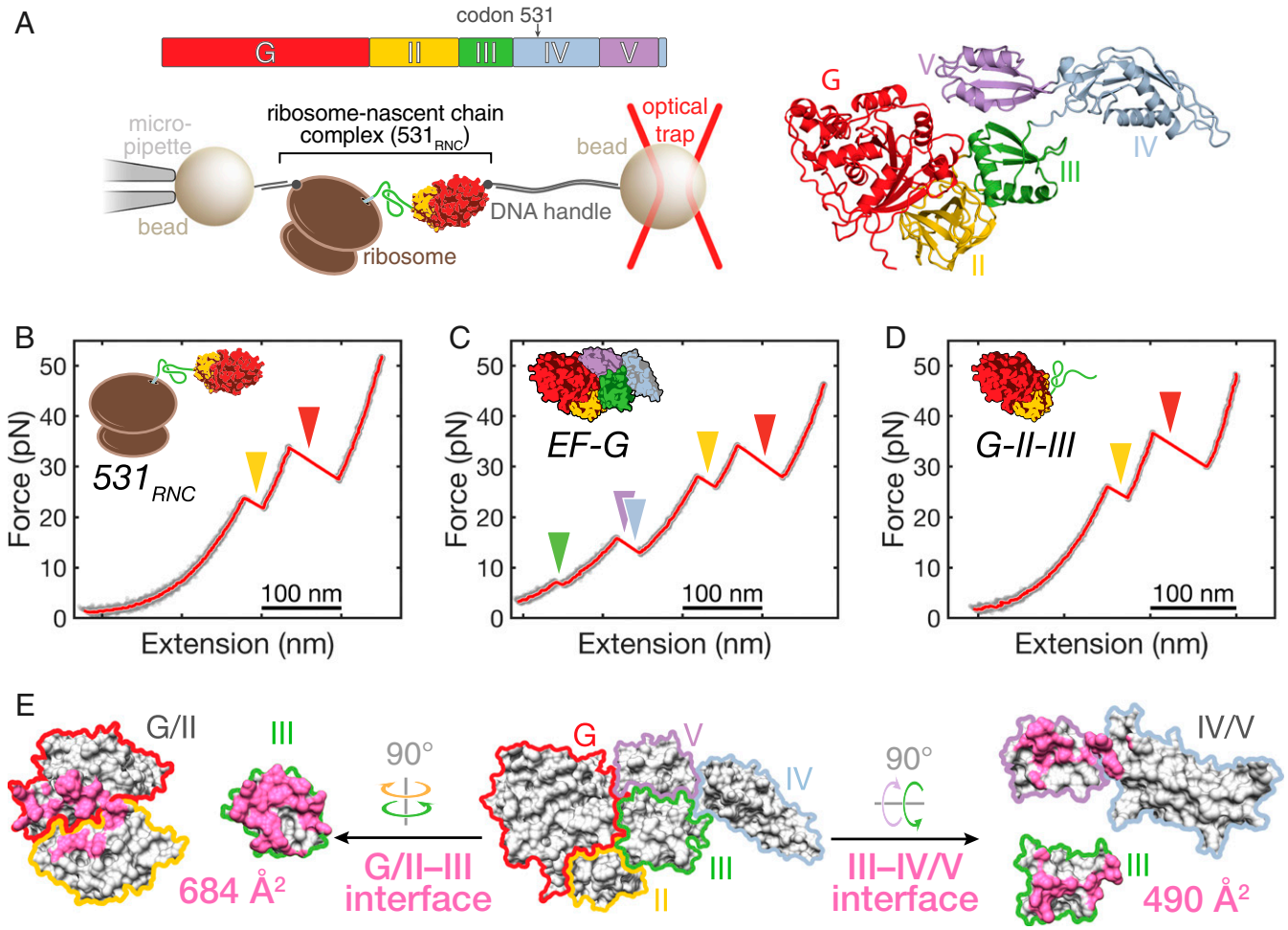

Fig. 1. Domain III does not fold when emerging from the ribosome. (*A*) Experimental setup for single-molecule folding measurements. Stalled RNCs are tethered to polystyrene beads for mechanical manipulation with optical tweezers. Stalling at codon 531 in the EF-G sequence allows the first 3 domains (G, II, and III) to fully emerge from the ribosome exit tunnel. For measurements with isolated proteins, the polypeptide is tethered by its termini. Domain diagram on *Top*; crystal structure (PDB code: 4v9p) on the *Right*. (*B–D*) Representative FECs showing the initial stretching of 531_{RNC}, *EF-G*, and *G–II–III* molecules. Unfolding transitions are indicated by arrowheads, colored by domain as in *A*. Unfolding of domain III is apparent only in *EF-G*, whereas 531_{RNC} and *G–II–III* exhibit clear transitions only for the G-domain and domain III. Gray dots, raw data (1,000 Hz); red line, filtered data (33 Hz). (*E*) Surface representation of the EF-G structure, with interaction surfaces colored in pink and the amount of buried surface area between domain III and either G/II or IV/V indicated.

folding in the context of native flanking domains. When we limit the applied force to 12 pN, domain III folds robustly during virtually every force ramp cycle in both the *III–IV–V* construct (Fig. 3*A*) and in *EF-G* (*SI Appendix*, Fig. S5). In contrast, native unfolding transitions are detected infrequently after refolding of the fully denatured *III–IV–V* construct, indicating that the molecule fails to adopt stable structure most of the time (Fig. 3*B*). The transitions that are occasionally detected differ from those of native domain III (Fig. 3*C*), showing on average longer contour length changes (Fig. 3*D*). The measured length change indicates how many amino acids are unfolded in the transition. The nonnative states therefore appear to be formed by more amino acids than those of just domain III, perhaps including residues from several domains. Taken together, domain III refolds robustly when the neighboring IV/V unit is stably structured. However, when the neighboring domains IV and V are unfolded, nonnative states are predominantly populated.

Using the force ramp protocol for selective unfolding, we compared unfolding events for domain III in *III–IV–V* (398 events) and *EF-G* (751 events). The unfolding transitions are very similar for the 2 constructs (*SI Appendix*, Fig. S5*C*), with observed contour length changes (24.0 ± 3.7 nm for *III–IV–V*, 24.6 ± 2.3 nm for *EF-G*) very close to the value calculated from the crystal structure (24.7 nm; *SI Appendix*, Table S2). The distributions of unfolding forces differ somewhat (see below) but fall into similar ranges (Fig. 3 and *SI Appendix*, Fig. S5). These measurements suggest that domain III adopts very similar structures in the *EF-G* and *III–IV–V* constructs.

Domain III Stabilization Is Derived Exclusively from C-Terminal Neighbors. How unfolding and refolding forces are distributed is determined by the kinetics of the underlying molecular processes. Utilizing the method developed by Dudko et al. (49), we converted the domain III unfolding and refolding force distributions (Fig. 4 *A* and *B*) into force-dependent rates. Fitting Bell's model (50) to the data yielded transition state distances as well as zero-force rates for folding and unfolding (Fig. 4 *C* and *D*). The parameters for refolding are very similar (*SI Appendix*, Fig. S6 and Table S3), suggesting similar pathways in *EF-G* and *III–IV–V*. Unfolding occurs at slightly lower rates in *EF-G* compared to *III–IV–V* (Fig. 4 *C* and *D* and *SI Appendix*, Fig. S6), and the transition state distance for unfolding is shorter. These effects are likely due to differences in how mechanical force is applied to domain III in the 2 constructs. For *III–IV–V*, we pull directly on the N terminus of

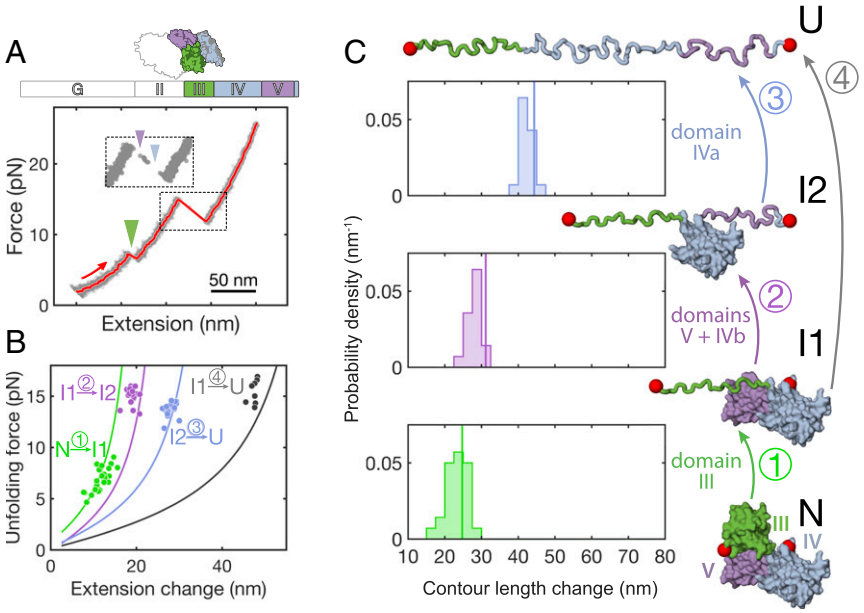

Fig. 2. Neighboring C-terminal domains stabilize domain III. (*A*) Initial FEC for a *III–IV–V* molecule. Unfolding of native domain III is indicated by a green arrowhead. The *Inset* shows unfolding of domains V and IVb (purple arrowhead), rapidly followed by domain IVa unfolding (blue arrowhead). (*B*) Unfolding events from fully folded *III–IV–V* molecules. Each dot represents an unfolding transition in the FECs. The lines represent worm-like chain models calculated for sequential unfolding of individual domains (see text for details). (*C*) Distributions of contour length changes for sequential domainwise unfolding of *III–IV–V*, colored by domain. The expected values, calculated from the EF-G crystal structure, are indicated by the vertical lines. The cartoons on the *Right* illustrate the structures populated during sequential unfolding. The positions of the N and C termini are indicated by red spheres.

domain III. In full-length *EF-G*, the force is transduced through the folded domains G and II, resulting in a different pulling axis.

Extrapolation of Bell's model to zero force yields estimates of the intrinsic folding and unfolding rates, which we find to be very similar for *EF-G* and *III–IV–V* (*SI Appendix*, Table S3). Thus, while the transition state distances for unfolding depend on sequence context (*EF-G* vs. *III–IV–V*), the intrinsic unfolding rates do not appear to depend on whether or not the N-terminal domains are present. From the intrinsic folding and unfolding rates, the thermodynamic stabilities of domain III in *III–IV–V* and *EF-G* can be estimated according to the following:

$$\Delta G = -k_B T ln\left(K_{eq}\right) = -k_B T ln\left(\frac{k_f^0}{k_u^0}\right).$$

Because the rates are very similar for both constructs, their stabilities obtained in this way are also very similar [$\Delta G(III_{EF\text{-}G}) = -7.2$ kcal/mol, $\Delta G(III_{III\text{-}IV\text{-}V}) = -6.9$ kcal/mol].

The results from our kinetics analysis suggest that, surprisingly, the increased stability of domain III is derived entirely from interactions with the C-terminal domains IV and V, despite the fact that the interface domain III forms with the N-terminal G/II unit is larger than that formed with the C-terminal IV/V unit (Fig. 1*E*). In order to measure the thermodynamic stability of domain III more directly and avoid the uncertainties associated with extrapolation of the kinetic data, we utilized Crooks' fluctuation theorem (CFT) (51), a statistical mechanics-based approach for extracting thermodynamic information from nonequilibrium experiments. CFT relates the distributions of irreversible work associated with mechanical folding and unfolding to the free energy of the underlying process. The intersection of the 2 distributions directly yields the thermodynamic stability. We applied CFT to the work distributions of domain III folding and unfolding in *EF-G* and the *III–IV–V* construct (Fig. 4 *E* and *F*) and obtained very similar values for the free energy that stabilizes domain III [$\Delta G_{CFT}(III_{EF\text{-}G}) = -6.2$ kcal/mol, ΔG_{CFT} $(III_{III\text{-}IV\text{-}V}) = -6.4$ kcal/mol]. Taken together, these results demonstrate that the N-terminal G/II unit does not contribute measurably to the stability of domain III, which derives its thermodynamic stabilization entirely from interactions with the C-terminal IV/V unit.

Misfolding Thwarts Efficient Folding of the C-Terminal Domains. The finding that domain III requires interactions with the C-terminal parts of EF-G to remain stably structured implies that it does not fold cotranslationally. As a consequence, the C-terminal domains likely accumulate as unfolded polypeptide, and stable folding is achieved only after synthesis of full-length EF-G is complete. EF-G extensively populates misfolded states after complete unfolding of all domains (16). Even among the 2 N-terminal domains (G and II), the population of misfolded, off-pathway species severely interferes with productive folding (17). To determine whether misfolding also takes place in the C-terminal half of the molecule, we fully unfolded the isolated *III–IV–V* polypeptide by stretching it to 25 pN, and then subjected the molecule to repeated force ramp cycles with a 5-s pause at 2 pN between successive cycles to allow for refolding. Full refolding is observed in only a small fraction of attempts (23 out of 213; *SI Appendix*, Fig. S7). Most traces (*SI Appendix*, Fig. S7) exhibit transitions from nonnative states (142 out of 213) or no measurable transitions at all (48 out of 213). The nonnative transitions are heterogeneous and mostly do not exhibit well-defined length changes, suggesting that they represent ensembles of several different states. It therefore appears that neither of the 3 domains refolds with high probability under these experimental conditions, presumably because the formation of off-pathway misfolded states competes with productive folding.

To follow folding directly, we carried out force clamp experiments (Fig. 5). We first unfolded *III–IV–V* completely at a constant force of 10 pN. Unfolding transitions (steps 1, 2, and 3, Fig. 5*A*, red trace) occur in the same order as in force ramp measurements (III, V/IVb, and IVa, Fig. 2). Domain III unfolds very rapidly at this force and is resolved in only some of the traces (*SI Appendix*, Fig. S8). After complete unfolding, we jumped the force to 3 pN and followed refolding (Fig. 5*A*, blue trace). The molecule

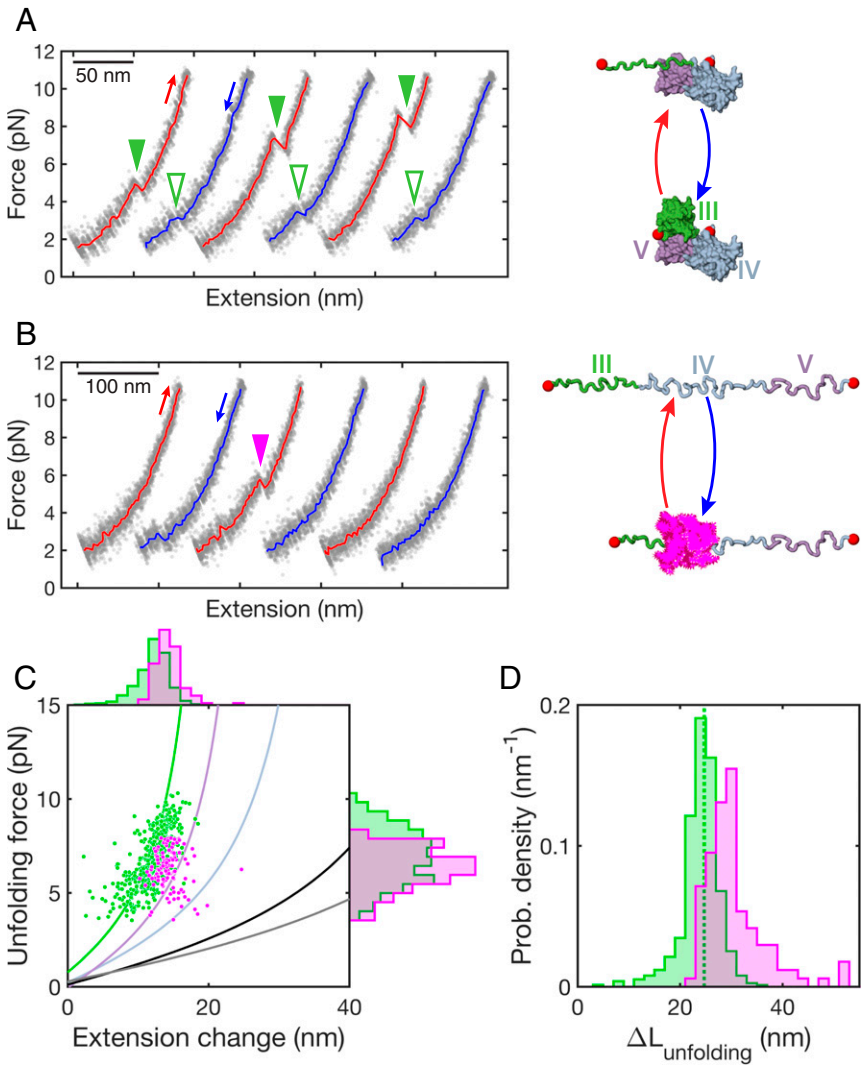

Fig. 3. Native neighboring domains are required for stable domain III folding. (*A*) FECs from successive pulling (red) and relaxation (blue) of a single *III–IV–V* molecule in the 2- to 12-pN force range. Domain III unfolds (closed arrowheads), whereas domains IV and V remain structured. Refolding of domain III (open arrowheads) is apparent in all relaxation traces. Traces are plotted with a horizontal offset for visual clarity. (*B*) Same as *A*, but after complete denaturation of all domains in *III–IV–V*. Under these conditions, some traces show transitions reminiscent of domain III unfolding (magenta arrowhead). Most traces do not exhibit clear unfolding transitions, indicating that none of the 3 domains acquires stable structure. (*C*) Extension changes and unfolding forces of unfolding events obtained after selective unfolding of domain III (green) or complete denaturation of *III–IV–V* (magenta). Each dot represents one transition; histograms show the distributions of extension changes (*Top*) and unfolding forces (*Right*). The distributions overlap, but are distinct, indicating the population of nonnative, misfolded states. The solid lines represent worm-like chain models for unfolding of individual domains (green, purple, blue), domains IV and V (black), and domains III, IV, and V (gray). (*D*) Distribution of contour lengths changes obtained from the transitions shown in *C*. The dotted line indicates the unfolding length for native domain III calculated from the crystal structure.

hops reversibly between the unfolded state and an intermediate state (termed I_M here; Fig. 5*A*). I_M is more compact than the unfolded state by 13.4 ± 3.4 nm, roughly the same as folding of domain IVa (calculated extension change of 12.2 nm at 3 pN). In a small fraction of attempts (6 out of 60), the molecule subsequently adopts its native state. In these cases, the molecule shows the expected compaction during refolding (Fig. 5*A*, black open arrowhead), and subsequently jumping the force back up to 10 pN shows the expected unfolding pattern (*SI Appendix*, Fig. S8). Most often, however, the native state is not reached within the 60-s time cutoff of the experiment (54 of 60 attempts). In these cases, unfolding transitions in the following high-force segment are not observed (*SI Appendix*, Fig. S9), indicating that none of the domains has fully folded.

The very slow refolding kinetics suggest that off-pathway states interfere with productive folding, i.e., that I_M might be a misfolded state. To dissect the refolding pathway, we carried out constant force refolding experiments with a modified experimental procedure (Fig. 5*B*). Instead of allowing all 3 domains to unfold, we lowered the force from 10 to 3 pN immediately after domains IVb and V unfolded, keeping domain IVa folded. Under these conditions, the molecule transitions back to the fully folded state very rapidly (Fig. 5*B*, open black arrowhead). Prior to folding, an intermediate is transiently populated (*SI Appendix*, Fig. S8). Although the extension of this intermediate is similar to that of I_M described above, its lifetime is clearly different from that of the I_M state described above. I_M therefore is different from the productive folding intermediate observed here (*P* value of 1×10^{-14}; *SI Appendix*, *Data Analysis*), showing that different states are populated depending on whether unfolding begins from the fully unfolded state (Fig. 5*A*) or from a partially unfolded state (Fig. 5*B*). Taken together, these results suggest that productive folding steps occur rapidly, but overall folding is greatly slowed down by the formation of misfolded species among the C-terminal domains of the protein.

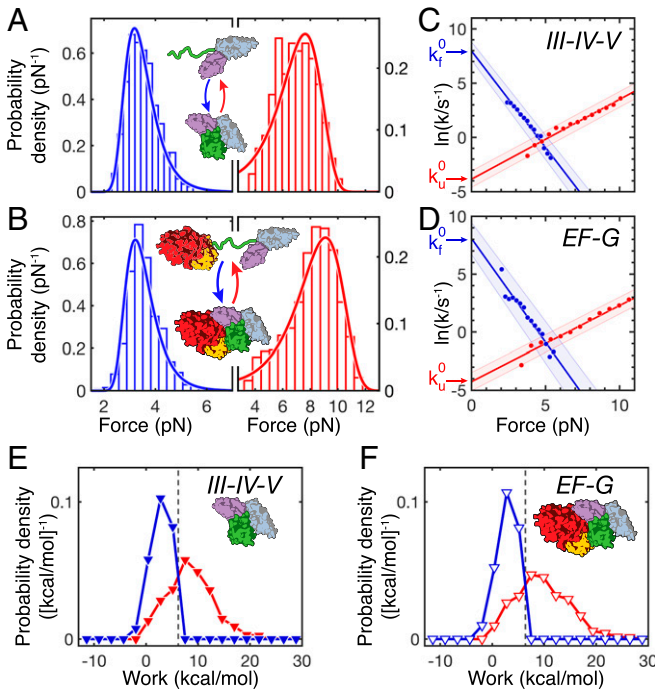

Fig. 4. Domain III stabilization is derived entirely from contacts with C-terminal neighbors. (*A*) Refolding (blue) and unfolding (red) force histograms for domain III in *III–IV–V*. The lines represent force distributions calculated from least-squares fits of the transformed histograms (*Materials and Methods*). (*B*) Same as *A* for *EF-G*. (*C*) Domain III folding and unfolding rates in *III–IV–V* as a function of force, determined from the data shown in *A*. The lines represent fits to Bell's model, extrapolated to zero force to obtain intrinsic folding and unfolding rates. (*D*) Same as *C* for *EF-G*. Rates are similar in *III–IV–V* and *EF-G*. (*E* and *F*) Work distributions for domain III folding (blue) and unfolding (red) in *III–IV–V* and *EF-G*. In both cases, the crossing point of the 2 distributions is near 6 kcal/mol, indicating that the thermodynamic stability of domain III is very similar in both constructs.

Discussion

We have used optical tweezers to dissect the folding and dynamics of the large multidomain protein EF-G, revealing intricate energetic coupling among the C-terminal domains of the molecule that is important for folding and, likely, structural flexibility, which are crucial aspects of EF-G function. The 5 EF-G domains form extensive contacts, collectively burying a total of more than 2,700 Å^2 of surface area across 7 domain–domain interfaces. Our measurements indicate that domain III constitutes the most dynamic part of EF-G, unfolding spontaneously approximately once per minute [estimated rate, $k^0_u(III_{EF\text{-}G}) \sim 0.014\ s^{-1}$; Fig. 4] (52). Refolding is robust and fast (estimated rate, $k^0_f \sim 3{,}200\ s^{-1}$). Domain III is therefore predominantly in the folded state. However, even small perturbations, such as a mechanical load of 3 pN, substantially accelerate unfolding [$k_u(III_{EF\text{-}G}, 3\ pN) \sim 0.22\ s^{-1}$; *SI Appendix*, Fig. S3]. The other domains are much more stable and remain fully folded over long periods of time. Single-molecule fluorescence polarization measurements have suggested that domain III moves relative to domain IV during translocation (7), which implies breaking of the domain interface between III and IV/V. The properties of domain III may be tuned such that it can act as a flexible hinge, enabling large-scale conformational motions without unfolding of additional domains.

In the absence of the C-terminal IV/V unit, domain III becomes markedly more dynamic, resulting in very low unfolding forces (Fig. 1*D* and *SI Appendix*, Fig. S1). In contrast, removal of the G/II unit has at most a subtle effect on domain III dynamics (Fig. 4 and *SI Appendix*, Figs. S4 and S6 and Tables S3 and S4). The energetic contributions from native interfaces have been suggested to scale with the amount of buried surface area (27, 28). Our measurements demonstrate that, counterintuitively, EF-G does not follow this trend. Two independent analyses indicate very similar free energies of domain III in the presence and absence of domains G and II (Fig. 4). Even though the interfaces of domain III with G/II and IV/V appear similar in terms of their overall hydrophobicity (*SI Appendix*, Fig. S10), the smaller of the 2 (with IV/V) is entirely responsible for providing additional stability, whereas the larger one (with G/II) is energetically neutral. Our results may help to define benchmarks for improving computational approaches for calculating the magnitude of energetic coupling among domains, which is very likely a crucial aspect of the stability, function, and regulation of many multidomain proteins (11).

The reduced stability of domain III in the absence of IV/V is mainly due to a ~100-fold increase in the unfolding rate (*SI Appendix*, *Data Analysis*). In contrast, domain III folding rates (*SI Appendix*, Figs. S2 and S4) are similarly high in *EF-G*, *G–II–III*, and *III–IV–V* (in the absence of any interdomain misfolding). Frustration (53) within domain III therefore appears to be low, and it is not increased by contacts with the neighboring native domains. This finding implies that stabilizing contacts are formed after folding is complete, and any interactions that might occur prior to that neither increase nor decrease the folding rate significantly. As such, the scenario resembles that of minimal frustration (54), which postulates that differences in thermodynamic stabilities of naturally occurring domains are largely due to changes in unfolding rates (55).

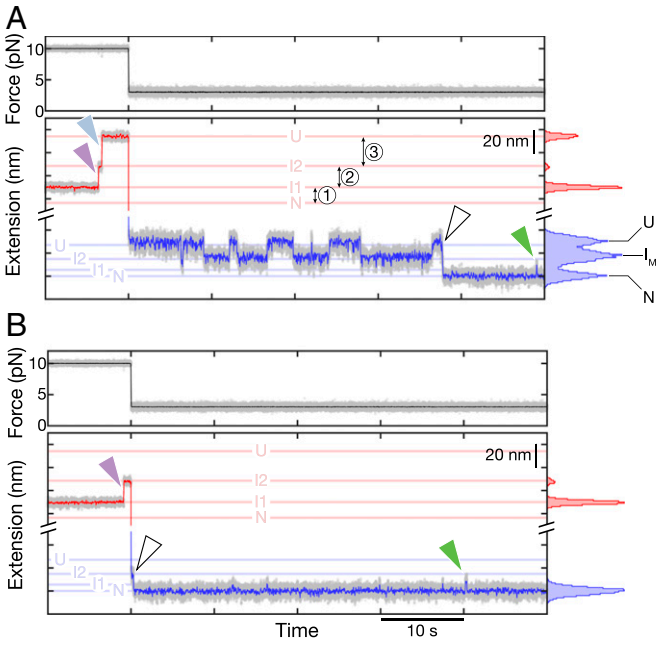

Fig. 5. Misfolding thwarts rapid productive folding. (*A*) Constant-force unfolding at 10 pN (red trace) and refolding at 3 pN (blue trace) of *III–IV–V*. The horizontal lines (pale red for unfolding; pale blue for refolding) indicate the calculated positions of native (N) and unfolded (U) *III–IV–V*, as well as 2 partially structured states (I1, I2). The spaces in between lines correspond to the extension changes upon folding or unfolding (1–3). States and steps are labeled as in Fig. 2*C*. Full unfolding at 10 pN occurs in several steps (purple and steel blue arrowheads). At 3 pN, the molecule hops between U and an intermediate state (I_M) before folding completely (black open arrowhead). (*B*) After partial unfolding of *III–IV–V* (red trace; purple arrowhead), domain IVa remains folded. When the force is lowered, the molecule rapidly refolds (black open arrowhead). The long-lived states populated in *A* therefore likely represent the population of off-pathway, misfolded states. In both examples, transient unfolding of domain III is visible after complete unfolding (green arrowheads).

Whereas domain III folds quickly and efficiently when the neighboring units are natively structured, predominantly nonnative structures are populated when domains IV and V are unfolded (Fig. 3). The presence of folded G/II does not reduce misfolding (*SI Appendix*, Fig. S5), consistent with our finding that it does not help to stabilize domain III energetically or increase its folding rate. Interdomain misfolding is prevalent after complete unfolding of EF-G (16) and other multidomain proteins (15, 56, 57). Constant force refolding experiments at 3 pN indicate that off-pathway nonnative structures are also populated during simultaneous refolding of domains III, IV, and V, competing with and slowing down productive folding (Fig. 5). Thus, while frustration is minimal within domain III, folding is frustrated by interactions among nonnative domains (Fig. 6). Application of mechanical force has been shown to reduce interdomain misfolding by disfavoring long-range nonnative contacts (15). It is therefore likely that misfolding among the C-terminal EF-G domains is even more severe in the absence of force.

Cotranslational folding facilitates the efficient folding of large proteins by reducing the complexity of the conformational search problem (19, 21, 22). In EF-G, this mechanism is important for the folding of the 2 N-terminal domains (17) (Fig. 6). The low stability of domain III that we observe here suggests that cotranslational folding does not continue beyond the G/II unit. The ribosome has been observed to stabilize structures of some nascent chains (58) and to destabilize others (34, 42–44, 59). In line with this trend, the ribosome further destabilizes the already dynamic domain III (Fig. 1*B* and *SI Appendix*, Fig. S1). As elongation continues, the effect is expected to decay as the spacing between the domain and the ribosome increases (17, 34, 42, 43). However, in the absence of stabilizing interactions formed with natively folded domains IV and V, domain III remains dynamic, with an estimated intrinsic unfolding rate of 1.4 s^{-1} (measured with the isolated *G–II–III* construct; *SI Appendix*, *Data Analysis*). Translation of the following domain (IVa), which is 127 amino acids long, requires at least 6 s [at elongation rates of 20 amino acids per s (60)], and completion of the full protein another ∼4 s (90 amino acids for domains V and IVb). The folding and unfolding rates measured with isolated EF-G fragments therefore indicate that domain III unfolds many times during the synthesis of the domains IV and V. Our constant-force measurements further suggest that the unfolded domains have a high propensity to form relatively long-lived misfolded states (Fig. 5), derailing productive cotranslational folding.

EF-G is highly conserved across all kingdoms of life. Efficient folding of the eukaryotic EF-G ortholog eEF2 requires a specialized chaperone, Hgh1, that likely binds cotranslationally, stabilizes domain III, and recruits additional chaperones, including the TRiC chaperonin and possibly Hsp90, to the nascent polypeptide (61, 62). Bacteria lack Hgh1 and likely use other chaperone systems to reduce misfolding during synthesis. Increased aggregation of EF-G upon deletion of the main nascent chain-binding chaperones in bacteria, trigger factor and DnaK, demonstrates their importance for its overall folding (63). It is intriguing to speculate that an as-yet-to-be-identified bacterial chaperone carries out a function similar to that of Hgh1, binding to the highly dynamic domain III and ensuring efficient folding of EF-G, which is highly abundant in *E. coli* (64).

While EF-G preferentially adopts an elongated state (65), as depicted in Fig. 1*A*, it assumes a compact conformation upon binding to ribosomes in the pretranslocated state (6). This dramatic conformational change is brought about by a rotation of domains III–V relative to G/II, which is essential for EF-G activity (8). Since contacts with IV/V are largely maintained, this transition does not perturb domain III stability, based on our findings. A series of conformational rearrangements in EF-G is then necessary to place domain IV into the ribosomal A-site and drive mRNA translocation forward. Transitions between rigid and flexible conformations have been proposed to be crucial for EF-G activity (5). Interestingly, mutants conferring resistance to the antibiotic fusidic acid, which locks EF-G in a state with high affinity for the ribosome and prevents dissociation after GTP hydrolysis, map primarily to domain III (66). The dynamics revealed by our measurements may reflect the intricate evolutionary tuning of domain III kinetics and energetics that enable EF-G to respond to GTP binding, hydrolysis, and release.

Taken together, our studies show that domain III is a highly flexible element in EF-G. Its energetic coupling to the C-terminal IV/V unit, which is likely important for EF-G activity, complicates folding, perhaps reflecting a trade-off between folding and function. Important details of the EF-G conformational cycle during tRNA translocation still remain to be determined. Regardless of what will turn out to best describe its mechanism of action, our characterization of domain III dynamics and folding might help to better understand the intricate motions and rearrangements in EF-G that are part of its duty cycle.

Materials and Methods

Proteins and RNC Constructs. Proteins (*EF-G*, *G–II–III*, and *III–IV–V*) were recombinantly expressed in *E. coli* BL21 cells and purified by affinity chromatography, followed by size exclusion chromatography. RNCs (531_{RNC}) were prepared using the PURExpress Δribosome kit (NEB) supplemented with in-house purified ribosomes and programmed with mRNAs generated by in vitro transcription. Experimental procedures for the preparation of proteins and RNCs are described in detail in *SI Appendix*.

Single-Molecule Measurements. Optical tweezers experiments were carried out using a home-built single-trap instrument (67) using functionalized 2-μm polystyrene beads as described previously (17). The trap stiffness was

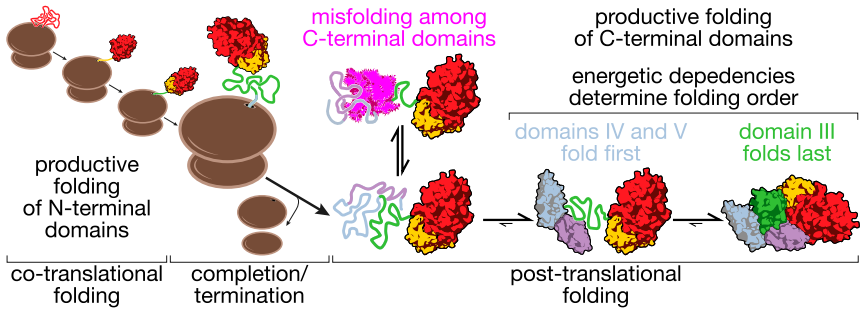

Fig. 6. Model for EF-G folding. The 2 N-terminal domains, G (red) and II (yellow), fold cotranslationally [Liu et al. (17)]. Domain III (green) is highly dynamic in the absence of fully folded domains IV (blue) and V (purple). As a consequence, cotranslational folding is interrupted. Accumulation of unfolded polypeptide results in the formation of misfolded species, sidetracking the molecule into nonproductive states and slowing down folding. Once domains IV and V have folded, domain III is stabilized. Chaperones are likely required to stabilize the C-terminal domains in a folding-competent state and prevent association of nonnative species into aggregates.

~0.1 pN/nm. The trap velocity for force ramp experiments was 100 nm/s. Experiments were carried out in buffer HKMβ (20 mM Hepes-KOH, 100 mM KCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, pH 7.4). Custom Matlab scripts were used for data analysis. Detailed experimental and data analysis procedures are provided in *SI Appendix*.

Data Availability Statement. All data are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS. This work was supported by a grant from the National Institutes of Health (5R01GM121567, to C.M.K.).

1. H. Yamamoto *et al*., EF-G and EF4: Translocation and back-translocation on the bacterial ribosome. *Nat. Rev. Microbiol.* **12**, 89–100 (2014).
2. J. Czworkowski, J. Wang, T. A. Steitz, P. B. Moore, The crystal structure of elongation factor G complexed with GDP, at 2.7 A resolution. *EMBO J.* **13**, 3661–3668 (1994).
3. A. AEvarsson *et al*., Three-dimensional structure of the ribosomal translocase: elongation factor G from *Thermus thermophilus*. *EMBO J.* **13**, 3669–3677 (1994).
4. P. Nissen *et al*., Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
5. A. Pulk, J. H. Cate, Control of ribosomal subunit rotation by elongation factor G. *Science* **340**, 1235970 (2013).
6. J. Lin, M. G. Gagnon, D. Bulkley, T. A. Steitz, Conformational changes of elongation factor G on the ribosome during tRNA translocation. *Cell* **160**, 219–227 (2015).
7. C. Chen *et al*., Elongation factor G initiates translocation through a power stroke. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7515–7520 (2016).
8. F. Peske, N. B. Matassova, A. Savelsbergh, M. V. Rodnina, W. Wintermeyer, Conformationally restricted elongation factor G retains GTPase activity but is inactive in translocation on the ribosome. *Mol. Cell* **6**, 501–505 (2000).
9. J. H. Han, S. Batey, A. A. Nickson, S. A. Teichmann, J. Clarke, The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* **8**, 319–330 (2007).
10. E. Braselmann, J. L. Chaney, P. L. Clark, Folding the proteome. *Trends Biochem. Sci.* **38**, 337–344 (2013).
11. Y. Levy, Protein assembly and building blocks: Beyond the limits of the LEGO brick metaphor. *Biochemistry* **56**, 5040–5048 (2017).
12. A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez, J. M. Fernandez, Single protein misfolding events captured by atomic force microscopy. *Nat. Struct. Biol.* **6**, 1025–1028 (1999).
13. M. B. Borgia *et al*., Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* **474**, 662–665 (2011).
14. Z. N. Scholl, W. Yang, P. E. Marszalek, Chaperones rescue luciferase folding by separating its domains. *J. Biol. Chem.* **289**, 28607–28618 (2014).
15. M. Jahn, J. Buchner, T. Hugel, M. Rief, Folding and assembly of the large molecular machine Hsp90 studied in single-molecule experiments. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1232–1237 (2016).
16. K. Liu, J. E. Rehfus, E. Mattson, C. M. Kaiser, The ribosome destabilizes native and non-native structures in a nascent multidomain protein. *Protein Sci.* **26**, 1439–1451 (2017).
17. K. Liu, K. Maciuba, C. M. Kaiser, The ribosome cooperates with a chaperone to guide multi-domain protein folding. *Mol. Cell.* **74**, 310–319.e7 (2019).
18. Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. U. Hartl, Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
19. M. Thommen, W. Holtkamp, M. V. Rodnina, Co-translational protein folding: Progress and methods. *Curr. Opin. Struct. Biol.* **42**, 83–89 (2017).
20. A. K. Sharma, E. P. O'Brien, Non-equilibrium coupling of protein structure and function to translation-elongation kinetics. *Curr. Opin. Struct. Biol.* **49**, 94–103 (2018).
21. K. C. Stein, J. Frydman, The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *J. Biol. Chem.* **294**, 2076–2084 (2019).
22. C. M. Kaiser, K. Liu, Folding up and moving on-nascent protein folding on the ribosome. *J. Mol. Biol.* **430**, 4580–4591 (2018).
23. S. Batey, J. Clarke, The folding pathway of a single domain in a multidomain protein is not affected by its neighbouring domain. *J. Mol. Biol.* **378**, 297–301 (2008).
24. O. B. Nilsson *et al*., Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **24**, 221–225 (2017).
25. S. Jones, A. Marin, J. M. Thornton, Protein domain interfaces: Characterization and comparison with oligomeric protein interfaces. *Protein Eng.* **13**, 77–82 (2000).
26. J. Chen, N. Sawyer, L. Regan, Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci.* **22**, 510–515 (2013).
27. R. M. Bhaskara, N. Srinivasan, Stability of domain structures in multi-domain proteins. *Sci. Rep.* **1**, 40 (2011).
28. S. Vishwanath, A. G. de Brevern, N. Srinivasan, Same but not alike: Structure, flexibility and energetics of domains in multi-domain proteins are influenced by the presence of other domains. *PLOS Comput. Biol.* **14**, e1006008 (2018).
29. S. Batey, A. A. Nickson, J. Clarke, Studying the folding of multidomain proteins. *HFSP J.* **2**, 365–377 (2008).
30. B. T. Walters, L. Mayne, J. R. Hinshaw, T. R. Sosnick, S. W. Englander, Folding of a large protein at high structural resolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18898–18903 (2013).
31. G. Zoldák, M. Rief, Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.* **23**, 48–57 (2013).
32. C. Cecconi, E. A. Shank, C. Bustamante, S. Marqusee, Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057–2060 (2005).
33. E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, C. Bustamante, The folding cooperativity of a protein is controlled by its chain topology. *Nature* **465**, 637–640 (2010).
34. C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, Jr, C. Bustamante, The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).
35. H. Yu *et al*., Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5283–5288 (2012).
36. A. Mashaghi *et al*., Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **500**, 98–101 (2013).
37. S. Zorman *et al*., Common intermediates and kinetics, but different energetics, in the assembly of SNARE proteins. *eLife* **3**, e03348 (2014).
38. K. Neupane *et al*., Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **352**, 239–242 (2016).
39. J. Stigler, F. Ziegler, A. Gieseke, J. C. Gebhardt, M. Rief, The complex folding network of single calmodulin molecules. *Science* **334**, 512–516 (2011).
40. P. O. Heidarsson *et al*., Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13069–13074 (2014).
41. S. Sen Mojumdar *et al*., Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat. Commun.* **8**, 1881 (2017).
42. C. A. Waudby *et al*., Systematic mapping of free energy landscapes of a growing filamin domain during biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9744–9749 (2018).
43. A. J. Samelson, M. K. Jensen, R. A. Soto, J. H. Cate, S. Marqusee, Quantitative determination of ribosome nascent chain stability. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13402–13407 (2016).
44. A. Hoffmann *et al*., Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell* **48**, 63–74 (2012).
45. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
46. J. A. Marsh, Buried and accessible surface area control intrinsic protein flexibility. *J. Mol. Biol.* **425**, 3250–3263 (2013).
47. C. Bustamante, J. F. Marko, E. D. Siggia, S. Smith, Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).
48. J. F. Marko, E. D. Siggia, Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
49. O. K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15755–15760 (2008).
50. G. I. Bell, Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).
51. G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **60**, 2721–2726 (1999).
52. I. Sgouralis, M. Whitmore, L. Lapidus, M. J. Comstock, S. Pressé, Single molecule force spectroscopy at high data acquisition: A Bayesian nonparametric analysis. *J. Chem. Phys.* **148**, 123320 (2018).
53. D. U. Ferreiro, E. A. Komives, P. G. Wolynes, Frustration, function and folding. *Curr. Opin. Struct. Biol.* **48**, 68–73 (2018).
54. J. D. Bryngelson, P. G. Wolynes, Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524–7528 (1987).
55. F. O. Tzul, D. Vasilchuk, G. I. Makhatadze, Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1627–E1632 (2017).
56. A. Mashaghi, S. Mashaghi, S. J. Tans, Misfolding of luciferase at the single-molecule level. *Angew. Chem. Int. Ed. Engl.* **53**, 10390–10393 (2014).
57. Z. N. Scholl, W. Yang, P. E. Marszalek, Competing pathways and multiple folding nuclei in a large multidomain protein, luciferase. *Biophys. J.* **112**, 1829–1840 (2017).
58. W. Holtkamp *et al*., Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104–1107 (2015).
59. L. D. Cabrita *et al*., A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* **23**, 278–285 (2016).
60. H. Bremer, P. P. Dennis, Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *Ecosal Plus* **3** (2008).
61. L. Mönkemeyer *et al*., Chaperone function of Hgh1 in the biogenesis of eukaryotic elongation factor 2. *Mol. Cell* **74**, 88–100.e9 (2019).
62. F. H. Schopf *et al*., The Co-chaperone Cns1 and the recruiter protein Hgh1 link Hsp90 to translation elongation via chaperoning elongation factor 2. *Mol. Cell* **74**, 73–87.e8 (2019).
63. S. Vorderwülbecke *et al*., Low temperature or GroEL/ES overproduction permits growth of Escherichia coli cells lacking trigger factor and DnaK. *FEBS Lett.* **559**, 181–187 (2004).
64. P. Lu, C. Vogel, R. Wang, X. Yao, E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124 (2007).
65. J. Czworkowski, P. B. Moore, The conformational properties of elongation factor G and the mechanism of translocation. *Biochemistry* **36**, 10327–10334 (1997).
66. M. Laurberg *et al*., Structure of a mutant EF-G reveals domain III and possibly the fusidic acid binding site. *J. Mol. Biol.* **303**, 593–603 (2000).
67. S. B. Smith, Y. Cui, C. Bustamante, Optical-trap force transducer that operates by direct measurement of light momentum. *Enzymol.* **361**, 134–162 (2003).