

# Translation and folding of single proteins in real time

Florian Wruck<sup>a,1</sup>, Alexandros Katranidis<sup>b,2</sup>, Knud H. Nierhaus<sup>c,3</sup>, Georg Büldt<sup>b,d</sup>, and Martin Hegner<sup>a,2</sup>

<sup>a</sup>Centre for Research on Adaptive Nanostructures and Nanodevices, School of Physics, Trinity College Dublin, Dublin 2, Ireland; <sup>b</sup>Institute of Complex Systems ICS-5, Forschungszentrum Jülich, 52425 Jülich, Germany; <sup>c</sup>Institute for Medical Physics and Biophysics, Charité–Universitätsmedizin Berlin, 10117 Berlin, Germany; and <sup>d</sup>Laboratory for Advanced Studies of Membrane Proteins, Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Russia

Edited by George H. Lorimer, University of Maryland, College Park, MD, and approved April 21, 2017 (received for review October 27, 2016)

**Protein biosynthesis is inherently coupled to cotranslational protein folding. Folding of the nascent chain already occurs during synthesis and is mediated by spatial constraints imposed by the ribosomal exit tunnel as well as self-interactions. The polypeptide's vectorial emergence from the ribosomal tunnel establishes the possible folding pathways leading to its native tertiary structure. How cotranslational protein folding and the rate of synthesis are linked to a protein's amino acid sequence is still not well defined. Here, we follow synthesis by individual ribosomes using dual-trap optical tweezers and observe simultaneous folding of the nascent polypeptide chain in real time. We show that observed stalling during translation correlates with slowed peptide bond formation at successive proline sequence positions and electrostatic interactions between positively charged amino acids and the ribosomal tunnel. We also determine possible cotranslational folding sites initiated by hydrophobic collapse for an unstructured and two globular proteins while directly measuring initial cotranslational folding forces. Our study elucidates the intricate relationship among a protein's amino acid sequence, its cotranslational nascent-chain elongation rate, and folding.**

ribosomes | cotranslational protein folding | protein synthesis | single molecule | optical tweezers

The overall rate of translation during synthesis is limited by decoding, peptidyl transfer, and translocation rates (1), and depends on several factors, such as the availability of tRNAs, amino acids, and translation factors, as well as the mRNA unwinding rate (2). In addition, interactions of the polypeptide with the ribosomal tunnel affect the speed of translation. If the rate of synthesis is greater than that of the exit tunnel emergence, bunching of the polypeptide occurs within the tunnel and the overall translation rate is reduced (3, 4). Folding of the nascent chain into its native structure is driven by free-energy minimization and occurs cotranslationally in a vectorial fashion, predominantly outside of the ribosomal tunnel (5, 6). Cotranslational folding within, as well as outside, the ribosomal exit tunnel exerts pulling forces on the nascent polypeptide that can prevent and even rescue translational stalling (7–9). During the early stages of peptide elongation, cotranslational folding compacts the nascent chain while it is still confined within the ribosome's exit tunnel (10). A polypeptide's amino acid sequence and interactions with itself and the solvent (solutes) determine its native folded structure. During synthesis, the spatial constraints within the ribosome's peptide tunnel define the number of possible cotranslational folding pathways (10). However, the connection of cotranslational folding and rate of synthesis to a protein's amino acid sequence is still not completely understood. The process of cotranslational protein synthesis and folding in the crowded environment of the cell is difficult to study due to the stochastic nature of mRNA translation and is usually investigated *in vitro*. These ensemble studies confirm that larger proteins collapse fast into a poly-globular conformation, fold through native-like intermediates in a distinct pathway, and emerge slowly as native structures (11, 12). With a few exceptions, stable tertiary structure formation requires the presence of hydrophobic amino acids (13, 14).

Single-molecule techniques have proven useful in the study of protein synthesis and subsequent folding (15–18), because these are asynchronous processes that are difficult to be observed using

ensemble methods. Optical tweezers have been used to observe stepping of motor proteins (19–23), DNA–protein complexes (24), as well as unfolding and refolding of RNA molecules and proteins (25, 26). This powerful single-molecule method has provided information on (i) the translation machinery by reporting on the strength of interactions between the ribosome and mRNA (27), (ii) its translocation along a short hairpin-forming mRNA molecule (28), as well as (iii) the release of an arrested nascent chain (7). The ribosome has been shown to modulate the folding rate of nascent chains (5, 26).

Dual-trap optical tweezers provide the necessary stability and sensitivity to isolate, confine, and measure the activity of individual macromolecules in their native environment, largely decoupling the system under study from environmental noise (29). Here, we observed synthesis and folding of single nascent polypeptides in the form of interbead distance variations using dual-trap optical tweezers in real time, while the individual nascent chains were held at various constant forces. Thus, the tension exerted on the nascent chain during the initial steps of cotranslational folding of synthesizing hydrophobic sequence stretches could be measured directly. It could also be shown that successive prolines caused transient translation pauses, while positively charged residues slowed the rate of synthesis under tension. Using this approach, we could observe the minimal hydrophobicity that was required by a nascent-chain section to overcome a certain applied tension,

## Significance

How proteins fold natively with efficient fidelity while being synthesized remains largely unexplored. Understanding protein synthesis on a single-molecule level is of particular interest to the life sciences and relevant for various diseases. Although protein synthesis and folding are well-studied subjects, cotranslational folding has been proven difficult to observe. Using optical tweezers, we measured the mechanics of synthesis and simultaneous folding in real time. We found that cotranslational folding occurs at predictable locations, exerting forces on the nascent polypeptide. Furthermore, we show that transient pauses and gradual slowing of translation occur in particular locations along the protein sequence, facilitating native secondary-structure formation. Thus, the rate of synthesis is inherently coupled to cotranslational folding, assuring reliable and fast native folding.

Author contributions: G.B. conceived the idea of the study; A.K., G.B., and M.H. designed research; F.W. and A.K. performed research; F.W. built the high-resolution optical tweezers and performed the tweezers experiments; A.K. designed and generated all constructs and performed the molecular biology; K.H.N., G.B., and M.H. supervised the study; F.W. and M.H. analyzed data; K.H.N. and G.B. contributed to the writing of the paper; and F.W., A.K., K.H.N., and M.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>Present address: Institute for Atomic and Molecular Physics (AMOLF), 1098 XG Amsterdam, The Netherlands.

<sup>2</sup>To whom correspondence may be addressed. Email: martin.hegner@tcd.ie or a.katranidis@fz-juelich.de.

<sup>3</sup>Deceased April 7, 2016.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617873114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617873114/-DCSupplemental).

initiating folding. It follows that this technique could be used to pinpoint the sequence location of a protein where cotranslational folding begins for a range of applied forces.

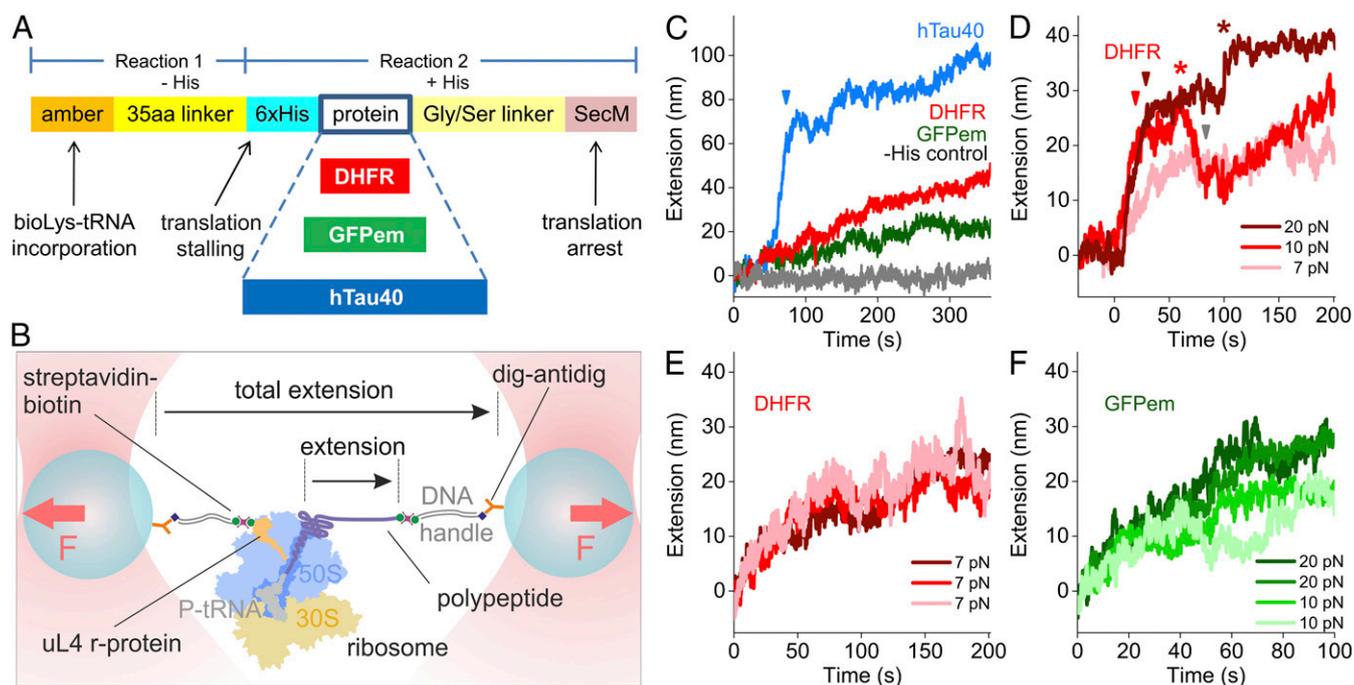
## Results

**Experimental Setup.** We used three DNA constructs for translation experiments, where the encoded protein under study was preceded by a N-terminal presequence and succeeded by a C-terminal linker. These constructs were synthesized by ribosomes biotinylated in vivo (30) at the uL4 ribosomal protein. The presequence consisted of an amber stop codon, followed by a 35-aa linker spanning the length of the ribosomal tunnel and six histidines (6xHis) (Fig. 1A). Adding a histidine-depleted cell-free transcription/translation system triggered synthesis of nascent chains up to the 6xHis tag, as confirmed in ensemble control experiments using the GFP variant Emerald (GFPem) construct (*SI Appendix, Fig. S1*). Concomitantly, biotin was cotranslationally incorporated at the N-terminal amber stop codon using the suppressor tRNA technique (31, 32) (Fig. 1A, reaction 1). Thus, the ribosome–nascent-chain complex (RNC) featured two biotin tags, one at the end of the stalled nascent chain, just appearing outside of the ribosomal tunnel and another biotin molecule linked to the uL4 ribosomal protein, as shown previously (*SI Appendix, Fig. S24*) (33). The RNC was tethered in situ between two optically trapped polystyrene beads (0.84  $\mu\text{m}$  in diameter) via two identical streptavidin-DNA handles (34) (0.33  $\mu\text{m}$  in length) in a microfluidic chamber (*SI Appendix, Fig. S3*). Both beads were held in orthogonally polarized optical traps of equal stiffness (35) ( $0.3 \pm 0.03$  pN/nm for all measurements) (Fig. 1B). Translation could be resumed by adding

a cell-free translation reaction mix containing His. The nascent polypeptide remained bound to the ribosome after synthesis due to the C-terminal SecM arrest peptide (AP) following the C-terminal linker (Fig. 1A, reaction 2).

We have previously shown that individual GFPem constructs folded natively while remaining bound to surface-tethered ribosomes (33). In addition, control measurements demonstrated that translation could be stalled efficiently at the His-tag, resuming as soon as the missing histidine was introduced with the transcription/translation mix (*SI Appendix, Fig. S1*). Further control experiments were performed to determine the influence of fluid flow on our measurements and to test His-tag stalling on the single-molecule level. Streptavidin-DNA handles were tested with and without stalled RNCs at a range of different forces (2–48 pN) (*SI Appendix, Fig. S4*). As demonstrated in these control experiments, injections had no impact on translation measurements due to the orthogonal fluid flow configuration. The RNCs introduced extra positional fluctuations, observable with the small bead size used here. These signature fluctuations were not seen in the DNA-only case. Thus, we could confirm whether or not a double tether with the correct length ( $\sim 680$  nm at 10 pN) included an RNC by performing a short constant force measurement before each translation measurement, comparing the measured positional noise to the two fits in *SI Appendix, Fig. S4B*.

**Full Protein Synthesis by Single Ribosomes Under Constant Force.** We followed the synthesis of an intrinsically disordered protein, hTau40 (36), and two globular proteins (*SI Appendix, Fig. S2 B*



**Fig. 1.** Experimental design and translation traces of proteins under different forces. (A) The three constructs used in this study were identically designed. Two structured proteins DHFR (187 aa, red) and GFPem (239 aa, green) and an intrinsically disordered protein hTau40 (441 aa, blue) were chosen. The SecM moiety ensured that the fully synthesized protein remained bound following translation. Reaction 1 incorporated biotin at the N-terminal amber stop codon halting synthesis at the His tag. (B) A constant-force  $F$  was applied on ribosome nascent-chain constructs, and reaction 2 was injected orthogonally to the measurement axis (*SI Appendix, Fig. S4*). The extension of the nascent polypeptide was measured during and after its synthesis. Not drawn to scale. (C) Comparison of typical translation traces for hTau40 (blue, 7 pN), DHFR (red, 10 pN), and GFPem (green, 10 pN), as well as a -His control (gray) (2 Hz). hTau40 showed the longest elongation free from stalling or folding (blue arrow) ( $n = 33$  hTau40 traces). (D) Comparison of translation trajectories at different forces of 7, 10, and 20 pN for the DHFR construct showing typical extensions between 20 and 40 nm. Red arrows mark a sudden change in extension rate during synthesis at 10 and 20 pN. Red stars mark partial unfolding (20 pN) and folding (10 pN) events. Upon partial folding under 10-pN tension, the DHFR construct acquires the same extension reached after synthesis at 7 pN (gray arrow) (2 Hz) ( $n = 35$  DHFR traces). (E) Reproducibility of translation trajectories of DHFR (7 pN). (F) Translation displacement traces (2 Hz) with the GFPem construct held at a constant force of 10 and 20 pN ( $n = 20$  GFPem traces) resulting in typical extensions of 15–30 nm.

and C), dihydrofolate reductase (DHFR) and the GFPem, under constant applied forces in the range of 7–20 pN (Fig. 1 C–F).

Measured displacement changes between the optically trapped beads during and after protein synthesis could be converted from elongation in nanometers to the number of translated amino acids for unstructured linear sections of a nascent chain by using the extensible worm-like chain model (eWLC) (37). The eWLC-derived nanometer-to-residue conversion is dependent on the applied forces (Table 1) and is valid for unstructured parts of the polypeptide chain.

During translation, hTau40's overall N- to C-terminal length gain was greater than that of DHFR and GFPem (all traces, Fig. 1C). Following injection of the reaction mix containing His, the His-tag stalled hTau40 construct exhibited an extension increase of ~98 nm in the measured interbead displacement at forces as low as 7 pN. The observed displacement changes corresponded to the expected length of the fully translated and unstructured hTau40 construct (491 residues 6×His-SecM), with an eWLC-derived nanometer-to-residue conversion factor of 0.2 nm/aa at 7-pN tension (Table 1).

On the contrary, the measured interbead displacement changes for both DHFR and GFPem were generally smaller than the expected length of the fully unfolded constructs. The DHFR construct showed displacement increases between  $19 \pm 5$  nm at 7 pN (95% of traces) and 30–40 nm (82%) up to  $76 \pm 3$  nm (18%) at 20-pN tension (Fig. 1D). DHFR's cotranslational N- to C-terminal extension was larger than the 15- to 30-nm displacement increase of the GFPem construct at forces between 10 and 20 pN (all traces, Fig. 1F), although DHFR's sequence is shorter than that of GFPem (SI Appendix, Fig. S5).

We conducted cotranslational measurements for the DHFR construct under different constant tensions between 7 and 20 pN and compared the observed bead displacement traces. Interestingly, one particular transition resulted in the same displacement change at different forces (SI Appendix, Fig. S6). At 7-pN tension, the nascent chain extended at a markedly lower rate of  $0.3 \pm 0.1$  nm/s than at higher forces, tapering off at ~20 nm from N- to C-terminal after ~80 s (Fig. 1E). For comparison, a fully folded DHFR protein is expected to have a N- to C-terminal extension of ~15 nm following translation, including the unstructured N-terminal and C-terminal linkers used here. By increasing the tension to 10–20 pN, the nascent chain extended at rates up to four times greater than at 7 pN and a sudden change in the measured extension rate took place at a later point, ~45 s after injection near sequence position 74 (red arrows in Fig. 1D and SI Appendix, Fig. S5). This was an indication that cotranslational folding was taking place continuously during synthesis at 7 pN, because increasing the tension to 10–20 pN appeared to counteract the folding forces. This allowed us to observe translation decoupled from folding for more than just a few seconds at higher applied

**Table 1. Fractional extension and length-to-residue conversion for extended polypeptides and dsDNA**

Force, pN	$x/L$ polypeptide	nm/aa polypeptide	$x/L$ dsDNA	nm/bp dsDNA
7	0.51	0.2	0.95	0.323
10	0.61	0.24	0.96	0.326
15	0.72	0.29	0.97	0.329
20	0.80	0.32	0.99	0.336

Here, the eWLC-derived fractional extension with the corresponding nanometer-to-residue conversion are shown for a number of different forces. These conversion factors are valid for unstructured parts of a polypeptide chain with a residue length of 0.4 nm/aa, a persistence length of 0.66 nm, and a stretch modulus of 200 pN (SI Appendix). The dsDNA mechanics are indicated with a base pair length of 0.34 nm/bp, a persistence length of 50 nm, and a stretch modulus of 1,000 pN (SI Appendix).

tensions. The sudden drop in extension rate between 20 and 25 nm during synthesis at 10- and 20-pN tension was followed by compaction at 10 pN and sudden partial unfolding at 20 pN (red stars in Fig. 1D). After compaction at 10 pN, DHFR reached a similar extension obtained after synthesis at 7 pN (gray arrow).

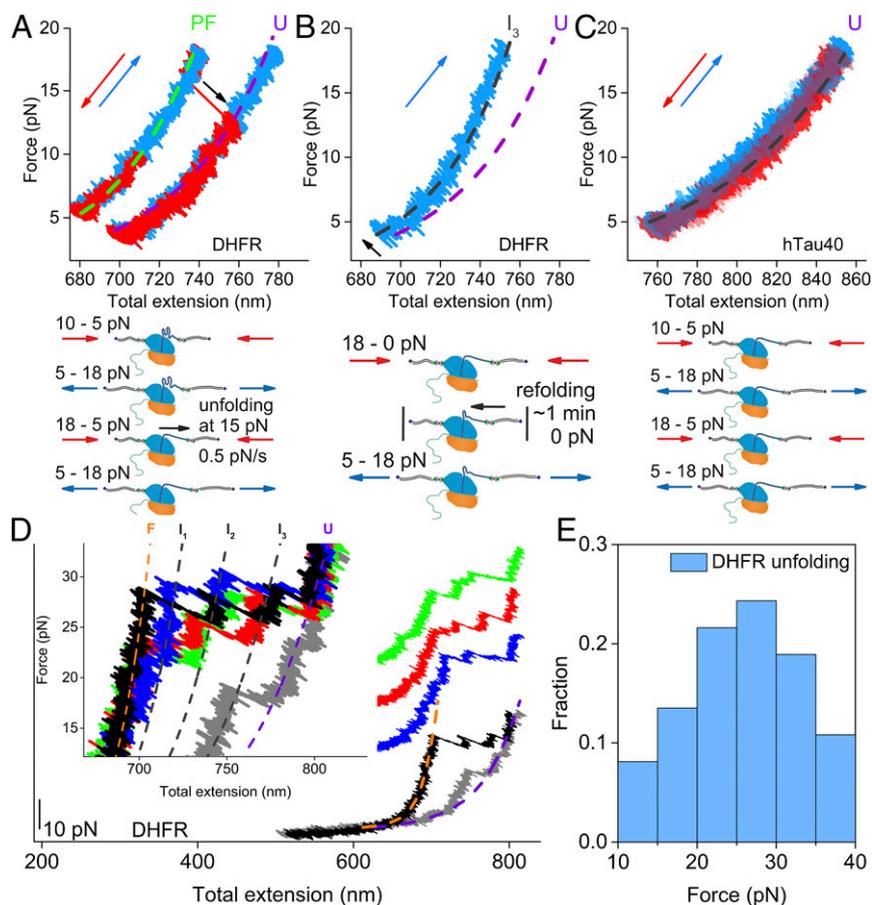
**Posttranslational Unfolding and Refolding of Proteins.** To investigate whether the DHFR construct was fully translated and partially folded during synthesis under tension, we conducted a number of posttranslational unfolding experiments immediately following synthesis under an applied force of 20 pN ( $N = 37$ ). By continuously varying the force between 0 and 40 pN with a loading rate of 6 pN/s, we observed unfolding from a partially folded state to DHFR's expected fully unfolded length (Fig. 2A and D), clearly indicating that DHFR partially folded while being fully translated. Relaxing the construct for ~1 min at 0 pN and subsequently increasing the force again revealed that the constructs refolded. The unfolding DHFR allowed resolving three intermediate unfolding steps, resulting in mean unfolding forces of ~26 pN ( $n = 37$ ; Fig. 2E). The intermediate states  $I_1$ ,  $I_2$ , and  $I_3$  correspond to N- to C-terminal extensions of 66, 128, and  $199 \pm 12$  aa, respectively. These displacement changes in the extension did not occur during our control experiments, where only DNA handles linked by a single streptavidin protein were present (SI Appendix, Fig. S7).

The fully unfolded DHFR construct, tethered by two 1-kbp dsDNA handles followed predictable eWLC behavior. This was demonstrated by overlaying the experimental force-extension data with a dsDNA eWLC function in series with a polypeptide eWLC function (Fig. 2, dashed purple curve and SI Appendix). Similar unfolding experiments were carried out after synthesis of the hTau40 construct ( $n = 17$ ). Here, we did not observe further unfolding, nor any refolding of the construct upon constant-rate force variations between 5 and 18 pN, because the protein was most likely already unfolded (Fig. 2C). A similar overlay of a combined eWLC function (Fig. 2C, dashed black curve, and SI Appendix) with an unstructured polypeptide contour length of 527 residues (hTau40 construct with linkers and SecM) demonstrated that it also matched the force-extension profile of the dsDNA-tethered, SecM-stalled, and unstructured hTau40 construct.

**Translation Rate and Stalling.** The longest observed translation pauses for unfolded segments of nascent polypeptides, lasting on the order of several seconds, coincided with sequence stretches featuring successive Pro residues downstream of several positively charged amino acids (Figs. 3 and 4 and SI Appendix, Figs. S8 and S9). In these instances, during incorporation of successive Pro, a number of positively charged Arg/Lys residues were confined inside the ribosomal tunnel, as illustrated in Fig. 3C and D. DHFR's fastest stretches of synthesis of 9–15 aa/s at tensions of 7- to 20-pN tension were observed within the first 10 s of translation (Figs. 3A and 4 D–F). The fastest observed translation rates of up to 16–22 aa/s were measured during synthesis of hTau40 constructs, at forces ranging from 7 to 20 pN (Figs. 3B and 4 A–C). In both constructs, they occurred along segments with low numbers of positively charged amino acids and Pro residues.

The nature of the amino acids preceding and following a successive proline sequence position also appeared to play a role. As shown in the Inset of Fig. 3B and SI Appendix, Fig. S8B, the first Pro–Pro incorporation into hTau40's sequence at <sup>158</sup>APPG<sup>161</sup> took longer than that of the following Pro–Pro incorporations in the linearly extended (unfolded) regime (white background in Fig. 4): <sup>175</sup>TPPA<sup>178</sup>, <sup>181</sup>TPPS<sup>184</sup>, <sup>187</sup>EPPK<sup>190</sup>, and <sup>217</sup>TPPT<sup>220</sup> (SI Appendix, Fig. S5).

**Cotranslational Nascent-Chain Compaction.** As seen in Fig. 3A and SI Appendix, Figs. S8A and S9, the first instance of cotranslational compaction for the DHFR construct, being synthesized at a tension of 10–20 pN, resulted in a sudden change in the rate of bead displacement at around sequence position 74, near a cluster



**Fig. 2.** Unfolding/refolding of newly synthesized proteins. (A) Partial unfolding of DHFR (directly after translation under 10-pN tension to an extension of  $136 \pm 12$  aa; green eWLC curve). Red data/arrows represent decreasing, and blue data/arrows denote increasing the tension on the partially folded intermediate. (0.5 pN/s, 1 kHz) As illustrated, the construct unfolded at 15 pN by  $142 \pm 12$  aa to  $272 \pm 12$  aa (black arrow, purple eWLC curve), corresponding to the fully unfolded length of DHFR (including Gly/Ser linker and SecM). Continued 0.5 pN/s force variation between 5 and 18 pN did not result in immediate refolding. (B) The tension on this unfolded construct was subsequently lowered to 0 pN. After  $\sim 1$  min at 0 pN, increasing the force (blue data/arrow, 10 pN/s) revealed partial refolding to an N- to C-terminal extension of  $206 \pm 12$  aa (gray eWLC curve  $I_3$ ). (C) hTau40 construct measurement initiated directly after translation at 10-pN tension to a total length of  $527 \pm 15$  aa (eWLC curve, including linkers and SecM), several minutes after reaction 2 (Fig. 1A) injection. The force was continuously varied between 5 and 18 pN (0.5 pN/s, 1 kHz) as illustrated. No folding or unfolding was observed for hTau40. (D) DHFR construct measurements starting directly after translation under tension of 20 pN. The force was continuously varied between 0 and 40 pN. The newly synthesized DHFR constructs were allowed to refold at 0 pN and subsequently extended at an extension rate of 6 pN/s, showing native-like unfolding at forces between 20 and 30 pN, featuring three intermediate states  $I_1$ – $I_3$ . These intermediate states correspond to N- to C-terminal extensions of 66, 128, and 199 aa, respectively. The fully unfolded state U has an extension of 272 aa. (E) Histogram of measured unfolding forces of newly synthesized, SecM-stalled, and refolded DHFR constructs, showing mean unfolding forces of 26 pN ( $n = 37$ ; bin size, 5 pN). Extensions shown here correspond to measured end-to-end distance of the newly synthesized proteins (including DNA handles).

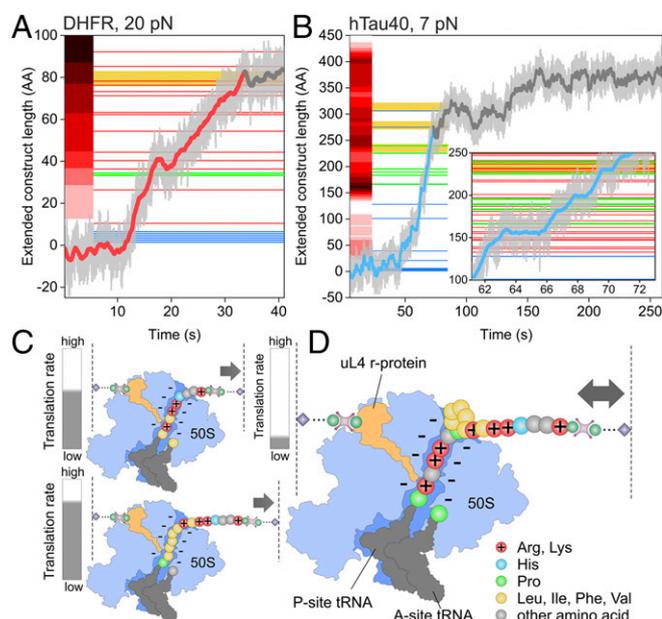
of hydrophobic residues composed of Leu, Ile, and Val. After partial cotranslational folding had occurred, we did not observe further fast increases of bead–bead displacements even though translation likely continued under the applied tension. The observed equilibrium position after synthesis of position 74 was probably the result of synchronous compaction during elongation due to folding (Fig. 3D), as demonstrated by posttranslational unfolding to the full length of the construct (Fig. 2A and D).

In the N-terminal half of hTau40's sequence, amino acids with lower hydrophobicity, such as Ala and Gly, form the bulk of all hydrophobic residues. The overall hydrophobicity of the C-terminal half of the hTau40 sequence starting at Val<sup>226</sup> is greater, because Ala and Gly become rarer. During translation of the hTau40 construct at a tension of 7–10 pN, we observed cotranslational compaction starting from sequence positions 228–307 (84% of traces at 7 pN, 71% at 10 pN) (Figs. 3B and 4 B and C, and *SI Appendix*, Figs. S8B and S10A).

In previous ensemble folding experiments, it had been shown that hydrophobic collapse is the first step in folding of proteins,

occurring before the formation of significant secondary structure (38, 39). A gaugeable definition of hydrophobic collapse probability as a function of sequence position was required to quantitatively describe the initiation of cotranslational compaction that was observed as a sudden change in the measured bead displacement during synthesis of highly hydrophobic segments (yellow horizontal bars, Fig. 3 and *SI Appendix*, Fig. S8) followed by slowed extension and/or shortening due to cotranslational nascent compaction. These compacted states could be unfolded posttranslationally as shown in Fig. 2.

The presence of a minimum number of closely spaced hydrophobic residues, from six to seven, is sufficient to induce cotranslational long-distance folding initiated by hydrophobic collapse in the absence of externally applied forces (14). Thus, the sequence positions where compaction is likely to take place depend on the number of hydrophobic residues along the sequence and their corresponding hydrophobicity, as well as the relative distances between them. The likelihood of cotranslational folding for a given sequence stretch can be expressed as the



**Fig. 3.** Translation trajectories of individual proteins. (A) DHFR translation trajectory at 20 pN (light gray, 1 kHz; red/gray overlay, 2 Hz; 0.32 nm/aa) showing stalling in regions with successive Pro–Pro (green lines) and slowing in regions with positively charged Arg, Lys (red lines). The red shaded zones depict the number of Arg, Lys confined within the ribosomal tunnel for an extended nascent chain (red curve, *SI Appendix*, Fig. S9). His (blue lines) did not significantly contribute to slowing here. Gray overlay depicts cotranslational compaction following translation of strongly hydrophobic segment indicated by the yellow horizontal bars (1 of  $n = 35$  DHFR traces). (B) Translation trace of hTau40 (7 pN; light gray, 1 kHz; blue/gray overlay, 2 Hz; 0.2 nm/aa) demonstrating stalling at Pro–Pro and subsequent slowing after multiple positively charged amino acids were incorporated (1 of  $n = 33$  hTau40 traces). Color coding as in A. The *Inset* magnifies the trajectory segment where Pro–Pro stalling occurred during translation. (C) Unfolded sequence stretches rich in positively charged residues within the ribosomal tunnel featured lower translation rates, resulting in slower observed increasing bead displacements (arrow) than segments with low Arg/Lys (His) ribosomal tunnel occupancy. (D) Prolonged stalling occurred during Pro–Pro incorporation concurring with a high positively charged residue ribosomal tunnel occupancy in DHFR and hTau40. Hydrophobic collapse followed by cotranslational compaction could cease or momentarily reverse observable increases in bead-to-bead displacements (arrow) despite polypeptide growth.

sequence-dependent hydrophobic collapse index (HCI) (*SI Appendix*) (14). The magnitude of the applied force on the growing nascent chain also influences cotranslational nascent chain compaction due to folding. The greater the applied force, the more closely spaced hydrophobic residues would be required to overcome it. A force-dependent threshold value,  $HCI_s(f)$ , predicting a sequence position,  $s$ , where cotranslational hydrophobic collapse is likely to occur, was estimated as shown in Fig. 5A. Segments with a HCI value greater than  $HCI_s(f)$  were identified as potential cotranslational folding sites, fitting with the positions where cotranslational compaction was observed.

**Translation Rate Model with Estimation of Force-Dependent Cotranslational Folding.** To describe the optical tweezers-observable translation rate  $\dot{x}_i(t)$ , as a function of sequence position  $i$ , we introduced the following simple model:

$$\dot{x}_i(t, f) = C_i(f) k \dot{x}_0(t) \exp \left[ l \rho_i + m N_i^{R,K} + n N_i^H \right], \quad [1]$$

with a scaled maximum translation rate  $k \dot{x}_0(t)$  and a nascent chain compaction factor  $C_i(f)$  (*SI Appendix*). The term  $\rho_i$  represents Pro–Pro positions along the sequence, whereas the terms  $N_i^{R,K}$

and  $N_i^H$  represent the approximate number of positively charged residues Arg/Lys and His, respectively, which can be found inside the ribosomal tunnel at position  $i$ , assuming a tunnel length of 30 extended residues (9, 40). Coefficients  $l$ ,  $m$ , and  $n$  determine the degree of stalling for  $\rho_i$ ,  $N_i^{R,K}$ , and  $N_i^H$ , respectively.

Fitting this simplified translation model to DHFR translation traces suggested that hydrophobic collapse and continued compaction could occur cotranslationally near sequence positions 9, 51, and 74 (Fig. 5B–D, yellow horizontal bars), where the respective HCI values peak at 0.52, 0.53, and 0.61, respectively (Fig. 5A). Below 10 pN, cotranslational folding could occur early on near position 9, resulting in GFPem-like translation traces (*SI Appendix*, Fig. S10B). At 10- and 20-pN tension, the construct extended in an unstructured fashion during synthesis of the first 70–80 residues, collapsing near the HCI peak of 0.61 at sequence position 74 (Fig. 5B and D). After that point, no further bead displacement changes decoupled from folding were observed, even though translation likely continued. In some cases, DHFR did not exhibit cotranslational collapse at 20-pN tension, extending to its full contour length (18% of traces, Fig. 5C). The cotranslational collapse at applied forces of 10–20 pN compacted the nascent chain at the strongly hydrophobic segment  $^{68}\text{LKDRINIVLSREL}^{80}$  (Fig. 5B and D, lowest dark yellow horizontal bar), which coincides with the sequence position of DHFR's fourth native  $\beta$ -sheet  $^{72}\text{INIVL}^{76}$ . At 20-pN tension, the construct shown in Fig. 5B partially unfolded posttranslationally after 93 s by  $23 \pm 9$  residues (3.2 Å/residue, *SI Appendix*). Because folding in GFPem already took place near sequence position 15 (HCI = 0.72, Fig. 5A), even at 20 pN, there were no segments decoupled from folding long enough for adequate fitting of the rate model.

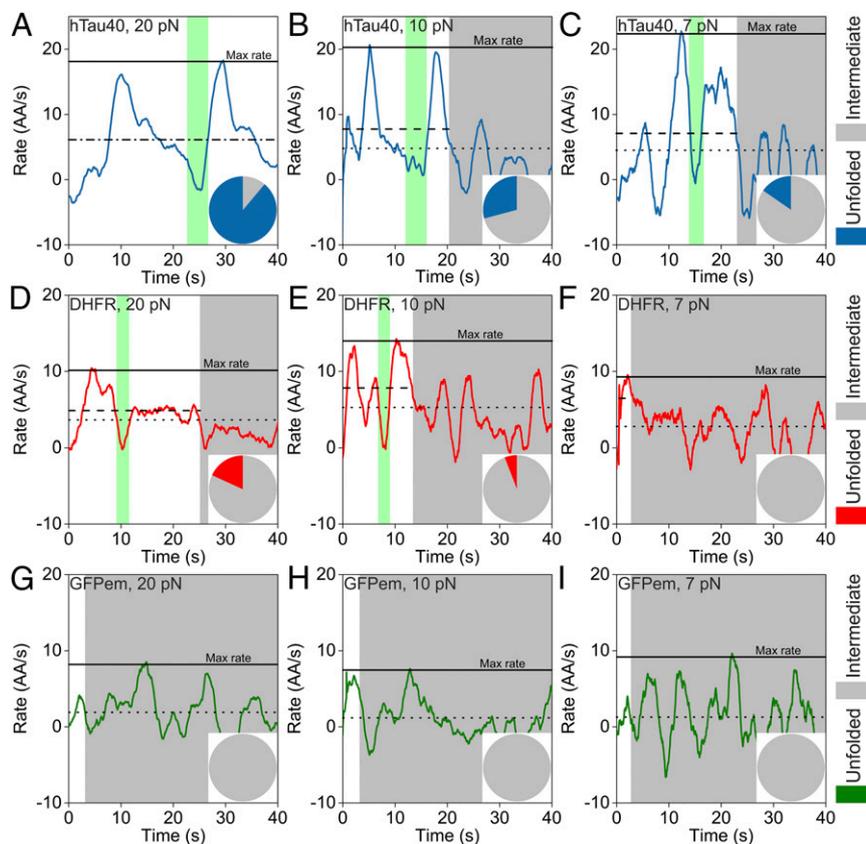
Fitting the same model to hTau40 translation traces indicated that nascent-chain compaction at 7 pN initiated between sequence positions 228–307 near the three HCI peaks of 0.48 (Fig. 5E and F, light-yellow horizontal bars), where transient  $\beta$ -sheet formation had been previously observed between positions 224 and 315 in NMR bulk studies (36) (*SI Appendix*, Fig. S5). Following translation, the construct unfolded to its full unstructured length under the applied tension, as demonstrated in Fig. 5E and *SI Appendix*, Fig. S10A. The lack of a strong hydrophobic core forming during translation coupled with a fast initial translation rate could prevent the occurrence of stable native contacts. This would explain the limited formation of secondary structure and the lack of a well-defined tertiary structure for hTau40 (36).

For the measured bead displacement traces of DHFR and hTau40, we numerically determined the negative stalling coefficients  $l$ ,  $m$ , and  $n$  of the model for the observed translation rate as a function of the sequence position  $i$ . In all fits, these coefficients were found to be negative, where  $|l| > |m| > |n|$  (Table 2). Thus, individual Pro–Pro positions had the greatest immediate retarding effect on the observed translation rate, followed by individual positive residues within the ribosomal tunnel, with His having only a minor contribution. Of course, these coefficients also depend on other situation-specific factors (*SI Appendix*).

## Discussion

The single-molecule experimental setup depicted in Fig. 1 allowed us to gently usher the nascent polypeptide out of the ribosomal tunnel during translation. Depending on the amino acid sequence being synthesized, the observed bead-to-bead displacement changes were a function of either nascent polypeptide growth, partial folding and unfolding of the newly translated polypeptide chains, or a combination of both synthesis and folding. Increasing the tension on the nascent chain decoupled synthesis from folding for initial segments of hTau40 and DHFR. The applied forces were low enough to preserve subtle dynamics in our measurements that would be hidden from bulk experiments.

We investigated three cases in our study where DHFR was best suited to the constant cotranslational force range of 7–20 pN used



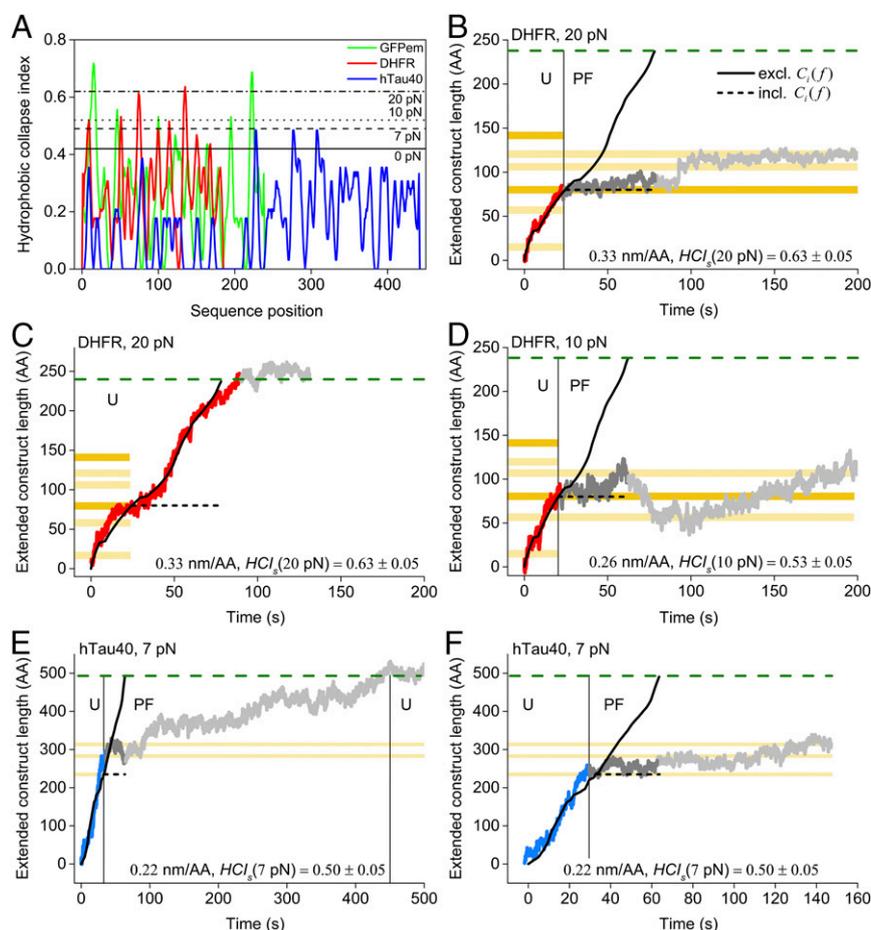
**Fig. 4.** The effect of force on translation rates. (A–I) The translation rate traces were determined by averaging the 1-kHz acquired bead displacement data to 0.5 Hz, and then taking the first derivative smoothed with a 2-s adjacent averaging window. The maximum translation rate (full horizontal line), average rate for translation before initial compaction (dashed line), and average rate over 40 s with compaction (dotted line) are shown. Pro–Pro stalling during translation free from compaction is highlighted with a green background. Pie charts depict the fraction of nascent chains remaining unfolded (colored) and folded (gray) during synthesis under tension. Increased tension on the nascent chain during synthesis reduces the average rate of translation before initial compaction for hTau40 and DHFR ( $n = 61$ ). Because GFPem already folds at the N-terminal under the tensions used here, only its maximum rate and overall average rate convoluted with folding/unfolding could be determined.

here. The other two constructs represented two extremes. GFPem remained compacted and never unfolded, whereas hTau40 remained mostly unfolded at these forces. DHFR, on the other hand, exhibited dynamic behavior, remaining compacted at 7 pN after synthesis of the first HCI peak (*SI Appendix, Fig. S10B*), while unfolding cotranslationally at 20 pN for instance.

Proteins with a well-defined hydrophobic core, such as DHFR and GFPem, showed a stronger tendency to fold cotranslationally than the intrinsically disordered hTau40 construct, a protein lacking in hydrophobic residues. The natively very compact ( $\beta$ -barrel) GFPem construct requires considerably high forces (100–600 pN) to unfold (41). During our measurements, it assumed a shorter, energetically more favorable conformation under tension than DHFR, despite being a longer construct. Under tension between 7 and 20 pN, the proteins never fully folded into their lowest native free-energy state, which would correspond to a total increase in the interbead displacement of  $\sim 15$  nm including the peptide linker (*SI Appendix, Fig. S5*). Instead, they folded partially, evident from the longer-than-native N- to C-terminal extension that they adopted during translation. Given enough time under tension, DHFR and hTau40 could be fully unfolded to their expected unstructured lengths (Figs. 2 and 5 *C* and *E*). Both SecM-stalled unstructured constructs followed eWLC chain behavior, evident from the overlays in Fig. 2. Posttranslational partial refolding was only possible for DHFR, although it took up to a minute without tension. This is in agreement with previous observations (26), showing that the fully translated nascent polypeptide is unable to refold quickly once

unfolded, probably due to its close proximity to the negatively charged rRNA backbone of the ribosome. Electrostatic interactions may slow long-range refolding times considerably (26).

Peptide bond formation between two consecutive Pro residues takes place at a much lower rate than between Pro and other amino acids (42). In addition, it had been shown previously that XPP motifs preceded by one of the following residues P, D, and A led to strong stalling. The same was also observed for PPX motifs followed by P, W, D, N, and G (43). Therefore, hTau40's  $^{158}\text{APPG}^{161}$  position should feature the strongest stalling, in full agreement with our observations (Fig. 3*B* and *SI Appendix, Fig. S8B*). Although the incorporation of individual successive Pro residues had a greater immediate retarding effect on the rate of synthesis than individual positively charged residues, apparent from  $|l| > |m| > |n|$  (Table 2), each positively charged residue contributed to slowing throughout its traversal of the ribosomal tunnel. Several successive positively charged residues along a sequence offered a synergistic, additive effect to the slowing of the overall translation rate over a relatively long sequence stretch. Codon use and moderately stable mRNA structures appeared to play a minor role in stalling (*SI Appendix, Fig. S11*); in accordance with Charneski and Hurst (3). The average rate of synthesis of hTau40 and DHFR before initial compaction decreased with greater applied tension on the nascent chain (Fig. 4). At an applied force between 7 and 10 pN, the average rate of synthesis was  $7 \pm 1$  aa/s for both constructs. When the tension was increased to 20 pN, however, the average rate before initial compaction



**Fig. 5.** (A) Prediction of hydrophobic collapse for DHFR (red), GFPem (green), and hTau40 (blue), considering only the hydrophobic amino acids Leu, Ile, Val, and Phe. Black horizontal lines highlight four estimated force-dependent values of  $HCl_s(f) = 10^{-2}f + 0.43 (\pm 0.05)$  at sequence positions, where hydrophobic collapse is likely to occur at 0, 7, 10, and 20 pN (*SI Appendix*). (B–F) Translation rate model fits (black curves, Eq. 1) for DHFR and hTau40 with (dashed) and without (full) the HCl term  $C_i(f)$ , considering positive residue–rRNA electrostatic interactions and successive Pro–Pro stalling. U indicates the unstructured state (colored curve; synthesis decoupled from folding), and PF indicates the partially folded (dark gray) state of the growing polypeptide. Light gray curve depicts unfolding of the SecM-stalled polypeptide. Dark-yellow and light-yellow bars represent unfolded segments with HCl values of 0.55–0.65 and 0.45–0.55, respectively. The dashed green line denotes the predicted length of the fully translated and unstructured construct under tension, as determined with eWLC model (*SI Appendix*). (B) Fitting Eq. 1 to a 20-pN DHFR trace predicted a translation time of  $\sim 78$  s. Compaction occurred cotranslationally after synthesis of a strongly hydrophobic segment (HCl peak 3 in A). (C) A DHFR translation extension curve at 20 pN without cotranslational compaction of the nascent chain. Translation terminates at the SecM AP after 81 s. (D) Overall predicted translation time of  $\sim 63$  s at 10 pN, featuring cotranslational compaction at HCl peak 3 (A and B). (E) hTau40 trace at 7 pN, fit resulted in an estimated overall translation time of  $\sim 64$  s. Cotranslational nascent chain compaction was observed after synthesis of strongly hydrophobic segment 228–307 (HCl peaks in A). (F) A synthesis time of  $\sim 65$  s for hTau40 at a tension of 7 pN resulted from fitting the model to a translation trace showing cotranslational compaction after synthesis of the sequence containing HCl peak at position 228.

dropped to  $5 \pm 1$  aa/s. Although not inhibiting peptide bond formation, the increased force exerted on the nascent chain applied here seems to slow synthesis. Similar forces would also occur during folding within the tunnel (8). This feedback

mechanism could serve as a folding aid, providing extra time for native contacts to form during synthesis.

DHFR folded cotranslationally at the same extremely hydrophobic segment (Figs. 3A and 5B and D, and *SI Appendix*, Figs. S8A and S9) at 10 and 20 pN, indicating that the observed folding was initiated by hydrophobic collapse. Because similar cotranslational folding at a very hydrophobic segment was also observed for GFPem (Figs. 1F and 4G–I) and hTau40 (Figs. 3B and 5E and F, and *SI Appendix*, Fig. S8B), it was evident that initial folding here was also driven by hydrophobic collapse. These observations agree with a multitude of previous ensemble studies citing hydrophobic collapse as the first step in folding (11, 12, 38, 39). Subsequently, the nascent chain can self-interact and seems to prefer to stay compact and close to the ribosomal exit tunnel rather than elongated, resulting in a “stalled” bead-to-bead distance. A pictogram providing a graphical interpretation of these different contributing factors is shown in Fig. 3C and D.

**Table 2. Fitting parameters for the translation rate model**

Fitting parameters		DHFR	hTau40
Pro–Pro dependent	$l$	$-2.6 \pm 0.5$	$-2.4 \pm 0.5$
Arg/Lys dependent	$m$	$-0.21 \pm 0.05$	$-0.17 \pm 0.04$
His dependent	$n$	$-0.05 \pm 0.02$	$-0.06 \pm 0.02$

Numerically determined negative coefficients  $l$  (stalling due to Pro–Pro peptide bond formation),  $m$  [stalling due to the number of Arg/Lys residues confined within the ribosomal tunnel (RT)], and  $n$  (stalling due to the number of His residues confined within the RT) for the DHFR and hTau40 translation trajectories (Fig. 4), assuming a RT length of 30 residues. A scaled maximum burst rate of  $k\dot{x}_0(t) = 5 \text{ nm}\cdot\text{s}^{-1}$  was chosen for all fits (*SI Appendix*).

We determined the HCI values for the GFPem sequence and identified position 15 as a probable initiation site for cotranslational folding (Fig. 5A). Increasing the tension to 20 pN was not enough to counteract the early occurring compaction of the nascent chain. Higher constant forces could not be used to avoid rupturing of the tethers (33). Nevertheless, we estimated that the HCI peak of 0.72 was high enough that cotranslational folding would take place even against a counteracting tension of  $\sim 30$  pN. Therefore, the interbead displacement trace of GFPem synthesis was never decoupled from folding.

At 7-pN tension, DHFR sequence positions 9 and 51 were recognized as potential cotranslational folding sites with HCI values of 0.52 and 0.53, respectively. These positions correspond to the first and third native  $\beta$ -sheets of DHFR. By increasing the tension to 10–20 pN, the site of first cotranslational nascent chain compaction was moved to position 74 with a HCI value peak at 0.61 (fourth native  $\beta$ -sheet of DHFR, Fig. 5B–D and *SI Appendix, Fig. S10B*). The transient Pro–Pro pause near position 27, observed during translation at 10- to 20-pN tension, could provide extra time for N-terminal secondary-structure formation after compaction of the hydrophobic segment with HCI peak at position 9. Likewise, the Pro–Pro stalling predicted at position 84 would also allow for more time to properly fold the N-terminal half after rapid translation and collapse of the segment containing HCI peaks 2 and 3 at positions 51 and 74, respectively. The positive residues in DHFR's N-terminal domain gradually slow translation (Fig. 3A), particularly after synthesis of the third HCI peak, enabling equilibrium-like sampling of the ribosomal tunnel-restricted conformational space (10, 44).

Similarly, HCI values determined for hTau40 gave three possible hydrophobic collapse sites. Interestingly, all three sites coincide with the positions of three transient  $\beta$ -sheets at the C-terminal half of hTau40. Following the initial rapid translation of the N-terminal domain, Pro–Pro stalling could facilitate the formation of long-range contacts between the N terminus and middle domain (*SI Appendix, Fig. S10A*). Our results corroborate that pauses during synthesis coordinate translation kinetics with the cotranslational folding of single domains (45).

Although the HCI values of a certain sequence stretch serve as an indicator for the likelihood of cotranslational folding along the sequence of a protein, it is not a definitive predictor. This is evident from the DHFR and hTau40 trajectories in Fig. 5 that show continued N- to C-terminal extension of the nascent polypeptide chain following synthesis of strongly hydrophobic sequence stretches. Thus, the HCI peaks determined for DHFR and hTau40 are just reaching the threshold hydrophobicity required for cotranslational folding to occur under tension between 7 and 20 pN. Because GFPem never unfolds under tension, even during posttranslational pulling experiments, its HCI peak is well above the threshold limit required for cotranslational folding to occur under tension at the applied range of constant forces.

Initial cotranslational hydrophobic collapse into a compact state likely occurs already inside the ribosomal tunnel, close to the exit vestibule (9, 10). The folding forces during initial compaction observed in our study were on the order of  $\sim 7$  pN for hTau40 during

collapse of its second-half hydrophobic patch and 10–20 pN for DHFR's two HCI peaks between sequence positions 50 and 100. GFPem's folding forces exceeded 20 pN. Although the forces applied to the GFPem nascent chain during translation were limited to 20 pN, the folding forces pulling on the nascent chain within the ribosomal tunnel could have been greater during the initial collapse of the N-terminal sequence stretch with HCI value of 0.72 (Fig. 5A). A minimal HCI value of  $0.43 \pm 0.05$  required for cotranslational folding without an applied tension was extrapolated from the apparent linear relationship between HCI and applied force (Fig. 5A). This would serve as a baseline for determining cotranslational folding sites of other proteins in vivo.

In summary, single-molecule analyses by dual-trap optical tweezers revealed cotranslational features of translation rate and protein folding, namely (i) the correlation of the number of positive amino acids and Pro–Pro locations with the translational rate, resulting in translation trajectories consisting of rapidly translating sequence stretches with intermittent pauses; (ii) evidence that stalling during synthesis provides extra time for the cotranslational formation of long-range contacts in N-terminal domains; (iii) the relationship between the density of strictly hydrophobic amino acids along the sequence and sites of possible cotranslational folding; and (iv) the magnitude of the forces exerted on the nascent chain during initial hydrophobic collapse. The forces exerted on the nascent chain during cotranslational folding also prevent stalling, while slowing the rate of synthesis at tensions above 10 pN. Not only does the genetic code contain the information for native folding, it dictates the relative speed of translation, assuring efficient cotranslational folding with high fidelity. Cotranslational folding itself applies a tension on the nascent chain within the ribosomal tunnel that in turn prevents unscheduled stalling events. This coupling of rate and folding would ensure optimal synchronized translation in a polysome complex.

## Methods

His-tag stalled RNCs were coupled to beads in a multistage reaction. First, ribosomes were biotinylated (46, 47) and coupled to streptavidin-DNA (1,000 bp)-covered beads (34). Synthesis of the biotinylated 35-aa linker was then carried out with the desired construct sequence using a customized version of the PURE cell-free transcription/translation system (48) (PURExpress  $\Delta$  amino acids, histidine, tRNA, ribosomes; NEB E3315Z; New England Biolabs). The experiments were conducted with a custom dual-trap optical tweezers instrument using backfocal interferometry with differential detection (29). Trap stiffness was kept at  $\sim 0.3$  pN/nm, and data were acquired at a rate of 62.5 kHz. The translation reaction occurred at  $\sim 26^\circ\text{C}$  in TICO buffer [20 mM Hepes-KOH, pH 7.6, 6 mM  $(\text{Ac}_2\text{Mg})$ , 30 mM  $\text{AcNH}_4$ , 4 mM  $\beta$ -mercaptoethanol]. A detailed description of the methods used is given in *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank Prof. E. Mandelkow (Deutsches Zentrum für Neurodegenerative Erkrankungen) for providing us with the gene for the hTau40 and Prof. M. P. Deutscher (Miller School of Medicine, University of Miami) for sending us the Can20/12E *E. coli* strain. We also thank Prof. S. Sen and Prof. P. Voorheis (Trinity College Dublin) for fruitful discussions and Carlos Bustamante (University of California, Berkeley) and his research team for critical reading and suggestions. The work was supported by the Helmholtz Association (Research Center Jülich) of German Research Centers (A.K. and G.B.), by Science Foundation Ireland under the principal investigator schemes SFI 09IN/1B2623 and 15/IA/3023 (for M.H. and F.W.), and from 5Top100 program of the Ministry for Science and Education of Russia (G.B.).

- Shoji S, Walker SE, Fredrick K (2009) Ribosomal translocation: One step closer to the molecular mechanism. *ACS Chem Biol* 4:93–107.
- Gloge F, Becker AH, Kramer G, Bukau B (2014) Co-translational mechanisms of protein maturation. *Curr Opin Struct Biol* 24:24–33.
- Charneski CA, Hurst LD (2013) Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol* 11:e1001508.
- Lu J, Deutsch C (2014) Regional discrimination and propagation of local rearrangements along the ribosomal exit tunnel. *J Mol Biol* 426:4061–4073.
- Hsu ST, et al. (2007) Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci USA* 104:16516–16521.
- Kim SJ, et al. (2015) Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science* 348:444–448.
- Goldman DH, et al. (2015) Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* 348:457–460.
- Nilsson OB, Müller-Lucks A, Kramer G, Bukau B, von Heijne G (2016) Trigger factor reduces the force exerted on the nascent chain by a cotranslationally folding protein. *J Mol Biol* 428:1356–1364.
- Nilsson OB, et al. (2015) Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep* 12:1533–1540.
- Holtkamp W, et al. (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science* 350:1104–1107.
- Udgaonkar JB (2013) Polypeptide chain collapse and protein folding. *Arch Biochem Biophys* 531:24–33.
- Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW (2013) Folding of a large protein at high structural resolution. *Proc Natl Acad Sci USA* 110:18898–18903.

13. Lodish H, Berk A, Zipursky SL, Matsudaira z, Baltimore D, Darnell J (2000) Hierarchical structure of proteins. *Molecular Cell Biology* (Freeman, New York), 4th Ed, Chap 3.1.
14. Sen S, Voorheis HP (2014) Protein folding: Understanding the role of water and the low Reynolds number environment as the peptide chain emerges from the ribosome and folds. *J Theor Biol* 363:169–187.
15. Rhoades E, Gussakovskiy E, Haran G (2003) Watching proteins fold one molecule at a time. *Proc Natl Acad Sci USA* 100:3197–3202.
16. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
17. Chen J, et al. (2014) Dynamic pathways of –1 translational frameshifting. *Nature* 512:328–332.
18. Kapanidis AN, Strick T (2009) Biology, one molecule at a time. *Trends Biochem Sci* 34:234–243.
19. Asbury CL, Fehr AN, Block SM (2003) Kinesin moves by an asymmetric hand-over-hand mechanism. *Science* 302:2130–2134.
20. Shaevitz JW, Abbondanzieri EA, Landick R, Block SM (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* 426:684–687.
21. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* 438:460–465.
22. Greenleaf WJ, Woodside MT, Block SM (2007) High-resolution, single-molecule measurements of biomolecular motion. *Annu Rev Biophys Biomol Struct* 36:171–190.
23. Svoboda K, Schmidt CF, Schnapp BJ, Block SM (1993) Direct observation of kinesin stepping by optical trapping interferometry. *Nature* 365:721–727.
24. Grange W, et al. (2008) VirE2: A unique ssDNA-compacting molecular machine. *PLoS Biol* 6:e44.
25. Liphardt J, Onoa B, Smith SB, Tinoco I, Jr, Bustamante C (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science* 292:733–737.
26. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, Jr, Bustamante C (2011) The ribosome modulates nascent protein folding. *Science* 334:1723–1727.
27. Uemura S, et al. (2007) Peptide bond formation destabilizes Shine-Dalgarno interaction on the ribosome. *Nature* 446:454–457.
28. Wen JD, et al. (2008) Following translation by single ribosomes one codon at a time. *Nature* 452:598–603.
29. Bustamante C, Chemla YR, Moffitt JR (2009) High-resolution dual-trap optical tweezers with differential detection: Instrument design. *Cold Spring Harb Protoc* 2009:ipb73.
30. Katranidis A, et al. (2009) Fast biosynthesis of GFP molecules: A single-molecule fluorescence study. *Angew Chem Int Ed Engl* 48:1758–1761.
31. Kobayashi T, et al. (2005) Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion. *Proc Natl Acad Sci USA* 102:1366–1371.
32. Ellman J, Mendel D, Anthony-Cahill S, Noren CJ, Schultz PG (1991) Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods Enzymol* 202:301–336.
33. Katranidis A, et al. (2011) Force measurements of the disruption of the nascent polypeptide chain from the ribosome by optical tweezers. *FEBS Lett* 585:1859–1863.
34. Jadhav VS, Brüggemann D, Wruck F, Hegner M (2016) Single-molecule mechanics of protein-labelled DNA handles. *Beilstein J Nanotechnol* 7:138–148.
35. Moffitt JR, Chemla YR, Izhaky D, Bustamante C (2006) Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc Natl Acad Sci USA* 103:9006–9011.
36. Mukrasch MD, et al. (2009) Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol* 7:e34.
37. Hegner M, Smith SB, Bustamante C (1999) Polymerization and mechanical properties of single RecA-DNA filaments. *Proc Natl Acad Sci USA* 96:10109–10114.
38. Sadqi M, Lapidus LJ, Muñoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci USA* 100:12117–12122.
39. Lapidus LJ, et al. (2007) Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. *Biophys J* 93:218–224.
40. Voss NR, Gerstein M, Steitz TA, Moore PB (2006) The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* 360:893–906.
41. Dietz H, Berkemeier F, Bertz M, Rief M (2006) Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci USA* 103:12724–12728.
42. Chevance FF, Le Guyon S, Hughes KT (2014) The effects of codon context on in vivo translation speed. *PLoS Genet* 10:e1004392.
43. Peil L, et al. (2013) Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc Natl Acad Sci USA* 110:15265–15270.
44. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM (2011) New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc* 133:513–526.
45. Zhang G, Hubalewska M, Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 16:274–280.
46. Rheinberger HJ, Geigenmüller U, Wedde M, Nierhaus KH (1988) Parameters for the preparation of *Escherichia coli* ribosomes and ribosomal subunits active in tRNA binding. *Methods Enzymol* 164:658–670.
47. Zaniewski R, Petkaitis E, Deutscher MP (1984) A multiple mutant of *Escherichia coli* lacking the exoribonucleases RNase II, RNase D, and RNase BN. *J Biol Chem* 259:11651–11653.
48. Ohashi H, Kanamori T, Shimizu Y, Ueda T (2010) A highly controllable reconstituted cell-free system—a breakthrough in protein synthesis research. *Curr Pharm Biotechnol* 11:267–271.

## Supplementary Information

### Translation and folding of single proteins in real time

Florian Wruck <sup>a,1</sup>, Alexandros Katranidis <sup>b,2</sup>, Knud H. Nierhaus <sup>c,3</sup>, Georg Büldt <sup>b,d</sup> and Martin Hegner <sup>a,2</sup>

<sup>a</sup> Centre for Research on Adaptive Nanostructures and Nanodevices (CRANN), School of Physics, Trinity College Dublin, Ireland.

<sup>1</sup> Institute for Atomic and Molecular Physics (AMOLF), Amsterdam, The Netherlands.

<sup>b</sup> Forschungszentrum Jülich, Institute of Complex Systems ICS-5, Jülich, Germany.

<sup>c</sup> Charité - Universitätsmedizin Berlin, Institute for Medical Physics and Biophysics, Berlin, Germany.

<sup>d</sup> Moscow Institute of Physics and Technology, Laboratory for Advanced Studies of Membrane Proteins, Dolgoprudny, Russia.

<sup>1</sup>Current address

<sup>2</sup>Correspondence to: *martin.hegner@tcd.ie* or *a.katranidis@fz-juelich.de*

<sup>3</sup>Deceased (April 7, 2016)

Classification: Biological Sciences, Biophysics and Computational Biology

Keywords: ribosomes, cotranslational protein folding, protein synthesis, single molecule, optical tweezers

## SI Methods

**Biotinylation of ribosomes in vivo.** Can20/12E (1), a RNase deficient *Escherichia coli* K-12 strain, was transformed with two plasmids, pBirAcm plasmid containing the birA gene to overexpress biotin ligase and pAN5 plasmid containing ribosomal protein uL4 with an N-terminal AviTag (Avidity, Denver, USA). The AviTag was biotinylated *in vivo* by the biotin ligase, after induction by 1 mM IPTG at  $A_{595} = 0.4$  and addition of 50  $\mu\text{M}$  biotin. Cells were incubated for an additional 1 h at 37°C before harvesting.

**Isolation of ribosomes and activity test.** Biotinylated ribosomes were isolated as previously described (2) and were checked for activity by synthesizing GFP emerald (GFPem) in an *in vitro* transcription/translation system (PURExpress  $\Delta$  amino acids, histidine, tRNA, ribosomes, NEB #E3315Z, New England Biolabs). GFPem fluorescence was monitored in a QM-7 spectrofluorometer (Photon Technology International, Birmingham, NJ).

**Plasmid construction.** The plasmid pRSET / EmGFP (Invitrogen), containing the gene of GFP Emerald (GFPem: S65T, S72A, N149K, M153T, I167T), was modified. An amber stop codon was inserted at the N-terminus, followed by a 35 amino acid linker and a sequence of 6 histidines (6 $\times$ His) upstream the gene of GFPem, as previously described (3). A C-terminal extension was added, but instead of 31 additional amino acids as also previously described (3), it was expressing a linker of 43 amino acids in total, comprising a region rich in Gly / Ser followed by the SecM AP (FSTPVWISQAQGIRAGP). The SecM AP induces translational arrest, by shifting the linkage between the nascent chain and the peptidyl-tRNA by 2 Å(4). The genes of human Tau40 and mouse DHFR were exchanged with the gene of GFPem, between the NcoI and HindIII sites, to generate the other two constructs used in this study. The total length of the GFPem protein segment is 239 amino acids, the one from DHFR 187 amino acids and the one from hTau40 441 amino acids, respectively.

**In vitro transcription/translation system.** The cell-free transcription/translation system used in this study is a customized version of the PURE system (5) without ribosomes (PURExpress  $\Delta$  ribosomes, New England Biolabs) that is additionally lacking the amino acid histidine.

**Coupling of streptavidin-DNA handles to beads.** 840 nm diameter protein G-functionalized polystyrene beads (Spherotech) were modified with anti-digoxigenin (anti-DIG, Roche), using the imidoester crosslinker dimethyl pimelimidate (DMP). dsDNA molecules were prepared by PCR amplification using digoxigenin (DIG) and biotin 5'-end-modified primers. The resulting 1000 bp PCR fragments (Bio-DNA-DIG) (0.34  $\mu$ m in length) reacted with streptavidin (from *Streptomyces avidinii*, Sigma) at a ratio of 100 streptavidin/DNA for 24 h at 4°C and were subsequently coupled to the anti-DIG beads at a reaction ratio of 150 streptavidin-DNA/bead for 2 h at room temperature. This coupling resulted in a final ratio of 1-3 freely accessible streptavidin-DNA handles per anti-DIG bead. The beads with the streptavidin-DNA handles were then washed several times in Tico buffer (20 mM Hepes-KOH pH 7.6, 6 mM (Ac)<sub>2</sub>Mg, 30 mM AcNH<sub>4</sub>, 4 mM  $\beta$ -mercaptoethanol) and split into two batches (6).

**Coupling of ribosomes to beads with DNA handles.** Biotinylated ribosomes were mixed with one batch of Tico-washed streptavidin-DNA modified beads at a ratio of 20k ribosomes/bead (~3000 ribosomes per accessible streptavidin binding pocket) and incubated at room temperature for 2 h. Excess of unbound ribosomes were removed by pelleting and washing the beads once with Tico buffer and resuspending the beads directly in the reaction 1 mix described below.

**Incorporation of biotin during synthesis.** Synthesis for all three constructs was initiated by mixing together the *in vitro* system (that lacked histidine), ribosomes coupled to beads via the DNA handles and 5.5 nM linearized plasmid, containing the appropriate target gene (reaction 1, Fig. 1A). Additionally, 10  $\mu$ M of a modified tRNA pre-charged with a biotinylated lysine was added to the mix and the reaction was carried out for 40 min at 37°C. The constructs were biotinylated cotranslationally at the N-terminal amber position using the suppressor tRNA technique, as described previously (3). Since the *in vitro* system lacked histidine, synthesis was halted upon reaching the 6 $\times$ His sequence. The stalled RNC complexes coupled to beads were kept on ice until injection into the microfluidic chamber.

**Translation reaction.** After coupling the N-terminal biotin of the construct to the second batch of beads with DNA handles (Fig. 1B) in TICO buffer (20 mM Hepes-KOH pH 7.6, 6 mM (Ac)<sub>2</sub>Mg, 30 mM AcNH<sub>4</sub>, 4 mM  $\beta$ -mercaptoethanol), synthesis was recommenced by adding a translation mix containing 150  $\mu$ M histidine (reaction 2, Fig. 1A). Prior to injection,

reaction 2 was gently mixed at 37°C for 5 min in order to allow some pre-charging of tRNA with amino acids by the aminoacyl-tRNA synthetases. It was then injected into inlet b of the microfluidic cell as depicted in Fig. S3 at speeds of up to 2 mm/s. The reaction took place at  $26 \pm 0.5^\circ\text{C}$  in the first compartment of the flow cell and translation was followed in real time. The concentration of the injected reaction components allowed synthesis under non-limiting conditions.

**Optical tweezers setup.** Two orthogonally polarized optical traps were formed deep inside the microfluidic chamber by overfilling an inverted microscope objective (UPLSAPO-60XW, Olympus) with two mostly overlapping orthogonally polarized beams from a single 4W single-mode NIR Nd:YVO4 diode-pumped solid-state laser (J20I-BL-106C, Spectra Physics) as described previously (7). The vertically plane-polarized beam could be steered relative to the horizontally plane-polarized beam with a feedback-stabilized piezo mirror (MTA2X, Mad City Labs) resulting in a steerable trap movement range of 15  $\mu\text{m}$  in the sample plane at a resolution of 0.3  $\text{\AA}$ . Back-focal-plane interferometry with differential detection was used to detect displacements and forces acting on the trapped polystyrene beads using two identical position sensitive detectors (DL100-7-PCBA3, First Sensor) (8). A custom power spectral calibration technique (9) was used to calibrate each bead before every measurement. This LabVIEW routine also took into account hydrodynamic effects and low-pass filtering effects of the detectors. The calibrated force signals orthogonal to the experiment measurement axis (parallel to injection flow) were also converted to fluid velocity and served as fluid flow indicators during measurements. Trap stiffness was kept at  $0.3 \pm 0.03$  pN/nm for both beads for all measurements. Data was sampled at a rate of 62.5 kHz with a PXI-6281 DAQ card (National Instruments), digitally low-pass filtered and subsequently decimated to 1 kHz using custom LabVIEW software routines (National Instruments). All measurements were performed in a custom-made microfluidic chamber (Fig. S3) that could be moved in 3 axes relative to the optical traps. All injections during measurements were carried out at orthogonal angles to the measurement and optical axes.

## Supplementary Text

**Intermediate folding state transitions.** During translation of all constructs, particularly DHFR and GFPem at lower forces (<15 pN), we observed gradual and also fast (ms) bead displacement changes of varying lengths (1-10 nm) continuously taking place, also long after translation would have ceased. As shown in Fig. 1D-F DHFR and GFPem constructs were not fully extended under the constant forces applied here. Immediately following translation at a constant applied force, these constructs would be in a non-native semi-folded state, extended in one dimension. Applied forces distort a tethered macromolecule's native free-energy landscape (10). Under these non-natural conditions many different structural conformations with similar free energy on the force-distorted free-energy landscape may be available that feature varying end-to-end distances along the measurement axis, populating quasi-stable intermediate folding states. The protein can be thought of as cycling through a number of conformational states with similar energies that are separated by small energy barriers, periodically overcome by random thermal fluctuations (Brownian motion) of the entire tethering system (11, 12). By keeping a constant tension over several minutes on the constructs their states were continuously, yet stochastically, pushed over non-trivial activation energies (or thermal barriers) into other non-native states, that mostly increased the constructs' end-to-end lengths in one dimension, along the direction of the applied force.

Following translation of the DHFR construct one particular transition resulted in the same bead displacement change of ~7.5 nm at different forces, occurring only once (irreversibly) at 20 pN and multiple times (hopping reversibly) at 10 pN, as shown in Fig. S6. This state transition was most likely due to a particular change between two intermediate folding states not seen during the unfolding experiments (Fig. 2A,D), an intermediate state  $I_x$  and a more 1-dimensionally elongated state  $I_y$ . At 10 pN both states were accessible, resulting in hopping between them, as well as other states with similar free energy and varying end-to-end distances. The greater constant force of 20 pN distorted the protein's energy landscape to such a degree as to prevent refolding of state  $I_y$  to  $I_x$ , allowing only the transition  $I_x$  to  $I_y$ . Due to the large number of possible conformations adopted by DHFR under constant tension, each with their own end-to-end distance, convoluted with the relatively high RNC fluctuation-induced position noise at low tensions (Fig. S4), it proved difficult to assign particular transitions to folding and unfolding of known secondary structural elements.

**Determining the number of translated amino acids.** To characterise the size of the unstructured polypeptides the force-extension curves shown in Fig. 2 were fitted with the extensible worm-like chain model (eWLC) (13). The eWLC model describes the force-extension behaviour of extensible semi-flexible polymers, dependent on the thermal energy  $k_B T$ , given by the equation:

$$\frac{x}{L} = 1 - \frac{1}{2} \left( \frac{k_B T}{F(x)A} \right)^{\frac{1}{2}} + \frac{F(x)}{S}$$

where  $k_B T = 4.1$  pN·nm at room temperature,  $F(x)$  is the applied force as a function of extension  $x$ ,  $A$  is the persistence length,  $S$  the stretch modulus and  $L$  is the contour length of the polymer. In particular, two combined eWLCs were fitted, one describing the semi-flexible behaviour of the dsDNA handles and the other describing that of the unstructured polypeptide. A persistence length of 50 nm and stretch modulus of 1000 pN were chosen for the two  $L = 340$  nm (1 kbp) dsDNA handles (6, 14). The unstructured nascent polypeptides were modelled with a persistence length of 0.66 nm (15, 16). Also, a single-residue contour length of  $\Delta L = 0.4$  nm was chosen (17). The total contour length of each construct was given by  $L = \sum_{i=1}^n \Delta L$ , where  $n$  is the total number of residues of the construct. Fixing these parameters, we obtained an unstructured polypeptide stretch modulus of about 200 pN, confirming earlier theoretical predictions (18). Using these parameters, the fractional extensions  $x/L$  of the polypeptide eWLC resulted in the nm-to-residue conversion factors given in table 1. Clearly, a nm-to-residue conversion using these parameters is only valid in the absence of nascent polypeptide folding.

**Hydrophobic collapse index (HCI) estimation.** The formation of partially folded intermediates just outside of the ribosomal tunnel (RT) is driven by a multitude of varied, rapid and continually occurring physical processes that induce sharp bending and twisting of the emerging polypeptide chain near hydrophobic collapse sites that lead to self-intersection, occurring at sequence stretches rich in hydrophobic amino acids. Polypeptide self-intersection is partly due to the entropically favorable minimization of polar solvent contact with non-polar hydrophobic residues (19).

As soon as self-intersection of the polypeptide takes place further self-interactions become possible, enabling secondary structure formation between formerly distant parts of the polypeptide chain for example. Only the highly hydrophobic amino acids Leu, Ile, Val and Phe were considered for the hydrophobic collapse site estimations (Fig. 5A). The

hydrophobic collapse index (HCI) values for each sequence were determined as previously described (19). Briefly, the HCI values for the sequences of hTau40, DHFR and GFPem were determined by first assigning a value of either 1 or 0 to each residue position along the protein sequences, resulting in three separate one-dimensional arrays whose total number of elements corresponded to the number of amino acids in each peptide sequence. A value of 1 was assigned to each sequence position containing one of the hydrophobic residues Leu, Ile, Val or Phe; all other residues of the sequence were assigned a value of 0. Adjacent-averaging each of these arrays with a 9-point trapezoid window function resulted in a HCI value for each sequence position. Values of HCI greater than 0.5 denote sequence regions where polypeptide self-intersection and thus hydrophobic collapse is likely to occur.

**Model of the observed translation rate.** The predicted observable elongation rate as a function of applied force  $f$  at sequence position  $i$  is given by the simple model  $\dot{x}_i(t, f) = C_i(f)k\dot{x}_0(t)\exp[l\rho_i + mN_i^{R,K} + nN_i^H]$ , for  $C_i(f) = 1$  if  $i < s$  and  $C_i = 0$  for  $i \geq s$ , where  $s$  is the first sequence position where the hydrophobic collapse index  $HCI_i = HCI_s(f)$ , and where  $HCI_s(f) = 10^{-2}f + 0.43$  ( $\pm 0.05$ ) is the estimated force-dependent value of HCI at which hydrophobic collapse is likely to occur (Fig. 5A). Here, the maximum burst rate  $\dot{x}_0(t)$  is scaled by a factor  $d$ , the temperature  $T$  and reaction mix concentration factor  $g$  dependent constant  $k = dgT$ . A scaled maximum burst rate of  $k\dot{x}_0(t) = 5 \text{ nm s}^{-1}$  was chosen for all fits. The proline-dependent term is given by  $\rho_i$ .  $N_i^{R,K}$  and  $N_i^H$  represent the number of positively charged residues and the number of histidines inside the RT at position  $i$ , assuming a tunnel length of 30 residues. Considering only successive Pro during fitting,  $\rho_i = 1$  at a sequence position containing a successive Pro residue and,  $\rho_i = 0$  everywhere else. The coefficients  $l$ ,  $m$ ,  $n$  were determined numerically by first fitting  $\sum_{i=1}^n \frac{\Delta x_i}{2} \left[ \frac{\Delta x_i}{\Delta \tau_i} + \frac{\Delta x_{i-1}}{\Delta \tau_{i-1}} \right]$  to translation trajectories, where  $\Delta x_i = x_i - x_{i-1}$  is the change in sequence position and  $\Delta \tau_i = \Delta t_0 + \Delta t_{\rho_i} + \Delta t_{N_i^{R,K}} + \Delta t_{N_i^H}$ , where  $\Delta t_i = \frac{\Delta x_i}{2} [t_i + t_{i-1}]$ . Here,  $\Delta t_0$  represents the constant burst rate time step,  $\Delta t_{\rho_i} = a\rho_i$ ,  $\Delta t_{N_i^{R,K}} = bN_i^{R,K}$  and  $\Delta t_{N_i^H} = cN_i^H$ . Fitting exponentials to the functions  $\frac{\Delta x_i}{\Delta t_0 + \Delta t_{\rho_i}}(\rho_i)$ ,  $\frac{\Delta x_i}{\Delta t_0 + \Delta t_{N_i^{R,K}}}(N_i^{R,K})$  and  $\frac{\Delta x_i}{\Delta t_0 + \Delta t_{N_i^H}}(N_i^H)$  gave  $l$ ,  $m$  and  $n$  respectively.

Also, the reaction mix concentration factor  $g$  is not completely constant over time during long measurements. In the high surface area-to-volume ratio environment of the microfluidic cell some loss to surfaces over time is expected. Laser heating-induced convection currents

and general diffusion contribute to mixing of buffer in the main buffer line with the reaction mix. These coefficients also depend on the relative concentration of aminoacyl-tRNA-EF-TuGTP ternary complexes, that depend partly on the concentration of individual amino acids and their corresponding tRNAs and aminoacyl-tRNA synthetase concentrations that can change between different reaction mix batches.

When fitted to sections of translation trajectories representing synthesis of unstructured parts of the nascent polypeptide where no cotranslational folding occurred, the simplified model given in equation 1 enabled us to determine the approximate total translation times for a given construct, even though cotranslational folding masked the endpoint of translation (Fig. 5B-F). Omitting the hydrophobic collapse term  $C_i(f)$  resulted in an extrapolated overall translation time of  $78 \pm 3$  s for the DHFR construct at 20 pN (Fig. 5B,C) and  $63 \pm 5$  s for the DHFR construct at a tension of 10 pN (Fig. 5D), giving us an overall average rate of synthesis of  $3.1 \pm 0.5$  AA/s at 20 pN and  $3.8 \pm 1$  AA/s at 10 pN. Similarly, the hTau40 construct in Fig. 5E,F was completed within a timeframe of  $69 \pm 6$  s, with an average rate of synthesis of  $6 \pm 1$  AA/s at 7 pN. The average translation rates determined by our model for DHFR and hTau40 (Fig. 5) are similar to the one observed for N5-glutamine methyltransferase HemK ( $\sim 3.6$  AA/s) in a recent study by W. Hotkamp *et al.* (20), as well as the rates observed during synthesis of the SFVP polyprotein (21). Previous bulk studies on the rate of translation in *E. coli* cite average rate of synthesis of 13-22 AA/s at 37°C (22). The lower average rate observed here may be due in part to the lower temperature of  $26 \pm 0.5^\circ\text{C}$  at which translation took place during experiments.

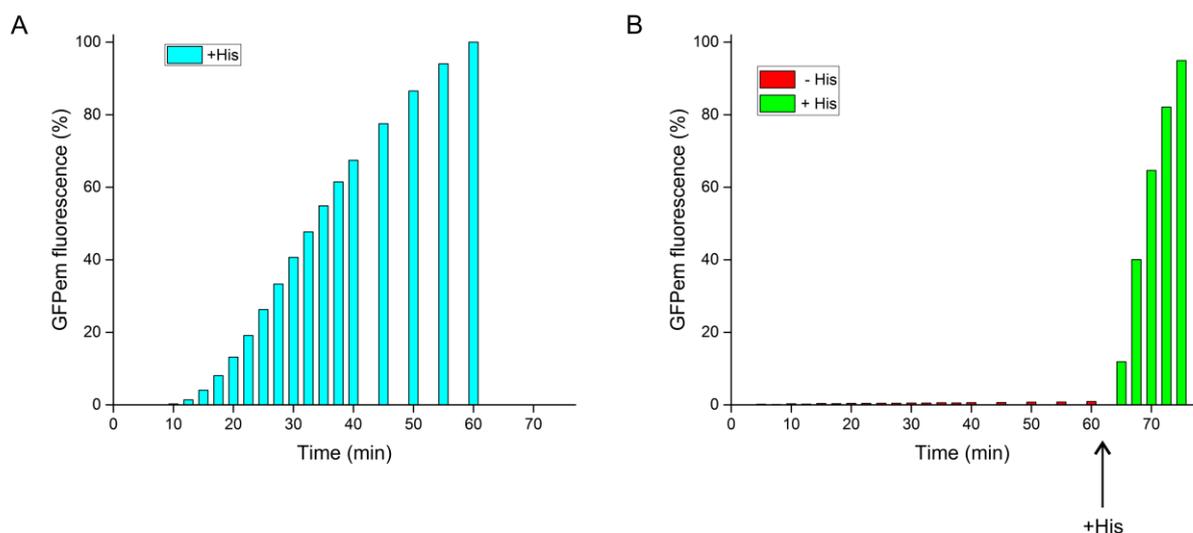
**mRNA secondary structure considerations.** Under physiological conditions RNA forms complex three-dimensional structures (23). As the ribosome translocates along the mRNA, it locally unwinds base-paired strands. RNA hairpins rich in GC base-pairs require more energy to dissociate than those with intermittent AU base-pairs (and of course those solely consisting of AU base-pairs), which can lead to a slowing of the translation rate along mRNA stretches containing consecutive CG base-pairing (24). Perhaps the variations in the translation rate and the observed stalling plateaus could be attributed, in part at least, to mRNA unwinding. However, it may not be possible to determine a definitive secondary structure of large RNA molecules, such as the mRNAs transcribed for the three constructs DHFR, GFPem and hTau40, since their secondary structure would probably not be static. Rather, it would most likely be undergoing conformational changes between states of similar energy before

translation and certainly during translation, as the ribosome perturbs the local free energy of the system while breaking hydrogen bonds. The Vienna RNA Package 2.0 (25) part of the Vienna RNA Websuite (26) was utilized to predict mRNA secondary structures for the three different constructs. Mean free energies were calculated using the widely used Turner model (27) to determine base-pairing probabilities and mRNA secondary structure for a temperature of 26°C, the measured equilibrium temperature of the objectives at a laser power of 200 mW (100 mW per trap). The results for all three constructs are depicted in Fig. S11. All constructs show several predicted hairpins with high base-pairing probabilities, some of which with several consecutive CG pairs. Unsurprisingly, hTau40's mRNA has the greatest number of hairpins as is expected, given its relative size to GFPem's or DHFR's mRNA. A careful analysis comparing the locations of highly probable CG-rich hair pins with sequence stretches featuring observed reduced translation speeds showed no obvious correlation in any of the three constructs.

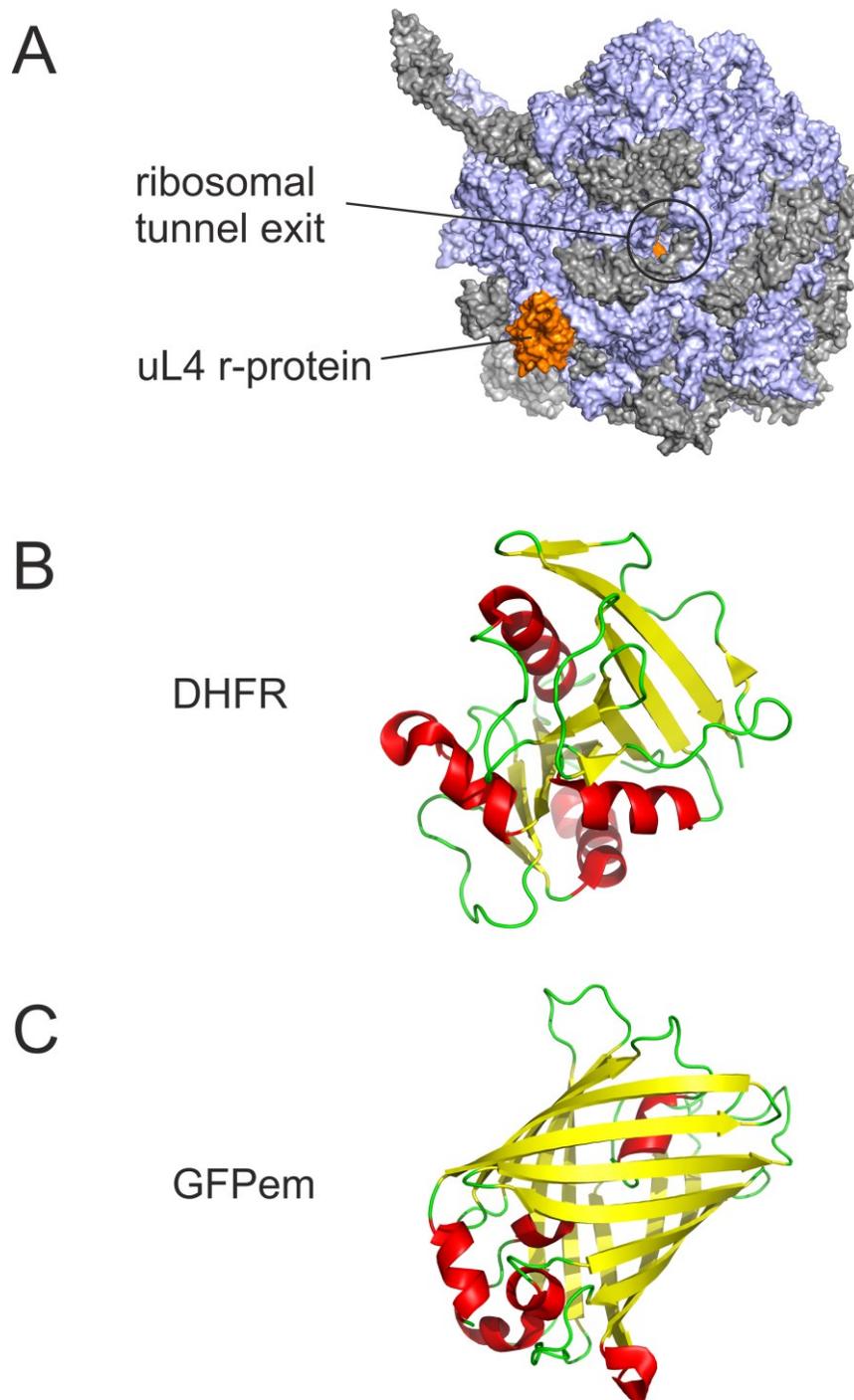
**Rare codons and tRNA abundance.** All 20 amino acids in the PURExpress translation reactions have the same concentration of 0.3 mM (5). However, the 46 individual tRNAs used in the reactions are purified as a mixture from *E. coli* (MRE600 strain). The relative concentrations of individual tRNAs can vary widely in organisms. A paper that describes the relative abundance of tRNAs in *E. coli* MRE600 (28) revealed a number of rarely occurring tRNAs. Even if all amino acids are equally available, a codon requiring a tRNA with low relative concentration (rare codon) could result in a longer average waiting time during the decoding part of the elongation cycle as the ribosome has to wait for it to bind to the mRNA at the A-site. A study by Dong *et. al* found that there is a correlation between the abundance of individual tRNAs in *E. Coli* and the frequency of occurrence of the matching mRNA codons in the bacteria's genes at all growth rates (29). In the measured elongation traces for hTau40 (Fig.'s 1C, 3B, 5E,F, S8 and S10) it was noted that most of the prolonged stalling took place during translation of sequence regions rich in proline (Pro). hTau40's sequence does contain a number of rare CCC codons, coding for Pro. However, comparing hTau40's codon sequence with the translation traces showed that in some instances there is a correlation with rare CCC codons and pauses and sometimes there is not. At the first long pause, at around codon 175, there are many prolines. However, most are not encoded by the rare codon CCC. At the second long pause at around codon 230-240, there is another cluster of Pro residues, including some encoded by the rare codon CCC. At the third pause at around 270, there are two rare prolines at positions 264 and 270. The next pause at around 310 also

has prolines encoded by the rare codon CCC at positions 311 and 313. Overall, however, the CCC codons are somewhat randomly distributed over the sequence and do not cause obvious stalling in sequence positions 6, 68, 111, 115, 142, 149 and 157 for instance. If prolines slow down translation because they are rare codons in *E. coli*, then there should be pauses at other rare codon positions as well. However, no correlation between rare Arg codons AGG, AGA and CGA at positions 1, 8, 51 in hTau40's sequence and stalling could be found. Evidently, rare codons by themselves cannot explain slowed elongation and stalled translation.

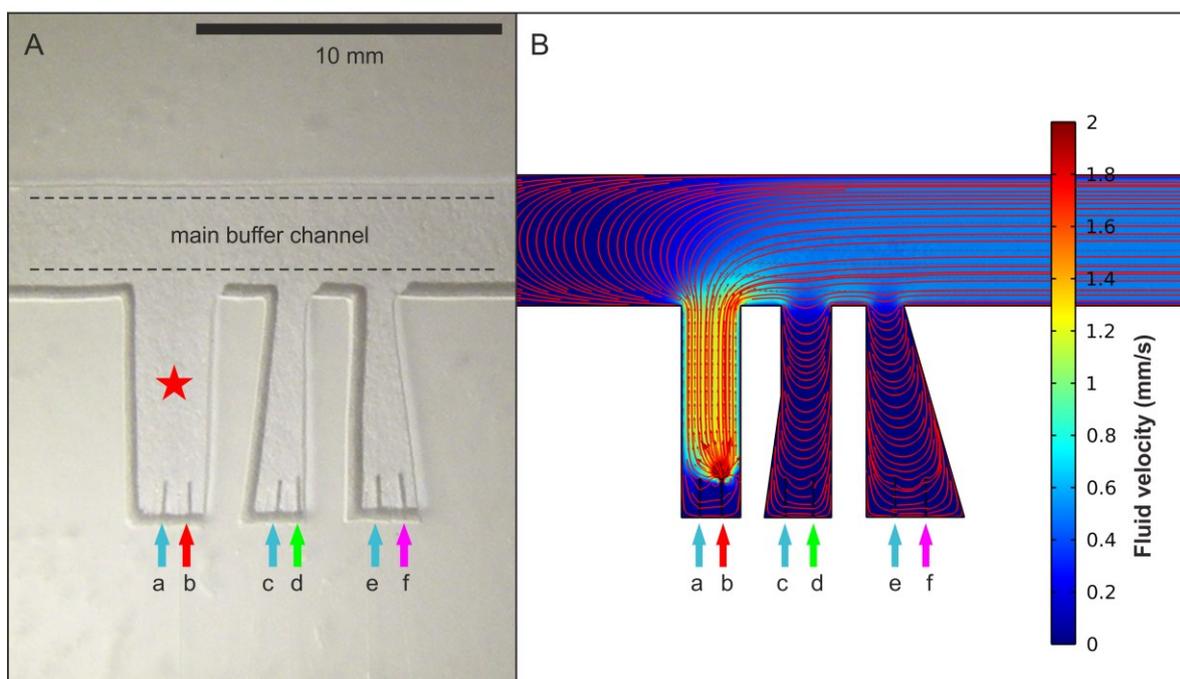
## Supplementary Figures



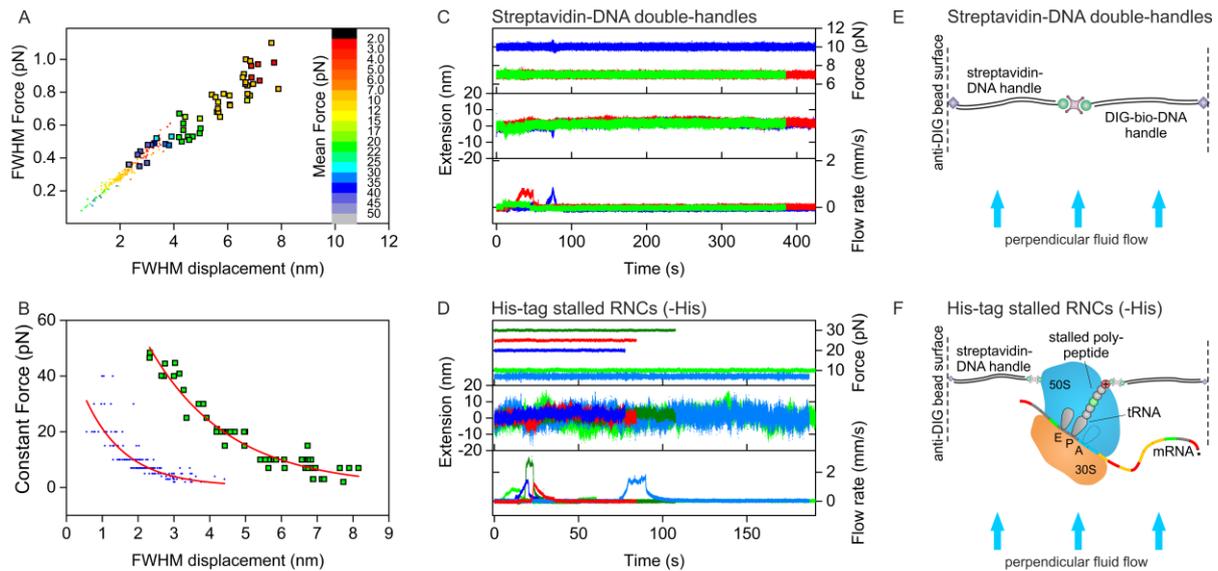
**Fig. S1.** Synthesis halts at the 6×His sequence. GFPem was synthesized and released from the ribosomes using the *in vitro* transcription/translation system and the fluorescence of the bulk was monitored with a QM-7 spectrofluorometer. (A) Fluorescence of GFPem increased continuously when histidine was present from the beginning (cyan bars) and the fluorescence intensity observed after 1 h was referred to as 100%. A certain time (15-20 min) was required until enough molecules of GFPem were present in the bulk, so that fluorescence could start to be detected. (B) When histidine was not present in the *in vitro* system, no significant fluorescence of GFPem was observed even after 1 h (red bars). When histidine was added after 1 h (black arrow), GFPem fluorescence intensity quickly increased, in 15 min it reached the level that the control showed after 1 h (green bars). A large number of ribosomes were halted at the 6×His sequence and thus could be synchronized, leading to a faster accumulation of fluorescing GFPem molecules after addition of histidine.



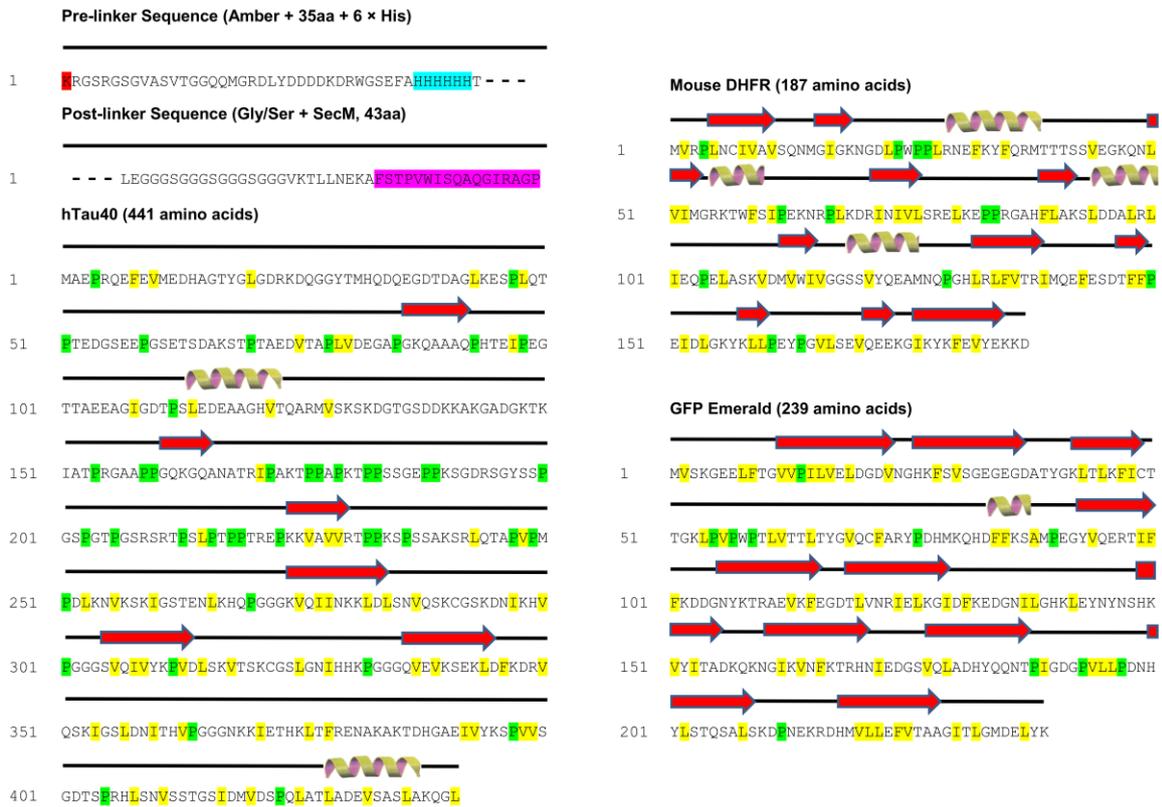
**Fig. S2.** (A) False-coloured surface electron density map of the 50S large ribosomal subunit (PDB ID 2AW4), highlighting the opening of the ribosomal exit tunnel, as well as the uL4 ribosomal protein (orange), utilized as anchoring point in our experiments. Ribosomal-proteins and ribosomal RNA are colored grey and purple respectively. Only the large ribosomal subunit is shown. (B) Crystal structure of natively folded DHFR (PDB ID 3K45) showing secondary structure. Red ribbons represent alpha helices and yellow ribbons denote beta sheets. (C) Secondary structure representation for the crystal structure of natively folded GFPem (PDB ID 4KW4), a particularly fast-folding mutant of GFP. Colour coding as in (B).



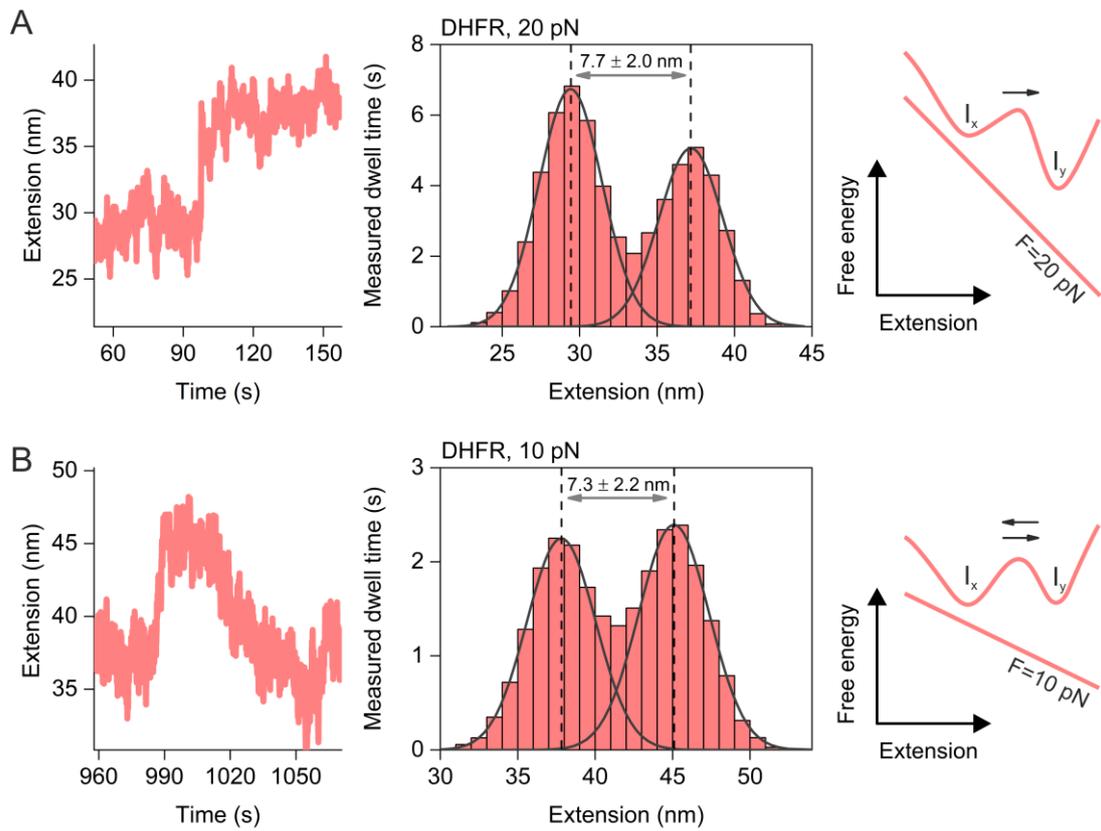
**Fig. S3.** Custom-made microfluidic cell. (A) The microfluidic cell consists of a main buffer channel and three smaller compartments that each have two separate inlets, consisting of 2 glass micropipettes with an inner diameter of 40  $\mu\text{m}$ . Inlets a, c and e were used as buffer lines. Inlet b was used for reaction 2 injection (Fig. 1A), inlet d for the injection of RNC-coupled streptavidin-DNA beads and inlet f for the injection of streptavidin-DNA beads. Pre-filling the bead compartments allowed time-efficient assembly of the experiment by simply moving the microfluidic cell relative to the two optical traps in order to collect single beads from each well in each optical trap. Calibration and measurements were performed in the centre of the first compartment (red star). Injections of the translation mix containing histidine (reaction 2, Fig. 1A) were performed at orthogonal angles to the measurement and optical axes. (B) This time-dependent laminar fluid flow COMSOL simulation of the fluid velocity field was calculated with modelled pure water initially at rest within a model of the microfluidic cell subjected to a pressure difference of 1 bar using the Navier-Stokes and continuity equations with constant viscosity and density (no-slip) at room temperature. The streamlines and arrows show the fluid velocity field directions. The colour-mapped slice shows the fluid velocity magnitude at the centre of the cell's interior, a distance of 90  $\mu\text{m}$  from both coverslips.



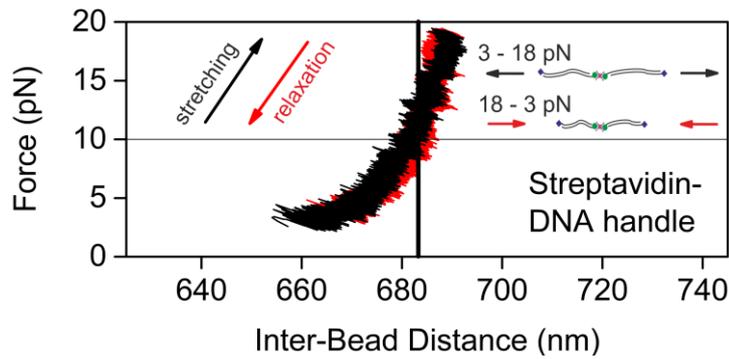
**Fig. S4.** Noise comparison of streptavidin-DNA handles with and without RNCs and negative controls. (A) FWHM of measured constant displacement and force data (30 s, 1 kHz) for  $N=176$  different streptavidin-DNA handles (control, small diamonds) and  $N=54$  different streptavidin-DNA-RNCs (His-tag stalled constructs, large squares) held at a range of constant tensions. The tethered stalled RNCs underwent rapid positional and conformational changes causing fast displacement alterations between both biotinylated anchoring points, resulting in the marked increase in position and force noise, especially at low forces. Since these fluctuations occur at faster timescales than our measurement bandwidth of 1 kHz, we could not fully resolve individual steps. (B) Plot of constant applied force as a function of FWHM of the measured displacement data in (A) (control, blue; constructs, green), each fitted with exponential functions (red lines). The overall lower displacement noise for the DNA double handles decreases more rapidly with increasing tension in the low force regime  $< 10$  pN. (C) Negative control examples of constant force measurements using double DNA-handles (streptavidin-biotin-DNA vs. biotin-DNA) without the RNCs (as illustrated in (E)). Here, 3 separate double handles (red, green, blue) are held under constant tensions for several minutes following injection of the reaction 2 mix. Shown here are the measured forces, displacements and injection flow rates for each case. The flow rate spikes below 100 s represent the measured fluid flow around the beads during injection. No elongation is observed. (D) Examples of negative control measurements using His-tag stalled RNC hook-ups with hTau40 (light/dark blue), DHFR (red) and GFPem (light/dark green), showing the measured forces, displacements and injection flow rates for each example (1 kHz). The reaction 2 mix without histidine is injected at varying flow rates at orthogonal angles to the measurement axis (as illustrated in (F)) and no translation of any of the three constructs (hTau40, GFPem, DHFR) is observed.



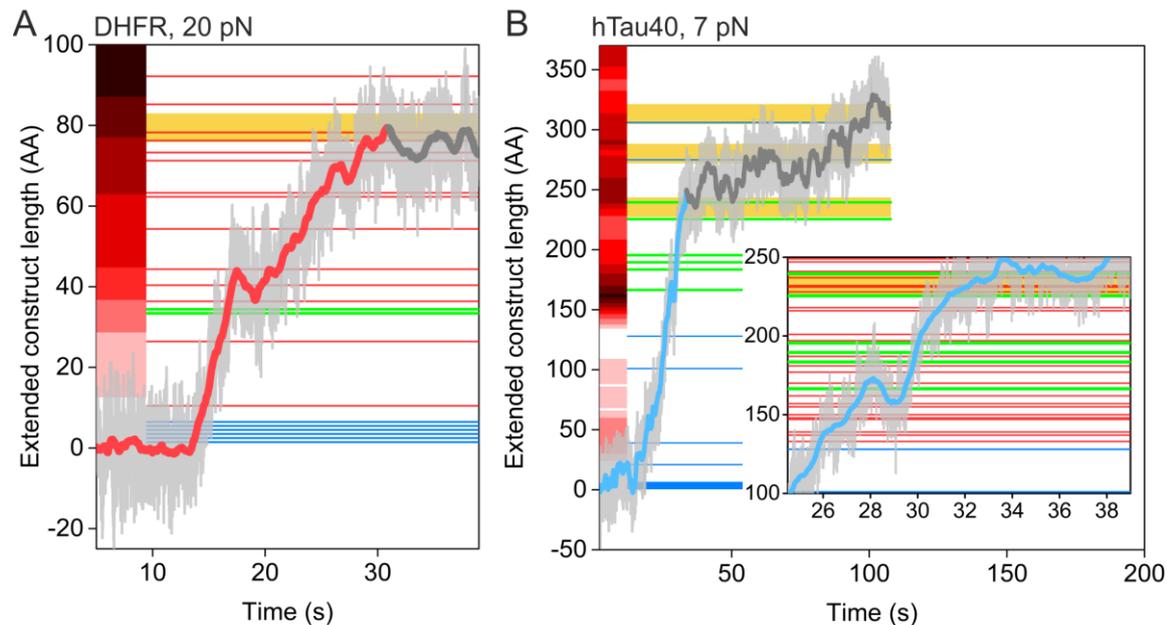
**Fig. S5.** Linker and primary sequences of the constructs with secondary structural elements. Linker sequences followed by the three primary sequences hTau40, DHFR and GFPem are shown (Fig. 1A). In the linker sequences, the amber stop codon is highlighted red, the His tag is highlighted cyan and the SecM sequence is highlighted pink. In all primary sequences hydrophobic amino acids (Leu, Ile, Phe, Val) (yellow) and Pro (green) are highlighted. Red arrows and helices indicate the position of beta strands and alpha helices, respectively.



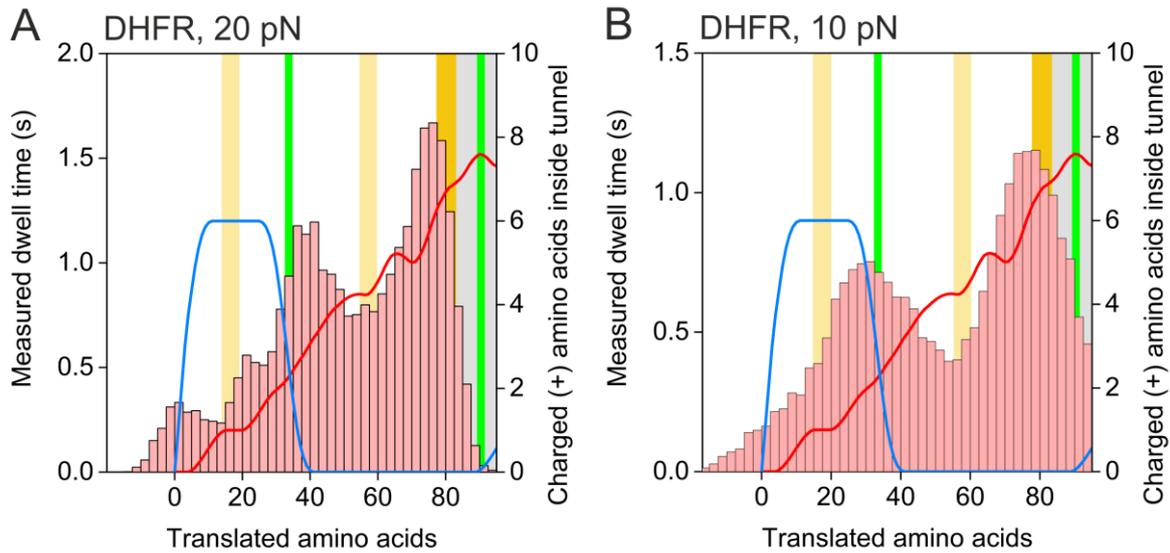
**Fig. S6.** (A) Measured trajectory and corresponding position histogram (with energy landscape illustration) of a  $7.7 \pm 2.0$  nm transition between two force-distorted intermediate states  $I_x$  and  $I_y$  of nascent DHFR at a force of 20 pN, due to a sudden unfolding of the partially folded intermediate  $\sim 90$  s after translation began. (B) Example of a similar instantaneous unfolding transition occurring at 10 pN multiple times (hopping). The lower force allowed for a slower refolding between similar intermediate states ( $7.3 \pm 2.2$  nm after 17 min), as illustrated in the force-perturbed energy landscape drawing, and transitions between other distinct states not observed at 20 pN tension.



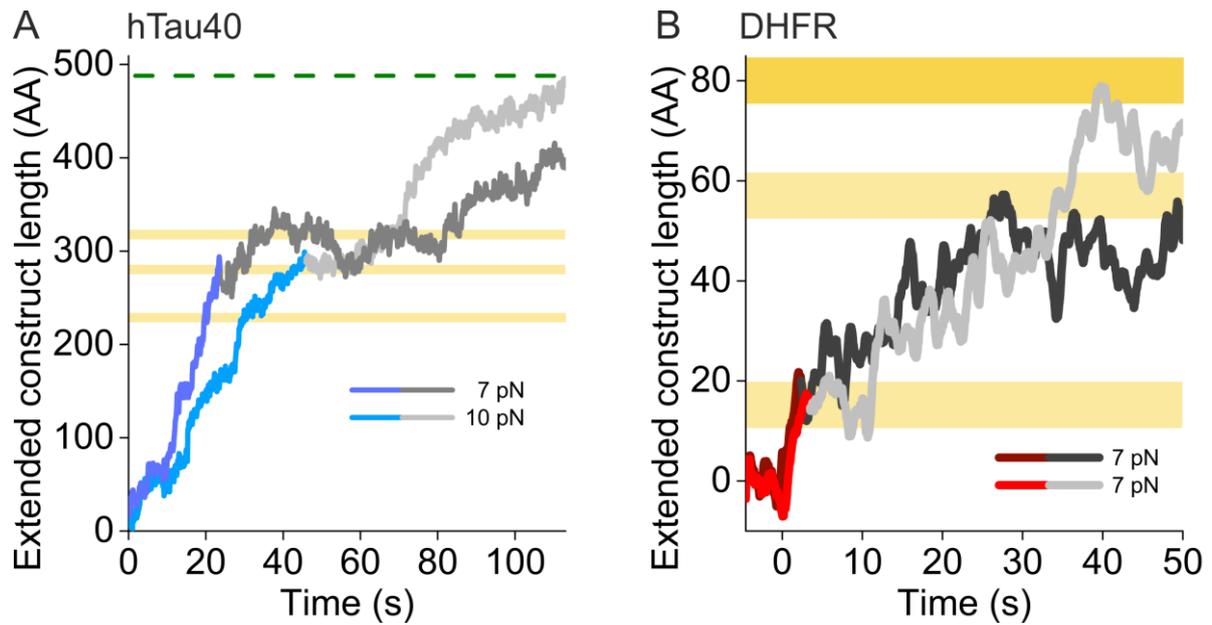
**Fig. S7.** Control experiments with DNA handles linked by a single streptavidin protein (0.5 pN/s, 1 kHz) as illustrated in S4E. Here, red data/arrows represent decreasing and black data/arrows denote increasing the applied tension. The extension shows the total measured length of the DNA double-handles.



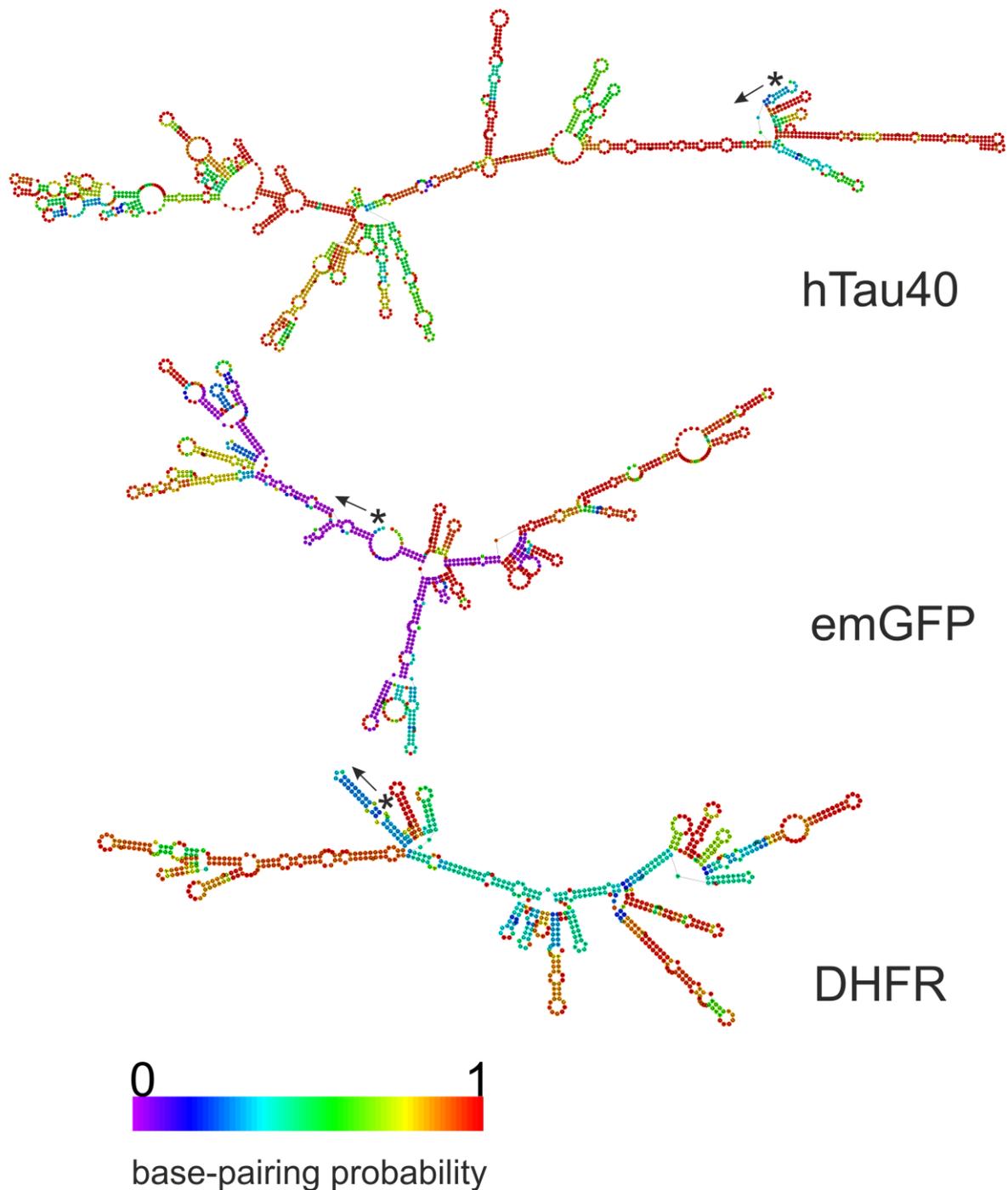
**Fig. S8.** Translation trajectories of individual proteins (*A*) DHFR translation trajectory at 20 pN (grey, 1 kHz; red overlay, 2 Hz; 0.33 nm/aa) showing stalling in regions with successive Pro-Pro (green lines) and slowing in regions with positively charged Arg, Lys (red lines). The red shaded zones depict the number of positively charged residues confined within the RT for an extended nascent chain (red curve in Fig. S9). His (blue lines) did not significantly contribute to slowing here. Grey overlay depicts cotranslational compaction following translation of strongly hydrophobic sequence stretches indicated by the yellow horizontal bars. (1 of  $N=35$  DHFR traces). (*B*) Translation trace of hTau40 (7 pN; grey, 1 kHz; blue overlay, 2 Hz; 0.22 nm/aa) demonstrating stalling at Pro-Pro and subsequent slowing after multiple positively charged amino acids were incorporated (1 of  $N=33$  hTau40 traces). The colour coding is the same as in (*A*), blue zones indicate tunnel-confined His. The inset magnifies the trajectory segment where Pro-Pro stalling occurred during translation.



**Fig. S9.** Translation rate correlation with Pro-Pro incorporation and positive residues confined within the RT. (A) Position histogram, from a 20 pN translation trajectory of DHFR (bin-size 2.5 nm, 0.32 nm/aa, SI Methods), showing stalling during Pro-Pro incorporation (green bars) and rate slowing at positions with more Arg/Lys (red curve), and to a lesser extent His (blue curve), inside the RT (assumed tunnel length of 30 residues) as illustrated in Fig. 3 C-D (30S subunit, mRNA not shown, not drawn to scale). Greater dwell times represent sequence regions with lower translation rate. Prolonged observable stalling occurred at the first HCI peak of 0.61 (dark yellow bar, Fig. 5D), afterwards nascent polypeptide growth was coupled to folding (grey background) as illustrated in 3D. Dark and light yellow bars represent sequence-segments with HCI values of 0.55-0.65 and 0.45-0.55, respectively. (B) Similar position histogram (bin-size 2.5 nm, 0.24 nm/aa, SI Methods) from a 10 pN DHFR translation trajectory. Comparing it to (A) shows similar dwell times, also with stalling at the Pro-Pro position and slowing due to Arg/Lys (His) (same colour coding as in (A)).



**Fig. S10.** Cotranslational compaction of hTau40 and DHFR translation traces. (A) Compaction of the hTau40 construct occurs during synthesis after translation of strongly hydrophobic sequence positions 228-307 at a tension of 7-10 pN. Colored traces denote synthesis decoupled from folding, while grey traces show synthesis coupled to folding. Dark- and light yellow horizontal bars represent sequence-segments with HCI values of 0.55-0.65 and 0.45-0.55, respectively (Fig. 5A). (B) At a tension of 7 pN cotranslational compaction is observed following synthesis of DHFR sequence position 9 (corresponding to translated position 15 when including the 6xHis tag). From this point on synthesis and folding are coupled and indistinguishable during mRNA translation under an applied force of 7 pN. Colour coding as in (A).



**Fig S11.** mRNA secondary structure comparison of the three constructs hTau40, GFPem and DHFR. The secondary structures and base-pairing probabilities were derived from mean free energy predictions (Turner model (27)) for the mRNA sequences of the three constructs at 26 °C. The position of the UAG amber stop codon at the start of the sequence is highlighted with a star and the direction of synthesis is indicated with an arrow for each construct. Hairpins containing multiple CG base pairs occur in all three constructs. No correlation between these predicted hairpins and translation rate could be found. All predictions were made using the Vienna RNA Websuite (25, 26).

## Supplementary References

1. Zaniewski R, Petkaitis E, & Deutscher MP (1984) A multiple mutant of Escherichia coli lacking the exoribonucleases RNase II, RNase D, and RNase BN. *J. Biol. Chem.* 259(19):11651-11653.
2. Rheinberger HJ, Geigenmuller U, Wedde M, & Nierhaus KH (1988) Parameters for the preparation of Escherichia coli ribosomes and ribosomal subunits active in tRNA binding. *Methods Enzymol.* 164:658-670.
3. Katranidis A, *et al.* (2011) Force measurements of the disruption of the nascent polypeptide chain from the ribosome by optical tweezers. *FEBS Lett.* 585(12):1859-1863.
4. Bhushan S, *et al.* (2011) SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* 9(1):e1000581.
5. Ohashi H, Kanamori T, Shimizu Y, & Ueda T (2010) A highly controllable reconstituted cell-free system--a breakthrough in protein synthesis research. *Curr. Pharm. Biotechnol.* 11(3):267-271.
6. Jadhav VS, Bruggemann D, Wruck F, & Hegner M (2016) Single-molecule mechanics of protein-labelled DNA handles. *Beilstein J. Nanotechnol.* 7:138-148.
7. Bustamante C, Chemla YR, & Moffitt JR (2009) High-resolution dual-trap optical tweezers with differential detection: instrument design. *Cold Spring Harb. Protoc.* 2009(10):pdb ip73.
8. Moffitt JR, Chemla YR, Izhaky D, & Bustamante C (2006) Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc. Natl. Acad. Sci. USA* 103(24):9006-9011.
9. Berg-Sorensen K & Flyvbjerg H (2004) Power spectrum analysis for optical tweezers. *Rev. Sci. Instrum.* 75(3):594-612.
10. Greenleaf WJ, Woodside MT, & Block SM (2007) High-resolution, single-molecule measurements of biomolecular motion. *Annu. Rev. Biophys. Biomol. Struct.* 36:171-190.
11. Onuchic JN & Wolynes PG (2004) Theory of protein folding. *Curr. Opin. Struct. Biol.* 14(1):70-75.
12. Caldarini M, *et al.* (2014) The complex folding behavior of HIV-1-protease monomer revealed by optical-tweezer single-molecule experiments and molecular dynamics simulations. *Biophys. Chem.* 195:32-42.
13. Hegner M, Smith SB, & Bustamante C (1999) Polymerization and mechanical properties of single RecA-DNA filaments. *Proc. Natl. Acad. Sci. USA* 96(18):10109-10114.
14. Bustamante C, Smith SB, Liphardt J, & Smith D (2000) Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* 10(3):279-285.
15. Lairez D, Pauthe E, & Pelta J (2003) Refolding of a high molecular weight protein: salt effect on collapse. *Biophys. J.* 84(6):3904-3916.
16. Cressiot B, *et al.* (2015) Dynamics and Energy Contributions for Transport of Unfolded Pertactin through a Protein Nanopore. *ACS Nano* 9(9):9050-9061.
17. Ainavarapu SR, *et al.* (2007) Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* 92(1):225-233.
18. Enomoto S & Krimm S (1962) Elastic Moduli of Helical Polypeptide Chain Structures. *Biophys. J.* 2(4):317-326.
19. Sen S & Voorheis HP (2014) Protein folding: understanding the role of water and the low Reynolds number environment as the peptide chain emerges from the ribosome and folds. *J. Theor. Biol.* 363:169-187.
20. Holtkamp W, *et al.* (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science* 350(6264):1104-1107.
21. Nissley DA, *et al.* (2016) Accurate prediction of cellular co-translational folding indicates proteins can switch from post- to co-translational folding. *Nat. Commun.* 7:10341.
22. Bremer H & Dennis PP (2008) Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* 3(1).
23. Batey RT (2006) Structures of regulatory elements in mRNAs. *Curr. Opin. Struct. Biol.* 16(3):299-306.
24. Wen JD, *et al.* (2008) Following translation by single ribosomes one codon at a time. *Nature* 452(7187):598-603.
25. Lorenz R, *et al.* (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26.
26. Gruber AR, Lorenz R, Bernhart SH, Neubock R, & Hofacker IL (2008) The Vienna RNA websuite. *Nucleic Acids Res.* 36(Web Server issue):W70-W74.
27. Mathews DH, *et al.* (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* 101(19):7287-7292.
28. Hossain M & Limbach PA (2007) Mass spectrometry-based detection of transfer RNAs by their signature endonuclease digestion products. *RNA* 13(2):295-303.

29. Dong H, Nilsson L, & Kurland CG (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 260(5):649-663.