

Structural determinants of co-translational protein complex assembly

Saurav Mallik^{1,5,*}, Johannes Venezian^{2,5}, Arseniy Lobov^{1,5}, Meta Heidenreich¹, Hector Garcia-Seisdedos³, Todd O. Yeates⁴, Ayala Shiber^{2,*}, Emmanuel D. Levy^{1,*}

Author Affiliations

¹ Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 7600001, Israel.

² Faculty of Biology, Technion-Israel Institute of Technology, Haifa 3200003, Israel.

³ Department of Structural and Molecular Biology, Institute of Molecular Biology, Barcelona 08028, Spain.

⁴ Department of Chemistry and Biochemistry, University of California, Los Angeles 90095, United States.

⁵ These authors contributed equally to the work.

* To whom the correspondence should be addressed.

Emails and ORCIDs

S.M.: saurav.mallik@weizmann.ac.il (0000-0002-9390-0715)

J.V.: venezian@campus.technion.ac.il (0009-0002-2507-3356)

A.L.: arseniy.lobov@weizmann.ac.il (0000-0002-0930-1171)

M.H.: meta.heidenreich@weizmann.ac.il (0000-0002-2095-830X)

H.G.S.: hgsbmc@ibmb.csic.es (0000-0001-7722-2793)

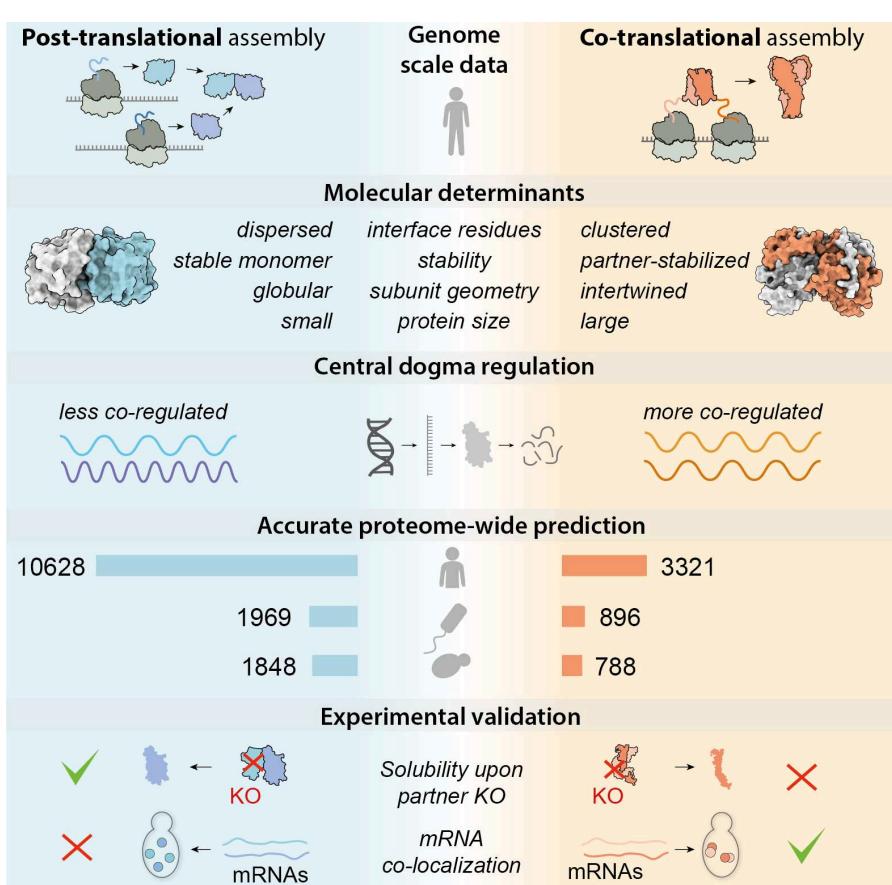
T.O.Y.: yeates@mbi.ucla.edu (0000-0001-5709-9839)

A.S.: ayalashiber@technion.ac.il (0000-0002-3167-1494)

E.D.L.: emmanuel.Levy@weizmann.ac.il (0000-0001-8959-7365)

One Sentence Summary. Protein complexes with topologically intertwined subunits require co-translational assembly and synchronized proteostasis of subunits, with implications in protein stability, mRNA localization, and evolution.

Graphical Abstract



ABSTRACT

The assembly of proteins into functional complexes is critical to life's processes. While textbooks depict complex assembly as occurring between fully synthesized proteins, we know today that thousands of proteins in the human proteome assemble co-translationally during their synthesis. Why this process takes place, however, remains unknown. We show that co-translational assembly is governed by biophysical and structural characteristics of the protein complex, and involves mutually stabilized, intertwined subunits. Consequently, these subunits are also co-regulated across the central dogma, from transcription to protein degradation. Leveraging structural signatures with AlphaFold2-based predictions enables us to accurately predict co-translational assembly on a proteome-wide scale, which we validated by ribosome profiling, genetic perturbations, and smFISH experiments. Notably, the latter showed that co-translationally assembling subunits exhibit co-localized mRNAs. This work unveils a fundamental connection between protein structure and the translation process, highlighting the overarching impact of three-dimensional structure on gene expression, mRNA localization, and proteostasis.

INTRODUCTION

The self-organization of biological macromolecules is essential for executing cellular functions. Among proteins, self-organization ranges from folding into specific three-dimensional structures^{1–6} to assembly into large multi-component complexes^{7–9}. Crucially, the ability to characterize these folding and assembly events in detail is intrinsically constrained by the suitability of proteins for biophysical analysis. Consequently, detailed biophysical studies of protein folding have predominantly concentrated on small, globular proteins due to their reversible folding properties⁶. However, these proteins represent less than 10% of the structural diversity within even a bacterial proteome¹⁰. Characterizing the remaining 90% poses a significant challenge, as these proteins are often more susceptible to misfolding or aggregation^{10–13}. This leaves the vast majority of the cellular machinery's folding and assembly mechanisms largely unexplored.

The analysis of protein complexes' assembly faces similar limitations. Indeed, the reconstitution of complexes can be notoriously challenging, necessitating specialized technologies for spatiotemporal control of gene expression^{11,12}. As a result, protein complex isolation often involves the purification of endogenously tagged copies and structural characterization, e.g., by cryo-electron microscopy (Cryo-EM) (16, 17). While Cryo-EM can provide insights into the assembly of complexes through snapshots acquired along their assembly pathway (12) or across different functional states (13, 14), such approaches are focused on specific systems so that molecular determinants of folding and assembly remain uncharacterized for the most part of proteomes. However, previous work has shown that static structural data can be leveraged in computational analyses to gain insights into protein folding^{4,5,14} and assembly mechanisms^{15–20}. This idea motivates us to leverage the wealth of structural data to gain insights into molecular determinants of the dynamic process that is co-translational assembly.

A key distinction in the assembly pathway of protein complexes lies in the post- or co-translational timing of assembly (**Fig. 1A**). In the latter, translation, folding, and assembly occur simultaneously, which can help funneling the assembly pathway of complexes, minimizing promiscuous interactions, or regulating orphan subunit degradation^{21,22}. Recent reports have shown that co-translational assembly is prevalent^{23–27}. In particular, Bertolini et al.²⁸ identified thousands of proteins undergoing co-translational assembly by sequencing footprints protected among isolated disomes, *i.e.*, pairs of ribosomes connected by their respective, interacting, nascent chains. Because disome footprints are sequenced over an entire lysate,

such an experiment can provide a list of co-translationally assembling protein subunits – here referred to as ‘coco’ subunits – but does not inform on interacting partners. These limitations motivated us to integrate these disome data with protein structures, both to uncover specific molecular signatures of co-translational assembly as well as identify specific interacting pairs undergoing this process in cells.

Our analysis uncovered that co-translational assembly is governed by structural and biophysical features of protein complexes and often involves intertwined and mutually stabilized subunits. This structural feature implies that complexes assembling co-translationally are more kinetically stable than complexes assembling post-translationally, which is manifested in longer subunit residence times and similar thermal denaturation temperatures of interacting subunits. Additionally, we find that the inter-dependence of interacting coco subunits necessitates their co-regulation, presumably to balance their stoichiometry. Utilizing these concepts, we developed a model relying on structural data to predict co- or post-translational assembly in both homo- and heteromeric complexes. We validated these predictions through established disome footprint data and targeted experiments. These findings pave the way to a better understanding of protein folding and assembly beyond the textbook knowledge of reversible folding and binding, and they highlight the overarching impact of protein structure on translation and expression regulation.

RESULTS

Subunit interaction geometry underlies co- and post-translational modes of assembly

We mapped the comprehensive dataset of human coco subunits from reference ²⁸ onto available 3D structures of homo- and hetero-oligomeric complexes ^{29,30}. We anticipated coco subunits to show distinct structural characteristics, notably intertwined subunits, which may suggest a concurrent folding-binding mechanism ^{31–33}. Indeed, intertwined subunits, stabilized through extensive contacts with their partner, are expected to be unstable as monomers and so could use co-translational assembly as a means to bypass an unstable monomeric state. To capture relevant geometric and structural features, we employed a combination of established and newly defined metrics (*Methods 1-2*).

We analyzed such signatures in coco and ‘post’ (referring to post-translationally assembling) subunits, using protein complexes of known structure. We initially focused on homo-oligomers as they alleviate the need to infer coco subunit pairs, which are not known. We identified 204 and 449 homo-oligomers of known structure experimentally classified as coco or post-subunits respectively.

Several structural properties of these complexes showed significant differences between the two groups (**Fig. 1 B-E**). Notably, interface residues appeared more clustered in the sequence among coco subunits than among post (**Fig. 1B**, Mann-Whitney U-test $p = 7E-38$). Such localized interaction sites may decrease the necessity to synthesize and fold long segments before binding, thereby promoting co-translational partner recognition. In a different analysis, we found the interface residues exhibited a higher ratio of inter/intra molecular contacts among coco subunits relative to post (MWU $p = 2E-32$, **Fig. 1C**). The pronounced inter-molecular contacts imply that coco subunits are equally stabilized by their partner (inter) as they are by their own tertiary structure (intra). Such partner-dependent stability is reminiscent of domain-swapping, which we analyzed through the V_m/V_d ratio, of convex-hull volumes enclosing the monomer (V_m) and dimer (V_d). Indeed, the larger ratios observed for coco-subunits relative to post (MWU $p = 4E-26$, **Fig. 1D**, *Methods 2.5*) indicate pronounced intertwined structures. Interestingly, large interface areas have previously been linked to co-translational assembly ³², and our findings suggest that the underlying determinant is the subunit interaction geometry (*Supplementary Note 1-2*). Finally, we observed

that coco subunits were encoded by longer polypeptide chains than post-subunits (MWU $p = 2E-19$, **Fig. 1E**), presumably promoting binding while the C-terminal region is still being synthesized.

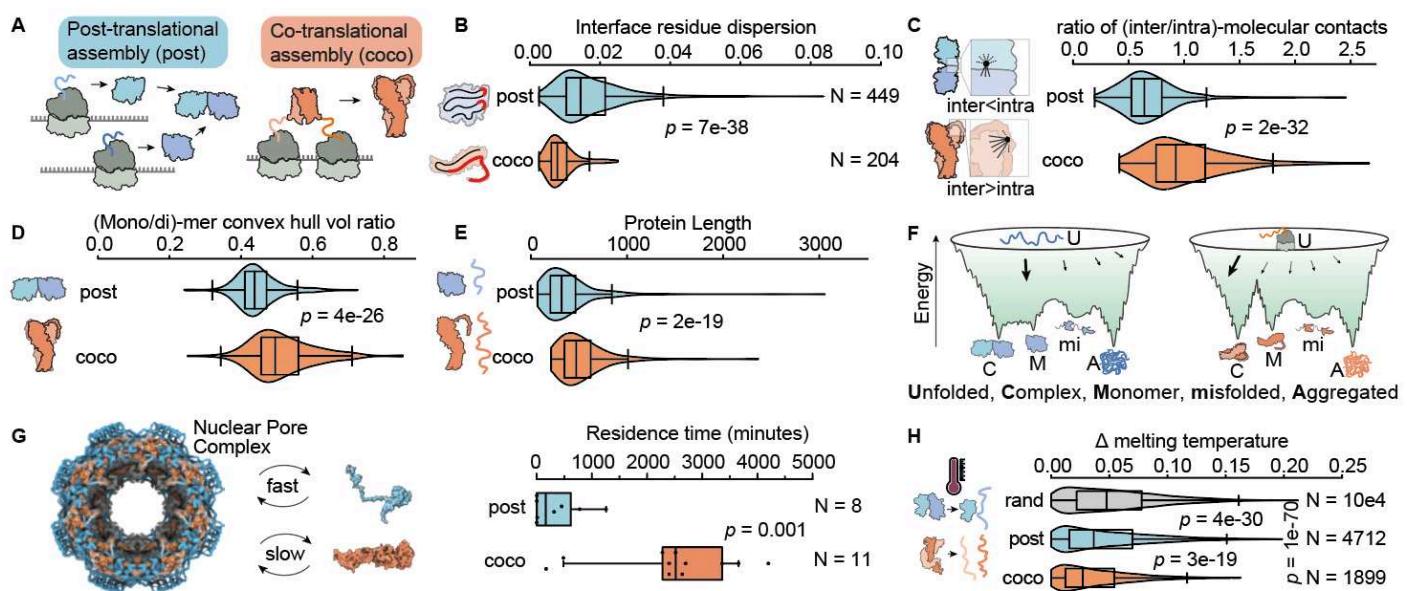


Fig. 1. Subunit interaction geometry underlies co- or post-translational modes of assembly. (A) Subunits are classically assumed to be translated and to fold independently before they interact (left, cyan). By contrast, in co-translational assembly (right, orange), subunits start interacting during their translation. (B) Violin plots comparing the dispersion of interface residues in the sequence among co- (orange) and post-translational (cyan) homomers of known structure. The boxes and whiskers indicate the 25th-75th and 5th-95th percentiles of the distributions. The central line shows the median. Differences between distributions are assessed by the Mann-Whitney U test. (C) Same as (B) for the ratios of inter- to intra-molecular atomic contacts of interface residues. (D) Same as (B) for the ratio of convex hull volume enclosing the monomer relative to that enclosing the dimer. (E) Same as (B) for the protein length. (F) Hypothetical energy landscapes associated with post- and co-translational assembly. The intertwined structure of coco-subunits implies that extensive conformational rearrangements are required for free monomers to bind to each other. Such rearrangements may be associated with a high activation energy, which co-translational assembly could circumvent by shunting the unbound, monomeric state. (G) We analyzed human nuclear pore complex (PDB code 7R5K³⁴) subunit exchange kinetics measured *in vivo*³⁵. We assigned subunits to the post (cyan) or coco (orange) category based on disome enrichment data. The boxplot shows residence time distributions for coco and post subunits. The coco subunits show slower exchange kinetics, in line with a high energy barrier separating their unbound and bound states. (H) Same as (B) for melting temperature deviations of random protein pairs (gray), post- (cyan), and coco-pairs (orange) from heterodimers. The procedure to assign heterodimers as post- or coco pairs is described in Methods 3.

These geometrical features, along with others shown in *Supplementary Note 1*, reflect a proto-typical coco-subunit as binding its partner through a local region in the sequence. Moreover, this binding region typically extends into the partner's structure, intertwines with it, and is thereby partner-stabilized. Such geometrical features imply that conformational rearrangements must occur upon binding, and we expect such rearrangements to create a high energy barrier that co-translational folding-binding helps overcome (**Fig. 1F**)^{33,36,37}. In that respect, we predict that complexes assembled co-translationally are more kinetically stable than those assembled post-translationally.

We investigated whether geometrical and kinetic observations were reproduced among heteromeric complexes. To this aim, we projected the disome enrichment data onto complexes of known structure (*Methods 2*). The differences in structural features were recapitulated among hetero-oligomers as well (*Supplementary Note 1*). To assess differences in kinetic stability, we analyzed the residence times of

subunits in the human nuclear pore complex, as previously measured *in vivo* by fluorescent labeling³⁵. Consistent with expectations, residence times of coco subunits were about five-fold longer than post subunits (**Fig. 1G**, MWU $p = 0.001$). Additionally, we hypothesized that coco protein pairs should melt at similar temperatures upon thermal denaturation. We analyzed a comprehensive dataset of T_m values measured for human proteins³⁸, revealing that coco subunits exhibit similar T_m values (low ΔT_m) when compared to post-subunits (coco < post < random, MWU $p < 3E-19$, **Fig. 1H**, also see *Supplementary Note 3*).

Taken together, these observations highlight that specific structural features introducing partner-dependent stability characterize coco subunits.

Proteostasis of coco subunit pairs is synchronized across the central dogma

The structural interdependence of coco subunit pairs suggests that their genes should be expressed together and that their mRNAs should be translated synchronously to yield 1:1 stoichiometry at the protein level. To assess such synchrony, we compared the similarity of expression and regulation between heteromeric coco and post pairs, across the central dogma (**Fig. 2A**). To this aim, we projected the disome enrichment data onto a dataset of human heterodimers integrated from multiple repositories (*Methods 3-4*), resulting in 19,855 pairs of which 4,016 and 15,839 were assigned as coco and post pairs respectively.

We initially analyzed transcription rates measured in human cells³⁹, revealing that coco pairs show more similar rates than post pairs (Δ transcription rate of coco < post < random, MWU $p < 5E-31$, **Fig. 2B**). Consistent with this observation, mRNA abundance levels from 273 RNA-seq experiments covering 53 human tissues⁴⁰ showed higher correlations for coco pairs than for post pairs (correlation of coco > post > random, MWU $p < 3E-28$, **Fig. 2C**).

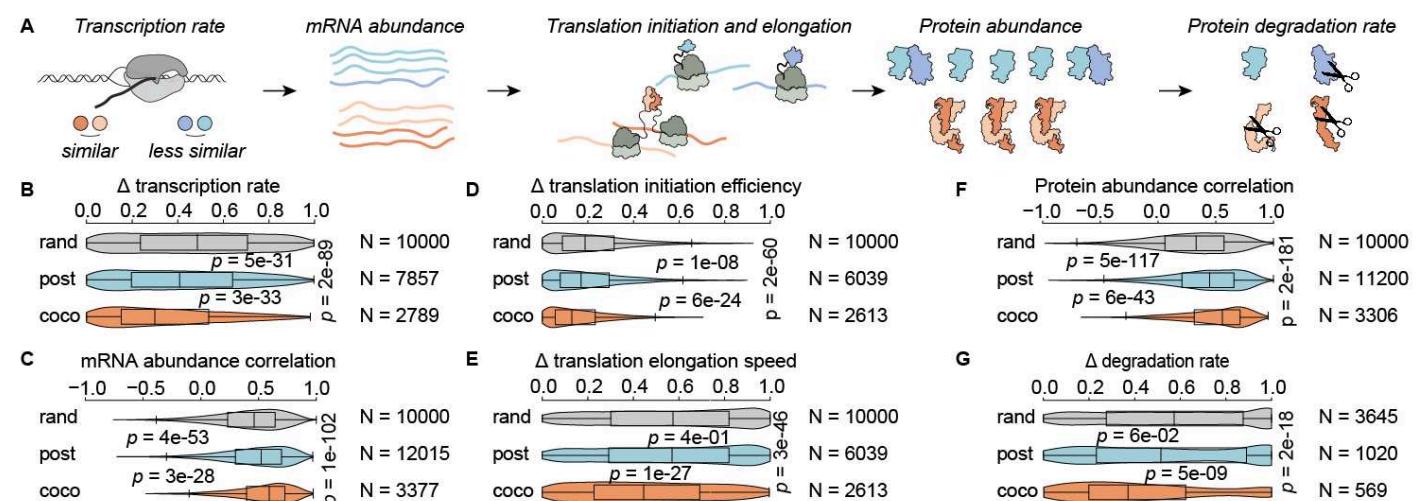


Fig. 2. Co-translational heteromeric pairs undergo synchronized regulation across the central dogma. (A) Schematic representation of the central dogma. (B) Violin plots comparing transcription rate deviations of co- (orange) and post-translational (cyan) heteromeric pairs. Random pairs are shown for reference (gray). Figure features follow Fig. 1B. (C) Similarity in mRNA abundance of coco or post subunit pairs measured as the Spearman correlation across 273 RNA-seq datasets from Bgee⁴⁰. (D) Same as (B) for mRNA translational initiation efficiency. (E) Same as (B) for mRNA translation elongation speed. (F) Same as (C) for protein abundance compiled from 156 proteomic datasets from PaxDB⁴¹. (G) Same as (B) for protein degradation rates.

At the translational level, we analyzed ribosome profiling experiments ⁴². Translation appeared more synchronized in coco pairs than in post, with the former exhibiting more similar initiation (Δ translation initiation efficiency of coco < post < random, MWU $p < 6E-24$, **Fig. 2D**) and elongation rates (Δ elongation rate of coco < post < random, MWU $p < 1E-27$, **Fig. 2E**). Furthermore, the variation of protein abundance measured across 153 datasets ⁴¹ showed higher correlation among coco subunits than among post, reflecting more consistent stoichiometries in the former type (correlation of coco > post > random, MWU $p < 6E-43$, **Fig. 2E**).

Beyond translation, we expected the partner-dependence for stability among coco pairs to be reflected in synchronized degradation of these subunits. Indeed, a dataset of degradation rates characterized for human proteins ⁴³, revealed more similar rates for coco pairs than for post (Δ degradation rate of coco < post < random, MWU $p < 5E-9$, **Fig. 2F**).

Taken together, these results depict how subunit interaction geometry underlies the mode of complex assembly, which in turn shapes the global regulatory landscape of the respective protein subunits.

Predicting co-translational assembly on a proteome-wide scale from structure

The marked geometric differences in the structures of coco and post subunits, prompted us to predict the mode of assembly of a complex solely based on its structure. We initially focused on homo-oligomers as these alleviate the need to infer coco pairs from the experimental disome enrichment data. Since Badonyi and Marsh ³² previously used Bertolini et al.'s ²⁸ coco/post definition to associate large interface areas with co-translational assembly, we asked how well a prediction model performs on their existing data, and whether our analyses would make any significant improvements.

We fitted a logistic regression model using Interface area for Badonyi and Marsh's dataset of 1351 homomeric structures ³², and assessed the prediction accuracy using a receiver operating characteristic (ROC) curve. Interface area recapitulated the experimental characterization of coco/post homomers with an area under the curve (AUC) of 0.629 (brown, **Fig. 3A**). Similar results were obtained when our structure dataset and coco/post definitions were used instead (green, AUC=0.751). Other orthogonal parameters such as ratio of inter/intra-molecular contacts captured the experimental data with a relatively higher accuracy (black, AUC=0.795), but no single parameter achieved highly accurate predictions (*Supplementary Note 2*). Remarkably, a dramatic improvement was obtained when eight structural features (each summarized as a single number) were integrated to capture the mode of assembly (AUC=0.940, Average Precision, AP=0.901, *Methods 5*). These highly accurate predictions motivated us to scale them up to entire proteomes.

Recent advances in machine learning have increased the coverage of protein tertiary structures ⁴⁴ and protein complexes ⁴⁵⁻⁴⁸ to entire proteomes. We integrated an atlas of 2012 homodimers (**Fig. 3A, Methods 5**), of which 672 and 1376 were assigned as coco and post respectively based on the experimental data. We fitted a logistic regression model integrating eight topological features, which together captured the mode of assembly accurately (AUC = 0.951; **Fig. 3B**; *Supplementary Note 2*). To assess the generality of this model, we retrieved disome-enrichment data for *E. coli* ²⁸, which served as a test set. Interestingly, these data were indirectly identified as part of an experiment analyzing disomes formed by a human protein heterologously expressed in *E. coli*, and were not analyzed before to describe complexes from *E. coli*. We mapped these disome-enrichment data to homo-oligomeric structures of *E. coli* and inferred a coco probability for each complex, based on the model trained on human proteins. These predictions

recapitulated well the experimental disome enrichment (AUC = 0.911; **Fig. 3C**), validating our model and its underlying assumption that co-translational assembly is primarily determined by the structure of a complex.

A remarkable evolutionary implication of our findings is that co-translational assembly should be conserved in evolution because protein structure is conserved itself. We provide such an example with two orthologous enzymes involved in purine biosynthesis: *purB* (*E. coli*) and *ADSL* (human). Their translation shifts from monosomes to disomes near the 200th codon, coinciding with the emergence of interface residues at the ribosome tunnel (**Fig. 3E**).

Accurately predicting homodimers' assembly mode from structure motivated us to apply the same strategy to heterodimers. While disome enrichment experiments do not provide specific interaction partners, we mapped putative coco and post partners in a two-step process (*Methods 5*). Briefly, we first selected interacting protein pairs of known structure with both subunits identified as either coco or post. We fitted a first predictive model based on this set using the same topological features, and kept only the highest-scoring complex for each protein. The rationale for this selection is that a protein may engage in multiple post-interactions (low scores), and we assumed there should be fewer (typically a single) co-translational interactions. The resulting set consisted of 679 and 2113 heterodimer structures assigned as coco and post respectively. Refitting our structural model onto this filtered set recapitulated the experimental data well (AUC = 0.886, **Fig. 3F**). Additionally, this model also generalized to heterodimers from *E. coli*, recapitulating these independent experimental data with an AUC value of 0.841 (*Supplementary Note 2*).

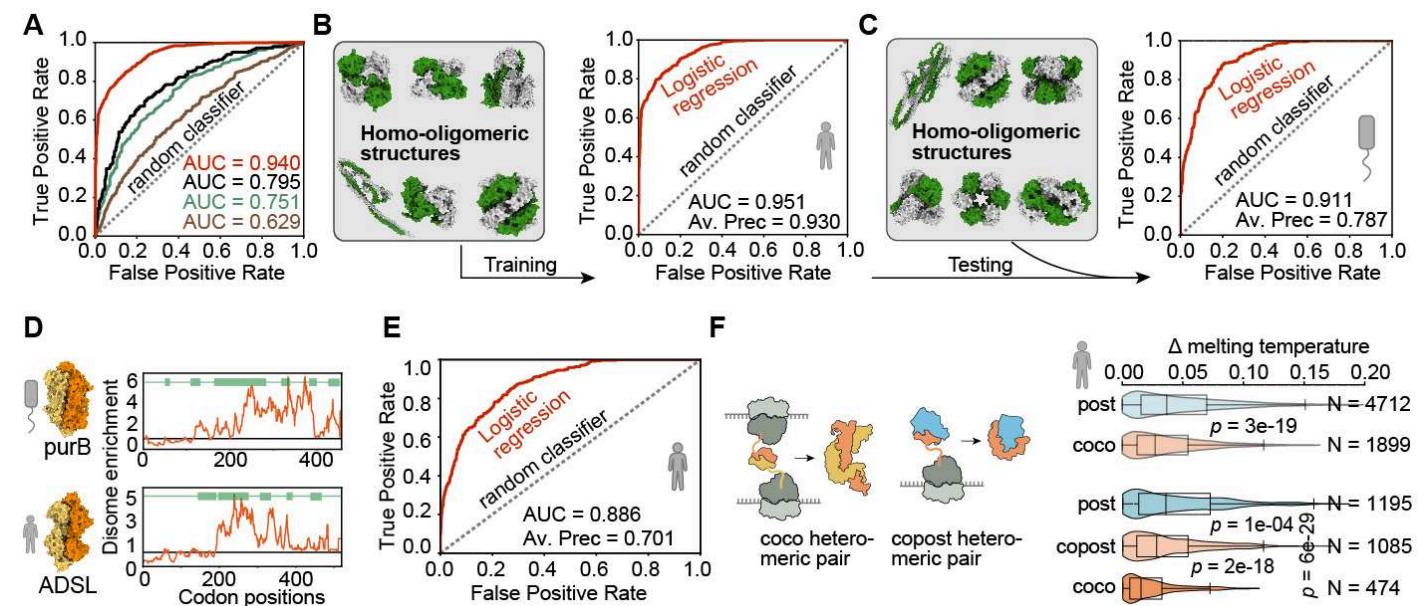


Fig. 3. Protein structure captures proteome-wide information on co- and post-translational assembly.

(A) The interface area is larger among coco complexes compared to post, but this criteria alone showed limited predictive power in a previous work (AUC=0.629, brown, (30)) as well as in our dataset (AUC=0.751, green). Other orthogonal structural features (e.g., ratio of inter/intra-molecular contacts) can capture the experimental data with higher accuracy (AUC=0.795, black), but no single parameter achieves high-accuracy predictions (*Supplementary Note 2*). Remarkably a logistic regression model utilizing eight structural features, each summarized as a single number, achieves high accuracy predictions (AUC=0.940, red). (B) Extending the dataset of 653 experimental structures used in panel A with 1395 homo-oligomeric structures predicted with AlphaFold2 ⁴⁴ (*Methods 2.1* and *2.2*) improves both the coverage and accuracy of our predictions using the same eight-parameter model (C) Same as (B), for the bacterium *Escherichia coli*. Prediction results of the model trained on human structures, now predicting the assembly mode of *E. coli* homomers. (D) The conservation of structure and co-translational assembly is depicted for

two orthologous enzymes: *E. coli* purB and human ADSL (PDB codes: 2PTQ⁴⁹ and 4FFX⁵⁰). We show the disome footprint enrichment relative to monosomes (red), smoothed over a 15-codon sliding window. Values above the line ($y=1$) represent a disome enrichment and thereby co-translational assembly. Interface residues are highlighted with green rectangles. **(E)** Same as (B), the ROC curve for predicting the co- and post-translational modes of heteromer assembly (*Methods 5*). **(F)** Co-translational heteromer assembly can be bidirectional (coco) or unidirectional (copost). Transparent violins replicate **Fig. 1H** and are compared to predictions after incorporating the uni- or bi-directionality of co-translational assembly based on structure.

Remarkably, because some of the structural features are directional, this model provides two distinct probabilities per complex, *i.e.*, one for each subunit. We applied this model to 7299 protein pairs of known structure, which we could now classify as post (3945 pairs, low score for both partners), coco (1246 pairs, high score for both partners), but also co-post (2108 pairs, only one partner with a high score). The latter type points to proteins that bind co-translationally to fully-translated partners.

As an additional validation of the model's predictive value, we examined whether it reproduced observations based on experimental data, specifically where coco pairs exhibited similar melting temperatures or co-regulation across the central dogma. Notably, the similarity in melting temperatures between proteins in the coco pairs predicted based on structure was even greater than that observed in coco pairs defined based on disome-enrichment data (**Fig. 3G**). Consistent with this observation, coco pairs predicted based on structure also showed greater synchrony across the central dogma than coco pairs defined based on experimental data (*Supplementary Note 4*).

These results show that co-translational assembly can be accurately predicted from structure, and imply that coco subunits and pairs can be inferred from any proteome for which structural information is available. This idea motivated us to gain a comprehensive overview of co-translational assembly across the *H. sapiens*, *E. coli*, and *S. cerevisiae* proteomes, although the latter had no systematic experimental information on co-translational assembly.

A comprehensive view of co-translational assembly from structure

We integrated structure-based predictions with experimental disome-enrichment data to gain an overview of co-translational assembly across the human proteome. In this integration, structure-based predictions added 561 proteins annotated as coco, yielding a total of 3073. Thus, about 17% of the proteome was estimated to assemble co-translationally, about 53% was annotated as post, and the remaining ~31% could not be annotated due to ambiguous or absent data (**Fig. 4A**). Focusing our analysis on 1616 heteromeric complexes of known structure, we predicted that 23%, 14%, and 63% of subunits assemble in coco, copost, and post manners, respectively. We reached similar conclusions for *E. coli*, where the proteome fraction of coco subunits was estimated at ~20% (**Fig. 4A**).

Although no disome enrichment information is available for the proteome of *S. cerevisiae*, we integrated structural information from the Protein Data Bank³⁰ as well as AlphaFold2-based structures for homo-⁴⁵ and hetero-oligomers⁴⁶. These structural data enabled us to derive proteome-wide predictions (**Fig. 4A**) that we subsequently evaluated with targeted experiments.

We first evaluated the predictions using selective ribosome profiling (SeRP)²⁴. We compared ribosome footprints from the total translatome to those co-purified with a C-terminally tagged protein (*Methods 6*). Thus, co-purified footprints point to a co-translationally interacting partner. We employed SeRP to analyze the Nat1/5 heterodimer (NatE complex) predicted as copost with translating Nat1 assembling onto full-length Nat5. Upon isolating tagged Nat5 we indeed co-purified ribosomes translating Nat1 mRNA. The footprint

density of Nat1 mRNA increased onwards of codon 480, coinciding with the ribosomal exit and exposure of its binding interface (**Fig. 4B**). Next we examined the Acc1 (Acetyl-CoA carboxylase) homodimer, also predicted to assemble co-translationally. SeRP uncovered that C-terminally tagged full-length Acc1 similarly co-purifies with ribosomes translating its own mRNA, and the increased footprint density onwards of codon 250 coincides with the ribosomal exit of its interface (**Fig. 4B**, also see *Supplementary Note 5*).

We next utilized RNA immunoprecipitation qPCR (RIP-qPCR)⁵¹ whereby translating mRNAs physically associated with the C-terminally tagged partner are quantified by qPCR after pull-down (*Methods 7*). Although RIP-qPCR primarily detects copost assembly, it could in principle also capture coco events when the tagged subunit is released by the ribosome before its partner (e.g., due to C-terminus proximity of the interface). We first examined a predicted coco pair Rpb1/2, forming the core of the RNA polymerase-II complex. We pulled down on tagged Rpb1, in conditions promoting ribosome stability (cycloheximide) or dissociation (EDTA), and quantified Rpb2 mRNA levels. We observed an enrichment in Rpb2 mRNA that was cycloheximide-dependent and also tag-dependent, indicating that Rpb2 binds Rpb1 co-translationally. These observations were mirrored when pulling down on tagged Rpb2, suggesting that this pair assembles co-translationally (**Fig. 4C**).

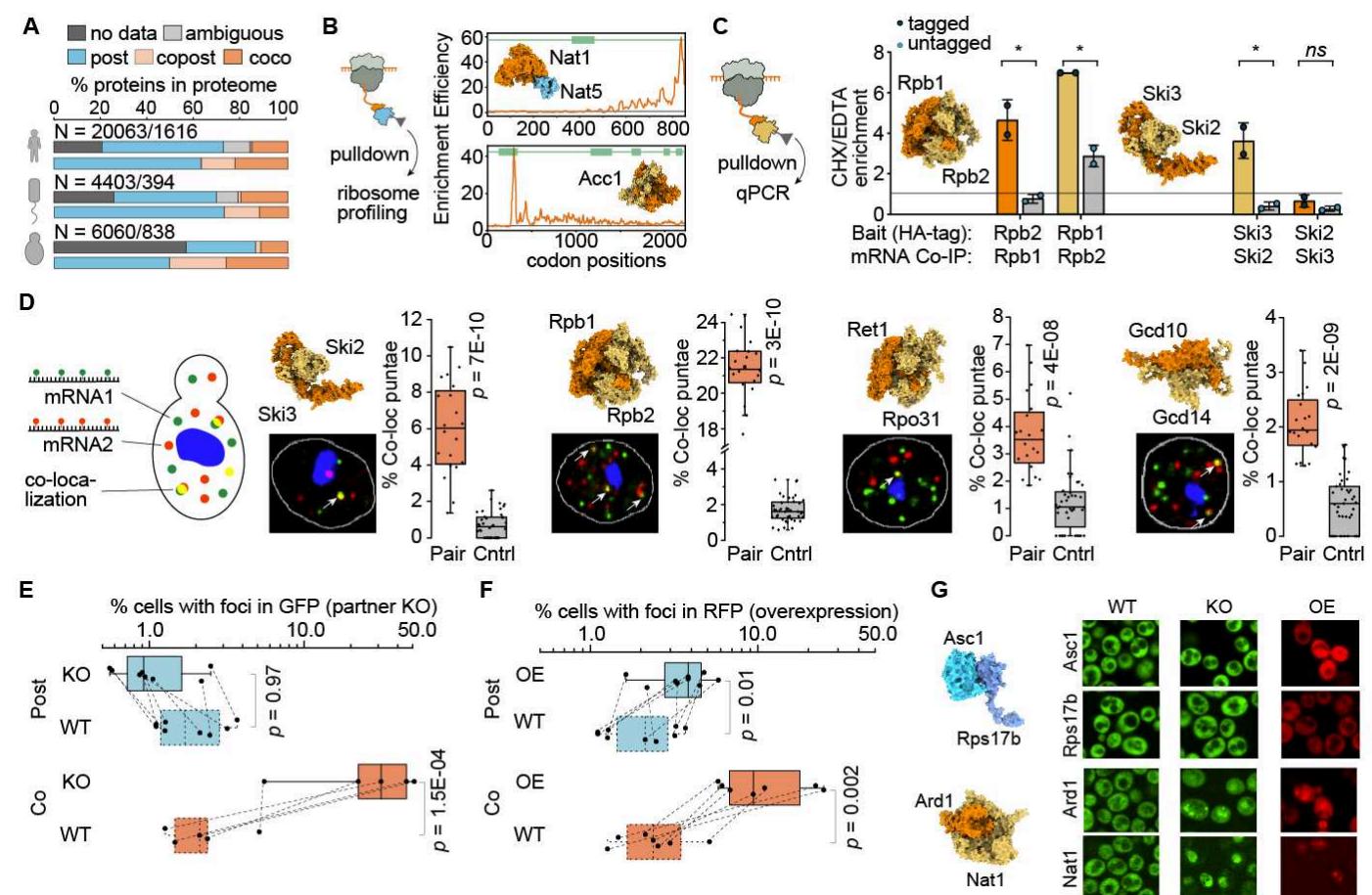


Fig. 4. De novo prediction of co-translational assembly across the yeast proteome. **(A)** Global statistics on co/post-translational assembly for the proteomes of human, *E. coli*, and *S. cerevisiae*. For each organism, the top and bottom bars show statistics for the reference proteome and for heteromeric complexes of known structure. **(B)** Quantifying co-translational assembly of Nat1/5 heteromer (top, PDB code: 6O07⁵²) and Acc1 homodimer (bottom, PDB code: 5CSL⁵³) by SeRP. The average ribosome footprint enrichment of two biological replicates is shown as an orange line. Interface residue positions are highlighted by green boxes. **(C)** Quantifying co-translational assembly by RIP-qPCR for Rpb1/2 (PDB code: 1Y1V,⁵⁴) and Ski2/3 (PDB code: 4BUJ⁵⁵). Barplots depict mRNA enrichment after pull-down following cycloheximide (CHX) vs. EDTA treatment. We show the enrichment after pull down of a HA-tagged (orange/yellow bars) or untagged (negative control, gray bars) protein partner. A horizontal line (y = 1) represents no enrichment. **(D)** Schematic of co-localization and puncta analysis. **(E)** Scatter plot of % cells with foci in GFP (partner KO) for Post (WT vs KO) and Co (KO vs WT) conditions. **(F)** Scatter plot of % cells with foci in RFP (overexpression) for Post (OE vs WT) and Co (OE vs WT) conditions. **(G)** Microscopy images for Asc1, Rps17b, Ard1, and Nat1 in WT, KO, and OE conditions.

enrichment; error bars show the standard deviation of two biological replicates. Statistical significance was assessed by two-sample t-tests: ns $p > 0.05$, * $p < 0.05$. (D) RNA-smFISH experiments in yeast cells reveal specific co-localization of mRNAs for four predicted coco pairs: Ski2/3, Rpb1/2, Ret1/Rpo31 (PDB code: 5M5W⁵⁶), and Gcd10/14 (ModelArchive code: ma-bak-cepc-0793⁴⁶). Fluorescent tags: blue, DAPI (nucleus); red, CAL Fluor Red 590; green, Quasar 670. Co-localizing spots are yellow. The box plots depict the distribution of percent co-localized punctae per field of view; negative controls consist of two mRNAs encoding unrelated proteins (*Methods 8, Supplementary Note 7*). Statistical significance was assessed by the MWU test. (E) We evaluated the tendency of coco (orange) or post (blue) subunits to become insoluble and form punctae upon partner knockout or (F) over-expression. In both panels, paired boxplots show the percentage of cells containing one or more punctae upon deletion of a partner (KO) or upon over-expressing the protein (OE); p -values are calculated using the MWU test. (G) Example structures of two candidate pairs (PDB codes: 6HD5⁵⁷, 7A1G⁵⁸) and corresponding images in KO or OE experiments.

We subsequently examined the Ski2/3 dimer of the RNA-helicase Ski complex, also predicted as coco. RIP-qPCR experiments indicated that translating Ski2 assembled with full-length Ski3, but the reverse interaction was not observed (**Fig. 4C**). This discrepancy may originate in the interface of Ski3 being closer to the C-terminus than Ski2 (*Supplementary Note 6*). Therefore, the release of the Ski2 C-terminus and its pull-down are unlikely to occur while Ski3 is being translated. A similar asymmetry was observed for the Rrn6/7 pair of the RNA polymerase-I core factor complex (*Supplementary Note 6*).

We reasoned that coco assembly is expected to yield co-localized mRNAs, motivating us to characterize the Ski2/3 complex further by single-molecule RNA-fluorescence *in situ* hybridization (smFISH)⁵⁹. Such mRNA co-localization has rarely been observed to our knowledge^{51,60}, perhaps due to the difficulty in reliably identifying coco subunit pairs. We designed FISH probes conjugated to two different colors, each specific to the mRNA encoding the subunit of a predicted coco pair. After hybridization, we visualized the subcellular localization of each mRNA by fluorescence microscopy and quantified the co-localization of each pair (*Methods 8, Supplementary Note 7*). The mRNAs encoding Ski2 and Ski3 showed significant co-localization when compared to two unrelated genes (MWU $p = 7E-10$, **Fig. 4D**), consistent with their predicted bidirectional coco assembly. We observed similar results for Rpb1/2, Ret1/Rpo31 (RNA polymerase-III complex), and Gcd10/14 (tRNA methyltransferase complex). Interestingly, the four pairs showed various extent of co-localization, from 2% to 21% (**Fig. 4D**), which may be explained by several factors that include the fraction of actively translated mRNAs, mRNA hybridization and detection efficiency, and potential alternative assembly pathways not involving co-translational assembly. Remarkably, the Gcd10/14 pair was predicted based on an AlphaFold2-predicted structure⁴⁶.

The structural and omics analyses from Figs. 1 and 2 show that coco subunits show partner-dependent stability and solubility, while post-subunits appear more independent. We examined such partner-dependence among yeast complexes predicted as coco and post. We tagged both types of subunits with a fluorescent protein and assessed their intracellular solubility based on the formation of puncta by fluorescence microscopy (*Methods 9*). We explored two conditions that interfere with 1:1 stoichiometry, namely partner knockout or overexpression. We found that coco subunits were more prone to form puncta than post subunits, both when their partner was knocked out, or when they were over-expressed themselves (**Fig. 4E-G**). Thus, both types of experiments suggest that dosage balance and stoichiometric expression of coco subunits are critical for proper cellular functions, an observation consistent with the synchrony seen across the proteostasis steps (**Fig. 2**).

Finally, we analyzed a shotgun proteomics dataset of thousands of knock-out strains⁶¹, which supported these observations further: while the abundance of proteins predicted to assemble post-translationally did not change significantly in the absence of their partner, the abundance of coco subunits dropped

significantly. Remarkably, for copost pairs, the drop in protein abundance was consistent with the predicted direction of assembly (*Supplementary Note 8*).

DISCUSSION AND CONCLUSIONS

We elucidated structural features facilitating co-translational assembly, including clustered interface residues, partner-stabilized intertwined interfaces, and prolonged translation, as reflected in longer protein sequence lengths. Additionally, coco subunits appeared structurally unstable as monomers and indeed, showed reduced solubility upon changes in partner stoichiometry. Accordingly, we saw that regulatory mechanisms ensure balanced stoichiometries in co-translational pairs, including synchronized transcription, translation, and degradation. In contrast, proteins that assemble post-translationally were typically stable, as indicated by their globular structures and their maintained solubility upon changes in partner stoichiometry.

Interestingly, the connection between co-translational assembly and structure suggests that two protein partners can start binding during synthesis if their emerging chains are in each other's vicinity. When considering homo-oligomers, this property is naturally met by polysomes. In this context, our results corroborate that biasing interface residues towards the C-terminus can ensure minimal interference between folding and binding⁶². Considering heteromers, mRNA co-localization is a prerequisite for co-translational assembly. Such co-localization could be achieved actively through zipcode sequences⁶³ or through partitioning into specific condensates⁶⁴. Alternatively, random diffusion could suffice to promote nascent chain recognition⁶⁵. Remarkably, the coco partners we predicted from structure showed marked mRNA co-localization as seen by smFISH, and thereby constitute a rich dataset to explore these alternative hypotheses.

Equally important, the conservation of structure over evolutionary timescales implies co-translational assembly to be conserved as well. In that respect, our predictive model and proteome-wide datasets offer solid grounds on which to base future comparative analyses of protein assembly.

We unveiled a fundamental connection between protein structure and the translation process, emphasizing its impact on gene expression, proteostasis, and mRNA localization. The resulting comprehensive description of protein co-translational assembly will serve as a valuable framework for future studies in protein interaction, regulation, and evolutionary biology, promoting a holistic understanding of proteome organization and functionality.

REFERENCES

1. Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1309–1314.
2. Sali, A., Shakhnovich, E., and Karplus, M. (1994). How does a protein fold? *Nature* **369**, 248–251.
3. Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.
4. Plaxco, K.W., Simons, K.T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
5. Fersht, A.R. (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1525–1529.
6. Schuler, B., Lipman, E.A., and Eaton, W.A. (2002). Probing the free-energy surface for protein folding with

- single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747.
7. Robinson, C.V., Sali, A., and Baumeister, W. (2007). The molecular sociology of the cell. *Nature* **450**, 973–982.
 8. Marsh, J.A., and Teichmann, S.A. (2015). Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84**, 551–575.
 9. Jones, S., and Thornton, J.M. (1996). Principles of protein–protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13–20.
 10. Braselmann, E., Chaney, J.L., and Clark, P.L. (2013). Folding the proteome. *Trends Biochem. Sci.* **38**, 337–344.
 11. Busso, D., Peleg, Y., Heidebrecht, T., Romier, C., Jacobovitch, Y., Dantes, A., Salim, L., Troesch, E., Schuetz, A., Heinemann, U., et al. (2011). Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: a benchmarking study. *J. Struct. Biol.* **175**, 159–170.
 12. Vonshak, O., Divon, Y., Förste, S., Garenne, D., Noireaux, V., Lipowsky, R., Rudorf, S., Daube, S.S., and Bar-Ziv, R.H. (2020). Programming multi-protein assembly by gene-brush patterns and two-dimensional compartment geometry. *Nat. Nanotechnol.* **15**, 783–791.
 13. To, P., Whitehead, B., Tarbox, H.E., and Fried, S.D. (2021). Nonrefoldability is Pervasive Across the Proteome. *J. Am. Chem. Soc.* **143**, 11435–11448.
 14. Bryngelson, J.D., Onuchic, J.N., Soccia, N.D., and Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195.
 15. Levy, E.D., Boeri Erba, E., Robinson, C.V., and Teichmann, S.A. (2008). Assembly reflects evolution of protein complexes. *Nature* **453**, 1262–1265.
 16. Marsh, J.A., Hernández, H., Hall, Z., Ahnert, S.E., Perica, T., Robinson, C.V., and Teichmann, S.A. (2013). Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461–470.
 17. Perica, T., Kondo, Y., Tiwari, S.P., McLaughlin, S.H., Kemplen, K.R., Zhang, X., Steward, A., Reuter, N., Clarke, J., and Teichmann, S.A. (2014). Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science* **346**, 1254346.
 18. Marsh, J.A., Rees, H.A., Ahnert, S.E., and Teichmann, S.A. (2015). Structural and evolutionary versatility in protein complexes with uneven stoichiometry. *Nat. Commun.* **6**, 6394.
 19. Ahnert, S.E., Marsh, J.A., Hernández, H., Robinson, C.V., and Teichmann, S.A. (2015). Principles of assembly reveal a periodic table of protein complexes. *Science* **350**, aaa2245.
 20. Schulz, L., Guo, Z., Zarzycki, J., Steinchen, W., Schuller, J.M., Heimerl, T., Prinz, S., Mueller-Cajar, O., Erb, T.J., and Hochberg, G.K.A. (2022). Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155–160.
 21. Schwarz, A., and Beck, M. (2019). The Benefits of Cotranslational Assembly: A Structural Perspective. *Trends Cell Biol.* **29**, 791–803.
 22. Morales-Polanco, F., Lee, J.H., Barbosa, N.M., and Frydman, J. (2022). Cotranslational Mechanisms of Protein Biogenesis and Complex Assembly in Eukaryotes. *Annu Rev Biomed Data Sci* **5**, 67–94.
 23. Duncan, C.D.S., and Mata, J. (2011). Widespread cotranslational formation of protein complexes. *PLoS Genet.* **7**, e1002398.
 24. Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., and Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* **561**, 268–272.
 25. Lautier, O., Penzo, A., Rouvière, J.O., Chevreux, G., Collet, L., Loïodice, I., Taddei, A., Devaux, F., Collart, M.A., and Palancade, B. (2021). Co-translational assembly and localized translation of nucleoporins in nuclear pore

- complex biogenesis. *Molecular Cell* 81, 2417–2427.e5. 10.1016/j.molcel.2021.03.030.
- 26. Seidel, M., Becker, A., Pereira, F., Landry, J.J.M., de Azevedo, N.T.D., Fusco, C.M., Kaindl, E., Romanov, N., Baumbach, J., Langer, J.D., et al. (2022). Co-translational assembly orchestrates competing biogenesis pathways. *Nat. Commun.* 13, 1224.
 - 27. Bernardini, A., Mukherjee, P., Scheer, E., Kamenova, I., Antonova, S., Mendoza Sanchez, P.K., Yayli, G., Morlet, B., Timmers, H.T.M., and Tora, L. (2023). Hierarchical TAF1-dependent co-translational assembly of the basal transcription factor TFIID. *Nat. Struct. Mol. Biol.* 30, 1141–1152.
 - 28. Bertolini, M., Fenzl, K., Kats, I., Wruck, F., Tippmann, F., Schmitt, J., Auburger, J.J., Tans, S., Bukau, B., and Kramer, G. (2021). Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science* 371, 57–64.
 - 29. Velankar, S., van Ginkel, G., Alhroub, Y., Battle, G.M., Berrisford, J.M., Conroy, M.J., Dana, J.M., Gore, S.P., Gutmanas, A., Haslam, P., et al. (2016). PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.* 44, D385–D395.
 - 30. Rose, P.W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Costanzo, L.D., Duarte, J.M., Dutta, S., Feng, Z., et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45, D271–D281.
 - 31. Levy, Y., Wolynes, P.G., and Onuchic, J.N. (2004). Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 101, 511–516.
 - 32. Badonyi, M., and Marsh, J.A. (2022). Large protein complex interfaces have evolved to promote cotranslational assembly. *Elife* 11. 10.7554/elife.79602.
 - 33. Bennett, M.J., Choe, S., and Eisenberg, D. (1994). Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. U. S. A.* 91, 3127–3131.
 - 34. Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Taniguchi, R., Turoňová, B., Zimmerli, C.E., Buczak, K., Schmidt, F.H., Margiotta, E., Mackmull, M.-T., et al. (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* 376, eabm9506.
 - 35. Rabut, G., Doye, V., and Ellenberg, J. (2004). Mapping the dynamic organization of the nuclear pore complex inside single living cells. *Nat. Cell Biol.* 6, 1114–1121.
 - 36. Rousseau, F., Schymkowitz, J.W.H., and Itzhaki, L.S. (2003). The unfolding story of three-dimensional domain swapping. *Structure* 11, 243–251.
 - 37. Yang, S., Cho, S.S., Levy, Y., Cheung, M.S., Levine, H., Wolynes, P.G., and Onuchic, J.N. (2004). Domain swapping is a consequence of minimal frustration. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13786–13791.
 - 38. Jarzab, A., Kurzawa, N., Hopf, T., Moerch, M., Zecha, J., Leijten, N., Bian, Y., Musiol, E., Maschberger, M., Stoehr, G., et al. (2020). Meltome atlas—thermal proteome stability across the tree of life. *Nature Methods* 17, 495–503. 10.1038/s41592-020-0801-4.
 - 39. Haussler, J., Mayo, A., Keren, L., and Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. *Nat. Commun.* 10, 68.
 - 40. Bastian, F.B., Roux, J., Niknejad, A., Comte, A., Fonseca Costa, S.S., de Farias, T.M., Moretti, S., Parmentier, G., de Laval, V.R., Rosikiewicz, M., et al. (2021). The Bgee suite: integrated curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Res.* 49, D831–D847.
 - 41. Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* 15, 3163–3168.
 - 42. Lian, X., Guo, J., Gu, W., Cui, Y., Zhong, J., Jin, J., He, Q.-Y., Wang, T., and Zhang, G. (2016). Genome-Wide and Experimental Resolution of Relative Translation Elongation Speed at Individual Gene Level in Human Cells. *PLoS Genet.* 12, e1005901.

43. Kristensen, A.R., Gsponer, J., and Foster, L.J. (2013). Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol. Syst. Biol.* **9**, 689.
44. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
45. Schweke, H., Levin, T., Pacesa, M., Goverde, C.A., Kumar, P., Duhoo, Y., Dornfeld, L.J., Dubreuil, B., Georgeon, S., Ovchinnikov, S., et al. (2023). An atlas of protein homo-oligomerization across domains of life. *bioRxiv*, 2023.06.09.544317. 10.1101/2023.06.09.544317.
46. Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., Zhang, J., Ness, T.J., Banjade, S., Bagde, S.R., et al. (2021). Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805.
47. Burke, D.F., Bryant, P., Barrio-Hernandez, I., Memon, D., Pozzati, G., Shenoy, A., Zhu, W., Dunham, A.S., Albanese, P., Keller, A., et al. (2023). Towards a structurally resolved human protein interaction network. *Nat. Struct. Mol. Biol.* **30**, 216–225.
48. Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265.
49. Tsai, M., Koo, J., Yip, P., Colman, R.F., Segall, M.L., and Howell, P.L. (2007). Substrate and product complexes of *Escherichia coli* adenylosuccinate lyase provide new insights into the enzymatic mechanism. *J. Mol. Biol.* **370**, 541–554.
50. Ray, S.P., Deaton, M.K., Capodagli, G.C., Calkins, L.A.F., Sawle, L., Ghosh, K., Patterson, D., and Pegan, S.D. (2012). Structural and biochemical characterization of human adenylosuccinate lyase (ADSL) and the R303C ADSL deficiency-associated mutation. *Biochemistry* **51**, 6701–6713.
51. Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., Garnier, J.-M., Dembele, D., Capponi, S., Timmers, H.T.M., et al. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. Commun.* **10**, 1740.
52. Deng, S., Magin, R.S., Wei, X., Pan, B., Petersson, E.J., and Marmorstein, R. (2019). Structure and Mechanism of Acetylation by the N-Terminal Dual Enzyme NatA/Naa50 Complex. *Structure* **27**, 1057–1070.e4.
53. Wei, J., and Tong, L. (2015). Crystal structure of the 500-kDa yeast acetyl-CoA carboxylase holoenzyme dimer. *Nature* **526**, 723–727.
54. Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965.
55. Halbach, F., Reichelt, P., Rode, M., and Conti, E. (2013). The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* **154**, 814–826.
56. Tafur, L., Sadian, Y., Hoffmann, N.A., Jakobi, A.J., Wetzel, R., Hagen, W.J.H., Sachse, C., and Müller, C.W. (2016). Molecular Structures of Transcribing RNA Polymerase I. *Mol. Cell* **64**, 1135–1143.
57. Knorr, A.G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B., and Beckmann, R. (2019). Ribosome-NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol.* **26**, 35–39.
58. Kratzat, H., Mackens-Kiani, T., Ameismeier, M., Potocnjak, M., Cheng, J., Dacheux, E., Namane, A., Berninghausen, O., Herzog, F., Fromont-Racine, M., et al. (2021). A structural inventory of native ribosomal ABCE1-43S pre-initiation complexes. *EMBO J.* **40**, e105179.
59. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879.
60. Panasenko, O.O., Somasekharan, S.P., Villanyi, Z., Zagatti, M., Bezrukov, F., Rashpa, R., Cornut, J., Iqbal, J., Longis, M., Carl, S.H., et al. (2019). Co-translational assembly of proteasome subunits in NOT1-containing

- assemblysomes. *Nat. Struct. Mol. Biol.* **26**, 110–120.
61. Messner, C.B., Demichev, V., Muenzner, J., Aulakh, S., Röhl, A., Herrera-Domínguez, L., Egger, A.-S., Kamrad, S., Lemke, O., Calvani, E., et al. The Proteomic Landscape of Genome-Wide Genetic Perturbations. 10.1101/2022.05.17.492318.
 62. Natan, E., Endoh, T., Haim-Vilmovsky, L., Flock, T., Chalancon, G., Hopper, J.T.S., Kintses, B., Horvath, P., Daruka, L., Fekete, G., et al. (2018). Cotranslational protein assembly imposes evolutionary constraints on homomeric proteins. *Nat. Struct. Mol. Biol.* **25**, 279–288.
 63. Mendonsa, S., von Kügelgen, N., Dantsuji, S., Ron, M., Breimann, L., Baranovskii, A., Lödige, I., Kirchner, M., Fischer, M., Zerna, N., et al. (2023). Massively parallel identification of mRNA localization elements in primary cortical neurons. *Nat. Neurosci.* **26**, 394–405.
 64. Allen, G., Weiss, B., Panasenko, O.O., Huch, S., Villanyi, Z., Albert, B., Dilg, D., Zagatti, M., Schaughency, P., Liao, S.E., et al. (2023). Not1 and Not4 inversely determine mRNA solubility that sets the dynamics of co-translational events. *Genome Biol.* **24**, 30.
 65. Heidenreich, M., Georgeson, J.M., Locatelli, E., Rovigatti, L., Nandi, S.K., Steinberg, A., Nadav, Y., Shimoni, E., Safran, S.A., Doye, J.P.K., et al. (2020). Designer protein assemblies with tunable phase diagrams in living cells. *Nat. Chem. Biol.* **16**, 939–945.

Acknowledgments. We thank Amnon Horovitz, Jeffrey Gerst, Yitzhak Pilpel, Daniel Dar, Sarah Teichmann, Judith Frydman, Bernd Bukau, Günter Kramer, Stephen Fried, Shimon Bershtain, Nir Fluman, Sudip Kundu, and all members of E.D.L and A.S.'s lab for helpful discussions; Jeffrey Gerst, Shahar Garin, and Gal Haimovich for advice and reagents on smFISH; Liat Alyagor and Roni Oren for advice on smFISH and help with testing different microscopes.

Funding. A.L. was supported by the Center for Integration in Science, the Ministry of Aliyah and Integration, the State of Israel. E.D.L. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 819318), by the Human Frontiers Science Program Organization (Ref. RGP0016/2022), by the Israel Science Foundation (grant no. 1452/18), and by the Abisch-Frenkel Foundation. A.S. acknowledges support from the European Research Council (ERC) Starting Grant (grant agreement No. 2031817), and the Israeli Science Foundation grant 2106/20.

Author contributions. S.M., A.S., and E.D.L. conceived and conceptualized the research; S.M. performed the structural and OMICs analysis shown in Figs. 1 and 2, and co-translational assembly predictions shown in Figs. 3 and 4A; J.V. performed the SeRP and RIP-qPCR experiments shown in Fig. 4B, 4C and Supplementary Note 5; A.L. performed the smFISH experiments shown in Fig. 4D and Supplementary Note 6; M.H. and H.G.S. performed the genetic perturbation experiments shown in Fig. 4E; A.S. supervised the SeRP and RIP-qPCR experiments; T.O.Y. conceived and guided S.M. with structural analyses on 'rigid body entanglement' and 'convex hull'; E.D.L. supervised the computational analyses, smFISH, and genetic perturbation experiments; S.M. prepared all the figures; E.D.L. edited the figures; S.M. and E.D.L. wrote the paper; A.S. and E.D.L. acquired funding; all authors discussed and commented on the results and edited versions of the paper.

Competing interests. The authors declare they have no competing interests.

Data and materials availability. All predictions and datasets are provided as Supplementary Material or are available on FigShare (<https://doi.org/10.6084/m9.figshare.24311917.v1>).

Code availability. Code and custom scripts used in this work are available from the authors upon request.