



Delft University of Technology

Arrest Peptide Profiling resolves co-translational folding pathways and chaperone interactions in vivo

Chen, Xiuqi; Hilser, Vincent J.; Kaiser, Christian M.

DOI

[10.1038/s41467-025-61398-6](https://doi.org/10.1038/s41467-025-61398-6)

Publication date

2025

Document Version

Final published version

Published in

Nature Communications

Citation (APA)

Chen, X., Hilser, V. J., & Kaiser, C. M. (2025). Arrest Peptide Profiling resolves co-translational folding pathways and chaperone interactions in vivo. *Nature Communications*, *16*(1), Article 6833. <https://doi.org/10.1038/s41467-025-61398-6>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.


Arrest Peptide Profiling resolves co-translational folding pathways and chaperone interactions in vivo

Received: 23 December 2024

Xiuqi Chen ^{1,2,4}, Vincent J. Hilser² & Christian M. Kaiser ^{2,3} 

Accepted: 18 June 2025

Published online: 24 July 2025

 Check for updates

Cytosolic proteins begin to fold co-translationally as soon as they emerge from the ribosome during translation. These early co-translational steps are crucial for overall folding and are guided by an intricate network of interactions with molecular chaperones. Because cellular co-translational folding is challenging to detect, its timing and progression remain largely elusive. To quantitatively define co-translational folding in live cells, we developed a high-throughput method that we term “Arrest Peptide Profiling” (AP Profiling). Combining AP Profiling with single-molecule experiments, we delineate co-translational folding for a set of GTPase domains with similar structures, defining how topology shapes folding pathways. Genetic ablation of nascent chain-binding chaperones results in discrete and localized folding changes, highlighting how functional redundancy among chaperones is achieved by distinct engagement with the nascent protein. Our work provides a window into cellular folding pathways of structurally intricate proteins and paves the way for systematic studies of nascent protein folding at exceptional resolution and throughput.

Commensurate with their vast functional diversity, natural proteins fold into a broad range of structures. Small proteins, which have principally been used to study folding mechanisms¹, fold reversibly and quickly in vitro. In contrast, multi-domain proteins exhibit a high propensity to misfold in vitro, highlighting the importance of early folding steps that begin before synthesis is complete². These co-translational folding steps are often modulated by interactions with cellular machinery, including the ribosome and molecular chaperones, to achieve robust and efficient folding in vivo^{3–6}.

Before emerging outside the ribosome, nascent proteins can already form structures inside the polypeptide exit tunnel that spans the large ribosomal subunit. Contacts with ribosomal proteins and ribosomal RNA lining the tunnel walls can stabilize such structures^{7,8}. These interactions can, in turn, affect ribosome activity, causing elongation arrest. Specialized sequences termed “arrest peptides” (APs) exploit this type of interaction to regulate ribosome activity in *cis*^{9,10}. More extensive folding takes place once the nascent chain is extruded from the tunnel. Interactions with the ribosome surface and

molecular chaperones then modulate folding rates^{11,12} and structural stability^{13–15}. Thus, during their slow and vectorial synthesis, nascent proteins transition through continuously changing folding environments. Modulation of folding kinetics may delay folding until sufficient polypeptide has been extruded for productive folding¹¹. Thermodynamic effects could destabilize misfolded species¹² or promote folding¹⁶. How these interactions benefit co-translational folding is just beginning to be understood. Experimental studies are needed to define principles of co-translational folding.

Biochemical and biophysical in vitro approaches have been invaluable for defining co-translational folding pathways. However, they cannot fully capture the networks of interactions that modulate nascent chain folding in the cytosol. Ribosome profiling experiments, combined with immunoprecipitation (IP), have been used to assess nascent chain folding in vivo^{17,18}, but may not capture dynamic chaperone-nascent chain interactions. NMR spectroscopy has been used to detect ribosome-nascent chain interactions in live cells¹⁹. However, the approach is technically very demanding, and it has not

¹CMDB Graduate Program, Johns Hopkins University, Baltimore, MD, USA. ²Department of Biology, Johns Hopkins University, Baltimore, MD, USA. ³Bijvoet Centre for Biomolecular Research, Department of Chemistry, Utrecht University, Utrecht, The Netherlands. ⁴Present address: Department of Bionanoscience, Kavli Institute of Nanoscience Delft, Delft University of Technology, Delft, The Netherlands. ✉ e-mail: c.m.kaiser@uu.nl

yet been possible to resolve nascent polypeptide folding on the ribosome in cells. New experimental approaches are needed to define co-translational folding pathways in the cellular environment. We have previously shown that mechanical force generated by nascent protein folding can be coupled to reporter expression through the use of a force-sensing AP^{20,21}.

Here, we combine AP-based detection of folding with cell sorting and deep sequencing into an assay that we term “Arrest Peptide Profiling” (AP Profiling). We show that related GTPase domains fold co-translationally along distinct pathways that are likely determined by differences in their topologies. Genetic ablation of nascent chain-binding chaperones reveals co-localized but distinct effects on co-translational folding, explaining how unrelated chaperone systems achieve functional redundancy. The approach resolves folding and chaperone action in individual domains and in large, multi-domain proteins. AP Profiling thus establishes a scalable platform for defining co-translational folding events in live cells at high throughput and codon resolution.

Results

Ratiometric fluorescence measurements report on co-translational folding in live cells

The 17 amino acid AP from the SecM protein arrests translation elongation by forming specific interactions with the polypeptide exit tunnel inside the ribosome^{9,10}. Mechanical force accelerates arrest release, and elongation resumes²⁰ (Fig. 1a). Coupling arrest release to translation of a downstream reporter thus affords an opportunity for sensitive and quantitative detection of nascent chain folding in cells^{20–22}. So far, studies employing AP-based nascent chain folding have been carried out with limited throughput, requiring the generation, isolation and individual expression of candidate constructs, one for each nascent chain length at which folding is assessed. To overcome this limitation, we developed “AP Profiling” to effectively scale up the AP reporter assay.

We designed AP Profiling constructs (Fig. 1b) that encode the candidate protein fused in frame to the SecM AP⁹ and monomeric superfolder GFP (msGFP)²³, which folds and matures quickly and robustly. Expression of this open reading frame (ORF) is controlled by an arabinose promoter (P_{araBAD})²⁴. On the same plasmid, expression of a separate ORF encoding mCherry²⁵ is driven by a separate P_{araBAD} promoter. Each individual AP Profiling plasmid thus encodes two fluorescent proteins: mCherry, which serves as an internal control for variation in overall expression levels, and msGFP, which reports on arrest release and, thus, on folding. This dual reporter system enables ratiometric fluorescence measurements in live cells for robust and sensitive measurements of nascent chain folding.

To validate the new reporter construct, we expressed individual constructs with the G-domain of EF-G (referred to here as G_{EF-G}) as the candidate protein. G_{EF-G} folding has previously been investigated in vitro^{12,26} and in vivo²¹. In AP-based measurements, maximal arrest release was observed at nascent chain lengths of around 330 amino acids (aa)²¹, in line with detection of the natively folded domain at a similar nascent chain length in single-molecule experiments¹². Shorter chain lengths are not expected to fold²¹. We individually cultured *E. coli* cells transformed with plasmids encoding two distinct nascent chain lengths, 212 and 331 amino acids, corresponding to arrest positions on the mRNA of $n_{\text{codons}} = 212$ and $n_{\text{codons}} = 331$ (Fig. 1c).

After expression, GFP and mCherry fluorescence of individual cells were measured by flow cytometry. The fully extruded G-domain ($n_{\text{codons}} = 331$; Fig. 1c, red) exhibited GFP intensities ~100-fold higher than those observed for the partially synthesized G_{EF-G} ($n_{\text{codons}} = 212$, Fig. 1c, orange), while both constructs exhibited similar mCherry intensities. This result is in line with expectations for these nascent chain lengths²¹, indicating that the reporter system developed here accurately reports on co-translational folding. To determine the upper limit of the dynamic range of the assay, we

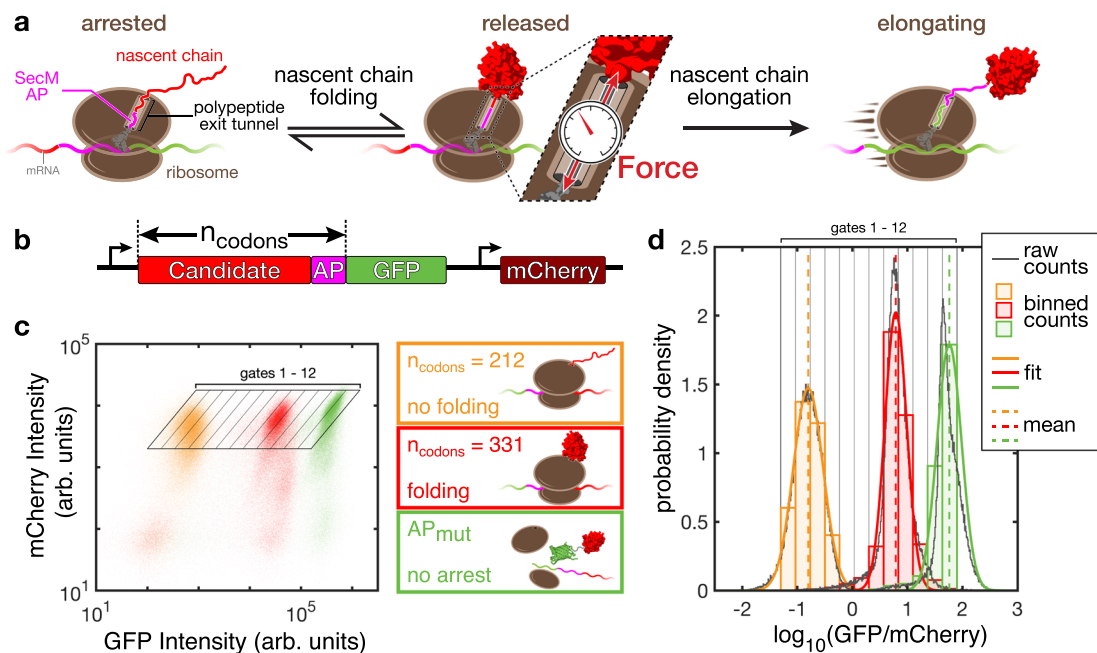


Fig. 1 | Arrest peptide-mediated detection of co-translational folding in cells. **a** Co-translational folding near the ribosome surface generates force and resumes translation arrested by the SecM arrest peptide (AP). **b** The candidate protein is fused to the SecM AP and GFP. Candidate folding results in increased GFP synthesis. The coding sequence of mCherry under control of a separate promoter on the same plasmid serves as an internal control. **c** Cytometry results of the individual reporter constructs. A non-folding truncation of EF-G ($n_{\text{codons}} = 212$) shows low GFP signal,

whereas a variant for which stable folding is expected ($n_{\text{codons}} = 331$) shows medium to high GFP signals. An arrest-defective AP variant (AP_{mut}) shows even higher GFP signal, delineating the upper boundary of the reporter readout. **d** Histogram of the GFP and mCherry signal ratios from the different variants. Imposing boundaries along constant $\log_{10}(\text{GFP}/\text{mCherry})$ ratios illustrates how the distribution of reporter signal can be reconstructed from cell sorting counts. Source data are provided as a Source Data file.

ablated arrest by replacing the wild-type SecM sequence with an arrest-defective mutant¹⁰ (AP_{mut}). Ablating arrest resulted in an approximately tenfold increase in GFP fluorescence compared to the $n_{\text{codons}} = 331$ construct (Fig. 1c, green). Therefore, arrest release by G_{EF-G} is well within the dynamic range of the assay, which spans about three orders of magnitude in GFP fluorescence intensity from the low ($n_{\text{codons}} = 212$) to the high (AP_{mut}) end.

Variations in experimental conditions, such as plasmid copy number or overall transcriptional activity, are likely to affect the overall accumulation of msGFP and mCherry. However, their ratio is expected to be relatively constant for a given construct. The logarithm of the measured fluorescence intensity ratio, $\log(I_{\text{GFP}}/I_{\text{mCherry}})$, is thus an appropriate measure of arrest release²⁷. For the individual constructs shown in Fig. 1c, raw counts of cells with a given $\log(I_{\text{GFP}}/I_{\text{mCherry}})$ ratio approximated a normal distribution (Fig. 1d, gray lines). We then grouped the experimental data into bins of equal $\log(I_{\text{GFP}}/I_{\text{mCherry}})$ intervals (diagonal gates 1–12 in Fig. 1c; bars in Fig. 1d). The cell counts across bins also approximated a normal distribution, with a mean very similar to that of the raw counts (colored lines in Fig. 1d). Thus, peak position can be retrieved from gated cell counts, and it should be possible to develop a multiplexed version of the AP assay by combining dual reporter expression with cell sorting and deep sequencing, as described in the following section.

AP Profiling quantifies co-translational folding with codon resolution

Based on the results described above, we designed an AP assay with increased throughput (Fig. 2a). We generated a library of G_{EF-G}

constructs by time-dependent exonuclease digestion²⁸. We optimized this approach to yield a uniform distribution of truncations along candidate ORFs. Out-of-frame constructs, which are expected to constitute approximately two-thirds of clones within the library, are terminated at stop codons before or within the SecM coding sequence and thus do not yield GFP-positive clones. Our sorting scheme eliminates these constructs (see Methods for details). Expressing the library in *E. coli* cells yielded GFP intensities spread over three orders of magnitude, whereas most mCherry intensities were more narrowly distributed (Fig. 2a). We subjected the cultures to fluorescence-activated cell sorting (FACS) using 12 sorting gates defined along boundaries of constant $\log(I_{\text{GFP}}/I_{\text{mCherry}})$ values (Fig. 2a and Methods). To determine the distribution of each construct from a mixture of clones in the sorting gates, we adopted a “sort-seq” scheme²⁷ to identify library members by their 3'-terminal sequence. Employing a custom deep-sequencing workflow, we determined the relative abundance of individual library members in each sorting gate. From the distribution of each truncation variant across sorting gates, we calculated an “AP score” for each clone in the library (see Methods). While short truncations are not well resolved (Methods), the approach yielded good coverage in the range $85 \leq n_{\text{codons}} \leq 400$ (Fig. S1).

AP Profiling analysis of G_{EF-G} revealed a strong folding event with a maximum in AP scores centered around $n_{\text{codons}} = 330$, a position where the complete domain has been extruded from the ribosome (Fig. 2b, red, region “ii”). In line with expectations²¹, short nascent chains (region “i”) and nascent chains beyond extrusion of the G-domain (region “iii”) do not exhibit high AP scores. However, we did observe substantially elevated AP scores in the region between $n_{\text{codons}} = 230$

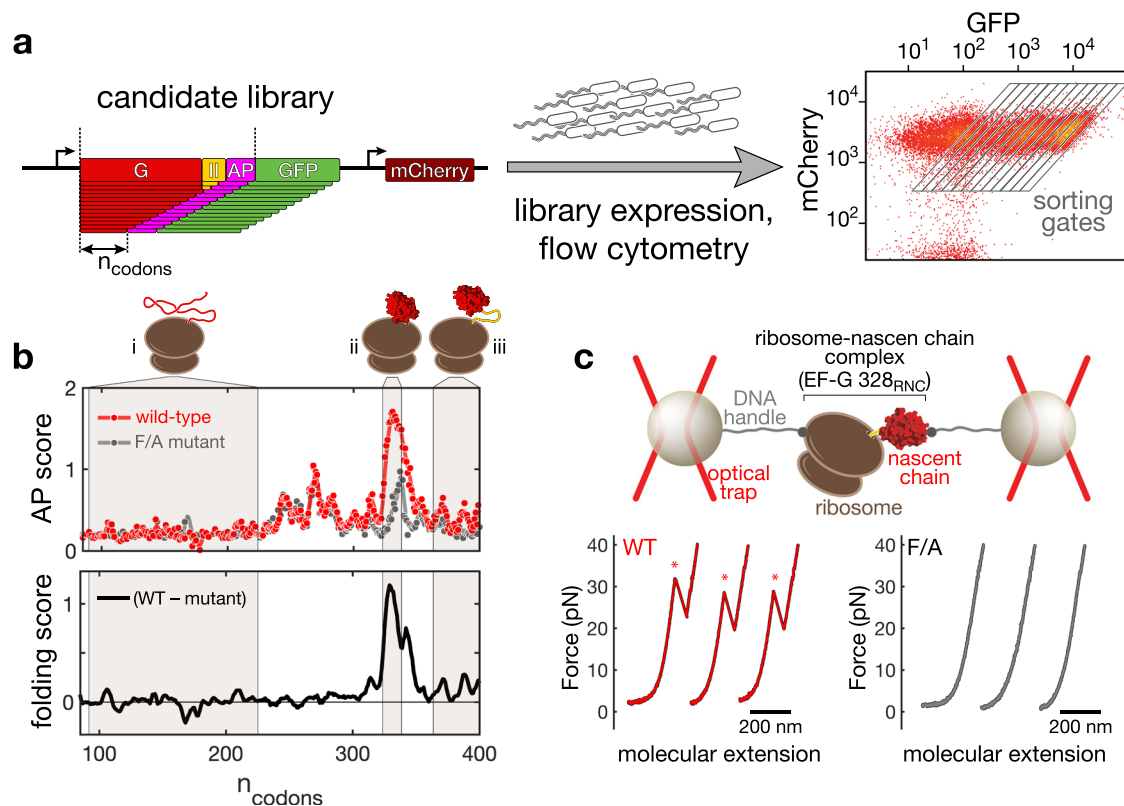


Fig. 2 | AP Profiling resolves folding at codon resolution. a A truncation library of the G-domain from EF-G (G_{EF-G}) is inserted into the AP Profiling construct. After library expression, cells are sorted into gates by their $\log_{10}(\text{GFP}/\text{mCherry})$ ratio. Deep sequencing of individual gates yields the distribution of each clone in the library. **b** AP scores (top panel) for wild-type G_{EF-G} (red) and a destabilized variant (F/A mutant, gray). AP scores are mostly similar, except for the region around $n_{\text{codons}} = 330$, where signals for the wild-type sequence are higher than for the

mutant, indicating folding into a stable tertiary structure. The folding score (bottom panel), calculated as the difference between wild-type and F/A mutant scores, reveals the position of tertiary folding around $n_{\text{codons}} = 330$. **c** Single-molecule optical tweezers experiments with ribosome-nascent chain complexes show the unfolding signature of wild-type G_{EF-G}, whereas the F/A mutant lacks a clear unfolding transition, indicating that it is not stably structured. Source data are provided as a Source Data file.

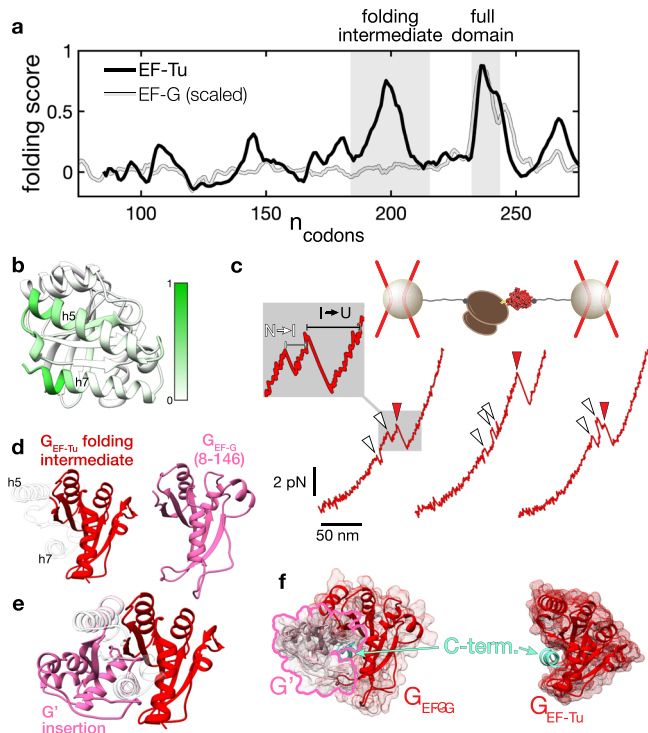


Fig. 3 | G_{EF-Tu} folds through a co-translational intermediate. **a** Folding scores obtained from AP Profiling experiments exhibit folding peaks for the partially (around $n_{\text{codons}} = 200$) and fully (around $n_{\text{codons}} = 230$) extruded G-domain of EF-Tu (G_{EF-Tu} , black line), indicating the population of a co-translational folding intermediate. The folding score of G_{EF-G} (scaled) is plotted for comparison (gray line). **b** Folding scores mapped onto the structure of G_{EF-Tu} indicate that the intermediate forms occur when helix 5 is being extruded from the ribosome. **c** Single-molecule optical tweezers recordings of G_{EF-Tu} nascent chain unfolding show reversible population of an on-pathway intermediate (white arrowhead) before complete unfolding (red arrowhead). **d** The structural region corresponding to the folding intermediate (red), based on length changes measured in optical tweezers experiments, comprises the N-terminus of G_{EF-Tu} up to helix 5. The unstructured part of the domain in the intermediate is shown as a transparent outline. The corresponding fragment from G_{EF-G} (pink), with identical topology and a very similar structure, is shown for comparison. **e** Position of the G' insertion in G_{EF-G} relative to the G_{EF-Tu} folding intermediate. The G' insertion is shown in pink, with flanking helices (corresponding to h5 and h6 in panel **d**) shown semi-transparent. **f** Comparison of G_{EF-G} and G_{EF-Tu} structures. G_{EF-Tu} lacks the G' subdomain present in G_{EF-G} that engulfs the last alpha-helix in the domain (cyan). pdb codes: 1efc (EF-Tu) and 4v9p (EF-G). Source data are provided as a Source Data file.

and $n_{\text{codons}} = 320$ (Fig. 2b, red), which had not been detected in previous experiments²¹. This apparent discrepancy likely results from higher sensitivity in AP Profiling (note that AP scores are calculated on a logarithmic scale), compared to a similar assay using a luciferase reporter²¹. Indeed, the linearly scaled $I_{GFP}/I_{mCherry}$ intensities resemble the folding profile reported previously using a luminescent reporter (Fig. S2).

Single-molecule force spectroscopy experiments with optical tweezers indicated that G_{EF-G} does not form a stable tertiary structure before extrusion of the complete domain²¹. To determine the origin of elevated AP scores preceding full domain extrusion, we generated a version of G_{EF-G} with perturbed stability. We changed all 8 phenylalanine residues in G_{EF-G} to alanine residues (F/A mutant). Replacing the bulky hydrophobic side chain of phenylalanine with the much smaller one of alanine disrupts packing of the hydrophobic core and greatly destabilizes tertiary structure²⁹. To confirm reduced stability of the G_{EF-G} F/A variant, we carried out single-molecule optical tweezers experiments³⁰ using ribosome-nascent chain complexes (RNCs)

(Fig. 2c), in which either the wild-type or the mutant domain were extruded from the ribosome¹². When we subjected RNCs with wild-type nascent chains to continuously increasing force (“force ramp”), we obtained force extension curves (FECs) that exhibit an unfolding transition (“rip”) characteristic for G_{EF-G} ^{12,26} (Fig. 2c, red traces). In contrast, the F/A mutant yielded smooth FECs (Fig. 2c, gray traces). The absence of the characteristic rip suggests that the mutant is not stably folded. These experiments confirm that the G_{EF-G} F/A mutant indeed lacks a stable tertiary structure.

We then analyzed the G_{EF-G} F/A mutant with AP Profiling (Fig. 2b, gray). The AP scores around $n_{\text{codons}} = 328$ (full G_{EF-G} extrusion) were greatly reduced, relative to the WT domain (Fig. 2b, region “ii”). In contrast, the minor peaks in the region of $230 \leq n_{\text{codons}} \leq 320$ remained largely unchanged (Fig. 2b, compare red and gray). Given that the F/A mutant is structurally unstable (Fig. 2c), these signals must originate from processes other than tertiary structure formation, perhaps including the formation of (transient) secondary structures. By subtracting the AP scores of the F/A mutant from those of the wild-type protein, we calculated a “folding score” (Fig. 2b, bottom panel, black line) that captures the formation of stable tertiary structures. Employing the destabilized mutant thus allows us to specifically detect co-translational formation of stable tertiary structures that represent major waypoints en route to fully folded proteins.

AP Profiling resolves a co-translational folding intermediate

EF-G belongs to a superfamily of translational GTPases (trGTPases) that share a conserved domain topology³¹. Elongation factor Tu (EF-Tu), a highly abundant and essential trGTPase^{32,33}, shares structural homology with EF-G. The G-domain of EF-Tu (G_{EF-Tu}) is much smaller than that of EF-G (200 vs 294 amino acids) because it lacks the G' insertion present in G_{EF-G} (Fig. S3).

To investigate how folding of G_{EF-Tu} differs from that of G_{EF-G} , we constructed AP Profiling libraries of wild-type (WT) G_{EF-Tu} and a folding-deficient F/A mutant. G_{EF-Tu} folding scores (Fig. 3a, black) exhibit several peaks along the coding sequence, in contrast to the single peak in G_{EF-G} scores (Fig. 3a, gray; see also Fig. S4). In addition to a strong peak near the position of full domain extrusion from the ribosome ($n_{\text{codons}} = 242$), G_{EF-Tu} exhibits a prominent signal at shorter chain lengths (centered at $n_{\text{codons}} = 198$). The AP Profiling analysis thus clearly indicates the formation of a stable tertiary structure before the domain is fully extruded from the ribosome exit tunnel, which sequesters ~40 amino acids. At position $n_{\text{codons}} = 198$, alpha-helix 5 (ending with Tyr160) has just exited the tunnel, whereas the high folding scores around $n_{\text{codons}} = 242$ map onto the end of the last alpha-helix in G_{EF-Tu} (Fig. 3b). AP Profiling therefore indicates that G_{EF-Tu} populates a co-translational folding intermediate, in contrast to G_{EF-G} .

From the AP Profiling results, it is not clear whether the observed folding intermediate represents a productive, on-pathway or a non-productive, potentially deleterious species. To distinguish between these two scenarios, we conducted single-molecule optical tweezers experiments with G_{EF-Tu} RNCs. Force extension curves of 240 amino acid long nascent chains reveal that G_{EF-Tu} partially unfolds in a near-equilibrium fashion before complete unfolding (Fig. 3c, white and red arrowheads, respectively). When combined, the two transitions add up to a contour length change of 71.8 ± 4.0 nm (SEM) (see Supplementary Methods and Fig. S5 for details), consistent with the expected value of 68.2 nm for unfolding natively structured G_{EF-Tu} . These results indicate that unfolding of fully folded G_{EF-Tu} proceeds through an obligatory intermediate.

The contour length of the intermediate state is 23.1 ± 1.2 nm (SEM) longer than the fully folded state (Fig. 3c and Fig. S5). This length change is consistent with the unfolding of ~60 amino acids from the C-terminus (Fig. S6). Our single-molecule analysis, therefore, suggests that the co-translational folding intermediate comprises the N-terminal ~140 amino acids (Fig. 3d). Importantly, G_{EF-Tu} never

transitions directly between the folded and the unfolded state, indicating that the intermediate is an obligatory species on the pathway between the unfolded state and native state.

The AP Profiling results suggest that the intermediate forms at an extruded length of around 160 amino acids, similar but not identical to the result from single-molecule experiments, which indicate folding at a length of around 140 amino acids (Fig. S6). This discrepancy may result from measurement uncertainties or differences between the environments for the *in vitro* single-molecule and the cellular AP Profiling experiments. Nevertheless, these experiments collectively suggest that G_{EF-Tu} populates an intermediate that is on-pathway to the natively folded domain.

The position at which the intermediate forms in G_{EF-Tu} corresponds to the site where the homologous G_{EF-G} domain contains a large insertion termed G' subdomain (Fig. 3e). A previous study found that G_{EF-G} folds only upon full domain extrusion from the ribosome, a finding corroborated by the data presented here (Fig. 2). This observation was rationalized by the presence of the G' insertion in G_{EF-G} that engulfs the helix at the extreme C-terminus of the domain (Fig. 3f). G_{EF-G} might thus have evolved a folding pathway in which the C-terminal helix folds prior to the G' insertion²¹. This interpretation is supported by the findings presented here. G_{EF-Tu} lacks the G'-insertion and the associated constraints on folding order. Our experiments show that the two homologous domains fold co-translationally along distinct pathways.

Functional redundancy of chaperone function in co-translational folding

Molecular chaperones guide co-translational folding³⁴. Even though their importance for efficient folding is firmly established, it is not well

understood how they promote folding and which steps along the folding pathways they affect. Individual nascent chain-binding chaperones have been the subject of mechanistic studies³. However, the presence of multiple chaperones with generally low binding specificity and moderate affinities results in competition for client binding in the cytosol³⁵. AP Profiling experiments, carried out in the presence of all cellular nascent chain-interacting proteins, may provide an opportunity to localize effects of individual nascent chain-binding chaperones, a first step toward understanding their contributions to nascent protein folding *in vivo*.

In *E. coli* and most other bacteria, trigger factor (TF) and DnaK, a member of the Hsp70 family, are the main chaperones that interact with growing nascent polypeptides, competing for nascent chain binding^{3,34} (Fig. 4a). The two chaperones share no homology and are regulated by distinct mechanisms. Individual deletion of TF or DnaK causes only relatively mild growth phenotypes³⁶, indicating functional redundancy³⁷, even though it is not clear whether they employ similar or distinct mechanisms to promote efficient *de novo* folding.

To detect specific chaperone effects during synthesis of G_{EF-G} and G_{EF-Tu} , we carried out AP Profiling experiments in TF and DnaK deletion strains³⁶ (Δtig and $\Delta dnaKJ$, respectively, both in an MC4100 background). For unknown reasons, libraries of *E. coli* G_{EF-Tu} exhibited growth defects upon expression in this strain background. We therefore chose to analyze the ortholog from *Bacillus subtilis*. The G-domain of *B. subtilis* EF-Tu shares 72% sequence identity, a virtually identical structure (RMSD of 0.55 Å) and very similar AP scores with the *E. coli* ortholog (Fig. S7), including the peaks for the folding intermediate (around $n_{codons} = 200$) and the complete domain (around $n_{codons} = 250$), respectively.

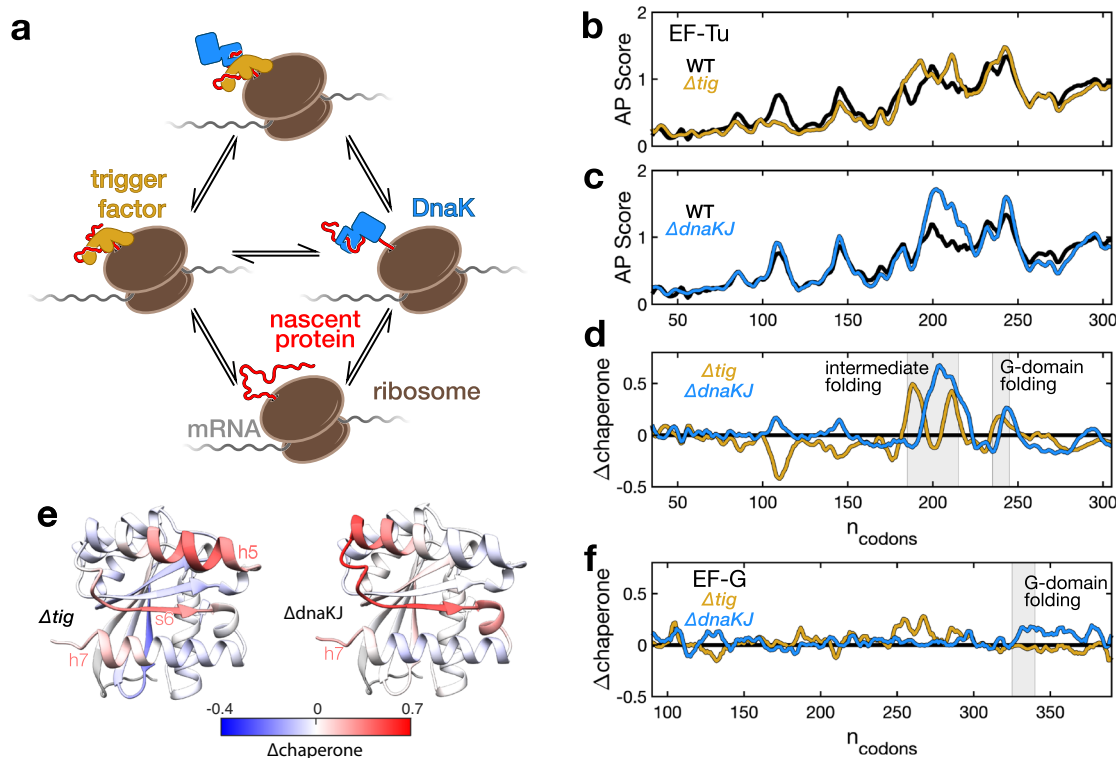


Fig. 4 | AP Profiling resolves chaperone effects on co-translational folding. **a** Trigger factor and DnaK compete for interactions with nascent chains on the ribosome. **b** Ablation of trigger factor (Δtig , gold) resulted in localized changes of AP scores compared to a wild-type background (WT, black). **c** Same as panel (b) for a DnaK deletion strain ($\Delta dnaKJ$, blue). **d** Δ chaperone scores, obtained as the difference in AP scores between wild-type and chaperone-knockout strains, exhibit

the most prominent change in the region of intermediate folding for both Δtig and $\Delta dnaKJ$. **e** Δ chaperone scores mapped onto the G_{EF-Tu} structure illustrate localization of chaperone deletion effects at distinct sites within the same region (around amino acid 200). **f** Chaperone deletion has only minor effects on G_{EF-G} scores, compared to G_{EF-Tu} . Source data are provided as a Source Data file.

The overall profiles of AP scores measured for G_{EF-Tu} are similar in Δtig and wild-type cells (Fig. 4b, yellow and black lines, respectively). However, specific regions showed markedly decreased (around $n_{codons} = 110$) or increased ($n_{codons} = 190$ and 210) scores upon TF ablation. Likewise, AP scores in the $\Delta dnaKJ$ strain were similar to those in wild-type cells, except for a prominent increase in the region around $n_{codons} = 210$ (Fig. 4c, blue and black lines, respectively). A minor effect is observed for both chaperones at position $n_{codons} \approx 240$, which corresponds to G-domain folding when the final alpha helix (h9) is extruded from the ribosome. Notably, neither TF nor DnaK ablation caused global changes in the measured AP scores, which is apparent when comparing the difference between wild-type and knockout backgrounds (Fig. 4d, $\Delta chaperone$).

The largest changes associated with chaperone ablation localize to the region where the co-translational folding intermediate is observed (see also Fig. 3), for both TF and DnaK. Interestingly, while both chaperones exert major effects in this region, the exact positions differ. Mapped onto the native G_{EF-Tu} structure, the most prominent effect of TF is apparent when helix h7 and strand s6 emerge from the ribosome, whereas the strongest DnaK effect is localized to the short segment in between these two elements (Fig. 4e). The chaperones therefore appear to engage nascent G_{EF-Tu} in distinct ways, yet in the same region. Very similar results were obtained for G_{EF-Tu} from *Thermus thermophilus* (Fig. S7), indicating conservation of these features. Our observations suggest that the functional redundancy between TF and DnaK might be achieved by directing these two chaperones to distinct sites within the same region.

Ablating TF or DnaKJ did not strongly affect folding signals along the ORF of G_{EF-G} (Fig. 4f), in contrast to the results with G_{EF-Tu} . The largest (although still minor) deviation from wild-type cells is seen for the Δtig strain in the region $240 \leq n_{codons} \leq 275$ of G_{EF-Tu} . In this region, corresponding roughly to the extrusion of the G' subdomain, we also observed high AP scores that were not due to tertiary structure formation (Fig. 2d). The $\Delta chaperone$ values are positive in this region, indicating that TF destabilizes these structures or interferes with their formation. Nevertheless, the AP Profiling results show that chaperone ablation does not profoundly affect G_{EF-G} folding.

AP Profiling for co-translational multi-domain protein folding

Most folding studies that used arrest peptides have employed reconstituted in vitro translation systems^{38–41} which not only lack the arsenal of nascent chain interactors present in the cytosol, but are also inefficient at synthesizing long polypeptides^{40,42}, restricting the approach to small proteins or domains. AP Profiling does not suffer from these limitations. We therefore utilized the approach to analyze how chaperones affect the folding of long multi-domain proteins, which constitute large fractions of all proteomes. For full-length EF-G, a 704-amino acid protein with five domains (Fig. S8a), we obtained good coverage throughout the entire open reading frame (Fig. S8b, red), which allowed us to calculate AP scores (Fig. S8c, red). The highest AP score maps to the position of complete G-domain at around $n_{codons} = 330$, a number of smaller local peaks around positions 437 (extrusion of domain II), 650 (extrusion of domain IV) and 750 (extrusion of the full-length protein, enabled by a C-terminal extension in the AP Profiling constructs).

To determine folding scores, we analyzed the destabilized F/A variant of full-length EF-G (Fig. S8d, Methods). The highest folding score around $n_{codons} = 330$ corresponds to co-translational G-domain folding. The second-highest peak is located at $n_{codons} = 437$, close to full extrusion of domain II, consistent with previously observed co-translational folding of domain II¹². Coverage for the F/A mutant dropped precipitously around $n_{codons} = 630$ (Fig. S8b) for unknown reasons. Long, destabilized EF-G fragments appear to interfere with cell growth, resulting in the elimination of these constructs from the library. The lack of F/A variant data does not allow us to draw firm

conclusions about the co-translational folding of the C-terminal part of EF-G. Nevertheless, the absence of high AP scores in this region (Fig. S8c) is consistent with in vitro experiments suggesting that this part of the protein completes folding post-translationally⁴³. Notably, one of the highest scores is observed when the full-length protein, tethered to the ribosome through a C-terminal linker, is extruded from the ribosome (Fig. S8c, around $n_{codons} = 750$).

To map chaperone effects on nascent EF-G, we next determined AP scores for wild-type, Δtig and $\Delta dnaKJ$ *E. coli* cells (Fig. S9). The chaperone deletion strains were created in an MC4100 background³⁶. To ensure that AP Profiling results are consistent among different strains, we compared results from wild-type MC4100 and Top10 cells. AP scores from MC4100 wild-type cells closely resembled those obtained from Top10 cells (Fig. S9a), with good coverage across the first 400 codons of the ORF. However, coverage in the C-terminal half of the protein was relatively poor in the wild-type and the chaperone deletion MC4100 strains (Fig. S9b), suggesting strain-specific growth defects. Expression of relatively long fragments of EF-G, a central component in protein synthesis, might pose challenges to cellular growth, similar to what we observed for a destabilized EF-G variant (Fig. S8).

The trGTPase superfamily contains several multi-domain proteins with strong homology to EF-G, including the six-domain translation elongation factor 4 (EF4) (Fig. S10). In contrast to EF-G, EF4 does not appear to be generally required for translation, instead functioning to facilitate translation elongation under specific stress conditions⁴⁴. To evaluate chaperone effects on co-translational folding of EF4, we expressed AP Profiling libraries of EF4 in wild-type and chaperone deletion strains (Fig. 5a and Fig. S11).

Ablation of Trigger factor resulted in AP score changes mainly at positions corresponding to extrusion of the first and third domains (G-domain and domain III) (Fig. 5a, b). The C-terminal domains remained largely unaffected by the deletion of either chaperone. As observed for the G-domain of EF-Tu (Fig. 4d), changes resulting from chaperone deletion are localized to defined positions. In the Δtig strain, both increased and decreased AP scores are observed (Fig. 5a). Deletion of the DnaK system caused changes that were small in magnitude compared to TF deletion (Fig. 5a, c).

The largest change in EF4 folding upon trigger factor deletion is localized to domain III (Fig. 5a, d). Around positions 386 and 398, AP scores in the Δtig strain are markedly decreased, before increasing around position 409, which corresponds to full extrusion of domain III (Fig. 5d). One possible explanation is that TF stabilizes smaller secondary structures, breaking up folding cooperativity. A similar trend is observed for domain III of another trGTPase, SelB (Fig. S12). Delineating chaperone effects on the folding of related multi-domain proteins may thus reveal recurring patterns of chaperone function.

Discussion

Co-translational folding is a crucial yet still poorly defined step in the biogenesis of functionally active proteins^{3,6,45}, particularly for large multi-domain proteins². We have combined AP-based folding detection in living cells with ratiometric flow cytometry, cell sorting and deep sequencing (Fig. 1), building a sensitive approach that maps co-translational folding with codon resolution in vivo. The approach not only yields high throughput and resolution, but also captures folding in the context of all the cytosolic factors that compete for nascent chain interaction. Thus, we were able to map major co-translational folding events and reveal chaperone effects on nascent chain folding using this approach, which we termed “Arrest Peptide Profiling” (AP Profiling).

Compared to previously reported AP-based folding measurements, which mostly rely on gel electrophoresis to detect arrest release, AP Profiling provides increased sensitivity. The AP score at each position reflects the average GFP/mCherry intensity ratio

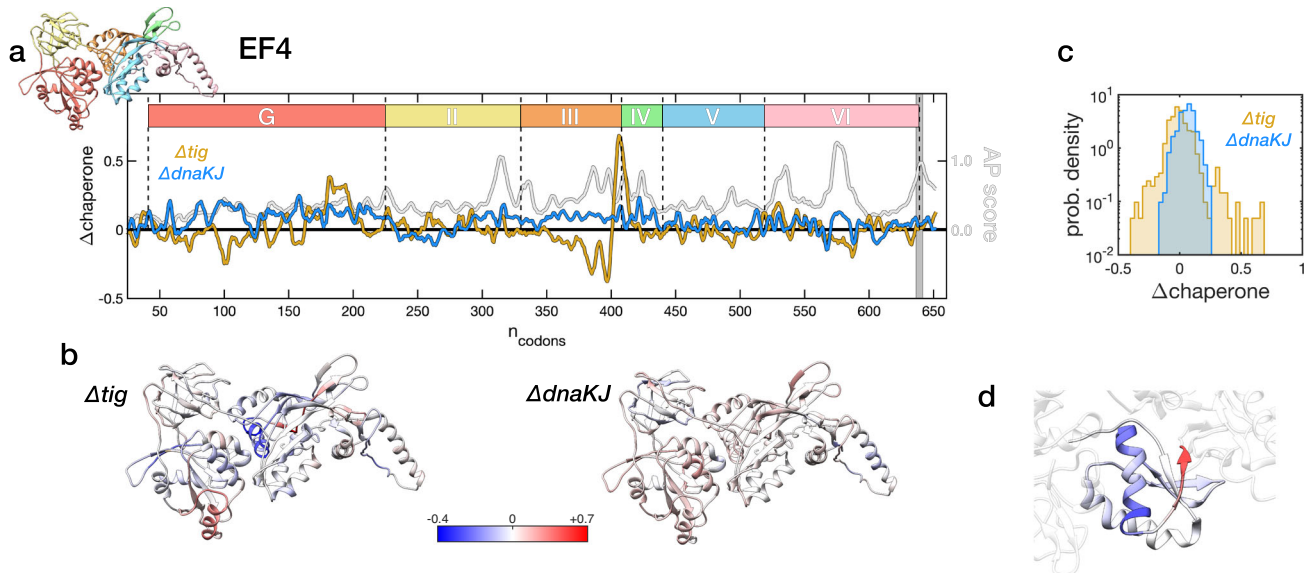


Fig. 5 | Localized chaperone effects on folding of a multi-domain protein. **a** The six-domain EF4 protein exhibits multiple AP score peaks (gray line), indicating several locations of co-translational folding. The vertical gray bar indicates the position of full-length protein extrusion from the ribosome. Chaperone deletions (Δ chaperone scores for Δ tig, gold, and Δ dnaKJ, blue) result in localized changes with both positive and negative signs. **b** Δ chaperone scores mapped onto the native

structure of full-length EF4. Trigger factor deletion elicits large effects on co-translational folding, most prominently for the G-domain and domain III. **c** Comparison of Δ chaperone scores for trigger factor and DnaK. Trigger factor causes larger changes to co-translational folding than DnaK. **d** Trigger factor causes opposite effects on neighboring alpha-helix and beta-sheet. Source data are provided as a Source Data file.

measured for dozens to hundreds of individual cells, yielding high signal-to-noise ratios. AP Profiling measurements show that the GTPase domain of EF-G (G_{EF-G}) does not fold into stable structures before the entire domain is extruded from the ribosome (Fig. 2), supporting earlier observations²¹. Even minor variations in AP scores along the G_{EF-G} ORF are resolved reproducibly (Fig. S13). At the same time, major co-translational events, such as full-length G_{EF-G} folding (around $n_{\text{codons}} = 330$), are still fully captured within the dynamic range of the assay (Fig. 1). These results illustrate the exquisite sensitivity and large dynamic range of the AP Profiling approach.

The high sensitivity of the AP Profiling assay revealed previously undetected co-translational folding signals during G_{EF-G} synthesis (compare Fig. 2b and Fig. S2). Experiments with destabilized candidate protein variants show that these well-defined increases in AP scores do not reflect tertiary structure formation. Rather, the formation of secondary structure, which can take place in ref. 46 and even be stabilized by ref. 8, the ribosome exit tunnel, may cause elevated AP scores. Alternatively, amino acid motifs adjacent to the SecM AP may affect arrest strength. In fact, even nascent chain segments outside the ribosome exit tunnel have been found to modulate arrest strength⁴⁷. It is not clear how either of these mechanisms would result in mechanical force on the SecM AP. As such, an arrest release profile may not represent a force profile.

Cells contain several quality control mechanisms that safeguard against potentially toxic effects of translation stalling and accumulation of structurally unstable proteins⁴⁸. It is possible that interactions of nascent chains with quality control machinery modulate the kinetics of SecM arrest release. Regardless of arrest release mechanism, distinguishing tertiary structure formation from other sources of signal AP-based folding measurements is important when defining co-translational folding pathways, e.g., with destabilized candidate variants (Fig. 2). Tertiary structures can clearly serve as waypoints for co-translational folding, whereas secondary structures may not be (kinetically) stable enough to serve as scaffolds for larger scale folding during slow polypeptide elongation.

In contrast to the GTPase domain of EF-G, G_{EF-Tu} exhibits a folding intermediate when -160 of its 200 amino acids have been

extruded (Fig. 3). Single-molecule optical tweezers experiments complementing AP Profiling measurements suggest that the intermediate corresponds to the core P-loop segment that is highly homologous in G_{EF-Tu} and G_{EF-G} . A common ancestor of the two domains may have folded through the co-translational intermediate suggested by our results. Insertion of the G' subdomain (Fig. 3f) may then have forced a complete remodeling of the folding energy landscape, resulting in distinct folding pathways of the homologous domains that start at the C-terminus for G_{EF-G} ²¹ and at the N-terminus for G_{EF-Tu} (this work). This interpretation would suggest that insertion of the G' domain resulted in changes to the G_{EF-G} sequence upstream of the insertion site that do not allow intermediate formation. Direct interactions between this part of the domain and the G' insertion cannot be responsible for the changed folding pathway, because the G' insertion is not yet synthesized at chain lengths where the intermediate would form.

The lack of co-translational folding intermediates suggests that G_{EF-G} accumulates in an unstructured or loosely structured conformation during synthesis, which is likely stabilized by nascent chain-binding chaperones against non-native folding or interactions. Consistent with this conjecture, the EF-G amino acid sequence exhibits normal hydrophobicity (see Fig. S14), a major determinant of chaperone binding⁴⁹. However, individually ablating trigger factor and DnaK, the two major nascent chain-binding chaperones in bacteria, caused only minor effects on the AP scores obtained with G_{EF-G} (Fig. 4). Chaperone binding itself therefore does not appear to have pronounced effects on AP scores in our experimental setup.

In contrast to G_{EF-G} , G_{EF-Tu} exhibits pronounced localized changes in AP scores when trigger factor or DnaK are removed (Fig. 4). The largest changes, expressed as Δ chaperone scores, are observed at a nascent chain length of ~200 amino acids, close to the position where the co-translational folding intermediate is detected. Notably, the effects of both chaperones localize to the same region, but distinct sites, indicating engagement with nascent G_{EF-Tu} in distinct ways. Notably, individual deletions of trigger factor and DnaK do not result in global changes to the folding scores we obtained with AP Profiling. These observations provide a molecular underpinning of the notion

that trigger factor and DnaK have overlapping binding sites and functions³⁶, even though they are unrelated.

Most of the pronounced Δ chaperone scores (Fig. 4) are positive, including those around $n_{\text{codons}}=200$ (intermediate folding) and $n_{\text{codons}}=240$ (full domain folding). A positive Δ chaperone score suggests that chaperone binding destabilizes folded structures, which is consistent with an earlier study of trigger factor effects on DHFR folding⁵⁰. However, trigger factor ablation also results in negative scores at some positions, including nascent chain lengths of 110 amino acids in $G_{\text{EF-Tu}}$ (Fig. 4) and 390 to 400 amino acids in EF4 (Fig. 5). Negative Δ chaperone scores might suggest that trigger factor promotes structure formation, perhaps within its binding cavity^{51–53}, or could be the result of particularly strong interactions of the chaperone with specific nascent chain sequences. The Δ chaperone score pattern observed for domain III in EF4 (Fig. 5) and SelB (Fig. S12), showing a dip followed by an increase, is consistent with both of these scenarios.

The combined action of trigger factor and DnaK likely has profound effects on the timing of co-translational folding. For instance, while domain II in EF-G folds immediately upon ribosome extrusion *in vitro*¹², its folding does not elicit high folding scores (Fig. S8), suggesting that the two co-translationally acting chaperones modulate its co-translational folding. Overall, DnaK ablation appears to mostly result in positive Δ chaperone scores (Figs. 4, 5), consistent with the canonical role of DnaK in stabilizing unstructured conformations of its clients. Δ chaperone scores also are, on average, smaller for DnaK ablation than for trigger factor ablation (Fig. 5 and Fig. S12). DnaK is believed to act downstream of trigger factor in chaperoning nascent chain clients¹⁶, which may result in fewer and weaker effects on co-translational folding near the ribosome. However, Δ chaperone scores of trigger factor and DnaK are of similar magnitude at specific positions (Fig. 4), illustrating that both chaperones can elicit strong effects on nascent chain folding.

AP Profiling enables co-translational folding measurements of large multi-domain proteins that cannot be efficiently synthesized in reconstituted *in vitro* translation systems. Here, we investigated EF-G (Fig. S8), EF4 (Fig. 5), and SelB (Fig. S12). The folding signals from EF-G AP Profiling pointed to co-translational folding of the N-terminal G and II domains and post-translational folding of domain III, consistent with the folding pathway determined by single-molecule optical tweezers^{12,43}. EF4 exhibited co-translational folding from all five domains at individual domain boundaries, with TF interaction detected at domain III (Fig. 5b). In contrast, SelB did not exhibit domain-wise folding at domains II and VI (Fig. S12). The observation that folding pathways diverge among homologous proteins highlights the intricacies of multi-domain protein folding and the need for more thorough and systematic examination. Likewise, interactions between domains in large globular proteins are ubiquitous and likely affect co-translational folding, even though few examples have been investigated in detail so far. AP Profiling experiments, combined with manipulations such as circular permutation or deletion of individual domains, will be a useful tool for guiding future investigations.

A limitation of the cellular AP Profiling approach is that expression of candidate truncations might result in growth defects and, consequently, lack of coverage (Figs. S8, 9). Reliable folding detection requires comparable cell viability and growth rates for all truncation fragments within the library. Expressing arrest peptides generates stalled ribosomes, which can interfere with growth⁵⁴. In addition, fragments of some proteins can cause toxicity⁵⁵. The N-terminal regions of candidate proteins are particularly impacted. They are expected to show low AP scores, resulting in their elimination from libraries alongside non-folding out-of-frame constructs (see Methods), and may themselves negatively affect cell growth⁵⁶. Careful optimization of expression conditions might alleviate some of these challenges. For instance, expression of an EF-G truncation library of *E. coli* EF-Tu yields marked differences in coverage upon expression in MC4100 and

Top10 cells, even though the overall AP score profile is preserved (Fig. S9). Likewise, orthologs from other species (here, we used *B. subtilis* and *T. thermophilus* EF-Tu) can circumvent expression problems (Fig. 3 and Fig. S7). Thus, the choice of expression strains and candidate orthologs for AP Profiling experiments can be tailored to achieve optimal coverage.

In summary, we describe here an effective methodology, termed AP Profiling, for mapping co-translational folding in live bacterial cells with high throughput. Our experiments revealed divergent co-translational folding pathways and can readily be extended to study other aspects of nascent chain biology, such as membrane insertion^{57,58}. The unparalleled throughput and sensitivity make AP Profiling ideally suited for *in vivo* studies of protein folding and membrane insertion and to dissect contributions from chaperones and other cellular machinery. Collectively, our advances provide much-needed tools for mechanistic investigations of protein biogenesis.

Methods

All commercially available enzymes were purchased from NEB unless stated otherwise. PCR reactions were carried out with Phusion high-fidelity DNA polymerase (Thermo Scientific, F530S). Chemicals were purchased from Sigma-Aldrich unless stated otherwise. DNA Tini spin column (Enzymax LLC, EZC106N) was used for DNA purification protocols. GeneJET plasmid miniprep kit (Thermo Scientific, K0503) was used for all plasmid miniprep steps. GeneJET PCR purification kit (Thermo Scientific, K0701) was used for all nucleotide purification. The binding buffer from K0701 was tested and used for quenching enzymatic reactions in AP Profiling at twice the reaction volume. *E. coli* TOP10 cells was from a commercial source (Invitrogen, C404010). MC4100 strains were from a previous study³⁶. Double-stranded DNA fragments (gBlock) and sequencing adapters are synthesized by IDT (Integrated DNA Technologies). See Supplementary Table S1 for candidate proteins used in this study.

AP Profiling vector design

The dual reporter design was inspired by a previous study⁵⁹. However, we encountered an unregulated sigma factor binding site in the mCherry coding sequence and decided to separate the two fluorescent proteins in different open reading frames. The AP Profiling vector (Fig. S15) is built by assembling multiple components from different sources (see Supplementary Methods for detailed descriptions). The various pieces of DNA fragments were obtained by PCR or commercial synthesis. The resulting vector allows for facile insertion of candidate proteins and unidirectional truncation library construction.

Truncation library construction

The truncation library is built by time-dependent exonuclease digestion, adapted from protocols described previously²⁸. The coding sequence of a candidate protein without the stop codon is cleared of conflicting endonuclease sites before being amplified with flanking digestion sites to be inserted between the BbsI sites in the AP Profiling vector. Fresh plasmid miniprep (more than 20 ml overnight culture for one protein) was performed to collect enough materials for enzymatic digestion. Purified plasmids were digested stepwise with SacI (NEB, R3156S) to prevent exoIII digestion on the arrest peptide side and then with AvrII (NEB, R0174S) to allow for exoIII digestion toward the candidate protein. Linearized dsDNA was purified and quantified by Qubit (1X dsDNA high sensitivity assay kit, Invitrogen, Q33231).

Incremental digestion by exoIII (Promega, M181I) was performed by diluting the linearized plasmids to a final concentration of 33 ng/ μ l before adding the exoIII amount to a final 0.1 U per nanogram of DNA (Qubit reading). We supplemented the reaction with NaCl at a final concentration of 30 mM. Based on our sequencing results, we estimated the digestion rate to be around 35 nt/min at 22 °C in the

digestion workflow, similar to the previously estimated rate⁶⁰. The reaction mixture was left in a heat block at 22 °C for 5 min before sampling every 30 s into a chilled quenching buffer (binding buffer from Thermo Scientific, K0701). Total digestion time depends on the length of the protein-coding sequence. The quenched DNA mixture was purified and quantified by Nanodrop (Thermo Scientific, ND-ONEC-W).

To blunt the digested DNA, mung bean nuclease (NEB, M0250S) was mixed with the purified DNA at a final concentration of 1.5 U per microgram of DNA (Nanodrop reading). In our workflow, 30 min reaction at 30 °C at this concentration is enough to blunt the DNA and digest the remaining RNA from miniprep without introducing non-specific cleavages in the dsDNA region. The DNA was purified and quantified by Nanodrop. Klenow fragment (NEB, M0210S) was added to the purified DNA for end-polishing at a final concentration of 1 U per microgram of DNA. Klenow reaction was performed at 25 °C for 15 min before purification and quantification by nanodrop.

The end-polished DNA was then diluted to around 2 ng/μl final concentration for blunt-end intramolecular ligation by T4 DNA ligase (NEB, M0202S). Ligation reaction was carried out at 16 °C for 16–20 h. The circularized DNA was then purified carefully by washing multiple times during a standard purification protocol and eluted to a final concentration around 40 ng/μl.

Transformation efficiency was tested by transforming 50–60 ng ligation product into TOP10 cells by electroporation. Importantly, all transformations were done with freshly cultured cells. One transformation reaction consists of 5 ml fresh culture harvested at OD₆₀₀ around 0.4. Cells were cultured in LB supplemented with 1% glucose at appropriate temperatures (37 °C for TOP10, 30 °C for MC4100 strains) to recover for 2 h at 220 rpm. Cultures were spun down and spread on agar plates supplemented with 1% glucose (to repress P_{BAD} expression) and carbenicillin plus strain-specific antibiotics³⁶. Plates were imaged for colony counting facilitated by openCFU⁶¹. Library coverage was calculated based on the potential truncation fragments (in-frame and out-of-frame constructs) and colony counts. Large-scale transformation was done for the truncation library to reach at least 1000X coverage. Colonies were then scraped and resuspended in fresh LB before being deposited as glycerol stocks for long-term storage.

Bacteria fluorescence-activated cell sorting (FACS)

Glycerol stocks of the truncation library in TOP10 cells were revived for overnight cultures in LB supplemented with 1% glucose. Cells carrying the truncation library were freshly diluted into glucose-free medium and grown for at least 2 hours to OD₆₀₀ = 0.2. The cultures were induced with arabinose at a final concentration of 0.05% (w/v) to express for 2 h.

The bacteria population was diluted in PBS and sorted with a MoFlo XDP equipment (Beckman Coulter) based on GFP and mCherry signals (488 nm laser, 514/30 band-pass filter for GFP detection; 561 nm laser 610/30 band-pass filter for mCherry detection). Based on our AP Profiling workflow and a similar bacteria sorting application⁶², the first sort of the population (enrichment sort) serves to enrich the high GFP population, which also serves to eliminate out-of-frame constructs which—due to the library construction workflow—are expected to constitute approximately two-thirds of clones. We arbitrarily set up four gates for high-mCherry events, which contain the expressing population, from low to high GFP signals. Sorting was done at around 14,000 events per second and typically proceeded for 10 min when the scarcest gate contains at least 20,000 sorted events (the most abundant gate normally has 400,000 events by this time). Sorted samples were grown in LB supplemented with 1% glucose (w/v) and carbenicillin plus strain-specific antibiotics overnight at 37 °C, 220 rpm. Plasmids and glycerol stocks were prepared from the overnight cultures.

For AP Profiling in TOP10 cells, glycerol stocks of TOP10 cells were revived for fresh culture before mixing in proportions to enrich high

GFP species. For AP Profiling to study chaperone effects, purified plasmids were mixed in proportions to enrich high GFP species before being electroporated into MC4100 strains (PG1 for WT, PG2 for ΔTF, and PG3 for ΔDnaKJ) with a minimum 1000X coverage. The new mixed cultures were then sorted again (quantification sort) with a more defined gating strategy (Fig. 2), similar to previous studies^{63,64}. Specifically, the log(GFP/mCherry) values linearly carve up the GFP-mCherry plot on the log scale during cytometry. The Summit (v5.5.0) software allows for manually adjusting gate boundaries according to a coordinate system. We then drew up 12 slanted (slope = 1) gates on the high-mCherry population, shoulder by shoulder, with a fixed gate width. Each sorting gate can now be represented by a single log(GFP/mCherry) value. The MoFlo XDP allows for sorting four gates simultaneously. We ran the sorter at around 14,000 events per second and recorded the exact numbers of detected events, sorted events and the total number of events in the high-mCherry cluster (for normalization across groups during analysis). Running the sorted samples back through the cytometry resulted in reproducible signal levels, verifying singlet bacteria sorting (Fig. S16). Each four-gate sorting lasts around 10 min to collect enough cells for sequencing detection. Sorted samples were grown in LB supplemented with 1% glucose (w/v) and carbenicillin plus strain-specific antibiotics overnight at appropriate temperatures (37 °C for TOP10, 30 °C for MC4100 strains), 220 rpm. Plasmids and glycerol stocks were prepared from the overnight cultures.

High-throughput sequencing

We developed a custom end targeting high-throughput sequencing scheme, inspired by Tail-seq⁶⁵. Plasmids extracted from truncation library constructions, transformations and sorting were used for sequencing library preparations. Briefly, the BsaI recognition site near the arrest peptide coding sequence was digested, and the overhang on the truncation fragment site was used to ligate with a biotin-modified P5 adapter containing the UMI and desired index. The single-stranded portion of the P5 adapter was then filled up by Klenow fragment before purification and quantification by Qubit. We subcloned the Tn5 transposase (from Addgene plasmid #60240) and purified the enzyme for DNA tagmentation⁶⁶. The ME-B fragment was transposed into the DNA samples before biotin enrichment by streptavidin-coated magnetic beads (M280 beads, Invitrogen, 11205D). The beads were vigorously washed with wash buffer according to the manufacturer's recommendations. The magnetic beads, bound with the biotin-labeled DNA fragments, were directly used as a template for a limited-cycle PCR (8–10 cycles) with Phusion DNA polymerase to attach the desired P7 adapter. PCR products were purified and subjected to size selection by AMPure beads (Beckman Coulter, A63880) to remove short DNA fragments (0.8X volume).

The resulting DNA architecture (Fig. S17) is compatible with Illumina sequencing platforms. The samples from this study were sequenced with HiSeq2500 (50 cycles single end) as well as NovaSeq6000 SP (100 cycles, single end) with additional index cycles to acquire UMI information.

AP Profiling data analysis

The sequencing data was processed by custom scripts chaining standard analysis tools. Briefly, short-read sequencing data were demultiplexed and de-duplicated based on index and UMI information. Reads were mapped to candidate protein sequences to locate the C-termini of the truncations, identifying the nascent chain lengths before translation arrest. The distribution of sequencing reads was correlated with cell sorting information to infer the distribution of the truncated protein across the sorting gates. Fitting a Gaussian curve over the distribution of cell counts produces an “AP score” (Fig. 1) for the truncated protein. The “folding score” refers to the value after subtracting the AP score from the F/A mutant of the same protein. A spline

curve is fitted for the scores along the open reading frame for visualization. Coverage was calculated as the fraction of codons with an AP score over a 10-codon sliding window. See Supplementary Methods for detailed processing steps and descriptions on folding score calculations. Because the nascent chain length is inferred from sequencing the template DNA, we report positions from AP Profiling experiments in terms of number of codons translated (n_{codons}), rather than nascent chain length in amino acids.

Flow cytometry of individual truncation constructs

Individual truncation constructs were generated by Gibson assembly. TOP10 *E. coli* cells were transformed with the truncation plasmids separately by heatshock. Individual colonies were used to inoculate fresh LB media supplemented with 1% glucose and carbenicillin for overnight cultures. Dilutions were made in fresh media without glucose and cultured for at least 2 h before induction by arabinose in the same way for FACS experiments. Cells were diluted in PBS before running through an Attune NxT Flow Cytometer (Invitrogen). FCS files were analyzed by custom Python scripts.

Single-molecule force spectroscopy of RNC

Ribosome nascent chain complexes (RNC) were generated and prepared as described in previous work³⁰. Briefly, protein-coding sequences with an N-terminal SpyTag⁶⁷ were amplified without a stop codon. Ribosomes with a SpyTag-labeled L17 subunit was used during the in vitro translation system (NEB, E3313S) to generate the stalled RNC. The reaction products were coupled to SpyCatcher conjugated double-stranded DNA that served as handles, which also carried affinity labels for binding to streptavidin or anti-digoxigenin coated polystyrene beads. The RNC/handle adducts were then tethered between two polystyrene beads with diameters of 2.1 μm (see ref. 30 for details). Optical tweezers force-ramp measurements were carried out with a dual-trap optical tweezers instrument (CTrap, Lumicks) at ambient temperature (22 °C). The trap stiffness was in the range of 0.4 to 0.45 pN/nm and was determined before each measurement. One of the traps was moved at a constant velocity of 100 nm/s for force-ramp measurements.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Illumina sequencing data has been deposited in the sequence read archive (SRA) under the BioProject accession code [PRJNA1212592](https://doi.org/10.5281/zenodo.14629149). Processed data and related files are deposited in Zenodo [<https://doi.org/10.5281/zenodo.14629149>]. Source data are provided with this paper.

Code availability

Custom code related to this work has been deposited in Zenodo [<https://doi.org/10.5281/zenodo.14629149>].

References

- Brasemann, E., Chaney, J. L. & Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* **38**, 337–344 (2013).
- Rajasekaran, N. & Kaiser, C. M. Navigating the complexities of multi-domain protein folding. *Curr. Opin. Struct. Biol.* **86**, 102790 (2024).
- Kramer, G., Shiber, A. & Bukau, B. Mechanisms of cotranslational maturation of newly synthesized proteins. *Annu. Rev. Biochem.* **88**, 337–364 (2019).
- Cassaignau, A. M. E., Cabrita, L. D. & Christodoulou, J. How does the ribosome fold the proteome?. *Annu. Rev. Biochem.* **89**, 389–415 (2020).
- Rajasekaran, N. & Kaiser, C. M. Co-Translational folding of multi-domain proteins. *Front. Mol. Biosci.* **9**, 869027 (2022).
- Samatova, E., Komar, A. A. & Rodnina, M. V. How the ribosome shapes cotranslational protein folding. *Curr. Opin. Struct. Biol.* **84**, 102740 (2024).
- Lu, J. & Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **12**, 1123–1129 (2005).
- Agirrezabala, X. et al. A switch from α -helical to β -strand conformation during co-translational protein folding. *EMBO J.* **41**, e109175 (2022).
- Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).
- Yap, M. N. & Bernstein, H. D. The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol. Cell* **34**, 201–211 (2009).
- Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).
- Liu, K., Maciuba, K. & Kaiser, C. M. The ribosome cooperates with a chaperone to guide multi-domain protein folding. *Mol. Cell* **74**, 310–319.e7 (2019).
- Hoffmann, A. et al. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell* **48**, 63–74 (2012).
- Chan, S. H. S. et al. The ribosome stabilizes partially folded intermediates of a nascent multi-domain protein. *Nat. Chem.* **14**, 1–9 (2022).
- Waudby, C. A. et al. Systematic mapping of free energy landscapes of a growing filamin domain during biosynthesis. *Proc. Natl Acad. Sci. USA* **115**, 9744–9749 (2018).
- Streit, J. O. et al. The ribosome lowers the entropic penalty of protein folding. *Nature* **633**, 1–8 (2024).
- Oh, E. et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308 (2011).
- Stein, K. C., Kriel, A. & Frydman, J. Nascent polypeptide domain topology and elongation rate direct the cotranslational hierarchy of Hsp70 and TRiC/CCT. *Mol. Cell* **75**, 1117–1130.e5 (2019).
- Deckert, A. et al. Common sequence motifs of nascent chains engage the ribosome surface and trigger factor. *Proc. Natl Acad. Sci. USA* **118**, e2103015118 (2021).
- Goldman, D. H. et al. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* **348**, 457–460 (2015).
- Chen, X., Rajasekaran, N., Liu, K. & Kaiser, C. M. Synthesis runs counter to directional folding of a nascent protein domain. *Nat. Commun.* **11**, 5096 (2020).
- Marino, J., von Heijne, G. & Beckmann, R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett.* **590**, 655–660 (2016).
- Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
- Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
- Shen, Y., Chen, Y., Wu, J., Shaner, N. C. & Campbell, R. E. Engineering of mCherry variants with long Stokes shift, red-shifted fluorescence, and low cytotoxicity. *PLoS ONE* **12**, 1–14 (2017).
- Liu, K., Rehfus, J. E., Mattson, E. & Kaiser, C. M. The ribosome destabilizes native and non-native structures in a nascent multi-domain protein. *Protein Sci.* **26**, 1439–1451 (2017).
- Peterman, N. & Levine, E. Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics* **17**, 1–17 (2016).

28. Patrick, W. M. & Gerth, M. L. ITCHY: Incremental Truncation for the Creation of Hybrid Enzymes. In *Directed Evolution Library Creation: Methods and Protocols* (eds. Gillam, E. M. J., Copp, J. N. & Ackerley, D.) Ch. 16 Vol 1179 (Springer, 2014) https://doi.org/10.1007/978-1-4939-1053-3_16.
29. Notari, L., Martínez-Carranza, M., Fariás-Rico, J. A., Stenmark, P. & von Heijne, G. Cotranslational folding of a pentarepeat β -helix protein. *J. Mol. Biol.* **430**, 5196–5206 (2018).
30. Maciuba, K., Zhang, F. & Kaiser, C. M. Facile tethering of stable and unstable proteins for optical tweezers experiments. *Biophys. J.* **120**, 2691–2700 (2021).
31. Maracci, C. & Rodnina, M. V. Review: translational GTPases. *Biopolymers* **105**, 463–475 (2016).
32. Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2016).
33. Nissen, P. et al. Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
34. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).
35. Hsieh, H.-H., Lee, J. H., Chandrasekar, S. & Shan, S. A ribosome-associated chaperone enables substrate triage in a cotranslational protein targeting complex. *Nat. Commun.* **11**, 5840 (2020).
36. Genevaux, P. et al. In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* **5**, 195–200 (2004).
37. Deuerling, E. et al. Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* **47**, 1317–1328 (2003).
38. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
39. Jensen, M. K., Samelson, A. J., Steward, A., Clarke, J. & Marqusee, S. The folding and unfolding behavior of ribonuclease H on the ribosome. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.RA120.013909> (2020).
40. Kemp, G., Kudva, R., de la Rosa, A. & von Heijne, G. Force-profile analysis of the cotranslational folding of HemK and filamin domains: comparison of biochemical and biophysical folding assays. *J. Mol. Biol.* **431**, 1308–1314 (2019).
41. Liutkute, M., Maiti, M., Samatova, E., Enderlein, J. & Rodnina, M. V. Gradual compaction of the nascent peptide during cotranslational folding on the ribosome. *eLife* **9**, 97–97 (2020).
42. Li, J. et al. Dissecting limiting factors of the protein synthesis using recombinant elements (PURE) system. *Translation* **5**, e1327006 (2017).
43. Liu, K., Chen, X. & Kaiser, C. M. Energetic dependencies dictate folding mechanism in a complex protein. *Proc. Natl Acad. Sci. USA* **116**, 25641–25648 (2019).
44. Pech, M. et al. Elongation factor 4 (EF4/LepA) accelerates protein synthesis at increased Mg²⁺ concentrations. *Proc. Natl Acad. Sci. USA* **108**, 3199–3203 (2011).
45. Waudby, C. A., Dobson, C. M. & Christodoulou, J. Nature and regulation of protein folding on the ribosome. *Trends Biochem. Sci.* **44**, 914–926 (2019).
46. Nilsson, O. B. et al. Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep.* **12**, 1533–1540 (2015).
47. Yang, Z., Iizuka, R. & Funatsu, T. Nascent SecM chain outside the ribosome reinforces translation arrest. *PLoS ONE* **10**, e0122017 (2015).
48. Filbeck, S., Cerullo, F., Pfeffer, S. & Joazeiro, C. A. P. Ribosome-associated quality-control mechanisms from bacteria to humans. *Mol. Cell* **82**, 1451–1466 (2022).
49. Rüdiger, S., Schneider-Mergener, J. & Bukau, B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J.* **20**, 1042–1050 (2001).
50. Nilsson, O. B., Müller-Lucks, A., Kramer, G., Bukau, B. & Von Heijne, G. Trigger factor reduces the force exerted on the nascent chain by a cotranslationally folding protein. *J. Mol. Biol.* **428**, 1356–1364 (2016).
51. Ferbitz, L. et al. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).
52. Rutkowska, A. et al. Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* **283**, 4124–4132 (2008).
53. Hoffmann, A., Bukau, B. & Kramer, G. Structure and function of the molecular chaperone trigger factor. *Biochim. Biophys. Acta* **1803**, 650–661 (2010).
54. Saito, K. et al. Ribosome collisions induce mRNA cleavage and ribosome rescue in bacteria. *Nature* **603**, 503–508 (2022).
55. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
56. Evans, M. S., Ugrinov, K. G., Frese, M.-A. & Clark, P. L. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat. Methods* **2**, 757–762 (2005).
57. Fujiwara, K., Katagi, Y., Ito, K. & Chiba, S. Proteome-wide capture of co-translational protein dynamics in *Bacillus subtilis* using TnDR, a transposable protein-dynamics Reporter. *Cell Rep.* **33**, 108250 (2020).
58. Nicolaus, F. et al. Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *eLife* **10**, e64302 (2021).
59. Levin-Karp, A. et al. Quantifying translational coupling in *E. coli* synthetic operons using RBS modulation and fluorescent reporters. *ACS Synth. Biol.* **2**, 327–336 (2013).
60. Ostermeier, M., Lutz, S. & Benkovic, S. Generation of protein fragment libraries by incremental truncation. *Protein-Protein Interact. Mol. Cloning Man.* **507**, 522 (2002).
61. Geissmann, Q. OpenCFU, a new free and open-source software to count cell colonies and other circular objects. *PLoS ONE* **8**, e54072 (2013).
62. Zutz, A. et al. A dual-reporter system for investigating and optimizing protein translation and folding in *E. coli*. *Nat. Commun.* **12**, 6093 (2021).
63. Kosuri, S. et al. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **110**, 14024–14029 (2013).
64. Levy, L. et al. A synthetic oligo library and sequencing approach reveals an insulation mechanism encoded within bacterial σ 54 promoters. *Cell Rep.* **21**, 845–858 (2017).
65. Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: Genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* **53**, 1044–1052 (2014).
66. Hennig, B. P. et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3* **8**, 79–89 (2018).
67. Min, D., Arbing, M. A., Jefferson, R. E. & Bowie, J. U. A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Sci.* **25**, 1535–1544 (2016).

Acknowledgements

This work was supported by a Grant from the National Institutes of Health (5R01GM121567) to C.M.K. We acknowledge Dr. Hao Zhang at the Cell Sorting Core Facility in JHMI for cell sorting operations, Dr. Will Ludington for generously sharing the flow cytometer, Dr. David Mohr at GRFC in JHMI for sequencing operations, and Advanced Research Computing at Hopkins (ARCH) core facility (supported by National Science Foundation grant number OAC1920103) for high-performance computation resources.

Author contributions

X.C. performed the experiments. X.C., V.J.H., and C.M.K. carried out the analysis. X.C. and C.M.K. wrote the manuscript, and all authors edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-61398-6>.

Correspondence and requests for materials should be addressed to Christian M. Kaiser.

Peer review information *Nature Communications* thanks Christian A. M. Wilson who co-reviewed with Francesca Burgos Bravo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025