Supplementary Information

The ribosome cooperates with a chaperone to guide multi-domain protein folding

Kaixian Liu[1,2], Kevin Maciuba[1,2], Christian M. Kaiser[2,3*]

[1]CMDB Graduate Program, Johns Hopkins University.
[2]Department of Biology, Johns Hopkins University.
[3]Department of Biophysics, Johns Hopkins University.
* Corresponding author: kaiser@jhu.edu

Contact for reagent and resource sharing.

Requests for further information and resources should be directed to and will be fulfilled by the lead contact (Christian Kaiser, kaiser@jhu.edu).

SUPPLEMENTARY FIGURES.

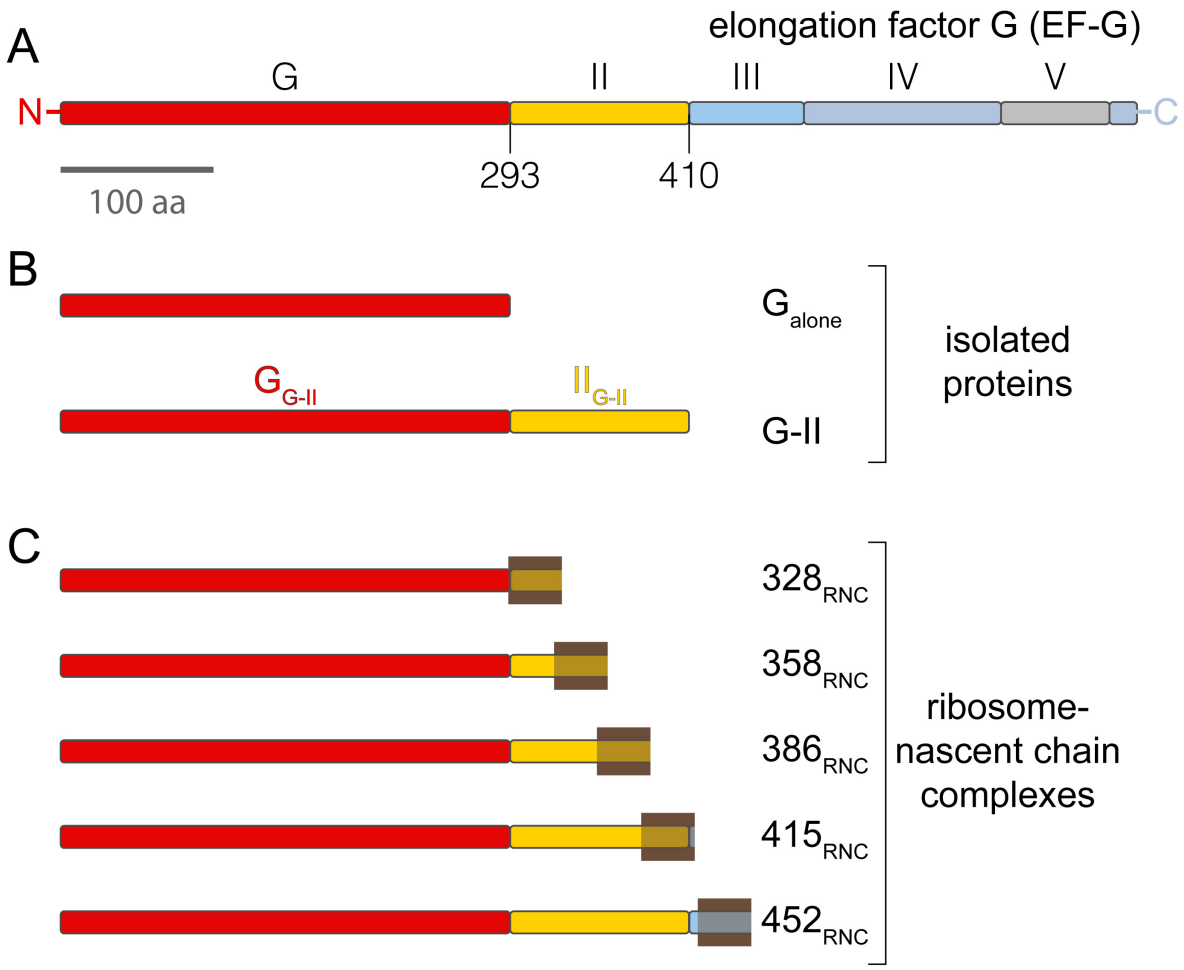

fig. S1 Overview of constructs used in the present study. Related to Figs. 1-6.
A. Domain diagram of EF-G, with the domain names indicated on top, and boundaries for the G-domain and domain II indicated on the bottom. **B.** Diagrams of the two EF-G constructs studied as isolated polypeptides in the absence of the ribosome. **C.** Diagrams of nascent chains studied here. The brown rectangle indicates the approximate position of the ribosome exit tunnel, i.e. the parts of the nascent EF-G polypeptides inside the boxes are sequestered within the exit tunnel (assuming that 35 amino acids are contained within the tunnel), the parts to the left of the boxes have fully emerged from the ribosome. In B. and C., the nomenclature for the respective construct used in the manuscript is indicated on the right.

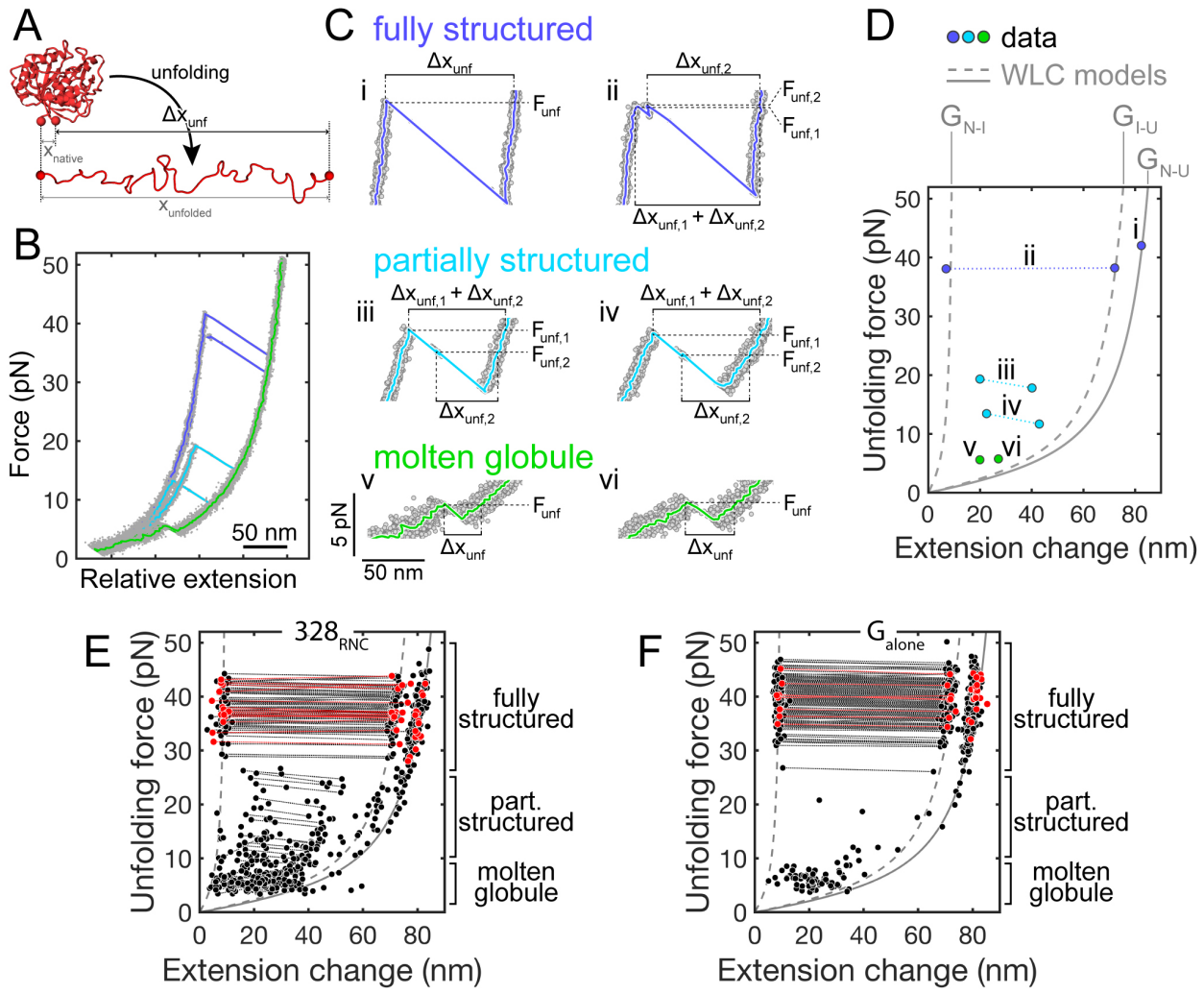

fig. S2 Characterization of states populated during G-domain refolding. Related to Fig. 1. **A.** Schematic of G-domain unfolding. The protein transitions from a compact folded state (with an end-to-end distance of x_{native}) to an extended unfolded state (with an end-to-end distance of $x_{unfolded}$). The observed extension change is $\Delta x_{unf} = x_{unfolded} - x_{native}$. $x_{unfolded}$ depends strongly on the applied force, as described by the worm-like chain model, whereas x_{native} does not. **B.** Overlay of six FECs recorded with a single *328_{RNC}* molecule. Grey dots: raw data (1000 Hz), colored lines: time-averaged date (33 Hz). The data was chosen to represent two examples each of molten globule-like (green), partially structured (cyan) and fully structured (blue) states. **C.** Magnified view of the unfolding transitions for the individual traces shown in A. Each transition is characterized by the unfolding force (F_{unf}) and extension change (Δx_{unf}), which is determined as described in STAR Methods: Quantitative and statistical analysis. Unfolding occurs either directly in one step (panels i, v, vi), or via an intermediate in two steps (ii, iii, iv). The number of unfolding steps, the unfolding force range and the extension change are characteristics of the protein structure that unfolds. **D.** Scatter plot of unfolding transitions. Each transition in represented by a circle. Transitions from two-step unfolding, resulting from the population of an unfolding intermediate, are connected by dotted lines. This representation enables visual comparison of a large number of unfolding events (as shown in E. and F.). Grey lines represent worm-like chain (WLC) models

(Bustamante et al., 1994) of one-step unfolding (contour length change $\Delta L(G_{N-U}) = 103.7$ nm, solid line) and two-step unfolding ($\Delta L(G_{N-I}) = 11$ nm, $\Delta L(G_{I-U}) = 92$ nm, dashed lines), which are characteristic for the fully folded G-domain of EF-G (Liu et al., 2017) (see STAR Methods: Quantitative and statistical analysis for details). A persistence length of 0.65 nm and contour length of 0.36 nm per amino acid residue were used for polypeptide WLC models. **E.** Scatter plot of unfolding transitions for 328_{RNC} (30 molecules). Red dots: initial pull, black dots: after refolding. Dotted lines connect transitions from two-step unfolding. Only ~36% of traces after refolding show transitions similar to the initial ones (labeled "fully structured"), indicating that the G-domain folds to the native structure in only about one third of attempts during the refolding pause (10 seconds). Transitions that do not match this pattern represent partially structured and molten globule-like state ("part. structured", "molten globule"). **F.** Scatter plot of unfolding transitions of G_{alone} (32 molecules). Most transitions fall into the "fully structured" region, indicating that G_{alone} reaches the native state with high probability during the 10 second refolding pause. Colors and labeling as in panel A. (Part of the data in panels A and B has previously been reported (Liu et al., 2017)). Dashed and solid grey lines represent WLC models for two-step and one-step unfolding.

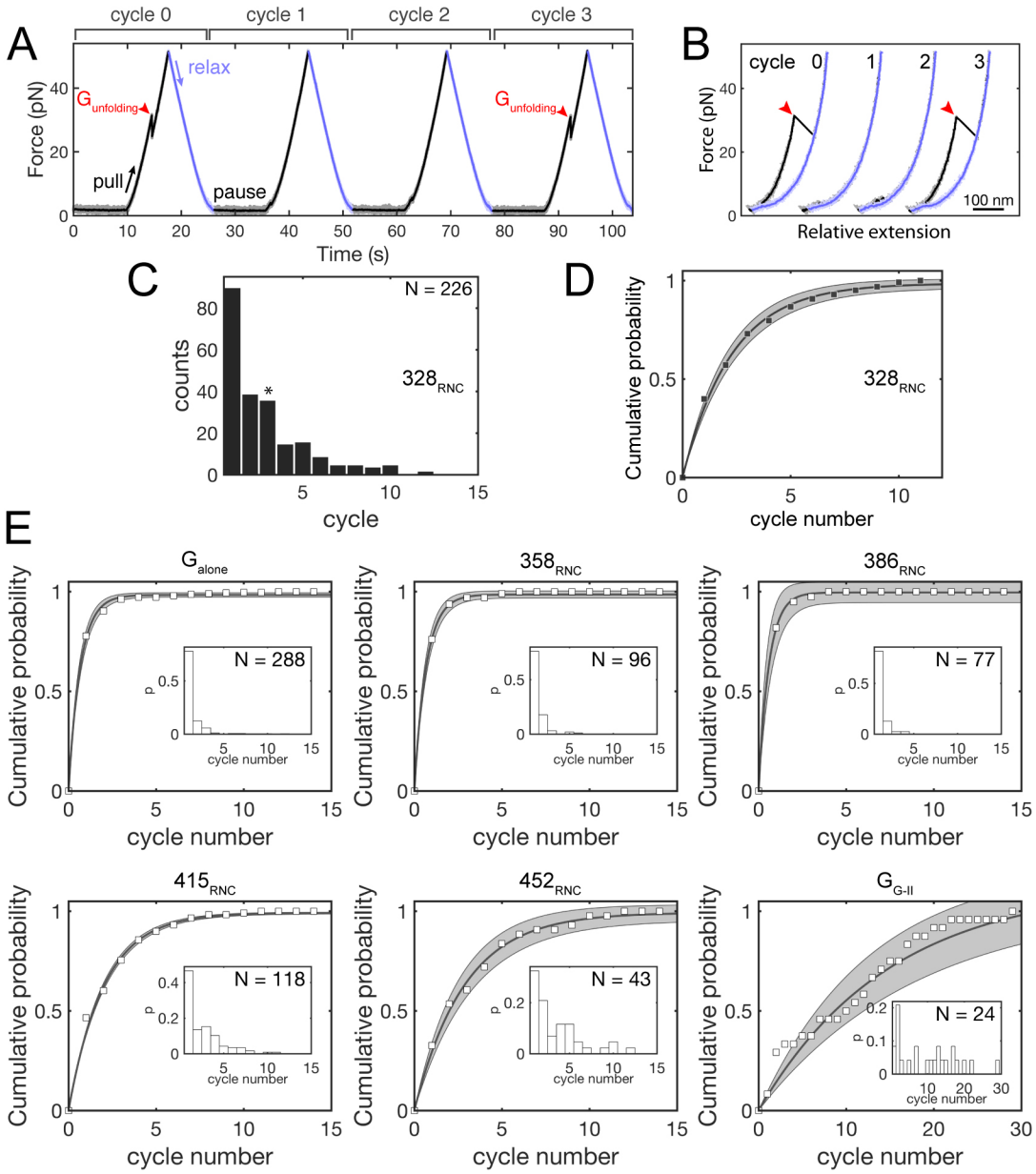

fig. S3 Folding cycle analysis of G-domain folding. Related to Fig. 1

A. Example plot of force *vs.* time for a 328_{RNC} molecule subjected to force ramp cycles. We define cycles as beginning at the 2 pN refolding pause and ending after the subsequent relaxation. Unfolding of the structured G-domain (red arrowheads) is apparent in the initial cycle and in cycle 3. Stretching is shown in black, relaxation in blue. **B.** The same cycles that are shown in A., represented as FECs. Cycles are offset along the x-axis for clarity. Arrowheads indicate the unfolding events of the fully structured G-domain that are also shown in A. **C.** Distribution of folding cycle (raw counts) for 328_{RNC} with total number of 226 native G-domain folding events. The asterisk indicates the bin to which the example in panel A contributed. **D.** Fit of a geometric distribution to G-domain folding probabilities. Folding probability (black squares) is obtained from folding cycle analysis (see STAR Methods: Quantitative and statistical analysis). The x-axis represents the cycle number in which unfolding occurs. Solid lines and shaded areas represent the

mean and 95% confidence intervals of the geometric distribution fit to the data. **E.** G-domain folding rates estimates for G_{alone}, 358_{RNC}, 386_{RNC}, 415_{RNC}, 452_{RNC} and *G-II*. The cumulative folding probability (white squares) is obtained from folding cycle analysis (see STAR Methods: Quantitative and statistical analysis) for the individual constructs. The x-axis represents the cycle number in which unfolding occurs. Insets show the distribution of raw folding cycle counts (similar to C.). The total number of G-domain unfolding events indicated. Solid lines and shaded areas represent the mean and confidence intervals of the geometric distribution fit to the data.

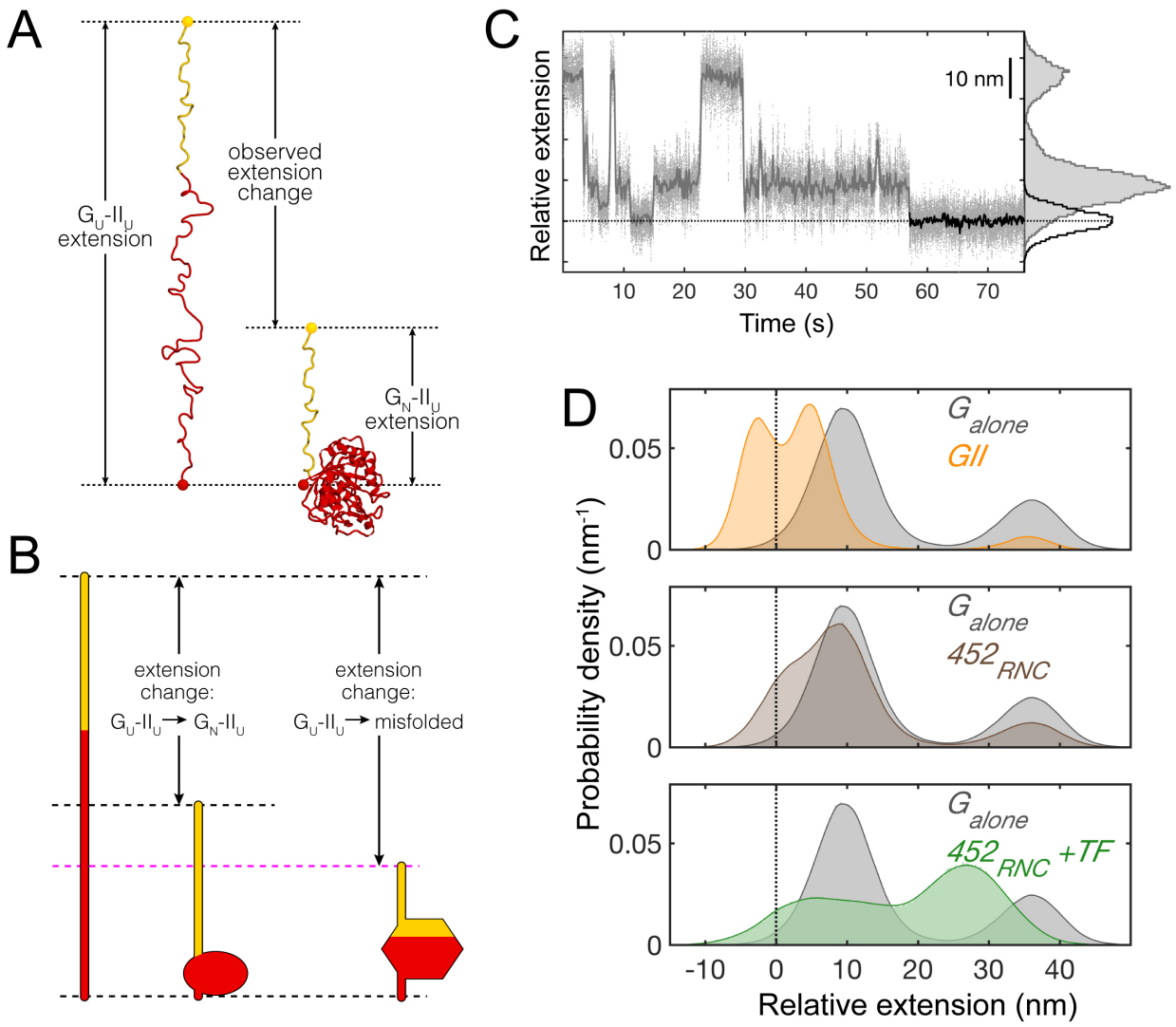

fig. S4 Interpretation of extension changes from force clamp experiments. Related to Fig. 2. **A.** The extension change observed during refolding at constant force is the difference in extension between the initial (fully unfolded; G_U-II_U) state and the final state of the molecule, in which the G-domain is folded (G_N-II_U). The folding intermediates observed in these experiments are not represented here. **B.** A shorter extension, indicated by the magenta line, is most likely due to the formation of a structure (hexagon) that contains elements from both the G-domain and domain II. See main text and STAR Methods: Quantitative and statistical analysis (*"Interpretation of force clamp refolding experiments"*) for details. **C.** Representative example force clamp trace of G_{alone} at 3.5 pN. Raw data (1000 Hz) is plotted as light grey dots, lines show time-averaged data (10 Hz). The equilibrium portion before G-domain folding of the trace is in dark grey, the portion after folding is in black. The histogram on the right shows the extension data for the segments before (grey) and after (black) G-domain folding. **D.** Probability densities from aggregated traces (same data as in Fig. 2E) for *G-II* (orange), 452_{RNC} (brown), and 452_{RNC} + TF (green), with data for G_{alone} shown in grey for comparison. The dotted black lines in C. and D. indicate the extension of the state in which the G-domain is folded, which was used as a reference for aligning traces and set to 0 nm.

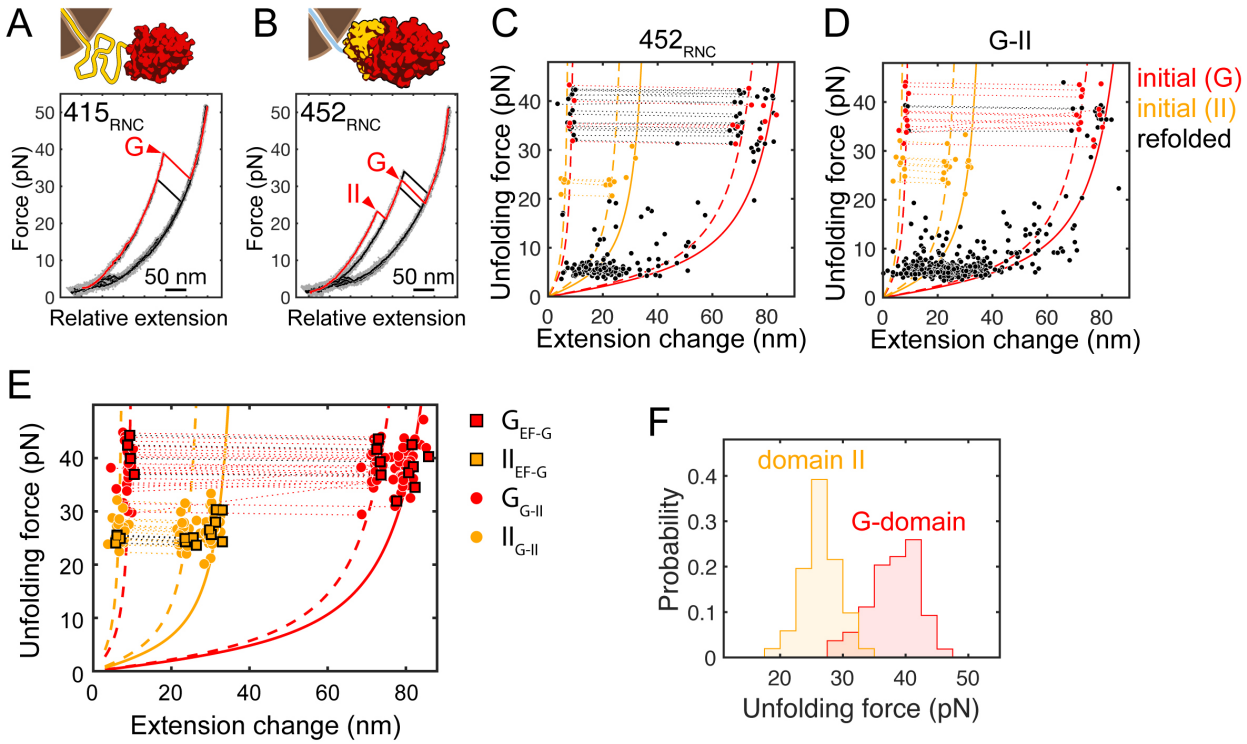

fig. S5 Folding signatures of the G-domain and domain II. Related to Fig. 3 and 4.
A., B. Example FECs for *415$_{RNC}$* and *452$_{RNC}$*. The initial unfolding trace (red) shows the signature of G-domain unfolding in both constructs ("G", arrowheads). The signature of domain II ("II", arrowhead) is apparent in *452$_{RNC}$*, but no transition that can be attributed to domain II is apparent in *415$_{RNC}$*. Thus, domain II folds only after fully emerging from the ribosome. **C.** Scatter plot showing unfolding transitions from *452$_{RNC}$* (9 molecules). Initial transitions are shown in red (G-domain) and yellow (domain II), transitions after refolding are shown in black. Pairs of transitions from two-step unfolding are connected by dotted lines. Most of the transitions after refolding do not resemble the initial transitions and likely represent partially structured or molten globule-like states. The G-domain refolds only sporadically, and unfolding of domain II is not observed. Lines represent WLC models for G-domain (red) and domain II (yellow) unfolding (solid lines: one-step unfolding, dashed lines: two-step unfolding; see fig. S2 and STAR Methods: Quantitative and statistical analysis, *"Unfolding transition analysis (force ramp experiments)"*). **D.** Scatter plot showing unfolding transitions from *G-II* (15 molecules). Colors and lines as in panel C. As with *452$_{RNC}$*, refolding of domain II is not observed, and refolding of the G-domain occurs only rarely, consistent with its low folding rate in this construct (see Fig. 1C). **E.** Scatter plot comparing initial unfolding transitions from *G-II* (circles; 54 molecules) and full-length EF-G (squares; 10 molecules). The unfolding transitions of the G-domain (red) and domain II (yellow) are very similar for both constructs. Additional transitions resulting from unfolding of domains III, IV and V that are apparent during unfolding of full-length EF-G (Liu et al., 2017) are not shown here. Lines as in panel C. (Data for full-length EF-G is from (Liu et al., 2017)). **F.** Comparison of unfolding force distributions from initial domain II (yellow) and G-domain (red) unfolding for 54 individual molecules of *G-II*. The G-domain unfolding forces are largely separate from those of domain II, with domain II being mechanically weaker and always unfolding first in force ramp experiments.

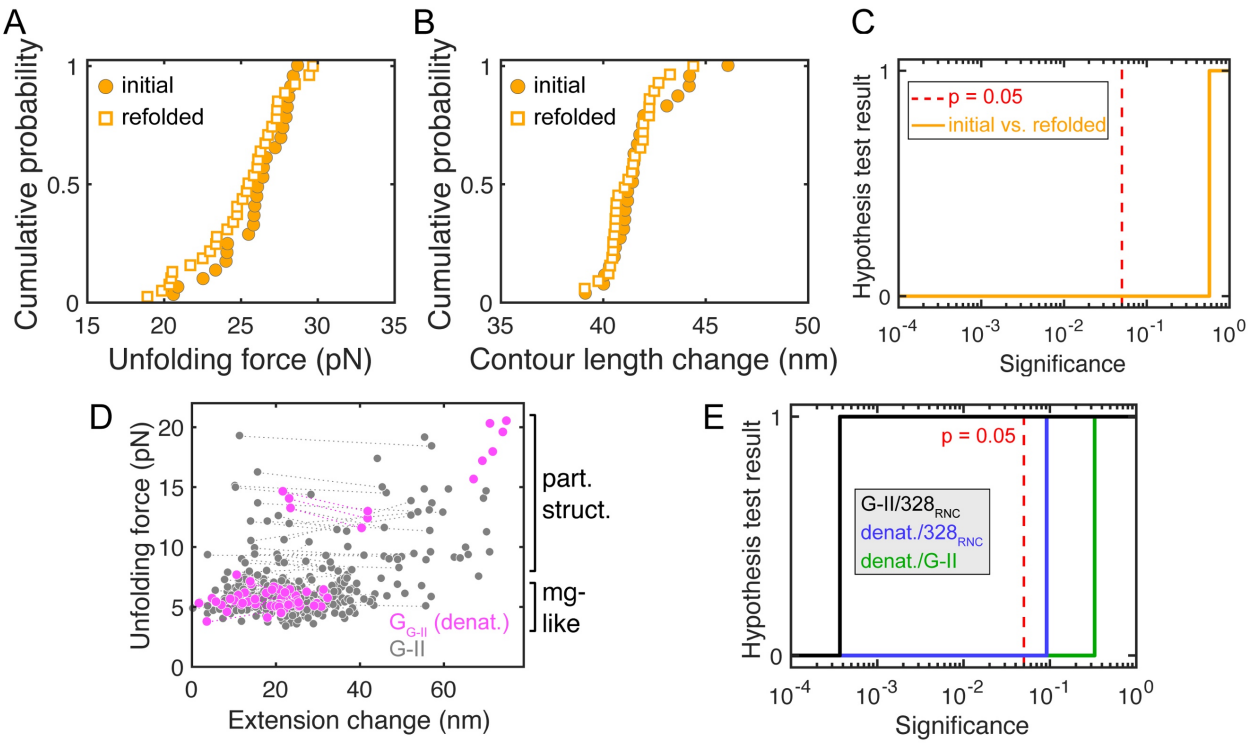

fig. S6 Domain II refolding and G-domain denaturation. Related to Fig. 4 and 5.
Cumulative distributions of unfolding forces (**A.**) and contour length changes (**B.**) of domain II from initial unfolding (circles) and unfolding after renaturation (squares). The G-domain was kept in the folded state by limiting the maximum force to 20 pN (see Fig. 4A). The similarity of the distributions indicates that domain II regains its native structure. This result is confirmed by one-dimensional two-sample Kolmogorov-Smirnov tests (Massey, 1951), which indicate that the respective data sets are from the same underlying distribution (p = 0.05; see STAR Methods: Quantitative and statistical analysis). **C.** Result from two-dimensional two-sample Kolmogorov-Smirnov test (Peacock, 1983), comparing unfolding transitions of initial and refolded domain II unfolding. The input data is shown in Fig. 4B. A test result of 1 indicates that the null hypothesis ("both data sets are drawn from the same distribution") can be rejected. The test was evaluated over a range of significance levels, and the result indicates that the two distributions are not significantly different (p = 0.57). **D.** Scatter plot of unfolding transitions after denaturation of the G-domain (magenta) and after full unfolding of *G-II* (grey). Transitions from two-step unfolding are connected by dotted lines. Both data sets show unfolding transitions indicating molten globule-like and partially structured states. See also Fig. 5C. **E.** Results of pairwise two-dimensional Kolmogorov-Smirnov tests (Peacock, 1983) comparing distributions of unfolding events, evaluated at different significance levels. A test result of 0 indicates that the null hypothesis ("data sets are from the same underlying distribution") cannot be rejected, a test result of 1 indicates that the null hypothesis can be rejected. Only unfolding events from non-native transitions are compared. The test indicates that events from *G-II* and 328_{RNC} are significantly different, whereas neither distribution differs significantly from unfolding of the denatured G-domain. This result indicates that unfolding transitions of the denatured G-domain have significant similarity to those of 328_{RNC} and *G-II*, and suggests that denaturation results in the population of partially structured states and, ultimately, complete unfolding.

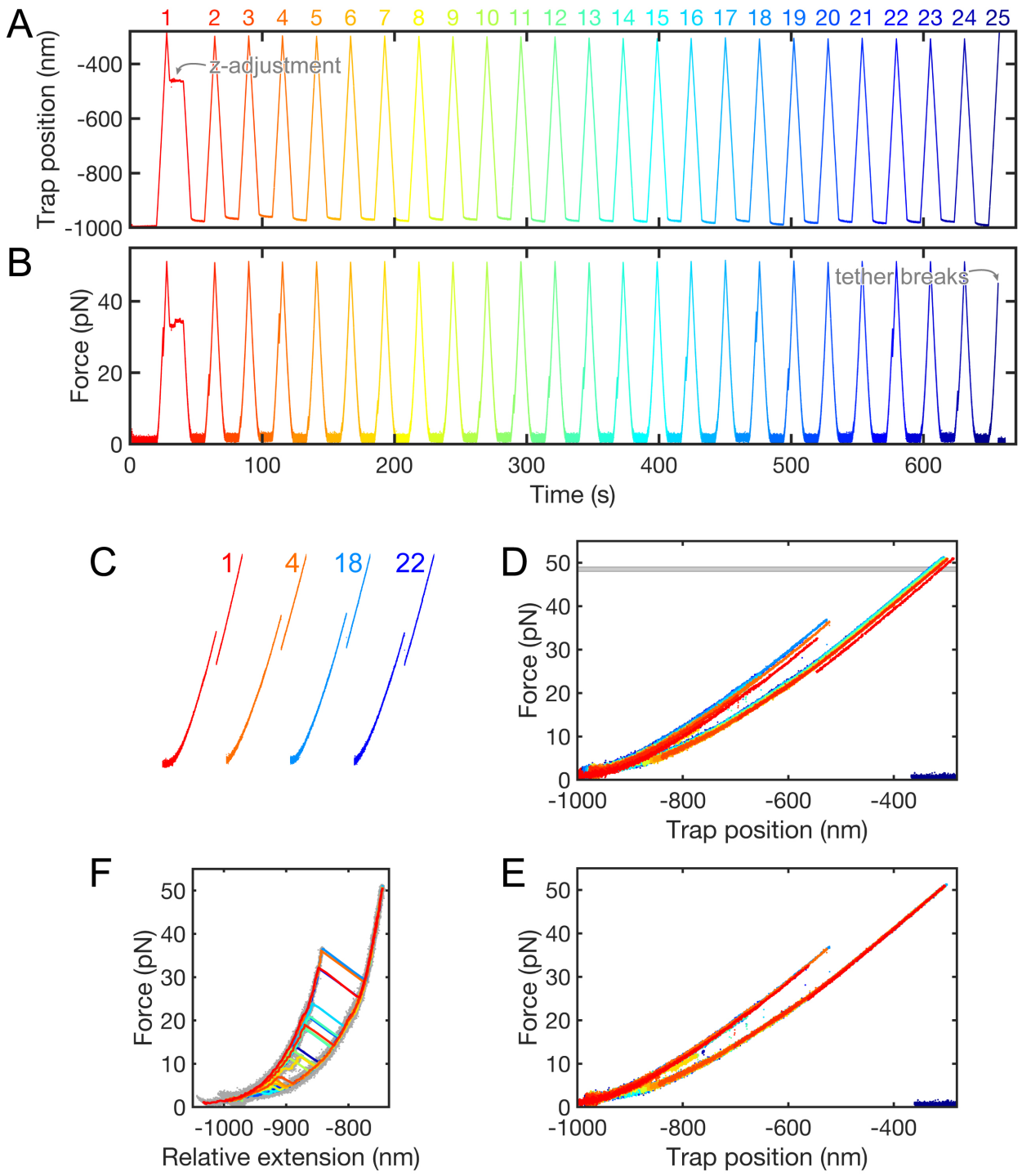

fig. S7 Raw data and data processing. Related to STAR Methods.
The figure illustrates the data processing steps to obtain force-extension curves from optical tweezers force ramp recordings. **A. and B.** Raw data of trap position (A) and force (B) vs. time for a representative molecule of 328_{RNC}, with cycle numbers indicated on top. The force ramp program is paused during cycle 1 to adjust the z-position of the micropipette, bringing the two beads into the same z-plane. The tether between the two beads breaks during cycle 25. **C.** Blow-up of the regions in the force-vs.-time trajectories that show full G-domain unfolding, with cycle

numbers indicated. **D.** Overlay of the raw data shown in A. and B. for all cycles without any data processing. The z-adjustment results in a relatively large offset from cycle 1 to the subsequent cycles. Slow mechanical drift results in additional offset during the course of the experiment. The region between 48 pN and 49 pN that was used to align the traces along the x-axis is shown as a grey box. **E.** Aligned raw data. The traces were shifted along the x-axis such that the mean trap positions are the same in the force region between 48 pN and 49 pN, where the G-domain is always in the unfolded state. **F.** Force extension curves, obtained by converting trap position (x_{trap}) to relative extension (x_{rel}) according to $x_{rel} = x_{trap} - F/k_{trap}$, with a trap stiffness of k_{trap} = 0.115 pN/nm. Only the stretching segments (regions of increasing force) are plotted for clarity. Raw data (1000 Hz) is plotted as grey dots, time-averaged data (33 Hz) is plotted as lines (colors as in A). The plotted data is the same as that shown in Fig. 1B.

SUPPLEMENTARY TABLES.

Table s1: DNA oligonucleotides for generating EF-G RNC templates. Related to STAR Methods.

Oligonucleotide name	sequences
T4Ls_T7_fw	5'-GTCGGCGATATAGGCGCCAG-3'
328RNC_rev	5'-CGGGTCGGTAGCGATTTTGAACG-3'
386RNC_rev	5'-GATAGCAGCAGCGATGTCGC-3'
415RNC_rev	5'-TACCGGCTCAGGGAATTCCATACG-3'
452RNC_rev	5'-GATTCTTCGTCAGTCCATACACGG-3'

Table s2: DNA oligonucleotides for generating DNA handles. Related to STAR Methods.

Oligonucleotide name	sequence	usage notes
NH-fw3	5'- /5AmMC6/CACCGAAACGCGCGAGGCAGCAG -3'	For cross-linking to CoA
cNH-fw3	5'- /5Phos/GTCGCTGCTGCCTCGCGCGTTTCGGTG -3'	For annealing to NH-fw3
NH-fw2	5'-/5AmMC6/CTGCTGGGGCAAACCAGCGTGGAC-3'	For cross-linking to polystyrene beads
cNH-fw2	5'-/5Phos/CGACGTCCACGCTGGTTTGCCCCAGCAG-3'	For annealing to NH-fw2