Spatiotemporal Regulation
of Nascent Protein Targeting

Thesis by

Zikun Zhu

In Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2024

(Defended April 23, 2024)

Zikun Zhu
ORCID: 0000-0001-5934-8368

ACKNOWLEDGEMENTS

My Ph.D. journey has been an adventure filled with exciting findings and formidable challenges, each teaching me valuable lessons through both successes and failures. I am deeply grateful for the tremendous support and help from my family, mentors, colleagues, and friends. Without their encouragement and guidance, I could not have reached this point.

First, I would like to thank my Ph.D. advisor, Shu-ou Shan, whose guidance and mentorship over the past five years have been invaluable. I am so fortunate to have had an advisor who was always available whenever I needed help. From her, I have learned to tackle problems with rigor and present findings effectively – important skills that I will carry forward in my academic career. She also provided me the freedom and resources to explore my own ideas, even when they diverged from the lab's main focus. This truly helped me in developing my ability to identify significant questions and grow up as an independent scientist.

I am also deeply thankful to my undergraduate advisor, Junjie Hu, who has long been my role model in science. Joining his lab introduced me to the field of molecular and cell biology, and significantly shaped my way of thinking about science. I also thank Xiangyang Guo, my mentor in the Hu lab. He helped me assimilate into the lab and taught me the principles of fundamental molecular biology techniques – knowledge that continues to benefit me today.

I extend my heartfelt thanks to my thesis committee: David Chan, Mitchell Guttman and Rebecca Voorhees for their invaluable advice, feedback and engaging discussions over the years. I want to especially thank David, with whom I had the privilege of working as a visiting student during my undergraduate study. David is not only an exceptional scientist, but also a remarkably kind and patient mentor. It has always been enjoyable to talk about science with him, and his insightful suggestions have consistently inspired me. I feel extremely fortunate to have had the opportunity to work with him, and I sincerely appreciate all his help both before and during my graduate studies at Caltech.

ABSTRACT

Proper protein targeting to the correct cellular compartments is essential for maintaining the functionality and organization of all cells. However, the mechanisms that ensure newly synthesized proteins are accurately and efficiently directed to their specific cellular destinations remain unclear. Moreover, how protein targeting is coordinated with protein folding and other cellular processes, both spatially and temporally, is largely unknown.

In my thesis, I first demonstrated the mechanism of a nascent protein transport pathway in prokaryotes, mediated by a conserved ATPase SecA. Using a combination of ribosome profiling methods, I revealed the essential roles of SecA in recognizing and resolving the widespread accumulation of large periplasmic loops of inner membrane proteins in the cytoplasm during their cotranslational translocation, and in the cotranslational transport of secretory proteins with highly hydrophobic signal sequences. I also uncovered a function of the chaperone trigger factor (TF) in temporally regulating SecA engagement on secretory proteins. These findings elucidate the principles of SecA-driven cotranslational protein translocation and reveal a hierarchical network of protein export pathways in bacteria (Chapter 2).

The second part of research focused on the more complex protein sorting systems of eukaryotes, where I comprehensively investigated the mitochondrial protein delivery from the cytosol using selective ribosome profiling in human cells. I found that the cotranslational protein targeting to mitochondria is initiated late during translation, directed by an N-terminal presequence and the exposure of a complex globular fold in the nascent protein. This pathway does not favor membrane proteins, but is predominantly used by large, multidomain and topologically complex proteins, whose import efficiency is enhanced when targeted cotranslationally. My results indicate that the cotranslational targeting of mitochondrial proteins is fundamentally different from that of the endoplasmic reticulum (ER) proteins, highlighting the diversity and specificity of protein targeting mechanisms across cellular systems (Chapter 3).

PUBLISHED CONTENT AND CONTRIBUTIONS

Zhu, Z.*, Wang, S.* & Shan, S. (2022). "Ribosome profiling reveals multiple roles of SecA in cotranslational protein export". In: *Nature Communications*. **13**, 3393. doi: 10.1038/s41467-022-31061-5. (* equal contribution)

> Z.Z. designed the experiments with S.W. and S.-o.S., performed all experiments with S.W., analyzed all data, and wrote the manuscript with S.-o.S..

Zhu, Z., Mallik, S., Stevens, T. A., Levy, E. D., Shan, S. (2024). "Timing and specificity of cotranslational mitochondrial protein import". *Submitted*

> Z.Z. designed the experiments with S.-o.S., performed all experiments and analyzed all data with help from S.M., T.S. and E.D.L. on contact order analysis and mitochondria purification, and wrote the manuscript with S.-o.S.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Tables   *Page*

NOMENCLATURE

ACO. Absolute contact order.

AID. Auxin-inducible degron.

CCCP. Carbonyl cyanide m-chlorophenyl hydrazone.

CHX. Cycloheximide.

DiSP. Selective disome profiling.

DSP. Dithiobis(succinimidyl propionate).

EDC. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide.

ER. Endoplasmic reticulum.

IMM. Inner mitochondrial membrane.

IMPs. Inner membrane proteins.

IMS. Intermembrane space.

MBP. Maltose-binding protein.

MTS. Mitochondrial targeting sequence.

NAC. Nascent polypeptide-associated complex.

OMM. Outer mitochondrial membrane.

OXPHOS. Oxidative phosphorylation.

PMF. Proton motive force.

PTC. Peptidyl transferase center.

ROC. Receiver operator characteristic.

RNCs. Ribosome-nascent chain complexes.

SeRP. Selective ribosome profiling.

SRP. Signal recognition particle.

TF. Trigger factor.

TMD. Transmembrane domain.

TOM. Translocase of the outer membrane.

ZnBD. Zinc binding domain.

C h a p t e r   1

INTRODUCTION

1.1 Ribosome profiling

Ribosome profiling is a technique based on deep sequencing of ribosome-protected mRNA fragments[1], which provides snapshots of ribosome positions on mRNA at a given moment, offering a high-resolution view of translation *in vivo*. Selective ribosome profiling[2,3] (SeRP), by coupling ribosome profiling with affinity purification of ribosomes bound to specific nascent chain-associated factors, provides information of specific subpopulations of translating ribosomes. This technique is particularly useful for profiling a protein's cotranslational interactions *in vivo*. In Chapter 2, we established SeRP of SecA in *Escherichia coli* to comprehensively investigate the roles of SecA in cotranslational protein export. To obtain the spatial information during protein targeting, we further developed fractionation-coupled ribosome profiling to monitor the arrival of nascent protein at the plasma membrane. In Chapter 3, we sought to investigate the protein targeting to mitochondria, a specific organelle in eukaryotes. By establishing SeRP of TOM complex in human cells, we were able to demonstrate the spatiotemporal regulation of nascent mitochondrial protein targeting.

1.2 Protein targeting in bacteria

In bacteria, inner membrane proteins (IMPs) and secretory proteins, which comprise a quarter of newly synthesized proteins, need to be transported onto or across the plasma membrane[4,5]. Previous studies have established two distinct targeting pathways for IMPs and secretory proteins: IMPs are cotranslationally recognized by signal recognition particle (SRP), typically via the first transmembrane domain (TMD), and targeted to the SecYEG or YidC translocation machineries at the inner membrane[6–9]; secretory proteins are targeted and translocated after they are released from ribosomes. SecA, an evolutionarily conserved and essential ATPase[10], drives the translocation of preproteins across SecYEG

posttranslationally[11–14]. However, recent studies showed that SecA also binds to ribosomes, challenging the strictly posttranslational mechanism of SecA-mediated protein targeting and translocation. The role of SecA-ribosome interaction remained unclear. Therefore, in Chapter 2, we comprehensively investigated SecA's cotranslational actions *in vivo* by generating proteome-wide cotranslational targeting and SecA interaction profiles at near-codon resolution. This allowed us to define the nascent substrate pool of SecA *in vivo* and identify determinants of the timing and specificity of these interactions.

In addition to the targeting factors such as SRP and SecA, nascent proteins may associate with cytosolic chaperones before reaching the membrane. Some secretory proteins are maintained in the unfolded state by the cytosolic chaperone SecB[15,16] during their delivery to membrane-bound SecA. The highly abundant ribosome-associated chaperone trigger factor (TF) was also suggested to act as a holdase during the targeting of secretory proteins[17,18]. How SecA, SecB, TF and SRP coordinated in space and time for ribosome binding remain poorly understood. Thus, we compared the SeRP profiles of SecA with those of SRP and TF, and examined the timing of SecA binding on nascent proteins in the absence of TF or SecB. Our findings in Chapter 2 revealed a hierarchical network of chaperones/targeting factors that ensure robust protein transport in bacteria.

1.3. Protein targeting to mitochondria

Mitochondria are eukaryotic organelles that play critical roles in various cellular processes, including ATP synthesis, calcium homeostasis and lipid metabolism[19]. Fundamental to these roles is the proper localization of mitochondrial proteins. Over 99% of mitochondrial proteins are encoded by the nuclear genome, synthesized in the cytosol, and must be imported into mitochondria. This dynamic interplay between the cytosol and mitochondria necessitates diverse and sophisticated protein targeting and translocation mechanisms[20,21]. In contrast to the well-studied protein translocation machineries in mitochondria, targeting steps in the cytosol that deliver proteins to mitochondria remain poorly understood[21,22]. While the classic view is that protein targeting to mitochondria occurs post-translationally, recent studies revealed the presence of mitochondrially localized mRNA[23,24], ribosomes [25],

and translation[26], suggesting that targeting may also occur cotranslationally. However, direct evidence supporting cotranslational protein import into mitochondria is missing, and the timing, physiological roles, and underlying principles of this process remain unclear. To address these questions and fill the gap between localized translation and protein import, we systematically investigated cotranslational protein targeting to mitochondria in human cells at near-codon resolution in Chapter 3.

Chapter 2

RIBOSOME PROFILING REVEALS MULTIPLE ROLES OF SECA IN COTRANSLATIONAL PROTEIN EXPORT

Adapted from:

2.1 Abstract

SecA, an ATPase known to posttranslationally translocate secretory proteins across the bacterial plasma membrane, also binds ribosomes, but the role of SecA's ribosome interaction has been unclear. Here, we used a combination of ribosome profiling methods to investigate the cotranslational actions of SecA. Our data reveal the widespread accumulation of large periplasmic loops of inner membrane proteins in the cytoplasm during their cotranslational translocation, which are specifically recognized and resolved by SecA in coordination with the proton motive force (PMF). Furthermore, SecA associates with 25% of secretory proteins with highly hydrophobic signal sequences at an early stage of translation and mediates their cotranslational transport. In contrast, the chaperone trigger factor (TF) delays SecA engagement on secretory proteins with weakly hydrophobic signal sequences, thus enforcing a posttranslational mode of their translocation. Our results elucidate the principles of SecA-driven cotranslational protein translocation and reveal a hierarchical network of protein export pathways in bacteria.

2.2 Introduction

Generation and maintenance of compartmentalization is essential for the proper functioning of all cells and requires that all newly synthesized proteins be localized to their correct cellular destinations. In bacteria, a quarter of newly synthesized proteins

need to be transported onto or across the plasma membrane[4,5]. Inner membrane proteins (IMPs) are cotranslationally recognized by signal recognition particle (SRP), typically via the first transmembrane domain (TMD), and targeted to the SecYEG or YidC translocation machineries at the inner membrane[6–9]. It was generally thought that most secretory proteins are targeted and translocated posttranslationally, after they are released from ribosomes. SecA, an evolutionarily conserved and essential ATPase[10], drives the translocation of preproteins across SecYEG[11–14]. Preprotein translocation can be further assisted by the proton-motive force (PMF)[11,27–29] and by SecDF, an ancillary complex that associates with SecYEG and ratchets translocating substrates into the periplasm using PMF[30,31].

Nascent secretory proteins may associate with cytosolic chaperones before reaching the membrane. Some secretory proteins are maintained in the unfolded state by the cytosolic chaperone SecB[15,16] during their delivery to membrane-bound SecA. The highly abundant ribosome-associated chaperone TF was also suggested to act as a holdase during the targeting of secretory proteins[17,18]. TF cotranslationally binds numerous secretory proteins beginning at a nascent chain length of ~100 amino acids[3]. Deletion of TF accelerates protein export[32] and results in more cotranslational translocation of some secretory proteins[3,33], but the mechanism and the physiological role of TF association with nascent secretory proteins remain unclear.

Recent studies challenge the strictly posttranslational mechanism of SecA-mediated protein targeting and translocation. First, SecA binds the ribosome on uL23 near the nascent polypeptide exit tunnel[34,35], and mutations that disrupt SecA's ribosome binding slowed the export of a classic SecB/A substrate, maltose-binding protein (MBP)[34]. In addition, SecA is required for the translocation of multiple IMPs, which must use a cotranslational mechanism of targeting and insertion mediated by SRP[36–40]. Finally, SecA recognizes ribosomes exposing the nascent chain of an IMP, RodZ, with high affinity (K_d < 1 nM) and is necessary and sufficient for the cotranslational targeting and translocation of RodZ in the absence of SRP[35,41,42]. These observations implicate a role of SecA in cotranslational protein transport.

However, many fundamental questions concerning the cotranslational actions of SecA remain unanswered. The spectrum of nascent proteins that cotranslationally interact with SecA is unclear, as is the precise role of these interactions in protein biogenesis. The timing of SecA-nascent chain association as well as the determinants of SecA recognition remain elusive. Moreover, the binding sites of SecA, ribosome, and SecYEG on one another overlap extensively[43–45], raising questions as to whether and how SecA engages ribosome-translocon complexes cotranslationally. Finally, how SecA coordinates with other ribosome-bound targeting factors and chaperones in bacteria to triage nascent membrane and secretory proteins is not well understood. Here, we decipher the cotranslational actions of SecA by globally profiling SecA-nascent chain interactions and monitoring the targeting of individual substrates *in vivo*. Our proteome-wide data set identifies the nascent substrate pool of SecA, elucidates the roles of SecA and its coordination with PMF in cotranslational protein translocation, underscores the role of TF in delaying the SecA-mediated transport of a subset of secretory proteins, and reveals a hierarchical network of protein targeting and translocation pathways in bacteria.

2.3 Results

2.3.1 Cotranslational engagement of SecA with nascent proteins

To understand the principle of SecA-nascent chain interactions, we carried out selective ribosome profiling (SeRP) of SecA in *Escherichia coli* (Fig. 2.1A). Chromosomal SecA was fused to a C-terminal thrombin-cleavable AviTag, which can be biotinylated by the endogenous biotin ligase, to facilitate the purification of SecA-bound ribosome-nascent chain complexes (RNCs). Very low amounts of ribosomes co-purified with SecA in initial experiments (fig. S2.1A, lane 3), which was attributed to the transient nature of SecA-RNC interactions. After extensive optimization, we chose to crosslink SecA to nascent chains in the lysate using EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide), a zero-length crosslinker that couples carboxyl groups with primary amines. Analysis of the elution products from affinity purification showed that addition of EDC significantly increased the yield of ribosomes, whereas no

ribosome was detected in parallel purifications from cells with untagged SecA (fig. S2.1A). Ribosome-protected mRNA fragments from the total ribosome population and from SecA-bound RNCs were then extracted and sequenced, generating information of the translatome and SecA-interactome (Fig. 2.1, fig. S2.1B,C and fig. S2.2).

To validate that our procedure correctly captured SecA-RNC complexes, we sorted all the detected proteins based on their localization and found that IMPs and a subset of secretory proteins are strongly enriched in the SecA interactome (Fig. 2.1B and Table S2.2), consistent with a specific role of SecA in protein transport. In addition, the SecA enrichment profile of its model cotranslational substrate, RodZ (Fig. 2.1C), showed a distinct peak beginning at a nascent chain length of 186 amino acids, indicating a cotranslational engagement by SecA, and further supporting the specificity of the affinity purification procedure in faithfully capturing SecA-nascent chain interactions.

We next determined the timing of cotranslational SecA engagement by performing metagene analyses of the total translatome and the SecA interactome on all Sec substrates, including IMPs and secretory proteins, aligned to their start and stop codons (Fig. 2.1D). The metagene translatome showed a uniform distribution of ribosome footprint read density along the transcript. The read density of the SecA interactome rises gradually during the translation of the first ~120 amino acids, reaches a plateau afterwards, and rises again near the end of translation (Fig. 2.1D). As described in the sections below, this metagene profile arises from the combination of the distinct timing of SecA interactions with different classes of substrates.
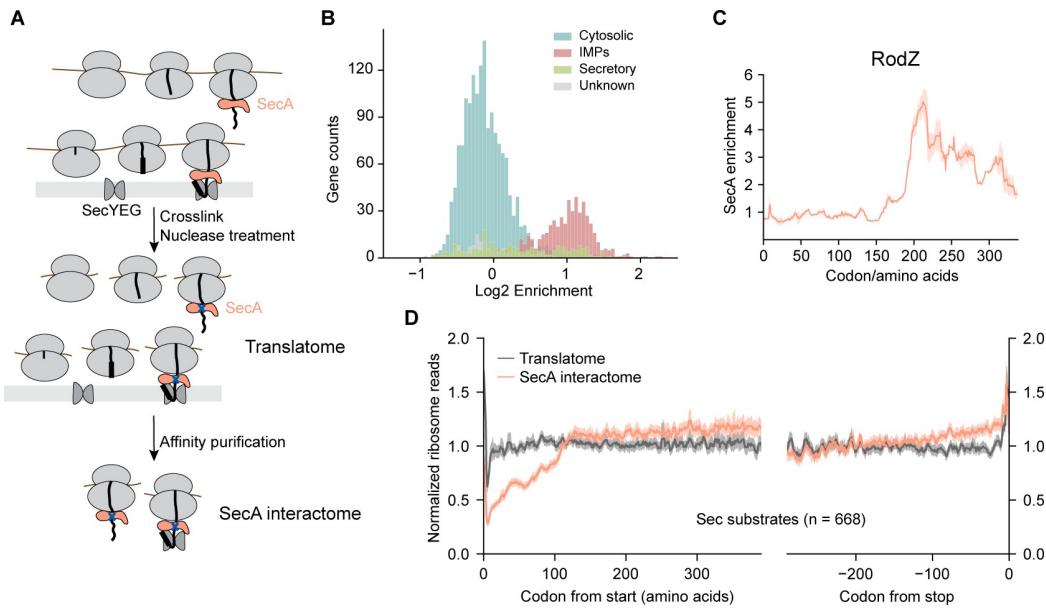
Fig. 2.1. Cotranslational engagement of SecA with RNCs

(**A**) Schematic of SeRP of SecA. Cells are harvested and cryogenically lysed, followed by crosslinking SecA-RNCs using EDC. Monosomes generated by nuclease digestion of polysomes are isolated by centrifugation through a sucrose cushion. SecA-RNCs are purified from total monosomes via the biotin moiety on Avi-tagged SecA and eluted by thrombin cleavage. Ribosome-protected fragments from the total ribosome population and from SecA-bound RNCs were extracted and sequenced, generating information on the translatome and SecA interactome. (**B**) Distribution of SecA enrichment on all genes (n = 2562) categorized by protein localization. (**C**) SecA interaction profile of the RodZ nascent chain. The solid line shows the mean values and shaded areas show the differences between two independent biological replicates. (**D**) Metagene translatome and SecA interactome profiles of Sec substrates aligned to the start and stop codon. The solid line shows the mean values and shaded areas show the 95% confidence interval (CI).

2.3.2 Fractionation-coupled ribosome profiling reveals the timing of membrane targeting

Considering the important role of SecA in protein translocation, we reasoned that SecA-RNC interaction may be correlated with the timing of nascent chain delivery to the membrane. To test this hypothesis, we developed a protocol, fractionation-coupled ribosome profiling, to monitor cotranslational protein targeting events (Fig. 2.2A and

fig. S2.3A). In this protocol, we carried out nuclease treatment before cell fractionation, so that only the RNCs stably bound to the membrane via protein-protein or protein-lipid interactions would be retained in the membrane fraction. In contrast, ribosomes on which the nascent chain does not stably engage with components on the membrane would be released into the cytosol after nuclease digestion and captured in the soluble fraction. Ribosome-protected reads from both fractions are then purified, sequenced, and mapped onto protein-coding sequences.

For IMPs, ribosome-protected reads in the soluble fraction declined sharply upon the emergence of the first TMD, with a corresponding rise in the read density of the same protein in the membrane fraction (Fig. 2.2B). To normalize for local variations in translation speed, we calculated the ratio of membrane reads to soluble reads at each codon, generating the ribosome localization profile for each protein (Fig. 2.2B). The metagene ribosome localization profile of all 517 IMPs, aligned to the N-terminus of their first TMD, revealed that IMPs on average are targeted to the membrane after their first TMD emerges from the ribosomal exit tunnel (Fig. 2.2D, lower panel), consistent with previous reports that SRP recognizes the first TMD of IMPs for targeting[7]. We also observed delayed targeting in the ribosome localization profiles of IMPs whose first TMD is skipped by SRP, which confirmed their SRP-dependent targeting (fig. S2.3B). In contrast, the ribosome localization profiles of cytosolic proteins remained unchanged throughout translation (fig. S2.3C,D).

We next compared the onset of SecA enrichment with the timing of membrane association for all IMPs (Fig. 2.2D). Except for a few IMPs that are too short to be cotranslationally targeted, 91% of all 517 IMPs detected display a gradual rise in SecA enrichment after the emergence of their first TMD, which coincide with their membrane association (Fig. 2.2D and fig. S2.3E). In the interaction profiles of individual IMPs, SecA binding increases concurrently with their membrane association but showed remarkably different patterns afterwards. On most IMPs, SecA association quickly plateaus at SecA enrichment of 2-3 fold and continues until the end of translation (Fig. 2.2C and fig. S2.3E). On a subset of IMPs, however, we observed strong SecA binding

peaks that rise above and then return to this plateau (Fig. 2.3A-C and described in the next section). Thus, the low and persistent level of SecA enrichment, which displays no specificity in the length or identity of the nascent chain, may represent transient interactions between membrane-bound SecA and the translocating RNCs until a specific and stable interaction is established.
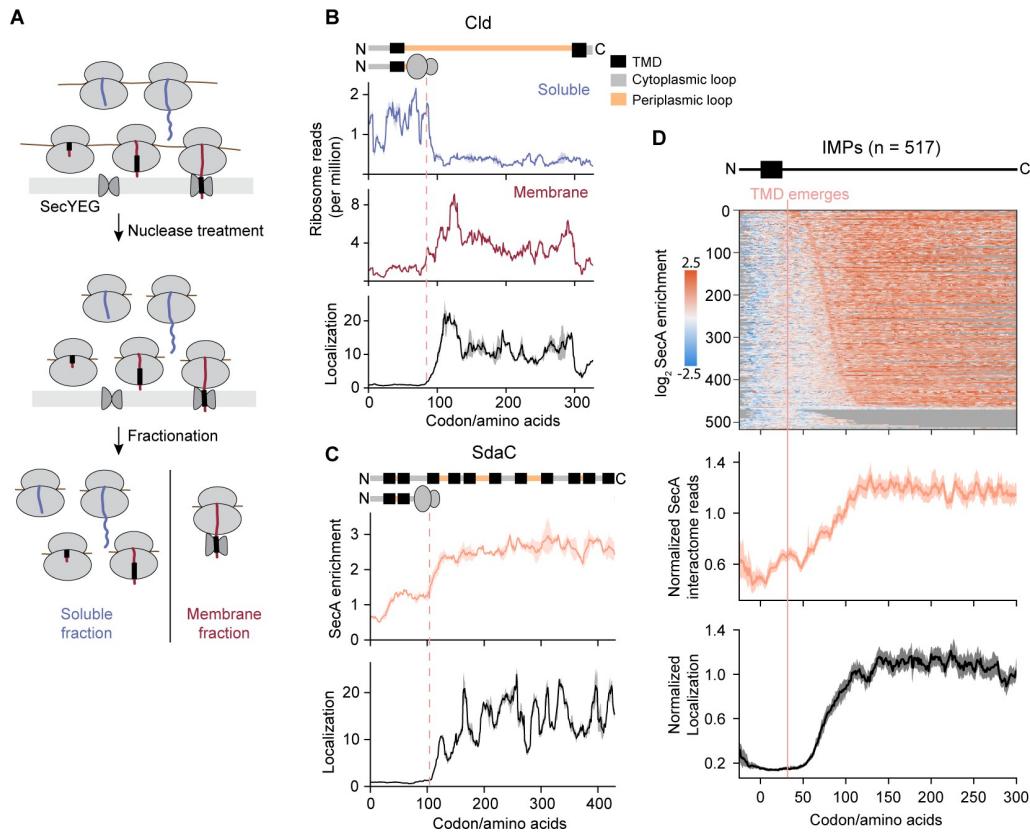
Fig. 2.2. Fractionation-coupled ribosome profiling reveals the timing of cotranslational membrane targeting

(**A**) Schematic of fractionation-coupled ribosome profiling. (**B**) Representative ribosome localization profile of an IMP. The localization score at each codon was calculated as the ratio of ribosome-protected reads in the membrane faction relative to those in the soluble fraction. Protein topology is shown above, with TMD in black, cytoplasmic loops in grey, and periplasmic loops in yellow. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**C**) Representative SecA interaction profile and ribosome

localization profile of an IMP. Protein topology is shown above and colored as in (**B**). Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D**) Metagene SecA interactome profile and ribosome localization profile of IMPs aligned to the N-terminus of the first TMD. Solid lines show the mean values and shaded areas show the 95% CI. The heatmap above shows the $\log_2$ SecA enrichment at each codon of all the IMPs used to derive the metagene SecA interactome profile, sorted by increasing distance from the TMD to the onset of SecA binding.

2.3.3 SecA cotranslationally translocates large periplasmic loops of inner membrane proteins

We further investigated the strong SecA binding peaks on IMPs, which appeared when part of a large periplasmic loop has emerged from the ribosomal exit tunnel (Fig. 2.3A-C and fig. S2.4A,B). These peaks typically persist for more than 30 codons and are independent of the number of TMDs in the protein (Fig. 2.3A-C and fig. S2.4A,B). Interestingly, for proteins with more than one large periplasmic loop, strong SecA interaction peaks could be observed multiple times, with binding peaks correlating with a partially exposed periplasmic loop (Fig. 2.3C). These observations agree with previous reports showing that the translocation of IMPs containing large periplasmic loops requires SecA[36–40], suggesting that SecA stably engages RNC-SecYEG complexes during the cotranslational translocation of large periplasmic loops.

To determine the timing of SecA engagement on the periplasmic loops, we generated a metagene SecA interactome profile for all the predicted periplasmic loops longer than 100 amino acids aligned to their N-terminus (Fig. 2.3D). Loops N-terminal to the first TMD, which are not targeted to the membrane during their synthesis, were excluded from this analysis. SecA on average initiates binding when the N-terminus of a large periplasmic loop reaches a distance of 75 amino acids from the peptidyl transferase center (PTC), and maximal binding is reached at a distance of 110 amino acids from the PTC (Fig. 2.3D). Assuming that the C-terminal ~30 amino acids of the nascent chain are buried in the ribosomal exit tunnel, the emergence of at least 45 amino acids of the loop from the ribosome is required for strong SecA engagement. Consistently, IMPs

containing a periplasmic loop longer than 45 amino acids are significantly enriched in the SecA interactome in a gene-level comparison (Fig. 2.3E).

To test whether the length or specific sequence features of a periplasmic loop determines SecA engagement, we studied how mutations introduced in the periplasmic loops of two proteins, OppC and AcrB, impact SecA binding (Fig. 2.3C,F). WT OppC has a periplasmic loop of 43 amino acids that does not cotranslationally recruit SecA (Fig. 2.3F). However, introduction of a tandem repeat of this stretch of 43 amino acids in OppC-43aa repeat, which only increased the length of the periplasmic loop, led to the appearance of a strong SecA binding peak after ~60 amino acids of the altered loop have emerged from the ribosome (Fig. 2.3F), in agreement with the timing of SecA engagement observed with the metagene SecA interactome (Fig. 2.3D). The second protein, AcrB, contains two large periplasmic loops that both strongly engage SecA (Fig. 2.3C). The first SecA binding peak ends at residue 105 of the first periplasmic loop. Considering that over 20 amino acids at the N-terminus of the loop are inserted into SecYEG for a TMD in type II topology[44], and at least 30 amino acids[46] at the C-terminus of the loop are still in the ribosomal exit tunnel, we reasoned that the emergence of residues 20-75 in this loop determined SecA association. If SecA engagement is sequence-dependent, it would be abolished with AcrB-71aa, in which residues 20-90 of this loop is deleted. Instead, the truncation led to a new SecA binding peak at residues 70-105 of the new periplasmic loop (Fig. 2.3C), at a protein sequence that showed no SecA binding in WT AcrB (Fig. 2.3C). These data indicate that high affinity SecA binding to RNC-SecYEG complexes is determined by the length, rather than specific amino acid sequence, of periplasmic loops of IMPs.

We next explored if SecA also interacts with other segments of nascent IMPs besides large periplasmic loops (e.g., TMDs of some proteins). We developed a peak detection algorithm to identify SecA peaks with at least 12 consecutive codons above a threshold of 3.5-fold enrichment. The stringent criteria ensures that only the strongest SecA interactors were detected and excludes the baseline levels of nonspecific SecA enrichment. Among the 63 substrates detected by this method, 58 of them expose a large

periplasmic loop outside the ribosome at the onset of SecA binding (Fig. 2.3G). A
closer inspection of the other 5 proteins, which lacks large periplasmic loops, reveals
only a higher baseline of SecA enrichment (fig. S2.4D). At the metagene level, the onset
of strong SecA binding occurs after the protein is completely targeted to the membrane
(fig. S2.4E), reinforcing SecA's involvement in protein translocation. Together, these
findings indicate that driving the translocation of large periplasmic loops is the primary
role of SecA in the cotranslational insertion of IMPs.
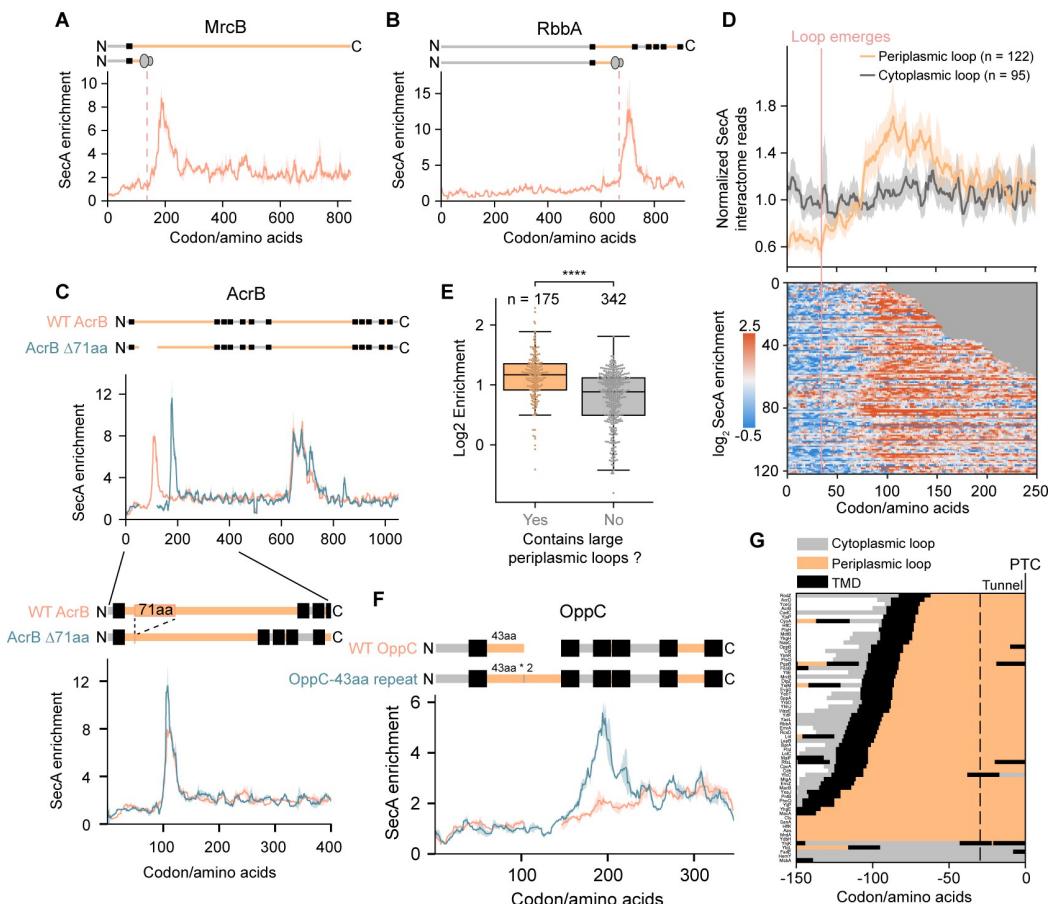
Fig. 2.3. SecA cotranslationally translocates large periplasmic loops on IMPs
(**A,B**) Representative SecA interaction profiles of IMPs with large periplasmic loops. Protein
topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and
shaded areas show the differences between two independent biological replicates. (**C**) Upper
panel, comparison of SecA interaction profile of WT AcrB and AcrB-Δ71aa. The topology of the

protein is shown above and colored as in Fig. 2.2B. The deleted region was skipped in the indicated protein topology and SecA interaction profile of AcrB-Δ71aa so that SecA enrichment at each codon corresponds to the same protein sequence between WT AcrB and AcrB-Δ71aa. Lower panel, zoom in of the first 400 codons. SecA enrichment is plotted as a function of ribosome position relative to the start codon for each protein. aa, amino acids. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D**) Metagene SecA interactome profile of periplasmic and cytoplasmic loops larger than 100 amino acids aligned to their N-terminus. Solid lines show the mean values and shaded areas show the 95% CI. The heatmap below shows the $\log_2$ SecA enrichment at each codon for all the periplasmic loops used to derive the metagene SecA interactome profile, sorted by the length of each periplasmic loop. (**E**) Comparison of gene-level enrichment between proteins with and without a large periplasmic loop. The centre line represents the median, the bounds of box represent the upper and lower quartiles, and the whiskers indicate 1.5x the interquartile range. $P = 8.326e^{-17}$, two-sided Wilcoxon rank-sum test. (**F**) Comparison of SecA interaction profile of WT OppC and OppC-43aa repeat, which contains two tandem copies of the 43aa periplasmic loop in WT OppC. Protein topology is shown above and colored as in Fig. 2.2B. The inserted region (43 amino acids) was skipped in the protein topology and SecA interaction profile of WT OppC so that the SecA enrichment at each codon corresponds to the same protein sequence between WT OppC and OppC-43aa repeat. aa, amino acids. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**G**) Topology of translated nascent chains of strong SecA interactors at the onset of SecA binding. The nascent chains are aligned to their C-terminus (position 0, at PTC) and the residues to the left of the dashed line are exposed outside the ribosomal tunnel exit.

2.3.4 Distinct modes of SecA engagement on nascent secretory proteins

We next investigated the interaction of SecA with nascent secretory proteins. To identify substrates of SecA among nascent secretory proteins, we used the peak detection algorithm to scan for regions with at least 12 codons above a threshold of 1.7-fold SecA enrichment. We identified 115 SecA interactors out of 233 reliably detected secretory proteins (Fig. 2.4A), indicating that approximately half of the nascent secretory proteome contact SecA before they are completely synthesized and released from the ribosome. Since the chaperone TF also interacts with nascent secretory proteins, we overlaid the TF

binding peaks[3] on the identified SecA substrates (Fig. 2.4B). SecA and TF share many substrates but seldom bind simultaneously on RNCs. Two classes of SecA substrates were then sorted based on which factor engages first.

Among the 115 nascent secretory substrates of SecA, 56 contact SecA before TF or do not engage TF throughout translation (referred to as SecA-first substrates; Fig. 2.4B,C and fig. S2.5A,B). The metagene SecA interactome profile of these SecA-first substrates showed that SecA binding begins when the nascent chain is ~100 amino acids long. SecA engagement persists for ~50 amino acids (Fig. 2.4D) and declines afterwards, long before protein synthesis is finished, which is analogous to the SecA interaction on the large periplasmic loops of translocating IMPs. Different from the IMPs (fig. S2.4E), however, there is no lag between the membrane association of these secretory proteins and the onset of SecA binding peak. Instead, SecA engagement precedes their membrane association (Fig. 2.4C,D and fig. S2.4A,B), suggesting a role of SecA in both the membrane targeting and initial translocation of these secretory proteins. TF seldom binds these substrates after SecA engagement (Fig. 2.4B). Collectively, these data suggest a SecA-mediated cotranslational export pathway utilized by ~25% of secretory proteins.

The other 59 SecA-interacting nascent secretory proteins first contact TF at a length of ~100 amino acids, and SecA does not bind these proteins until after TF dissociation (referred to as TF-first substrates; Fig. 2.4B,E and fig. S2.5C,D). The interaction of SecA with TF-first substrates rises with increasing nascent chain length and lasts until the end of translation at both the metagene level and for individual proteins (Fig. 2.4B,E,F and fig. S2.5C,D). Considering the ability of SecA to bind secretory proteins post-translationally, SecA may continue to engage these substrates after they complete translation and are released from the ribosome. The membrane targeting of these TF-first substrates was also relatively late, occurring near the end of translation (Fig. 2.4E,F and fig. S2.5C,D). Thus, the observations on TF-first substrates may reflect a temporally cotranslational but mechanistically posttranslational mechanism of export[47] mediated by SecA.
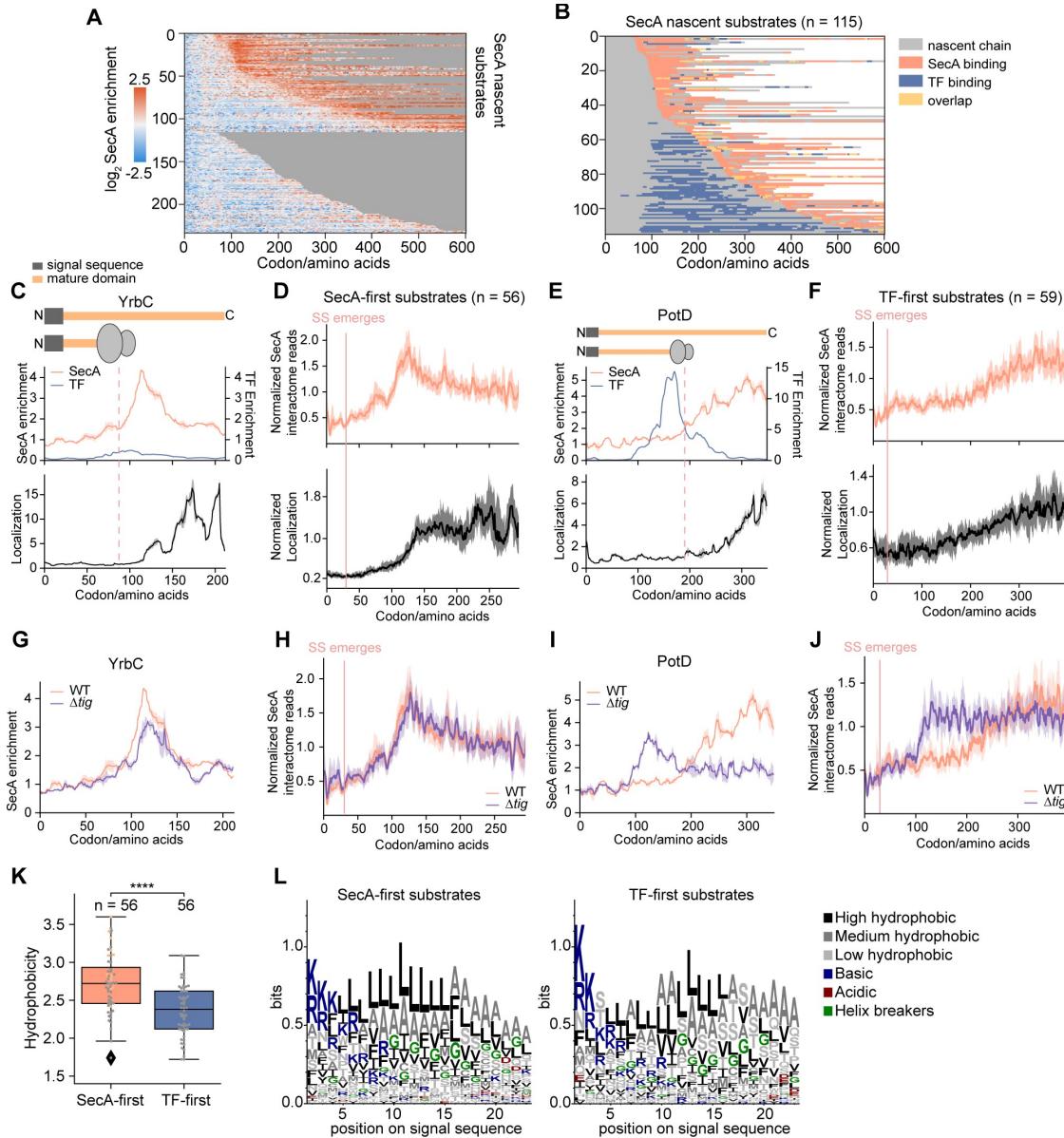
Fig. 2.4. SecA engagement on nascent secretory proteins is delayed by trigger factor

(**A**) Heatmap of $\log_2$ SecA enrichment at each codon of all secretory proteins. (**B**) Heatmap of SecA and TF binding[3] on SecA-interacting nascent secretory proteins, sorted by increasing distance from the start codon to the onset of SecA binding. (**C, E**) Representative SecA and TF[3] interaction profiles and ribosome localization profiles of a SecA-first (**C**) and a TF-first (**E**) substrate. Protein topology is shown above, with signal sequence in dark grey and mature domain in yellow. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D, F**) Metagene SecA interactome profile and ribosome

localization profile of SecA-first (**D**) and TF-first (**F**) substrates aligned to the N-terminus of signal sequences. Solid lines show the mean values and shaded areas show the 95% CI. (**G, I**) Representative SecA interaction profiles of a SecA-first (**G**) and a TF-first (**I**) substrate. Protein topology is shown above and colored as in (**C**). Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**H, J**) Metagene SecA interactome profile of SecA-first (**H**) and TF-first (**J**) substrates aligned to the N-terminus of signal sequences. Solid lines show the mean values and shaded areas show the 95% CI. (**K**) Hydrophobicity of the signal sequences of SecA-first and TF-first substrates. Yellow dots indicate substrates that also engage with SRP. The signal sequences of Ppk, SpeA and HybC cannot be faithfully predicted and are excluded from the analysis. The centre line represents the median, the bounds of box represent the upper and lower quartiles, and the whiskers indicate 1.5x the interquartile range. $P = 3.716e^{-6}$, two-sided Wilcoxon rank-sum test. Source data are provided in the Source Data file. (**L**) WebLogo representations of the amino acid compositions of the signal sequences of SecA-first and TF-first substrates aligned to the second amino acid.

2.3.5 TF delays SecA engagement on nascent secretory proteins

To further understand the molecular interplay between SecA and TF, we analyzed the SecA interactome in cells lacking the *tig* gene encoding TF. The timing of SecA engagement on SecA-first substrates was unaffected by *tig* deletion, although the enrichment was slightly lower for individual genes (Fig. 2.4G,H and fig. S2.6A,B). In contrast, we observed much earlier SecA interactions with TF-first substrates in the Δ*tig* strain, starting at a nascent chain length of ~100 amino acids (Fig. 2.4I,J and fig. S2.6C,D). In addition, deletion of TF resulted in the cotranslational engagement of SecA with secretory proteins that, in WT cells, interacted with TF throughout translation and did not cotranslationally bind SecA (fig. S2.6E-H). These results suggest that SecA by itself does not distinguish and sort secretory proteins into co- or post-translational pathways. Instead, TF selectively interacts with a subset of secretory proteins and delays their early engagement with SecA, thereby enforcing a posttranslational mode of their export.

To explore what determines the early recognition of TF versus SecA on nascent secretory proteins, we compared the Kyte-Doolittle hydrophobicity score of the signal

sequences of the two classes of SecA substrates (Fig. 2.4K), since only the signal sequence and the following ~50 amino acids are exposed outside the ribosome at the onset of SecA or TF binding. The signal sequences of TF-first SecA substrates are significantly less hydrophobic than those of the SecA-first substrates. Furthermore, thirteen of the SecA-first substrates recruit SRP cotranslationally[7], consistent with the higher hydrophobicity of their signal sequences (fig. S2.6I). We next performed a position-wise analysis of sequence logos for the signal sequences[48] and found that SecA-first substrates are more enriched in hydrophobic residues, especially leucine (Fig. 2.4L). The hydrophobicity of signal sequences of non-SecA substrates, most of which are recognized by TF, are similar to that of TF-first SecA substrates (Fig. 2.4L and fig. S2.6J,K). Thus, hydrophobicity of the signal sequence is one of the factors that dictates differential recognition of preproteins by TF versus SecA. TF preferentially recognizes secretory proteins with less hydrophobic signal sequences and delays their interaction with SecA.

We further asked what feature necessitates the use of a cotranslational pathway of export for SecA-first substrates. Considering that the narrow SecYEG pore only allows the passage of unfolded proteins[49,50], one possibility is that cotranslational translocation ensures the successful export of secretory proteins that fold rapidly and stably, which may otherwise block the translocon in a posttranslational export pathway[51,52]. To test this hypothesis, we calculated the absolute contact order for all SecA-interacting nascent substrates. The absolute contact order is a measure of the total distance between residues that form native contacts in the folded protein and was shown to accurately predict protein folding rates[53]. Interestingly, SecA-first substrates have significantly lower absolute contact order compared to TF-first substrates (fig. S2.6L), indicating their propensity to fold more quickly. This may explain, in part, the cotranslational engagement of these nascent proteins with translocation machineries.

2.3.6 Impact of SecB on SecA-nascent chain interaction

SecB is an interaction partner of SecA during the posttranslational export of preproteins and has the ability to engage nascent polypeptides[54,55]. We therefore explored whether SecB impacts the cotranslational interaction of SecA with nascent proteins. To this end, we performed SeRP in two mutant *E. coli* strains: Δ*secB*, in which the *secB* gene is deleted; and *secAΔZnBD*, in which we chromosomally deleted the zinc binding domain (ZnBD; aa 885-896) of SecA required for SecA-SecB binding[56] to minimize cell stress responses that may be triggered by SecB deletion[57].

Unlike the observations with the Δ*tig* strain, the timing of SecA engagement with nascent secretory proteins was unaffected in both the Δ*secB* and *secAΔZnBD* strains (Fig. 2.5A,B and fig. S2.7A,B), indicating that SecA binding on most nascent secretory proteins is independent of SecB. Furthermore, the peak detection analyses showed a heavy overlap of nascent secretory protein substrates of WT SecA and SecAΔZnBD, with only 20 proteins that interacted exclusively with WT SecA and 8 that interacted exclusively with SecAΔZnBD (Fig. 2.5C). A closer inspection of these 28 proteins showed that only SecA engagement on nascent MBP, a classic SecB substrate, is clearly dependent on the SecA-SecB interaction (Fig. 2.5D). The other 27 proteins have a modest difference in the enrichment score of the SecA binding peaks, such that they fell below or rose above the criteria in our peak detection algorithm (fig. S2.7C). A comparison of the gene-level enrichment of the overlapping substrates further identified 7 proteins whose interaction with SecA were severely impaired by the ΔZnBD mutation (fig. S2.7D,E). These defects were also observed in SecB deletion cells (Fig. 2.5D and fig. S2.7C-E). Thus, our results indicate that SecB may facilitate but is not strictly required for SecA recognition on nascent secretory proteins.

Notably, disruption of the SecA-SecB interaction or deletion of SecB led to a reduction in the cotranslational binding of SecA on the large periplasmic loops of IMPs (Fig. 2.5E,F and fig. S2.7F). This reduction was not observed with SecA interaction on nascent secretory proteins (fig. S2.7G,H), indicating that it did not arise from experimental variations. These results suggest that, although SecA can cotranslationally

engage with the large periplasmic loops of IMPs in the absence of SecB, SecB helps stabilize the formation of high affinity RNC-SecY-SecA complexes during translocation.
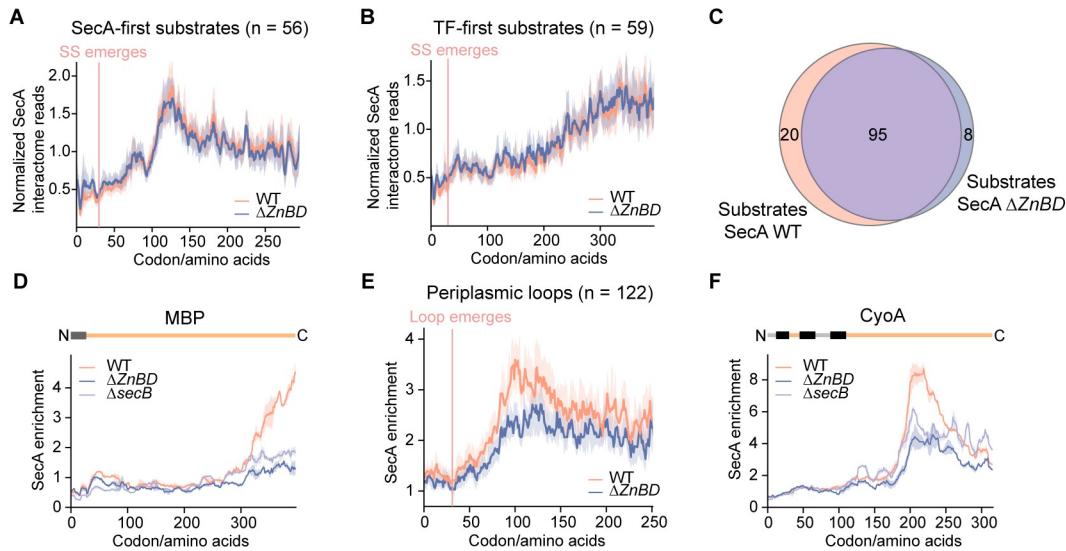
Fig. 2.5. Impact of SecB on cotranslational SecA interactions

(**A, B**) Metagene SecA interactome profile of SecA-first (**A**) and TF-first (**B**) substrates aligned to the N-terminus of signal sequences. Solid lines show the mean values and shaded areas show the 95% CI. (**C**) Venn diagrams showing the overlap of nascent secretory substrates of SecA and SecAΔZnBD. (**D, F**) Representative SecA interaction profiles of a secretory protein (**D**) and an IMP (**F**) on which SecA binding is compromised. Protein topology is shown above and colored as in Fig. 2.2B and Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**E**) Metagene SecA enrichment of large periplasmic loops aligned to their N-terminus. The median values of SecA enrichment in WT strain and in *ΔZnBD* strain are compared. Shaded areas show the 95% CI.

2.3.7 Proper cotranslational SecA interaction requires the proton motive force

An unexpected feature of cotranslational SecA engagement is that most SecA binding events peak within ~50 amino acids after its onset, and decline long before the periplasmic loop is completely translated. These observations suggest that other driving forces may also be involved in protein translocation. One candidate is PMF, which stimulates the translocation of multiple model substrate proteins[27–29]. Disruption of PMF

also affects the localization of SecA within the cell[58]. We therefore analyzed the SecA interactome in cells treated with carbonyl cyanide m-chlorophenyl hydrazone (CCCP), which dissipates the PMF.

We first used the peak detection algorithm to identify the strongest SecA-nascent IMP interactions upon CCCP treatment. SecA engagement on all except two substrates started after the emergence of a large periplasmic loop (fig. S2.7A), suggesting that SecA retains its role in translocating large periplasmic loops under conditions of PMF dissipation. However, the metagene SecA interactome of large periplasmic loops of IMPs aligned to their N-terminus showed that PMF dissipation profoundly alters the timing of SecA engagement with RNCs (Fig. 2.6A).

Inspection of the SecA interaction profiles of individual proteins revealed two distinct effects of PMF dissipation. Half of the 122 large periplasmic loops displayed prolonged SecA engagement (Fig. 2.6B and fig. S2.8B). Sustained SecA binding was also observed on some of the nascent secretory proteins upon CCCP treatment (Fig. 2.6B and fig. S2.8C). This is consistent with the results of a recent single molecule imaging study, which showed repeated SecA localization at specific membrane loci upon the dissipation of PMF[58]. In contrast, the other half of large periplasmic loops showed significantly reduced SecA engagement or even lost SecA interaction (Fig. 2.6D and fig. S2.8D). In addition, SecA enrichment on IMPs without a large periplasmic loop, which represents transient encounters of SecA as it diffuses on the inner membrane, is also significantly reduced in the absence of PMF (Fig. 2.6E,F). These results are in agreement with the reduced SecA diffusion on the plasma membrane upon CCCP treatment[58]. To test if this loss of SecA engagement is due to a failure of the membrane targeting of substrate proteins, we carried out fractionation-coupled ribosome profiling in CCCP-treated cells. Comparison of the ribosome localization profiles in CCCP-treated and untreated cells revealed that cotranslational protein targeting is unaffected by PMF dissipation (fig. S2.8E-H), excluding this possibility. Together, these results show that the proper timing and specificity of SecA's cotranslational interaction with IMPs and secretory proteins require an intact PMF.
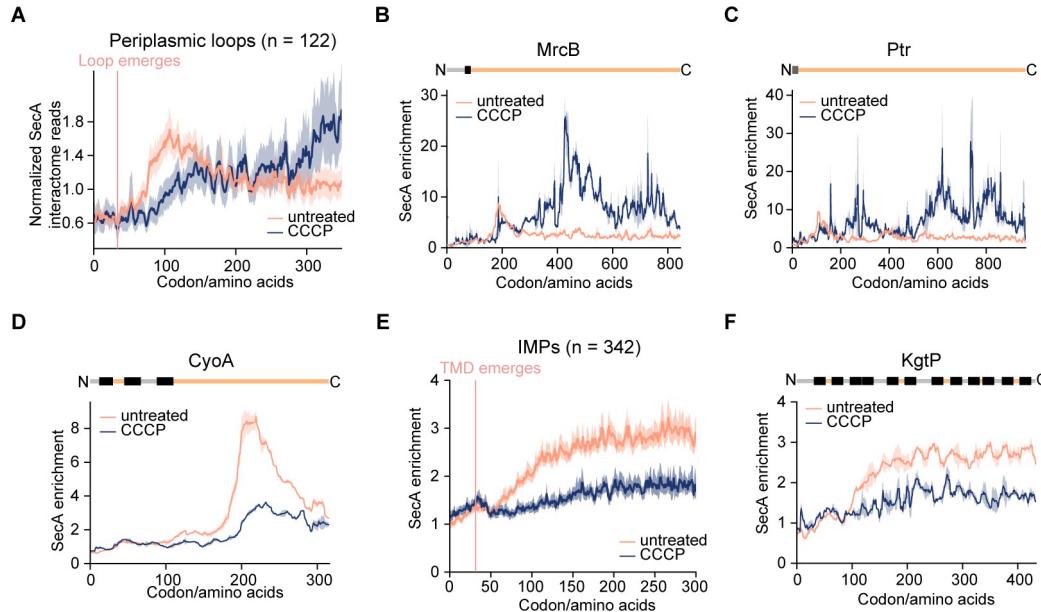
Fig. 2.6. PMF regulates cotranslational SecA interactions

(**A**) Metagene SecA interactome profile of periplasmic loops larger than 100 amino acids aligned to their N-terminus. Solid lines show the mean values and shaded areas show the 95% CI. (**B, C**) Representative SecA interaction profile of an IMP with a large periplasmic loop (**B**) and a secretory protein (**C**) that showed persistent SecA association upon CCCP treatment. Protein topology is shown above and colored as in Fig. 2.2B and Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D**) Representative SecA interaction profile of an IMP with a large periplasmic loop that showed significantly reduced SecA association upon CCCP treatment. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**E**) Metagene SecA enrichment of IMPs with no periplasmic loops longer than 45 amino acids aligned to the N-terminus of their first TMD. The mean SecA enrichment values in CCCP-treated and untreated cells are compared. Solid lines show the mean values and shaded areas show the 95% CI. (**F**) Representative SecA interaction profile of an IMP without a large periplasmic loop in untreated and CCCP-treated cells. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates.

2.4 Discussion

Despite recent progress in elucidating the molecular details of SecA-nascent chain interactions on the ribosome, the role of SecA's cotranslational actions *in vivo* and its relationship with other protein export pathways in bacteria remain elusive. Here, we used SeRP and fractionation-coupled ribosome profiling to address these questions. Our approaches generated proteome-wide cotranslational targeting and SecA interaction profiles at near-codon resolution in *Escherichia coli*, which allows us to define the nascent substrate pool of SecA *in vivo* and identify determinants of the timing and specificity of these interactions. Our results uncover the role of SecA in resolving topological problems during the cotranslational translocation of large periplasmic loops. Moreover, comparison of the SeRP profiles of SecA with those of SRP and TF, coupled with genetic perturbations, reveal a hierarchical network of chaperones/targeting factors that ensure robust protein transport in bacteria.

At the gene level, the SecA interactome is strongly enriched in IMPs (Fig. 2.1B), consistent with a microarray analysis of mRNAs copurified with SecA[59]. However, the modest SecA enrichment level along the transcripts of most IMPs and the lack of specificity of these interactions may reflect SecA in a scanning mode, in which it diffuses along the membrane to sample RNC-SecYEG complexes. In agreement with this interpretation, a recent single molecule imaging study showed that SecA is predominantly localized at and rapidly diffuses on the plasma membrane[58]. Given the heavy overlap of SecA and ribosome binding sites on SecYEG[43–45,60], SecA interaction with SecYEG or RNC is likely transient and quickly displaced by the RNC-SecYEG interaction. Nevertheless, this scanning mode might enable SecA to sense its substrates on SecYEG in a timely manner, allowing it to initiate more stable engagement with specific substrates.

Notably, SecA strongly and specifically engages translocating periplasmic loops of IMPs after ~45 amino acids of the loop emerge from the ribosome (Fig. 2.3). Crystal structure of a translationally stalled RNC-SecY complex showed that the mature domain

of DsbA forms a loop on the cytosolic surface of SecY, rather than entering the channel directly, during initiation of cotranslational translocation[44]. Our data agree with this structural observation and suggest that, for proteins with large periplasmic loops, cytosolic accumulation and exposure of the periplasmic sequence is a ubiquitous phenomenon during their cotranslational translocation *in vivo*, which could finally disrupt the ribosome-SecYEG junction. SecA specifically recognizes these species as soon as the topological problems emerge, and likely uses its ATPase cycle to drive the translocation of the periplasmic loops across SecYEG[30,61,62]. This provides a molecular model to explain why IMPs with large periplasmic loops tend to display a strong dependence on SecA for translocation[36–40].

Interestingly, most of these strong SecA interactions persist for only 50-100 amino acids and declines long before the periplasmic loop is completely translated (Fig. 2.3A-D), suggesting that SecA is only responsible for initial translocation. One possibility is that SecA-driven translocation across SecYEG is faster than translation elongation[5,63], which quickly resolves the cytosolic accumulation of the periplasmic region. The ribosome is then able to gain close approach to SecYEG to recover the ribosome-translocon junction, outcompeting SecA during this process. Subsequent translocation could be driven by other forces including PMF[27–29], PMF-driven pulling by SecD/F[31], folding of the periplasmic domains[64,65], or their binding by periplasmic chaperones[66,67]. Efficient continuous translocation prevents further accumulation of the periplasmic loop in the cytosol, thus alleviating the need for SecA during later stages of translocation. In support of this hypothesis, dissipation of the PMF leads to prolonged SecA association with large periplasmic loops of a subset of IMPs (Fig. 2.6A,B) which might reflect repeated attempts of SecA to reinitiate translocation in the absence of a driving force for continued translocation. Together, our findings reveal the role of SecA in resolving topological problems encountered by SecYEG during the cotranslational translocation of proteins with large periplasmic loops.

SecA also cotranslationally engages ~50% of the secretory proteins in two distinct modes depending on their signal sequence (Fig. 2.4). For secretory proteins harboring

highly hydrophobic signal sequences, SecA association occurs early, long before protein synthesis is finished. Their membrane association occurs following SecA engagement, suggesting that SecA can cotranslationally target and translocate these substrates (Fig. 2.4C,D,K). An example is DsbA, a periplasmic protein with a hydrophobic signal sequence known to drive cotranslational protein export[52]. DsbA was assumed to be an SRP substrate[52] but did not associate with SRP in a ribosome profiling study[7]. The early cotranslational association of DsbA with SecA (fig. S2.5B) resolves these inconsistencies and explains why DsbA export showed a stronger dependence on SecA than on SRP[52]. In addition, the rapid and stable folding of the mature domain of DsbA in the cytosol was suggested to necessitate the cotranslational mechanism of its export[52,68]. Here we found that this feature is enriched among SecA-first substrates, suggesting that signal sequences co-evolved with the biophysical properties of the nascent protein to optimize their export efficiency. Taken together, our results suggest that ~25% of secretory proteins in bacteria can use a SecA-mediated cotranslational export pathway.

In contrast, secretory proteins with weakly hydrophobic signal sequences associate with TF at a nascent chain length of ~100 amino acids and display a gradual rise in SecA enrichment and membrane association only after the dissociation of TF (Fig. 2.4E,F,K). We speculate that the translocation of these proteins is only temporally cotranslational but not mechanistically coupled to translation elongation[69,70], which is distinct from the SRP- and SecA-mediated cotranslational transport discussed earlier. However, disruption of SecA-ribosome interaction *in vivo* leads to accumulation of the precursor form of a TF-first substrate, MBP (Fig. 2.5D)[34], suggesting that although the mechanism of their translocation is analogous to a posttranslational pathway, the ribosome interaction of SecA facilitates their export.

The chaperone SecB engages nascent secretory proteins and is implicated in SecA-mediated posttranslational targeting[54,55]. Nevertheless, neither disruption of the SecB-binding domain in SecA nor deletion of the *secB* gene substantially affected the substrate pool of SecA (Fig. 2.5C) or the timing of cotranslational SecA engagement on secretory

proteins (Fig. 2.5A,B and fig. S2.7A,B), consistent with previous studies that only identified a small number of strictly SecB-dependent preproteins[17,57]. In addition, as we did not observe prolonged SecA binding after deletion of SecB or disruption of the SecA-SecB interaction (Fig. 2.5A,B and fig. S2.7A,B), our data do not support a previous model in which SecB releases SecA from the nascent chain[59]. Together, these results indicate that cotranslational SecA engagement on most secretory proteins is largely independent of SecB. On the other hand, SecA binding on long periplasmic loops of translocating IMPs is reduced upon the deletion of SecB or disruption of its binding with SecA, suggesting that SecB could facilitate the formation of high affinity RNC-SecYEG-SecA complexes during cotranslational IMP translocation.

TF is an abundant and conserved cotranslational chaperone that forms a molecular cradle for nascent polypeptides at the ribosome exit site and, together with DnaJ/K[71], ensures the proper folding of cytosolic proteins in bacteria. TF also associates with nascent secretory proteins, and its deletion leads to accelerated preprotein export and increased cotranslational targeting[3,33]. The mechanism behind these observations and the precise role of TF in protein secretion is not well understood. Here we found that TF deletion leads to premature cotranslational associations of SecA with nascent secretory proteins that otherwise bind TF first (Fig. 2.4J), including those that are devoid of cotranslational SecA binding in wildtype cells (fig. S2.6E-H). This suggests a competition between TF and SecA in RNC binding *in vivo*, consistent with their overlapping docking sites on the ribosome[34,35,72]. On the other hand, SecA engagement with the large periplasmic loops of IMPs is compromised in the absence of TF (fig. S2.6M,N). These observations suggest a role of TF as a master regulator that delays SecA engagement on secretory proteins with weakly hydrophobic signal sequences and thus enforces the posttranslational mode of their targeting. This prioritizes the limited pool of SecA for substrates that require a strictly cotranslational mode of translocation. TF is the only cotranslational protein biogenesis factor whose concentration is stoichiometric with the ribosome[73] and may therefore be particularly suited to serve this role. In addition, cotranslational protein translocation is limited by translation rate and thought to be much

slower than posttranslational translocation[5,63]. Considering the limited number of SecYEG in the cell[73], TF-induced enforcement of posttranslational translocation may also reduce preprotein residence time on SecYEG, which can be beneficial to cells.

Previous work showed that SRP strongly prefers IMPs with hydrophobic TMDs as the targeting signal as well as secretory proteins with highly hydrophobic signal sequences[7]. Our analyses here further show that secretory proteins are targeted via the SecA-first cotranslational targeting pathway or the TF-first posttranslational pathway based on the hydrophobicity of the signal sequence. These observations support a model in which non-cytosolic substrates are triaged into distinct targeting routes based on the hydrophobicity of the targeting signal. Additional features of the nascent chain, such as helical propensity and potential targeting signals in the mature domain[74], could also play important roles in substrate triage. Nevertheless, considering that only the signal sequence and the following ~50 amino acids are exposed at the time of TF/SecA engagement, the signal sequence may be one of the most important determinants for nascent protein triage.

On the other hand, overlap in the specificity of SecA with that of SRP and TF enables SecA to provide a backup targeting pathway when the other targeting factor or protein chaperone is deleted. The extensive early association of SecA on TF-first secretory proteins upon TF deletion (Fig. 2.4J) and the faster export observed in the absence of TF[32] provides a salient example of this redundancy. Another example is observed with the IMP RodZ. The SeRP profiles (fig. S2.4C) showed that SRP binds the nascent RodZ TMD as soon as it emerges from the ribosome, followed by SecA binding at the membrane after SRP dissociation, demonstrating SRP-dependent targeting and SecA-dependent translocation of RodZ in WT cells. However, RodZ biogenesis *in vivo* is weakly dependent on SRP and independent of SRP receptor[42], and SecA alone is sufficient to reconstitute RodZ insertion into the membrane[41], suggesting that SecA provides a backup targeting and translocation route for RodZ in the absence of SRP[75]. Taken together, these overlapping specificities enable the distinct targeting pathways to

form a robust network that minimizes protein biogenesis defects when one of the pathways is disrupted.

In summary, our work reveals multiple roles of SecA in co- and post-translational protein transport pathways (Fig. 2.7). Proteins with a highly hydrophobic N-terminal targeting signal, which include most IMPs and a few secretory proteins, are recognized and targeted to the plasma membrane by SRP upon emergence of their first TMD or signal sequence (Fig. 2.7, left). While the cotranslational insertion of TMDs and short periplasmic loops occurs independently of SecA, translation elongation appears insufficient to drive the translocation of large periplasmic loops across SecYEG, leading to the accumulation of these regions at the cytoplasmic surface of the translocating complex and disruption of the ribosome-SecYEG junction. SecA, which rapidly diffuses on the membrane and scans translocating RNCs, recognizes these topologically constrained translocating complexes, likely via a combination of interactions with the exposed periplasmic regions, the ribosome, and SecYEG. Using its ATPase cycle, SecA drives the translocation of the periplasmic regions across SecYEG. This restores the ribosome-SecYEG junction and allows continued translocation to be driven by PMF, translation elongation, and other forces. In contrast, secretory proteins harboring a weakly hydrophobic signal sequence cotranslationally associate with TF after ~100 amino acids are synthesized (Fig. 2.7, right). TF binding delays the engagement of SecA on these substrates until synthesis of the nascent polypeptide is close to complete, thus enforcing a posttranslational mode of their targeting and translocation to the plasma membrane. The other ~25% of secretory proteins with a more hydrophobic signal sequence cotranslationally engage SecA at a nascent chain length of ~100 amino acids, and likely are targeted and translocated by SecA in a manner analogous to the translocation of large periplasmic loops of IMPs (Fig. 2.7, middle). Our model highlights the hierarchical organization of protein transport in bacteria, in which multiple chaperones and targeting factors form a robust network of transport routes for nascent proteins harboring diverse targeting signals.
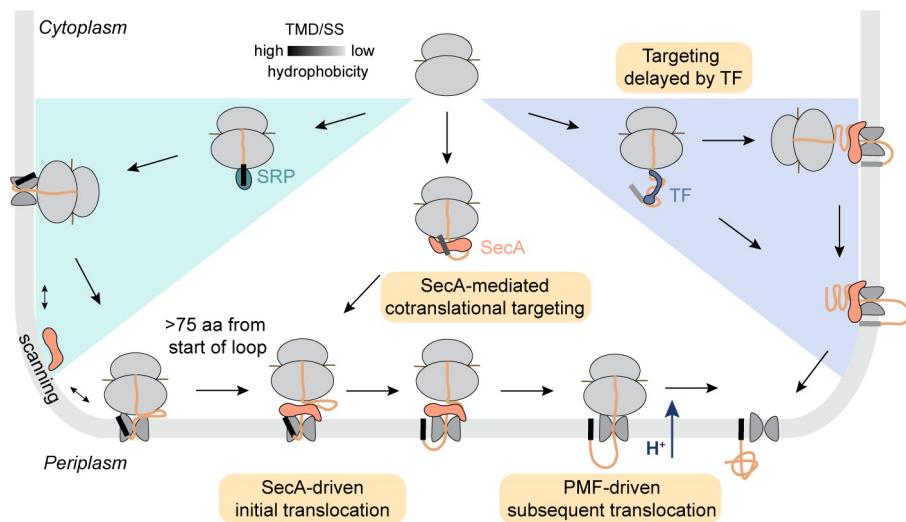
Fig. 2.7. Model of cotranslational SecA engagement during protein transport

Proteins with highly hydrophobic N-terminal signals, including most IMPs and a few secretory proteins, are recognized and targeted to the membrane by SRP. Large periplasmic loops on IMPs cannot be translocated across SecYEG, and thus accumulate at the cytoplasmic surface and disrupts the ribosome-SecYEG junction when >75 amino acids of the periplasmic loop is translated. SecA then gains access to these stalled translocating complexes, uses its ATPase cycle to drive the initial translocation of the loop across SecYEG and restores the ribosome-SecYEG junction. Subsequence translocation is driven by PMF or other forces. Secretory proteins with weakly hydrophobic signal sequences are recognized by TF at a nascent chain length of ~100 amino acids. TF delays their engagement by SecA until protein synthesis is close to complete, thus enforcing a mechanistically posttranslational mode of their export. The other ~25% of secretory proteins with more hydrophobic signal sequences engage with SecA at a nascent chain length of ~100 amino acids. They are possibly targeted to the membrane by SecA, and are translocated across SecYEG in a manner analogous to the translocation of large periplasmic loops on IMPs.

2.5 Materials and Methods

Strain construction

E. coli K-12 strain W3110 was used for this study. A thrombin-Avi tag (GLVPRGSGSGSGLNDIFEAQKIEWHE) was chromosomally fused to the C-terminus of SecA using CRISPR [76]. For gene deletions, a cassette containing kanamycin resistance marker was amplified from pKD4 plasmid and inserted into the target locus to replace the WT SecA gene [77]. For construction of a strain harboring OppC-43aa repeat, a silently mutated DNA fragment (AGTCAATTCGCGTATGATGACACGGATTGGGCCATGATGAGCTCTGCGCCAG ACATGGAAAGCGGCCATTATTTCGGCACCGATAGCAGCGGCCGTGATTTGCT GGTCCGTGTGGCCATCGGTGGCCGA) encoding the first periplasmic loop of OppC (bp 178-306) was synthesized and chromosomally inserted into OppC using CRISPR [76]. For construction of AcrB-Δ71aa, WT AcrB was cloned into a plasmid and AcrB 50-120 amino acids was deleted by amplifying the plasmid using oligonucleotides 5'GATCTCCGCCTCCTACGAAGTTCAGCAGCAAGGGGT3' and 5'CTTGCTGCTGAACTTCGTAGGAGGCGGAGATCGTTA3'. Genomic AcrB was then replaced with AcrB-Δ71aa using CRISPR [76]. All strains were confirmed by amplification of genomic DNA and sequencing.

Purification of SecA-RNCs for SeRP

400 mL of *E. coli* cells were grown in LB medium at 37 °C to an OD600 of 0.4 and harvested by rapid filtration through nitrocellulose membranes of 0.2 mm in pore size, followed by flash freezing in liquid nitrogen. For CCCP-treated cells, CCCP (Sigma) was added into the cell culture to a final concentration of 50 µM immediately before filtering such that the treatment time is approximately 2.5 min. Frozen cells were mixed with 1 mL frozen Lysis Buffer (50 mM HEPES pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM chloramphenicol, 1 mM PMSF, 50 U/ml DNaseI (recombinant DNaseI, Roche)) and lysed by mixer milling (2 min, 25 Hz, Retsch). To increase the yield of purified SecA-RNCs, 1.6 L of cells were cultured and lysed in four batches. Frozen lysed

cell powder was mixed, and half of the powder was added in batches to 4 mL of Lysis Buffer EDC (50 mM HEPES pH 4.5, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM chloramphenicol, 1 mM PMSF) supplemented with 20 mM EDC with constant stirring at room temperature. Once the cell powder was completely dissolved, EDC was added again. The other half of the frozen cell powder was then added as described above. The final concentration of EDC is 20 mM. The thawed lysate was stirred for an additional 5 min after the powder was completely dissolved. The crosslinking reaction was quenched with 20 mM glycine, 100 mM Tris pH 8.0, 4 mM $NaHCO_3$ and incubated for 5 min at room temperature with constant stirring. After quenching, the lysate was further incubated with 1% Triton X-100 for 5 min and the concentration of $MgCl_2$ and $CaCl_2$ was adjusted to 10 mM and 5 mM, respectively, to account for the volume increase. The RNA concentration was determined, and polysomes supplemented with 100 U/mL SUPERase*In (Ambion) were digested using MNase (3750 U/1 mg RNA) for 15 min at room temperature. The reaction was terminated by addition of 6 mM EGTA and chilling on ice. Unbroken cells were removed by centrifugation at 3,000 rpm for 6 min at 4 °C. Monosomes were purified by centrifugation through sucrose cushion (30% sucrose, 50 mM Tris pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, 1x protease inhibitors (Complete EDTA-free, Roche), 0.4% Triton X-100, 0.1% NP-40) in a TLA100.3 rotor at 80,000 rpm for 100 min at 4 °C. Pellets were washed once and resuspended in Wash Buffer (50 mM Tris pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, 1 mM PMSF, 0.4% Triton X-100, 0.1% NP-40).

20 μg of total RNA were removed from resuspended monosomes for ribosome profiling of the total translatome. To the rest, 100 μL Pierce Streptavidin Magnetic Beads (pre-washed for 3 times in Wash Buffer) was added per 3 mg RNA, and the suspension was rotated for 1 hr at 4 °C. Bead were washed 2x 10 min at 4 °C and 3x 5 min at room temperature in Wash Buffer. After the fifth wash, beads were re-equilibrated in Cleavage Buffer (50 mM Tris pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, 0.01% Triton X-100). Thrombin cleavage was performed by mixing the beads with 400

µL of Cleavage Buffer plus 8 U of thrombin (Cytiva) and incubating on a nutator for 30 min at room temperature. The eluate was used for subsequent RNA extraction.

Library preparation

Libraries are prepared as described in [78] with modifications. All steps were performed in non-stick, RNase free microfuge tubes (Ambion). RNA from total monosomes and SecA-bound monosomes was extracted using Direct-zol kit (Zymo) according to the manufacturer's instructions. 3 µg of RNA from each sample was mixed with 2x Novex TBE-Urea sample buffer (Invitrogen) and denatured for 2 min at 80 °C. Samples were loaded onto a home-made 15% polyacrylamide TBE-Urea gel and run for 50 min at 200V in 1x TBE. The gel was stained for 5 min with SYBR-gold (Invitrogen) and the ribosome footprints were isolated by excising the 15 nt - 45 nt region as indicated by the 10 bp ladder (Invotrogen) and a 45 nt oligo from the gel. The gel pieces were placed into 0.5 mL gel breaker tubes (IST Engineering), nested into a 1.5 mL tube and centrifuged for 5 min at 20,000xg. 500 µL of RNA extraction buffer (300 mM NaOAc pH 5.5, 1 mM EDTA pH 8.0) supplemented with 100 U/mL SUPERase*In was added to each sample, and the tubes were shaken overnight in a thermomixer at 1,400 rpm, 4 °C. Gel pieces were transferred to a Spin-X cellulose acetate column (Fisher) and centrifuged at 20,000xg for 5 min. The flow through was mixed with 1.5 µL Glycoblue and 600 µL isopropanol and incubated for 1 hr on dry ice to precipitate RNA. The samples were centrifuged for 30 min at 20,000xg, 4 °C and the pellets were washed with ice-cold 80% ethanal followed by resuspended in 4 µL of 10 mM Tris pH 8.0. For dephosphorylation, 3.5 µL of sample was denatured for 2 min at 80°C followed by quick cooling on ice, mixed with 0.5 µL 10x T4 polynucleotide kinase buffer (NEB), 0.5 µL T4 polynucleotide kinase (NEB) and 0.5 µL SUPERase*In, and incubated for 1 hr at 37 °C. For linker ligation, 3.5 µL 50% w/v PEG-8000, 0.5 µL 10x T4 RNA ligase 2 buffer (NEB), 0.5 µL 20 µM Universal miRNA cloning linker (NEB) and 0.5 µL T4 RNA ligase 2 truncated (NEB) were added into each sample and incubated for 3 hrs at 30 °C. To deplete unligated linker, 0.5 µL Yeast 5'-deadenylase (NEB) and 0.75 µL RecJ exonuclease (Epicentre) were added after linker ligation and the reaction was further incubated for 45

min at 30 °C. The ligation products was purified with Oligo and Concentration Kit (Zymo) and eluted in 10 μl of nuclease-free H_2O. For reverse transcription, the sample was denatured at 80 °C for 2 min and mixed with 4 μL 5x RT buffer (Invitrogen), 1 μL 10 mM dNTPs, 1 μL 0.1 M DTT, 1 μL SUPERase*In and 1 μL Superscript III RT (Invitrogen). Samples were incubated for 30 min at 50 °C followed by addition of 2.3 μL of 1 M NaOH and incubation for 20 min at 70 °C to hydrolyze RNA. Samples were loaded onto a home-made 10% polyacrylamide TBE-Urea gel and run for 60 min at 200 V in 1x TBE. The gel was stained for 5 min with SYBR-gold, the bands between 111 nt – 141 nt were excised, and nucleic acids were extracted as described earlier, except that DNA extraction buffer was used (300 mM NaCl, 1 mM EDTA pH 8.0, 10 mM Tris pH 8), and the DNA was resuspended in 15 μL of 10 mM Tris pH 8.0. For circularization, 2 μL 10x CircLigase buffer (Epicentre), 1 μL 1 mM ATP, 1 μL 50 mM $MnCl_2$ and 1 μL CircLigase (Epicentre) were added and incubated for 1hr at 60 °C. To inactivate the enzyme, the reaction was further incubated for 10min at 80°C. For PCR amplification, 4 μL of circularized DNA was mixed with 4 μL 10 μM barcoded primer, 49.4 μL nuclease-free water, 16 μL 5x HF buffer (NEB), 1.6 μL of 10 mM dNTPs, 4 μL forward primer and 1 μL Phusion polymerase (NEB), and the PCR reaction was run for 6-10 cycles. Samples were loaded onto a home-made 8% polyacrylamide native TBE gel and run for 1 hr at 120V in 1x TBE. The gel was stained for 5 min with SYBR-gold, and the bands at ~170 bp were excised. The DNA product are extracted as described earlier followed by resuspension in 10 μL of 10 mM Tris pH 8. The libraries were quantified by Qubit and sequenced on a Hiseq 2500.

Fractionation-coupled ribosome profiling

400 mL of E.coli cells were harvested and lysed as described earlier. Frozen cell powder was thawed in 1 mL of Lysis Buffer (50 mM HEPES pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM chloramphenicol, 1 mM PMSF) at room temperature and immediately centrifuged at 3,000 rpm for 6 min at 4 °C to remove unbroken cells. After quantifying RNA concentration, MNase digestion was performed as described earlier followed by centrifugation in a TLA120.2 rotor at 30,000 rpm for 20 min at 4 °C. The

supernatant was saved as cytosolic fraction. The pellet was washed and resuspended with a dounce homogenizer in Resuspension Buffer (50 mM HEPES pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM chloramphenicol, 1 mM PMSF, 1% Triton X-100, 6 mM EGTA). Insoluble material was removed by centrifugation in a TLA120.2 rotor at 30,000 rpm for 20 min at 4 °C, and the supernatant was saved as the membrane faction. Monosomes from both the cytosolic and membrane factions were purified, RNA was extracted, and sequence libraries were generated and sequenced as described earlier.

Western blotting

For affinity purification analysis, total monosomes were collected, and affinity purification was performed as described earlier except that the cleavage step was omitted. To concentrate the SecA-RNCs, the beads were mixed with a small volume of 5x SDS sample buffer after washing and boiled for 5 min to elute SecA-RNCs. Total monosomes were mixed with 5x SDS sample buffer and boiled for 5 min. Samples were resolved on 12.5% or 15% Tris-glycine gels, transferred to nitrocellulose membrane (Bio-Rad), and probed with rabbit anti-SecA antibody (a gift from Tom A. Rapoport, 1:1000 dilution) and mouse anti-S13 antibody (DSHB, 1:3000 dilution). Primary antibodies were incubated with IRDye 800CW Goat anti-Mouse and Goat anti-Rabbit secondary antibodies (LI-COR, 1:15000 dilution) for detection. The membranes were scanned by Odyssey Imager (LI-COR) and the images were processed using ImageJ.

For cell fractionation analysis, cell fractionation was performed as described earlier. Total lysate, supernatant and resuspended pellet were collected and resolved on 12.5% or 15% Tris-glycine gels. YidC and DnaK were detected by western blot using rabbit anti-YidC antibody (a gift from Ross E. Dalbey, 1:5000 dilution) and mouse anti-DnaK antibody (Abcam, 1:3000 dilution). Purified 70S ribosomes were subjected to centrifugation at different speeds indicated in fig. S2.2A. Ribosomal proteins were detected by silver staining.

Purification of MNase

MNase purification was performed following the protocol described in [2]. Briefly, *E. coli* cells overexpressing mature His$_6$-tagged MNase were lysed in lysis buffer (50 mM Tris pH 7.5, 25 mM NaCl, 1 mM CaCl2, 5 mM imidazole pH 8.0, 1 mM PMSF, 5% glycerol, 1x Halt protease inhibitor) by French press followed by centrifugation at 31,000xg for 30 min in JA20 rotor. Clarified lysates were loaded onto Ni-sepharose resin column equilibrated in Ni buffer (50 mM Tris pH 7.5, 25 mM NaCl, 1 mM CaCl$_2$, 5 mM imidazole pH 8.0, 5% glycerol). The resin was extensively washed using washing buffer (50 mM Tris pH 7.5, 250 mM NaCl, 10 mM imidazole pH 8.0, 5% glycerol) with a final wash in Ni buffer to adjust salt concentration. The protein was eluted using elution buffer (50 mM Tris pH 7.5, 25 mM NaCl, 250 mM imidazole pH 8.0, 5% glycerol) and dialyzed overnight against dialysis buffer (50 mM Tris pH 7.5, 25 mM NaCl, 1 mM EDTA, 5% glycerol). His6-tagged MNase was concentrated to ~14 mg/mL and the MNase activity was determined by performing the MNase activity assay as previously described [2]. Aliquots were stored at -80°C.

Data analysis

Adaptor sequences were trimmed from sequencing reads using Cutadapt. Reads were mapped to bacterial genome using Bowtie after discarding the reads mapping to ribosomal RNAs (Table S2.1). The *E.coli* W3110 reference genome assembly (ASM1024v1) was downloaded from EnsemblBacteria (https://bacteria.ensembl.org). Ribosome density was assigned to 14-nt upstream of the 3'-end of reads using reads with size range 15–45 nt as described elsewhere [79] to reach single-codon resolution. Nucleotide reads at each codon were then summed and used for all additional analyses.

Gene-level enrichment

For each gene, the sum of raw reads and RPM-normalized reads at each codon, excluding the first five and last five codons, were calculated for both translatome and SecA interactome. Only the genes with greater than 100 reads in both biological replicates of translatome and SecA interactome were included. SecA enrichment on each gene was

calculated as the ratio of RPM-normalized reads from SecA interactome to that from translatome.

Positional enrichment

For each gene, the enrichment at each codon was calculated as the ratio of the normalized reads over a window of 7 residues in the SecA interactome over that in the translatome. For the heatmaps in Fig. 2.2D and Fig. 2.4A, log2 enrichment was calculated at each codon. The localization score at each codon was calculated as the ratio over a window of 11 residues of reads in membrane fraction and that in soluble fraction to normalize for local variations in translation speed.

For metagene analyses of SecA enrichment, the first five and last five codons were excluded and only the genes that had an average reads per codon > 0.5 in both translatome and SecA interactome were used. Reads at each codon are first smoothed using a 5 residue rolling average and genes were normalized to their expression level by dividing the reads at each codon by the average reads per codon of the respective gene. ORFs from specific subsets (e.g., Sec substrates) are then aligned to the start codon or to the N-terminus of initial TMD/signal sequence as indicated and the mean of normalized reads and bootstrapped 95% CI were calculated at each position. For metagene analyses of ribosome membrane association, localization score at each codon is divided by the mean localization score of the respective gene. ORFs are aligned as indicated and the median of normalized localization scores and bootstrapped 95% CI were calculated at each position.

Loop analysis

The length of each periplasmic loop and cytoplasmic loop was calculated based on the predicted topology of each inner membrane protein. Periplasmic or cytoplasmic loops longer than 100 amino acids are aligned to their N-terminus, except for the loops N-terminal to the first TMD, which usually are not targeted to the membrane during their synthesis. After excluding the loops from the genes with a low read coverage (average

reads per codon < 0.5), we continued with 122 large periplasmic loops and 95 large cytoplasmic loops. For both replicates of translatome and SecA interactome, reads at each codon on the loops are smoothed using a 5 residue rolling average and divided by the average reads per codon of the gene harboring the loop, to normalize for the expression level of the gene. The mean of normalized reads and bootstrapped 95% CI were calculated at each position. When comparing the SecA enrichment after the emergence of a large periplasmic loop between CCCP-treated and untreated cells (Fig. 2.6A), we included the residues downstream of the periplasmic loops in the metagene analyses of both cells because of the observed prolonged SecA binding after CCCP treatment, by aligning the genes to the N-terminus of its predicted periplasmic loops and calculating the mean of normalized reads and bootstrapped 95% CI at each position.

Peak detection

To identify SecA binding peaks, we developed an algorithm to scan for the stretches of residues with a high SecA enrichment. We first excluded the genes with a low read coverage or a low correlation between replicates by applying three thresholds. Only the genes that passed the following thresholds after excluding the first five and last five codons were considered for further peak detection analysis. First, the sum of reads in both replicates of translatome and SecA interactome > 100. Second, the average reads per codon > 0.5 in both replicates of translatome and SecA interactome. Third, the Pearsons's correlation coefficient (Pearson's r) between replicates > 0.5 in both translatome and SecA interactome.

We proceeded to peak detection analysis with genes fulfilling all these requirements. The ratio over a window of 5 residues of normalized reads in SecA interactome and translatome for both replicates was calculated. We defined the SecA binding peaks as the regions that met the following criteria: (1) SecA enrichment >= 1.7-fold for at least 12 consecutive codons in both replicates. (2) The overlap of the peaks from two replicates >= 6 codons. (3) The position of the peak > 30 codons to avoid the detection of anomalous peaks caused by the known ribosome profiling technicalities and because we

found SecA binding requires the emergence of the nascent chain. When identifying the strong SecA binding peaks (Fig. 2.3G and fig. S2.7A), we increased the threshold of SecA enrichment from 1.7-fold to 3.5-fold to avoid the detection of SecA binding peaks induced by SecA scanning on the membrane.

WebLogo analysis

WebLogo analyses to compare the amino acid compositions of signal sequences of SecA-first substrates, TF-first substrates and non-SecA substrates were performed using the WebLogo3 (http://weblogo.threeplusone.com)38, with the signal sequences aligned to the second amino acid from their N-terminus.

Gene categorization and protein topology predictions

Genes were categorized based on their GO (gene ontology) annotations taken from Ecocyc71. Secretory proteins are composed of periplasmic proteins, outer membrane proteins, lipoproteins and extracellular proteins. The topology of inner membrane proteins was predicted by TOPCONS72 and TMHMM73 unless it has been experimentally verified and deposited in Uniprot74. The signal sequences of secretory proteins were predicted by SignalP75.

Hydrophobicity determination

The hydrophobicity of signal sequences was assessed by calculating the average Kyte-Doolittle hydrophobicity with a 10 amino acids rolling window. The Kyte-Doolittle value of the most hydrophobic window was recorded as the hydrophobicity of the signal sequence.

Contact order calculation

The predicted structures of SecA-first and TF-first substrates were downloaded from AlphaFold Protein Structure Database (https:// alphafold.ebi.ac.uk/)76,77. The absolute contact order for each protein is calculated based on the structure, as previously described43.

Quantification and statistical analysis

All analysis was performed in python. Statistical significance in comparing the distributions of SecA enrichment (Fig. 2.3E) and hydrophobicity (Fig. 2.4K and fig. S2.6J) was determined using Wilcoxon rank-sum tests. None of the experiments involved blinding or randomization. The number of independent biological replicates used for an experiment and *p*-vales are indicated in the figure legends.

2.6 Acknowledgements

Data and materials availability: The accession number for the data reported in this paper is GSE185572. The protein structures used to calculate absolute contact order were downloaded from AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/).

2.7 Supplementary figures
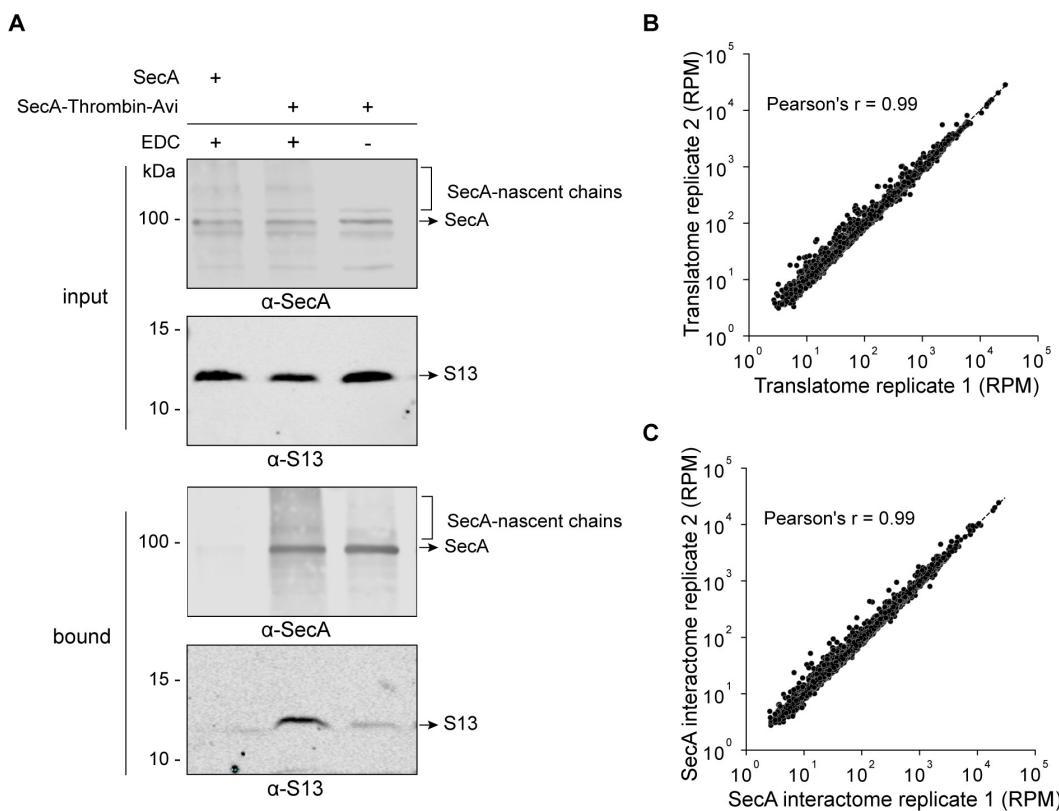
fig. S2.1. Selective ribosome profiling of SecA.

(**A**) Affinity purification of SecA-RNC complexes analyzed by western blot using antibodies against SecA and the ribosomal protein S13. Representative results from two independent experiments are shown. Source data are provided in the Source Data file. (**B, C**) Reproducibility of the translatome and SecA interactome data sets from two biological replicates.
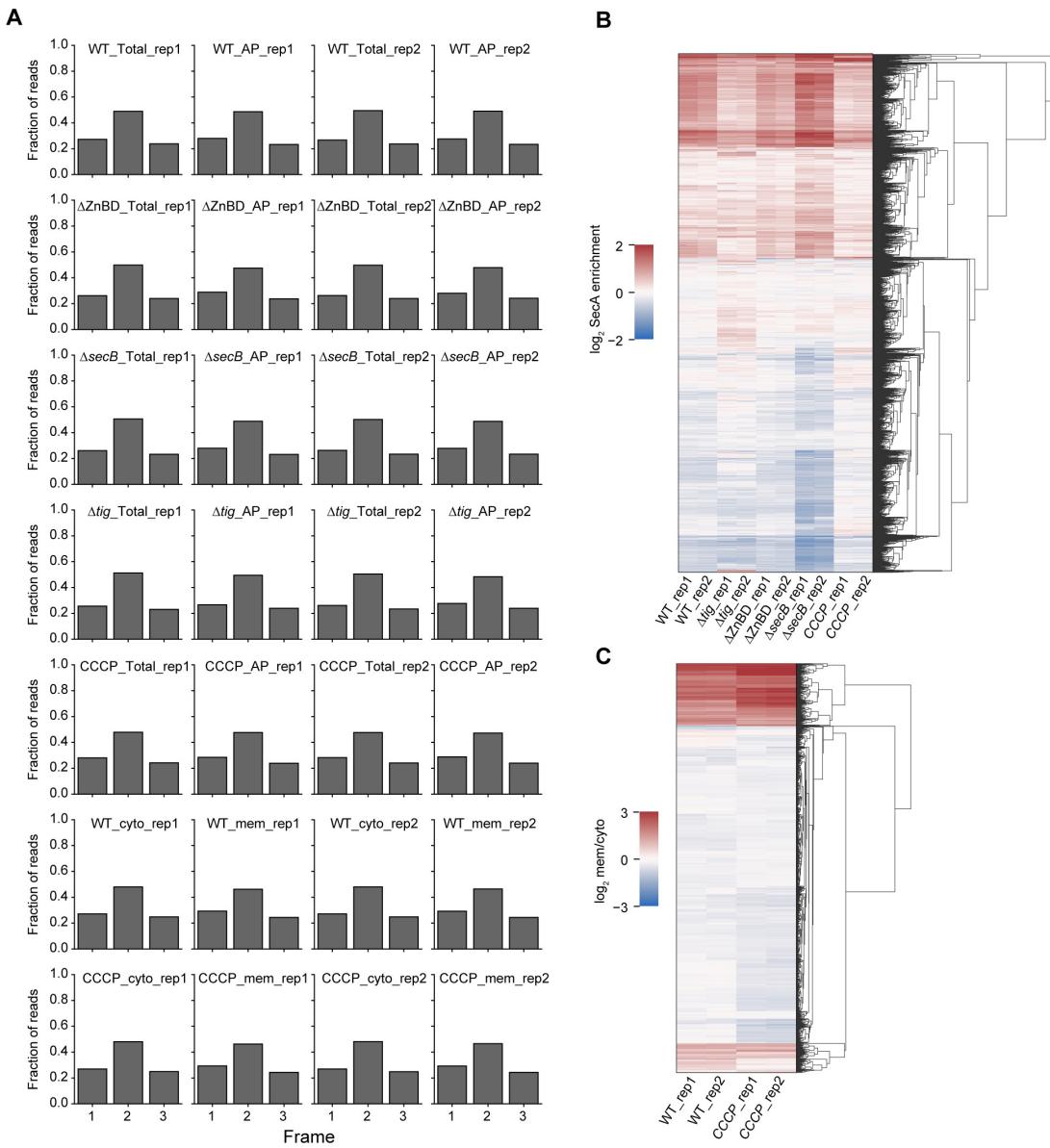
fig. S2.2. Comparison of all datasets generated in this study.

(**A**) The fraction of reads with P-site at three possible frames. (**B**) Hierarchical clustering of gene-level enrichment for all selective ribosome profiling experiments. (**C**) Hierarchical clustering of gene-level enrichment for all fractionation-coupled ribosome profiling experiments.
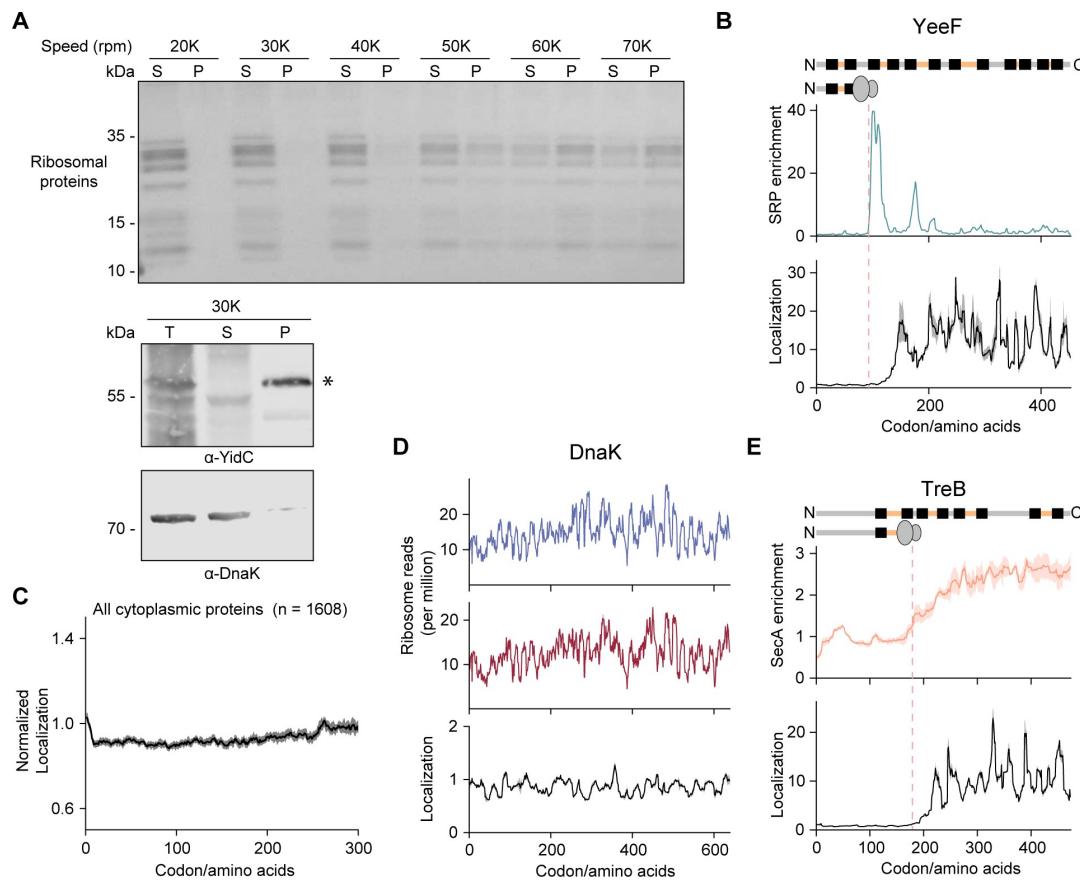
fig. S2.3. Fractionation-coupled ribosome profiling.

(**A**) Upper panel, centrifugation of purified 70S ribosomes using TLA120.2 rotor at different speeds analyzed by silver stain. The purified 70S ribosomes pelleted at centrifugation speeds above 40K rpm. To avoid the pelleting of ribosomes in cytosolic fraction during cell fractionation, we used 30K rpm, 20 min. S, supernatant; P, pellet. Lower panel, the centrifugation condition we used in fractionation experiment is sufficient to separate cytosolic and membrane fractions as analyzed by western blot using antibodies against YidC (an IMP) and DnaK (a cytosolic protein). The asterisk denotes the YidC protein. T, total lysate; S, supernatant; P, pellet. Representative results from two independent experiments are shown. Source data are provided in the Source Data file. (**B**) The SRP interaction profile and ribosome localization profile of YeeF, an IMP whose first TMD was skipped by SRP[4]. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**C**) Metagene ribosome localization profile of all cytoplasmic proteins aligned to the start codon. Solid lines show the mean values and shaded

areas show the 95% CI. (**D**) Representative ribosome localization profile of a cytosolic protein. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**E**) Representative SecA interaction profile and ribosome localization profile of an IMP. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates.
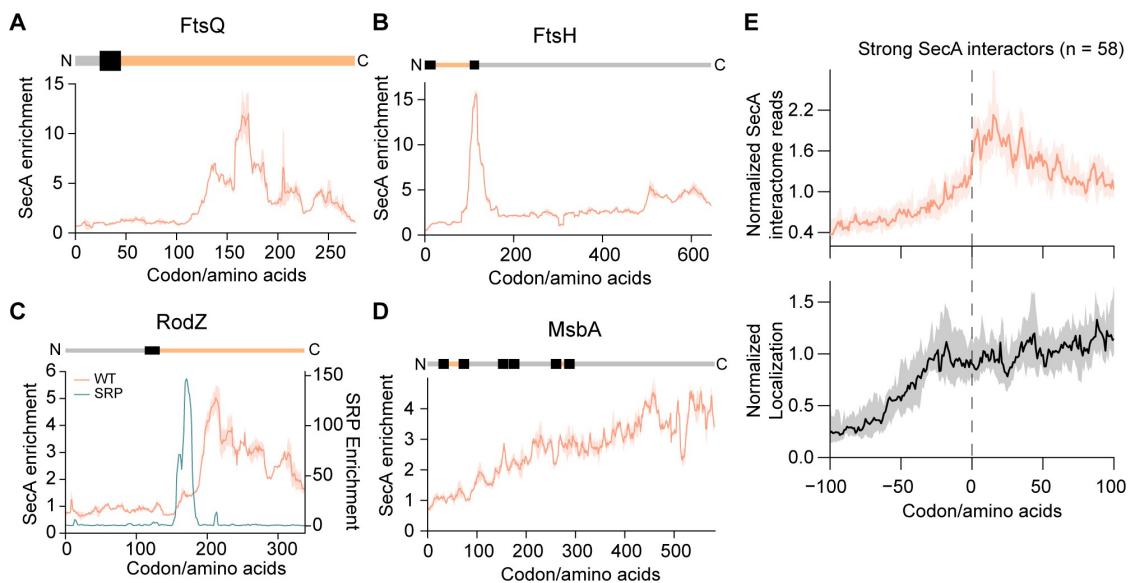
fig. S2.4. SecA interacts with IMPs containing large periplasmic loops.

(**A-C**) Representative SecA interaction profiles of IMPs with large periplasmic loops. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D**) SecA interaction profile of a detected SecA strong interactor lacking a periplasmic loop. Its detection is likely due to a higher baseline of SecA enrichment caused by the scanning mode. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**E**) Metagene SecA interactome profile and ribosome localization profile of 58 SecA strong interactors aligned to the onset of SecA binding peaks (position 0). Solid lines show the mean values and shaded areas show the 95% CI.
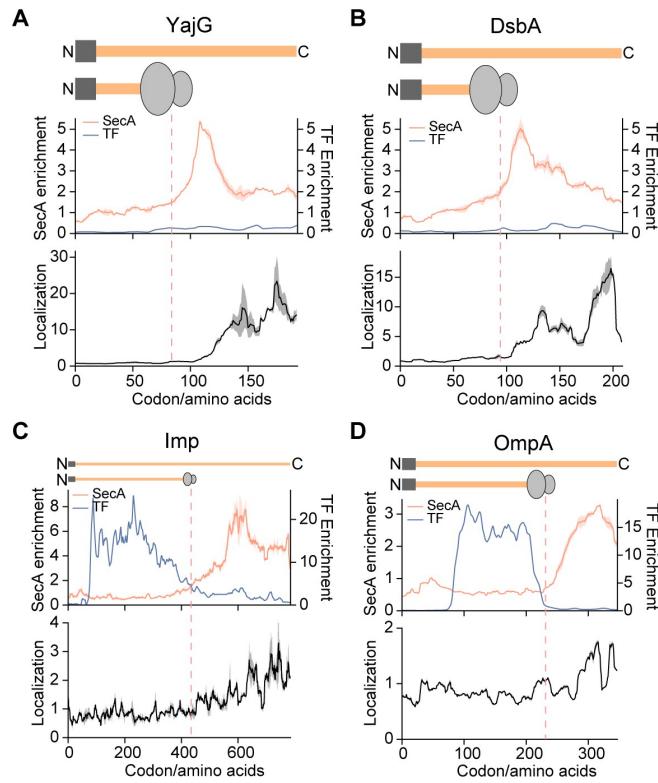
fig. S2.5. Temporal separation of SecA and TF binding on ribosomes translating secretory proteins.

(**A, B**) Representative SecA and TF interaction profiles, and ribosome localization profiles of SecA-first substrates. Protein topology is shown above and colored as in Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**C, D**) Representative SecA and TF interaction profiles, and ribosome localization profiles of TF-first substrates. Protein topology is shown above and colored as in Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates.
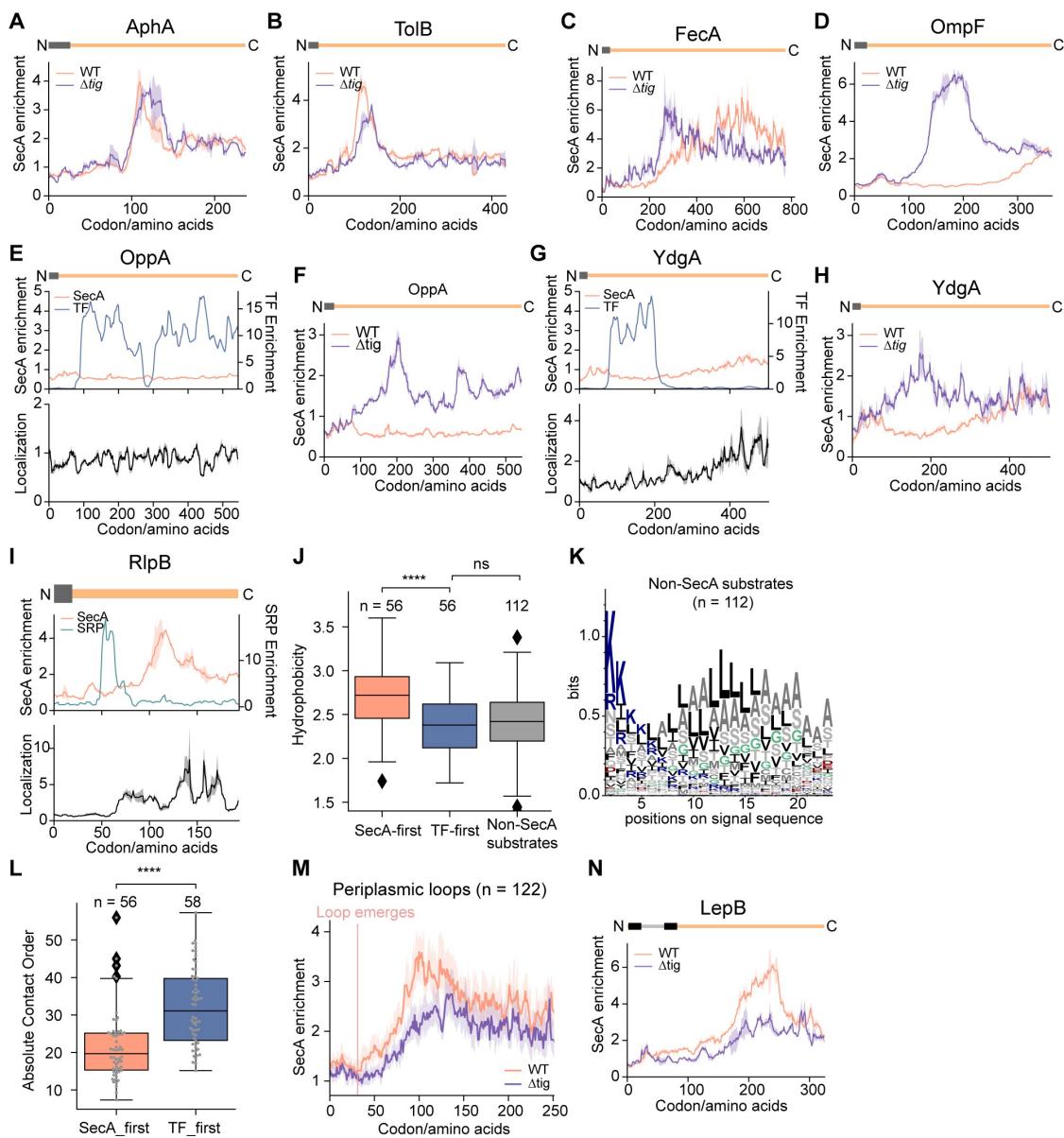
fig. S2.6. TF deletion leads to earlier SecA engagement on nascent secretory proteins and decreased SecA binding on IMPs.

(**A-H**) Representative SecA interaction profiles of SecA-first substrates (**A, B**), TF-first SecA substrates (**C, D**), and non-SecA substrates (**E-H**). Protein topology is shown above and colored as in Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**I**) Representative SecA and SRP interaction profile, and ribosome localization profile of a SecA-first substrate that recruit SRP cotranslationally[4]. Protein topology is shown above and colored as in Fig. 2.4C. Solid lines show the mean values and

shaded areas show the differences between two independent biological replicates. (**J**) Hydrophobicity of signal sequences of SecA-first, TF-first, and non-SecA substrates. The centre line represents the median, the bounds of box represent the upper and lower quartiles, and the whiskers indicate 1.5x the interquartile range. SecA-first vs. TF-first: $P = 3.716e^{-6}$; TF-first vs. non-SecA substrates: $P = 4.078e^{-1}$, Wilcoxon rank-sum test (**K**) WebLogo representations of the amino acid compositions of signal sequences of non-SecA substrates aligned to the second amino acid. (**L**) Absolute contact order calculated for SecA-first and TF-first substrates based on the AlphaFold-predicted structures of each protein. The structure of FdoG is not deposited in the database and is excluded from the analysis. The centre line represents the median, the bounds of box represent the upper and lower quartiles, and the whiskers indicate 1.5x the interquartile range. $P = 5.042e^{-8}$, two-sided Wilcoxon rank-sum test. Source data are provided in the Source Data file. (**M**) Metagene SecA enrichment of large periplasmic loops aligned to their N-terminus. The median values of SecA enrichment in WT strain and in Δ*tig* strain are compared. Shaded areas show the 95% CI. (**N**) Representative SecA interaction profile of an IMP with a large periplasmic loop on which SecA binding is compromised upon TF deletion. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates.
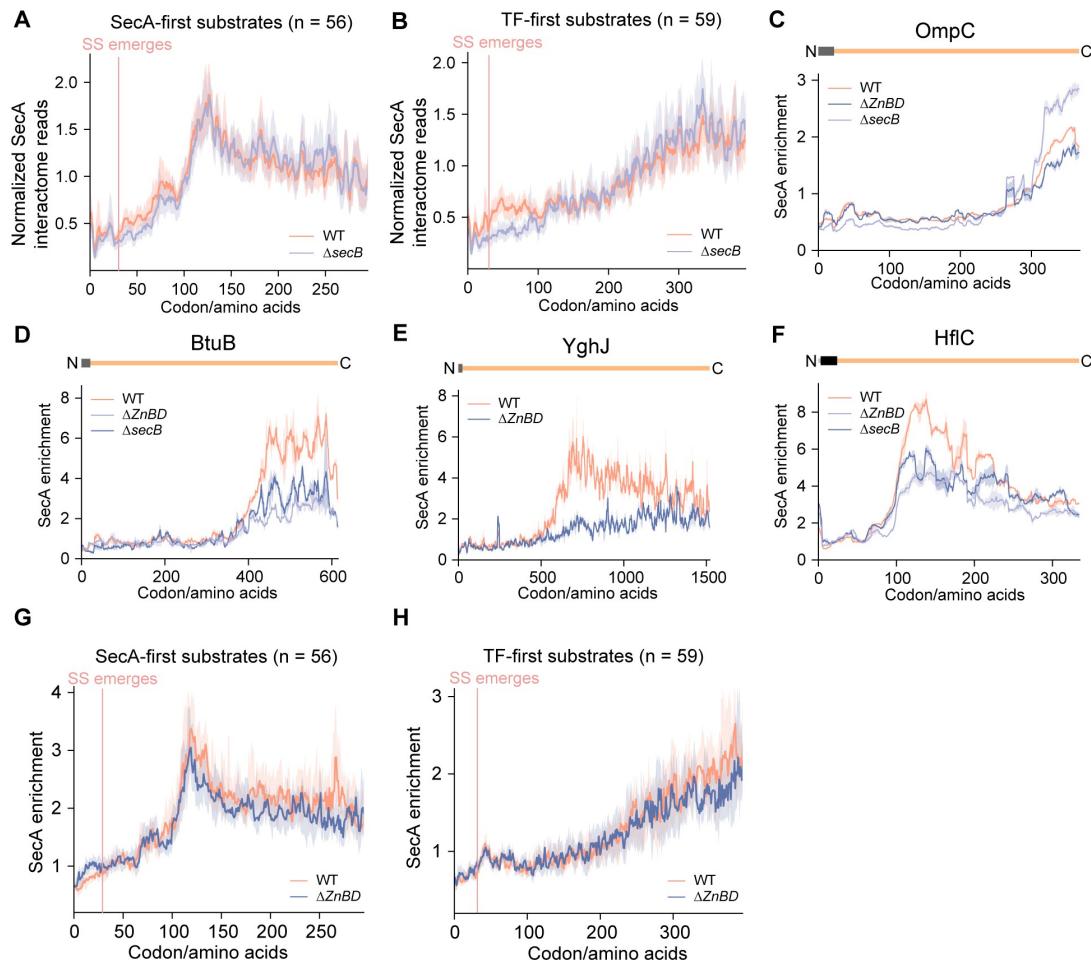
fig. S2.7. Role of SecB in cotranslational SecA interactions.

(**A, B**) Metagene SecA interactome profile of SecA-first (**A**) and TF-first (**B**) substrates aligned to the N-terminus of signal sequences. Solid lines show the mean values and shaded areas show the 95% CI. (**C**) Representative SecA interaction profile of a secretory protein on which the SecA binding peak showed modest changes upon SecB deletion. Protein topology is shown above and colored as in Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D, E**) Representative SecA interaction profiles of secretory proteins on which SecA binding peaks are severely compromised by the SecB deletion or the SecAΔZnBD mutation. The SecA interaction profile for YghJ from the Δ*secB* strain is noisy and therefore not shown in (**E**). Protein topology is shown above and colored as in Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**F**) Representative SecA interaction profile of an IMP on

which SecA binding is compromised. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**G, H**) Metagene SecA enrichment of SecA-first (**G**) and TF-first (**H**) substrates aligned to the N-terminus of signal sequences. The median values of SecA enrichment in WT strain and in Δ*ZnBD* strain are compared. Shaded areas show the 95% CI.
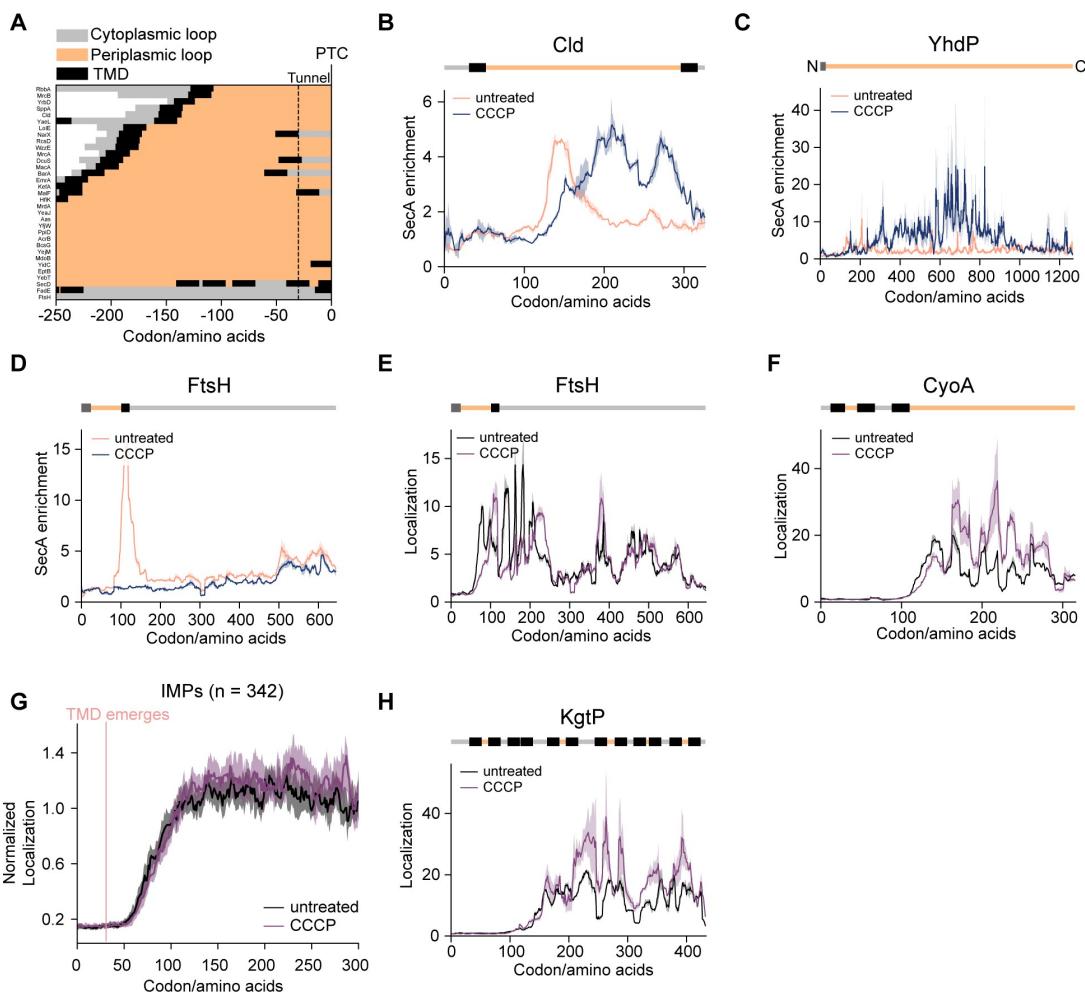
fig. S2.8. Loss of PMF leads to dysregulated SecA binding.

(**A**) Topology of translated nascent chains of strong SecA interactors, aligned to their C-terminus (position 0, at PTC) at the onset of SecA binding. The residues to the left of the dashed line are exposed outside the ribosomal tunnel exit. (**B, C**) Representative SecA interaction profile of an IMP with a large periplasmic loop (**B**) and a secretory protein (**C**) that showed persistent SecA association upon CCCP treatment. Protein topology is shown above and colored as in Fig. 2.2B and Fig. 2.4C. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**D**) Representative SecA interaction profile of an IMP with a large periplasmic loop that showed significantly reduced SecA association upon CCCP treatment. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological

replicates. (**E, F**) Representative ribosome localization profiles of IMPs with a large periplasmic loop that showed significantly reduced SecA association upon CCCP treatment. Corresponding SecA interaction profiles shown in Fig. 2.6D and fig. S2.8D. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates. (**G**) Metagene ribosome localization profile of IMPs with no periplasmic loops larger than 45 amino acids, aligned to the N-terminus of their first TMD, in untreated and CCCP-treated cells. Solid lines show the mean values and shaded areas show the 95% CI. (**H**) Representative ribosome localization profiles of an IMP without a large periplasmic loop in untreated and CCCP-treated cells. Corresponding SecA interaction profile shown in Fig. 2.6F. Protein topology is shown above and colored as in Fig. 2.2B. Solid lines show the mean values and shaded areas show the differences between two independent biological replicates.

Table S2.1 Coverage of datasets generated in this study

Sample name	Uniquely aligned reads	Reads aligned to CDS
WT SecA interactome, rep1	34804789	30825565
WT SecA interactome, rep2	21788144	18970768
WT total translatome, rep1	38264052	32904519
WT total translatome, rep2	21240554	18188618
Δ*tig* SecA interactome, rep1	18356171	16289432
Δ*tig* SecA interactome, rep2	13406407	11886252
Δ*tig* total translatome, rep1	16774681	14282487
Δ*tig* total translatome, rep2	18691996	15850378
secAΔZnBD SecA interactome, rep1	31926390	28064636
secAΔZnBD SecA interactome, rep2	19381780	16553781
secAΔZnBD total translatome, rep1	42745477	36632557
secAΔZnBD total translatome, rep2	19980672	17155952
Δ*secB* SecA interactome, rep1	13629065	11996830
Δ*secB* SecA interactome, rep2	12963717	11397043
Δ*secB* total translatome, rep1	15436331	13228815
Δ*secB* total translatome, rep2	13572397	11655639
WT SecA interactome, rep1, CCCP	9399204	8240630
WT SecA interactome, rep2, CCCP	15525893	13483727
WT total translatome, rep1, CCCP	21846422	18660959
WT total translatome, rep2, CCCP	11769273	10012956
WT cytosolic monosomes, rep1	23024694	19551026
WT cytosolic monosomes, rep2	11718022	9951046
WT membrane monosomes, rep1	11362514	9916473
WT membrane monosomes, rep2	7295535	6361487
WT cytosolic monosomes, rep1, CCCP	17337810	15148035
WT cytosolic monosomes, rep2, CCCP	14996202	13067180
WT membrane monosomes, rep1, CCCP	25583593	23464722
WT membrane monosomes, rep2, CCCP	22731118	20803452

Table S2.2 Selective ribosome profiling data analysis

The table is available online at

Zhu, Z.*, Wang, S.* & Shan, S. (2022). "Ribosome profiling reveals multiple roles of SecA in cotranslational protein export". In: *Nature Communications.* **13**, 3393. doi: 10.1038/s41467-022-31061-5. (* equal contribution)

Chapter 3

TIMING AND SPECIFICITY OF COTRANSLATIONAL MITOCHONDRIAL PROTEIN IMPORT

3.1 Abstract

The biogenesis of nearly all mitochondrial proteins begins with translation on cytosolic ribosomes. How these proteins are subsequently delivered to mitochondria remains poorly understood. Here, we comprehensively investigated the coupling of mitochondrial protein translation and import using selective ribosome profiling in human cells. Cotranslational targeting requires an N-terminal presequence on the nascent protein and contributes to mRNA localization at the mitochondrial surface. This pathway is predominantly used by large, multidomain and topologically complex proteins, whose import efficiency is enhanced when targeted cotranslationally. In contrast to protein targeting to the endoplasmic reticulum (ER), cotranslational mitochondrial import does not favor membrane proteins and initiates late during translation, specifically upon the exposure of a complex globular fold in the nascent protein. Our findings reveal a multi-layered protein sorting system that recognizes both the targeting signal and protein folding status during translation.

3.2 Introduction

Mitochondria are eukaryotic organelles that play critical roles in various cellular processes, including ATP synthesis, calcium homeostasis and lipid metabolism[19]. Fundamental to these roles is the proper localization of mitochondrial proteins. Over 99% of mitochondrial proteins are encoded by the nuclear genome, synthesized in the cytosol, and must be imported into mitochondria. This dynamic interplay between the cytosol and mitochondria necessitates diverse and sophisticated protein targeting and translocation mechanisms[20,21]. Almost all mitochondrial precursor proteins initiate their import at the

translocase of the outer membrane (TOM) complex and are subsequently sorted into distinct mitochondrial subcompartments via different pathways[80–82]. In contrast to the well-studied protein translocation machineries in mitochondria, targeting steps in the cytosol that deliver mitochondrial proteins to the TOM complex remain poorly understood[21,22]. While the classic view is that protein targeting to mitochondria occurs post-translationally, recent studies revealed the presence of mitochondrially localized mRNA[23,24], ribosomes [25], and translation [26], suggesting that targeting may also occur cotranslationally. However, direct evidence supporting cotranslational protein import into mitochondria is missing, and the timing, physiological roles, and underlying principles of this process remain unclear. To address these questions and fill the gap between localized translation and protein import, we used selective ribosome profiling[3] (SeRP) to systematically investigate cotranslational protein targeting to mitochondria in human cells at near-codon resolution.

3.3 Results

3.3.1 The N-terminal mitochondrial targeting sequence (MTS) directs cotranslational protein targeting to mitochondria

To unambiguously identify the proteins that undergo cotranslational import into mitochondria and exclude those that are translated on the mitochondrial surface but imported post-translationally, we reasoned that only the former would produce ribosome-nascent chain complexes (RNCs) that physically interact with the TOM complex. The presence of such RNCs was supported by the co-sedimentation of ribosomes with TOM40 and TOM22 (fig. S3.1A), subunits of the core TOM complex[83,84]. No ribosome association was observed with TOM70, a receptor on the outer mitochondrial membrane (OMM) that dynamically associates with the core TOM complex. Thus, isolating TOM-RNC complexes after nuclease treatment would allow us to identify the nascent proteins that are being cotranslationally imported into mitochondria (Fig. 3.1A). To enable purification of the TOM complex and the associated RNCs, we fused chromosomal *TOMM22* to a C-terminal Twin-strep tag[83], which is located in the mitochondrial

intermembrane space (IMS) and avoids disruption of TOM-RNC interactions. Following affinity purification, the ribosome-protected mRNA fragments (i.e., ribosome footprints) from the total ribosome population and from TOM-bound RNCs were extracted and sequenced, generating information on the total and TOM-bound translatomes (Fig. 3.1A and fig. S3.1, B and C).

Gene-level analysis showed that a subset of mitochondrial proteins was specifically enriched in the TOM-bound translatome, whereas proteins with other subcellular localizations were almost completely excluded (Fig. 3.1B and fig. S3.1D). This selective enrichment indicates that we successfully captured a snapshot of nascent proteins that are being cotranslationally imported into mitochondria. The interaction profiles of individual genes revealed that TOM enrichment rose sharply during translation and remained steady until the end of the coding sequence (Fig. 3.1C), suggesting that mitochondrial proteins complete the targeting process and initiate import into mitochondria at a certain point during translation, following which they are committed to continuous translocation through the TOM complex (Fig. 3.1C).

To identify all proteins that are cotranslationally targeted to mitochondria, we developed a peak detection algorithm with stringent criteria to scan for regions with at least 7 codons above a threshold of 2.5-fold TOM enrichment. All mitochondrial proteins that exhibit gene-level enrichment above 2-fold were detected, supporting the robustness of this method (Fig. 3.1D). It additionally captured 38 proteins with reproducibly strong but transient TOM interactions. This analysis showed that, out of the 746 mitochondrial proteins that were confidently identified in our dataset, 137 (18.4%) could be targeted cotranslationally (Fig. 3.1D), which we thereafter refer to as cotranslational TOM substrates. Although these findings do not exclude the possibility that a fraction of these substrates could also be imported post-translationally, they provide definitive evidence that a cotranslational import pathway exists for mitochondrial proteins *in vivo*.

We next examined the relationship between cotranslational import and mRNA localization of mitochondrial proteins. A strong correlation was observed between the

enrichment of nascent proteins at the TOM complex in our study and the enrichment of mRNAs localized at mitochondrial surface previously detected by APEX-seq[23] (Fig. 3.1E). The proteins encoded by the mRNAs that associate with mitochondria in a translation-dependent manner were specifically enriched in our dataset (Fig. 3.1F and fig. S3.1E), suggesting that the cotranslational targeting process brings the associated mRNAs to the vicinity of mitochondria and tethers the mRNA on the mitochondrial surface through interactions between the nascent chain and the TOM complex.

Functional characterization of cotranslational TOM substrates revealed their diverse and critical roles within mitochondria. Particularly enriched functional classes include mitochondrial protein degradation pathways, tRNA synthetases, ABC transporters, and various metabolic processes (fig. S3.1, F and G). Mitochondrial nucleoid-associated proteins involved in the maintenance and expression of the mitochondrial genome are also overrepresented (fig. S3.1G). Intriguingly, DELE1 and PINK1, whose import deficiency could activate the integrated stress response and mitophagy, respectively[85–87], were identified as cotranslational TOM substrates (fig. S3.1H). The role of cotranslational import in mitochondrial quality control awaits to be tested. In contrast, components of mitochondrial ribosomes and oxidative phosphorylation (OXPHOS) complex, whose assembly requires coordinated cytosolic and mitochondrial translation, were encoded by mRNAs that associate with the OMM in a translation-independent fashion and were rarely cotranslationally imported (fig. S3.1G).

We next explored the specific signals that direct cotranslational import. The TOM complex mediates the import of four major classes of mitochondrial proteins with distinct targeting signals, including preproteins with a cleavable N-terminal MTS (also termed presequence), carrier proteins on the inner mitochondrial membrane (IMM), β-barrel proteins on the OMM, and cysteine-rich proteins in the IMS[81] (Fig. 3.1G). Virtually all the cotranslational TOM substrates contain an N-terminal MTS (Fig. 3.1H and fig. S3.1I). The majority of these proteins (90 out of 137) reside in the mitochondrial matrix and comprise 25.5% of the matrix proteome (Fig. 3.1I). Among the 41 cotranslational TOM substrates that are annotated as IMM proteins (Fig. 3.1I), more than half lack a transmembrane

domain (TMD) and are matrix-localized subunits of protein complexes on the IMM. The remainder of IMM proteins and the two IMS proteins that exhibit TOM enrichment (Fig. 3.1I) also harbor an N-terminal MTS, which is removed by the mitochondrial processing peptidase in the matrix. The hydrophobic stop-transfer signals on these proteins mediate their release from translocase of the inner membrane (TIM23) into the IMM or IMS. Only one OMM protein, GPAM, showed significant TOM enrichment (fig. S3.1J). The mRNA encoding GPAM was also found to localize to mitochondria in a translation-dependent manner[23]. GPAM does not contain a cleavable MTS, but has an N-terminal loop-helix region that serves as a mitochondrial targeting signal[88].

Taken together, SeRP of the TOM complex reveals that a cotranslational import pathway is used by nearly 20% of the mitochondrial proteome. These proteins are predominantly substrates of the presequence pathway (Fig. 3.1H), indicating that the N-terminal MTS directs their cotranslational import into mitochondria.
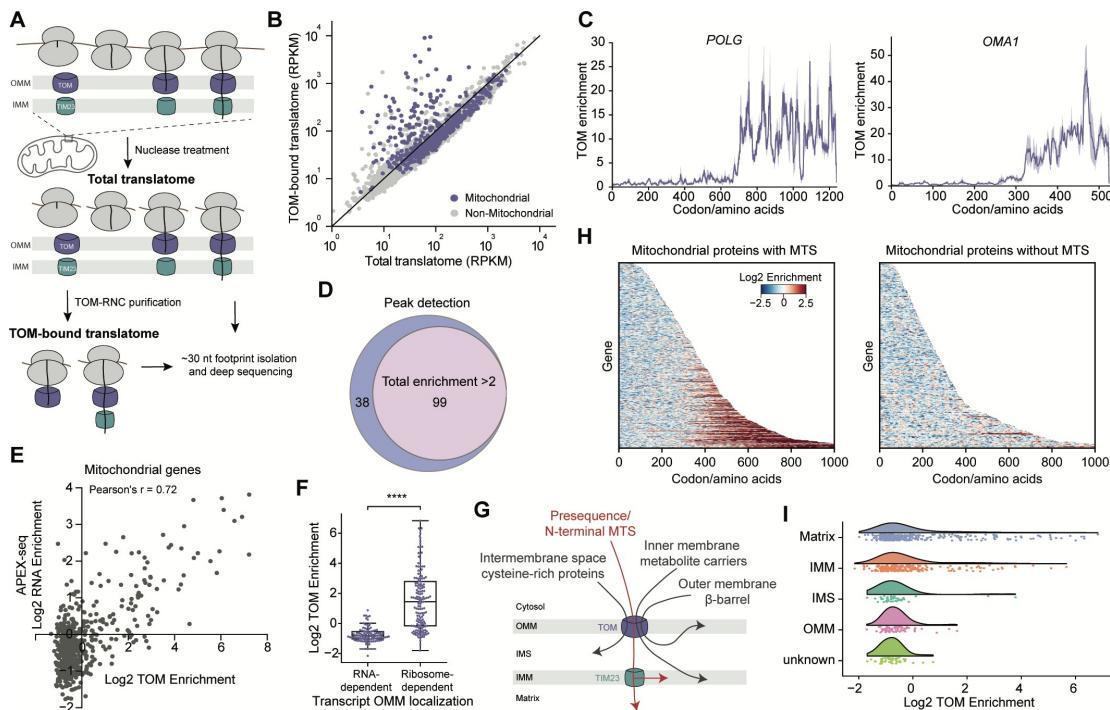
Fig. 3.1. SeRP of the TOM complex reveals cotranslationally imported mitochondrial proteins.

(**A**) Schematic of SeRP of the TOM complex. (**B**) Comparison of the total translatome and TOM-bound translatome footprint density of all detected genes. RPKM, reads per kilobase per million reads. (**C**) Representative TOM interaction profiles of mitochondrial proteins. TOM enrichment is calculated as the ratio of footprint density in the TOM-bound translatome over that in the total translatome. Solid lines show the mean values, and shaded areas show the range of data from two biological replicates. (**D**) Venn diagrams showing the overlap of cotranslational TOM substrates detected by the two methods. (**E**) Correlation between gene-level TOM enrichment from this study and the RNA enrichment from APEX-seq study [23] for all mitochondrial genes. (**F**) Comparison of the TOM enrichment between mitochondrial genes whose mRNAs are localized to mitochondria in the RNA- or ribosome-dependent manner. ****, p < 0.0001. (**G**) Schematic showing the four major biogenesis pathways for protein import through the TOM complex. ~60% of the mitochondrial proteome contains a cleavable N-terminal MTS and is further translocated across or inserted into the IMM by the TIM23 complex (the presequence pathway in *red*). Additional pathways mediate the insertion and folding of mitochondrial proteins without a cleavable MTS (*grey*), including cysteine-rich proteins in the IMS, carrier proteins at the IMM, and β-barrel proteins at the OMM. (**H**) Heatmap of log2 TOM enrichment at each codon for all detected mitochondrial proteins. Proteins are categorized by the presence of an N-terminal MTS and sorted by protein length. (**I**) Raincloud plot comparing the distribution of TOM enrichment on mitochondrial proteins based on the annotated localization in mitochondrial subcompartments.

3.3.2 Timing of cotranslational protein import into mitochondria

To further understand the coupling between the translation and translocation of mitochondrial proteins, we analyzed the timing of protein targeting to mitochondria. A metagene profile of the total and TOM-bound translatomes read density showed that, on average, engagement with the TOM complex began when the nascent protein is ~350 amino acids in length (Fig. 3.2A). This onset ranged from 200 to 800 amino acids for individual proteins (Fig. 3.1, C and H and Fig. 3.2B), with the majority of proteins initiating their TOM interaction at ~400 amino acids (Fig. 3.2B). Considering that the N-terminal MTS is only 10–100 amino acids long and that 50 amino acids are sufficient to span the translocases on both the mitochondrial outer and inner membranes[89,90], this observation indicates that import does not begin immediately upon exposure of the MTS

but rather, occurs relatively late during translation. SeRP in the presence of DSP (dithiobis(succinimidyl propionate)), which crosslinks TOM40 to nascent mitochondrial proteins, did not alter the onset of import (fig. S3.2), suggesting that the observed late initiation of mitochondrial import was not due to the loss of transient early interactions between translating ribosomes and the TOM complex.

Additional support for the late onset of mitochondrial import was obtained by analysis of the selective disome profiling (DiSP) data (Fig. 3.2C), which was previously performed to reveal the cotranslational assembly of nascent protein complexes (denoted as co-co interactions) in human cells[91]. We found widespread co-co interactions for nascent mitochondrial proteins in the published dataset (fig. S3.3, A and B). Nearly all cotranslational TOM substrates exhibited co-co interactions (Fig. 3.2D), which began at a nascent chain length of ~150 amino acids (fig. S3.3C). Notably, these co-co interactions decreased immediately before the onset of TOM engagement (Fig. 3.2, E and F). This was observed on individual proteins (Fig. 3.2E) as well as in a metagene analysis (Fig. 3.2F). In contrast, co-co interaction persisted until the end of translation for post-translationally targeted proteins (fig. S3.3D). Considering that the interaction of a nascent mitochondrial protein with another nascent chain is incompatible with its passage through the TOM and TIM23 complexes (Fig. 3.2C), which requires a largely unfolded polypeptide, the observed co-co interactions likely reflect transient interactions between nascent mitochondrial proteins in the cytosol that must be dissolved before their import. The chronological alignment of the two independent datasets strongly suggests that our data accurately captured the timing of cotranslational mitochondrial protein import.

These observations prompted us to examine a prevalent hypothesis that nascent polypeptide-associated complex (NAC) mediates cotranslational protein targeting to mitochondria (fig. S3.4A). NAC is an abundant ribosome-associated factor conserved across eukaryotic organisms[92] and plays a key role in preventing the mistargeting of mitochondrial proteins to the endoplasmic reticulum (ER)[93–95]. In yeast, NAC facilitates the recruitment of RNCs to mitochondria via interaction with Om14 on the OMM surface[96]. However, Om14 lacks a homolog in higher eukaryotes, raising questions about

the role of NAC in cotranslational protein targeting to mitochondria in human cells. In addition, NAC starts engaging ribosome with high affinity early during translation, before the nascent protein reaches 20 amino acids in length[97]. This contrasts with the late onset of protein import into mitochondria that we observed and argues against a direct involvement of NAC in this process.

To directly test the role of NAC in cotranslational protein targeting to mitochondria, we monitored the targeting process after NAC depletion. Because NAC is essential in higher eukaryotic cells and involved in numerous protein biogenesis events, we employed the plant-derived auxin-inducible degron (AID) system[98] to rapidly degrade NAC and thus minimize pleiotropic responses as well as cellular adaption. As NAC is a heterodimeric complex composed of α and β subunits[95,99], we engineered two cell lines for the acute depletion of NACα or NACβ. Following auxin addition, we observed a complete loss of NACα within 1.5 hours and NACβ within 4.5 hours (Fig. 3.2G).

We first used a split-GFP-based reporter system[100] to assess the effect of NAC depletion on mitochondrial protein import. C-terminally GFP11-tagged TRAP1, a cotranslational TOM substrate identified in our dataset, was co-expressed with matrix-targeted GFP1-10. Successful import of TRAP1-GFP11 generated fluorescence signal in mitochondria via GFP complementation (fig. S3.4B), whereas TRAP1-GFP11 without an MTS did not (fig. S3.4C). The depletion of neither NACα nor NACβ altered the fluorescence signal of TRAP1-GFP11 (fig. S3.4B), indicating that its import into mitochondria is not strictly dependent on NAC.

To more broadly examine the role of NAC in cotranslational protein targeting, we carried out SeRP of the TOM complex after acute depletion of NACβ, which mediates the ribosome association of the NAC complex[94]. In agreement with the results from the split-GFP-based reporter assay, acute depletion of NACβ did not significantly impact the association of nascent proteins with the TOM complex. We detected no significant decrease in gene-level enrichment for all cotranslational TOM substrates (Fig. 3.2H), nor changes in the timing of their import (Fig. 3.2, I and J). Collectively, these findings show

that NAC is not required for timely cotranslational protein targeting to the TOM complex in human cells.
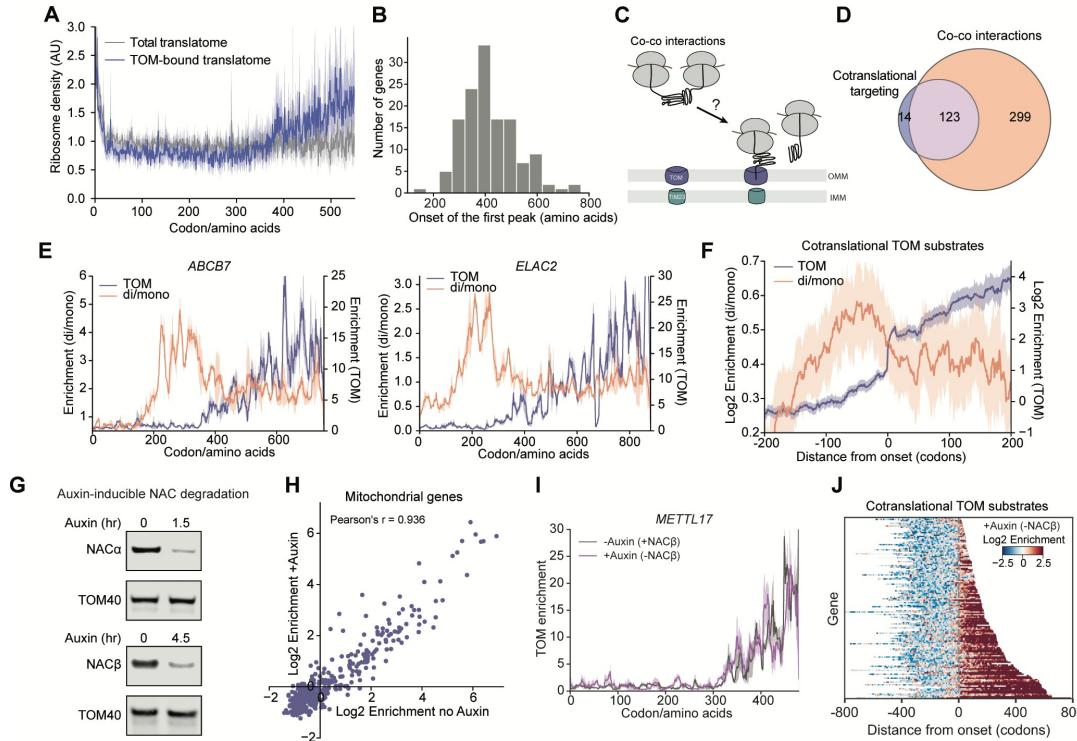
Fig. 3.2. Cotranslational protein targeting to mitochondria occurs late and independently of NAC.

(**A**) Metagene total translatome and TOM-bound translatome profiles of all mitochondrial genes aligned to the start codon. (**B**) Distribution of the onset of TOM interaction for cotranslationally targeted proteins. (**C**) Schematic showing the incompatibility of co-co interaction with cotranslational protein import into mitochondria. (**D**) Venn diagrams showing the overlap of cotranslational TOM substrates with mitochondrial proteins displaying co-co interactions. *MMUT, PRORP* are two cotranslational TOM substrates which were not detected in the co-co interaction dataset and were excluded from the comparison. (**E**) TOM interaction and co-co interaction profiles of representative cotranslational TOM substrates. (**F**) Metagene TOM enrichment and co-co interaction (disome/monosome) profiles of all cotranslational TOM substrates, aligned to the onset of TOM interaction peaks. (**G**) Western blot showing the depletion of NACα (upper) and NACβ (lower) in NACα-AID and NACβ-AID cell lines, respectively, upon auxin addition. (**H**) Correlation of gene-level TOM enrichment for all mitochondrial genes

without and with a 4.5 hr auxin treatment in the NACβ-AID cell line. (**I**) Representative TOM interaction profile of a mitochondrial gene before and after auxin addition (4.5 hr) in NACβ-AID cells. (**J**) Heatmap of log2 TOM enrichment at each codon for cotranslational TOM substrates after a 4.5 hr auxin treatment in the NACβ-AID cell line. Proteins are aligned to the onset of TOM interaction detected under standard conditions and are sorted by the distance from the onset to the stop codon. In (A) and (F), solid lines show the mean values, and shaded areas show the 95% CI. In (E) and (I), solid lines show the mean values, and shaded areas show the range of data from two biological replicates.

3.3.3 Cotranslational mitochondrial import initiates upon the exposure of a complex protein fold on the nascent chain

We next asked what dictates the late onset of protein import into mitochondria. The following models were considered: (i) targeting is nascent chain length-dependent: the targeting machinery becomes activated only after a sufficient length of the nascent protein has been synthesized; (ii) targeting is time-dependent: targeting is initiated after the emergence of the N-terminal MTS but takes ~60 seconds to complete, during which ~300 amino acids of the nascent chain has been translated given a translation rate of ~5 amino acids/second [101]; (iii) targeting initiates only when specific sequences or features of the nascent protein emerged from the ribosome. To distinguish between these models, we asked how altering the length of the N-terminal unstructured region of a cotranslational TOM substrate impacts the timing of its mitochondrial targeting. Models (i) and (ii) predict that the onset of TOM engagement will occur at the same nascent chain length regardless of the N-terminal sequence, whereas model (iii) predicts that nascent proteins with a longer N-terminal sequence will be delayed in targeting.

The mitochondrial import of COQ3 initiated when 281 amino acids had been translated (Fig. 3.3A). The insertion of a tandem repeat of a 44-amino acid sequence after its N-terminal MTS delayed targeting by ~50 amino acids, occurring when 330 amino acids have been synthesized (Fig. 3.3A and fig. S3.5A). The protein sequence exposed at the ribosome exit at the onset of TOM engagement were the same (Fig. 3.3A), arguing against models

(i) and (ii), and suggesting that a molecular signal in the nascent protein is needed to initiate import.

We further tested this model by comparing the timing of protein targeting among mitochondrial paralogs. COQ8A and COQ8B share highly conserved core regions that adopt similar structures, but evolved divergent N-terminal unstructured regions that differ ~120 amino acids in length (fig. S3.5, B and C). In agreement with the results from lengthening the unstructured N-terminus (Fig. 3.3A), the onset of cotranslational TOM engagement for these two proteins also differed by ~120 amino acids, occurring when homologous sequences in a specific region of their structured core emerged from the ribosome (Fig. 3.3B).

Finally, we probed the kinetics of protein targeting to mitochondria by examining the effects of the translation elongation inhibitor, cycloheximide (CHX). By stalling translation, CHX treatment provides more time for targeting-competent RNCs to engage with the TOM complex. If an intrinsically slow targeting process is involved in the late import of mitochondrial proteins, CHX treatment would lead to an earlier onset of TOM interaction. However, the timing of TOM engagement was insensitive to CHX treatment (fig. S3.5, D and E), contrary to this assumption. Together, our results indicate that cotranslational targeting to mitochondria is not determined by time or nascent chain length, but rather, begins upon the emergence of specific features in the mature region of the nascent mitochondrial protein.

To identify the features that are responsible for initiating import into mitochondria, we first analyzed the properties of exposed amino acids at the onset of TOM engagement (fig. S3.6). This analysis did not reveal any features at the sequence level, such as conserved sequence motifs (fig. S3.6A) or clusters of charged or hydrophobic residues (fig. S3.6, B and C), that emerge prior to import. The propensity to form an α helix or β sheet also showed no notable changes before the onset of import (fig. S3.6, D and E). In addition, we examined a potential role of internal MTS (iMTS) in the mature region of the protein, which is important for the targeting of some mitochondrial proteins lacking an N-terminal

MTS[102,103]. However, the sequences exposed at the onset of import did not exhibit iMTS-like properties either (fig. S3.6F). These findings suggest that cotranslational import into mitochondria is not mediated by primary protein sequence or secondary structural features.

We then asked whether the emergence of tertiary structural features initiates cotranslational protein import into mitochondria. As individual protein domains are folding units, we reasoned that proteins that share a conserved domain would display similar structural features during their synthesis and may thus be similar in the timing of their import. To test this hypothesis, we classified cotranslational TOM substrates based on the shared protein domains that have been translated prior to TOM engagement (fig. S3.7A). P-loop like, Rossmann-like, ALDH-like domains and TIM barrel folds are most prevalent among the domains detected in this analysis (fig. S3.7A). By aligning proteins harboring the same domains to the onset of their import, we found that import generally began near or after the complete exposure of these large and topologically complex domains (Fig. 3.3, C and D, and fig. S3.7B). This pattern was conserved across cotranslational TOM substrates despite the diversity of the protein domains exposed (fig. S3.7B), suggesting that a relatively complex globular fold is required to initiate mitochondrial import.

In agreement with this hypothesis, the exposure of small N-terminal domains with fewer than ~150 amino acids during translation did not correlate with the onset of mitochondrial import (Fig. 3.3E and fig. S3.7, C and D). One example is DLAT, which contains three independently folding small N-terminal domains spanning almost 400 amino acids followed by a large C-terminal domain of ~220 amino acids (Fig. 3.3E). The import of DLAT did not initiate until ~600 amino acids have been synthesized, when its large C-terminal domain has almost completely emerged from the ribosome (Fig. 3.3E). In support of this late import, co-co interactions were observed during the translation of this large C-terminal domain (fig. S3.7E). Taken together, these observations indicate that cotranslational import into mitochondria initiates specifically upon the exposure of a topologically complex protein folding unit on the nascent protein.
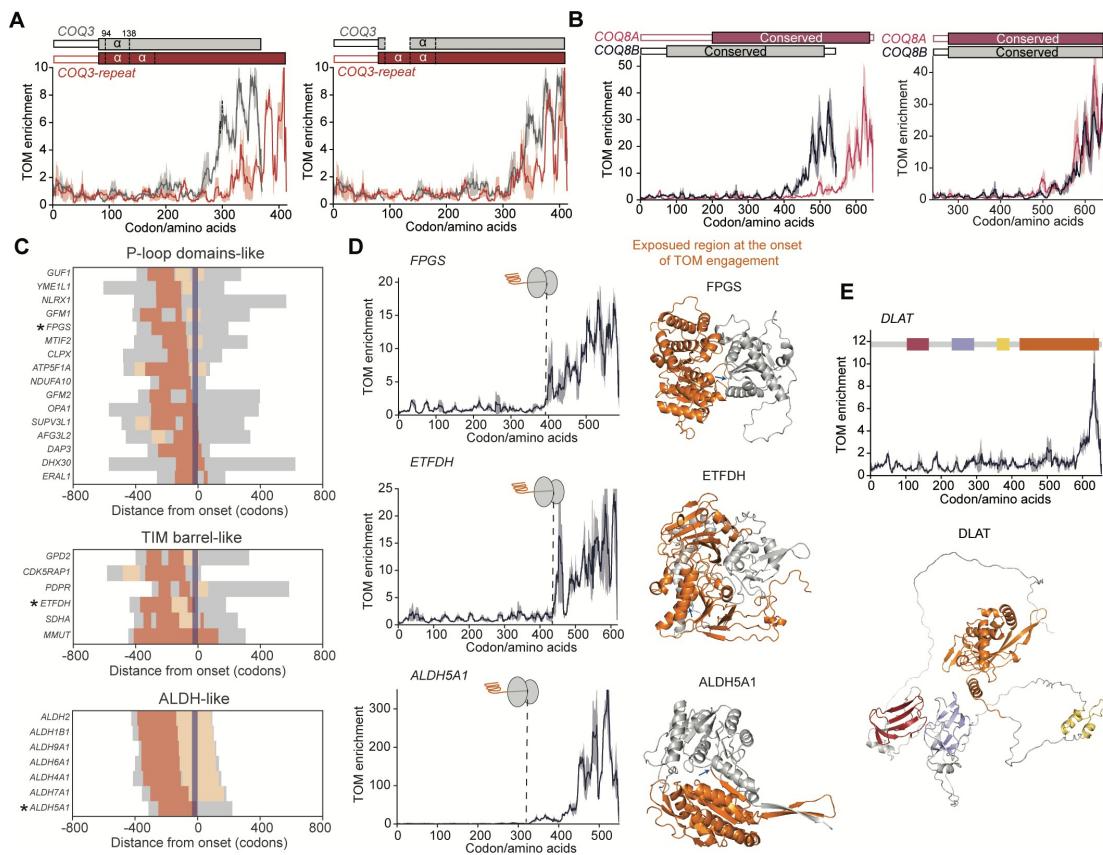
Fig. 3.3. Cotranslational mitochondrial import initiates upon the exposure of a complex globular protein fold.

(**A**) Comparison of the TOM interaction profile of WT *COQ3* and *COQ3-repeat*, which contains a tandem repeat of residues 94-138 (a helix-turn-helix motif, labeled as α) in WT *COQ3*. The scheme of the proteins above depicts the MTS regions as open bars and mature regions as solid bars. Proteins are aligned to the start codon (left) or by consensus amino acid sequence (right). (**B**) Comparison of the TOM interaction profile of *COQ8A* and *COQ8B*. The scheme of the proteins above depicts the unstructured regions as open bars and the conserved structured regions as solid bars. Proteins are aligned to the start codon (left) or to the N-terminus of their conserved structured regions (right). (**C**) Proteins are grouped by shared domains and aligned to the onset of TOM interaction, with the shared domain in each group in *orange*, additional domains that have initiated translation at the onset of import in *yellow*, and the remainder of the protein in *grey*. Residues to the left of the purple shaded area, which indicates the ~35 amino acids in the ribosome exit tunnel, are exposed at the onset of TOM interaction. (**D**) TOM enrichment profiles (left) and AlphaFold-predicted structures (right) of representative proteins from each group in (C) (marked by asterisk

(*)). The exposed regions at the onset of TOM engagement are in *orange* and the onset of TOM interaction is marked by a *blue* arrow in the structural models. The flexible MTS of each protein was not shown for clarity. (**E**) TOM enrichment profile (top) and AlphaFold-predicted structure (bottom) of DLAT. The scheme of the protein above depicts unstructured regions in grey and the individual folding domains as solid bars in different colors. In (A), (B), (D) and (E), solid lines show the mean values, and shaded areas show the range of data from two biological replicates.

3.3.4 Folding/unfolding challenges of large, multidomain mitochondrial proteins necessitate cotranslational import

The majority of mitochondrial proteins can be imported post-translationally, raising questions about what necessitates the evolution of a cotranslational targeting pathway for a specific subset of mitochondrial proteins. IMM proteins were previously found to be enriched in mitochondrially-localized translation in yeast, and the toxicity associated with the cytosolic exposure of the TMDs on these proteins was hypothesized to necessitate their cotranslational import[26]. However, among the 240 IMM proteins detected in our dataset, only 119 contain predicted TMDs, and cotranslational targeting did not exhibit a preference for TMD-containing proteins (Fig. 3.4A). In addition, most of the metabolite transporters, which are the most aggregation-prone multipass membrane proteins on the IMM, were not imported cotranslationally (fig. S3.8, A and B). The only exceptions are the ABC transporters, which contain an N-terminal MTS (fig. S3.8, A and B). For the 17 TMD-containing IMM proteins that cotranslationally engaged TOM, the onset of their import did not correlate with the exposure of the first TMD or the number of TMDs exposed (fig. S3.8C). These results suggest that the prevention of TMDs from aggregation in the cytosol is not the principle function of cotranslational protein targeting to mitochondria in human cells.

We therefore searched for alternative biophysical features of mitochondrial proteins that correlate with cotranslational targeting. We first noted that all the cotranslational TOM substrates are relatively large proteins (Fig. 3.4B). In addition to harboring an N-terminal MTS, having a protein size exceeding 300 amino acids was also an effective marker to distinguish cotranslational TOM substrates from other mitochondrial proteins (Fig. 3.4B).

We then trained a logistic regression classification algorithm using N-terminal MTS and protein length as the sole parameters. This classifier demonstrated remarkable accuracy in assigning mitochondrial proteins to either the co- or post-translational targeting pathways (Fig. 3.4C), suggesting that a large protein size is a key feature of cotranslationally targeted mitochondrial proteins.

A larger protein size tends to increase the challenges in both folding and unfolding[104,105]. To investigate if processes associated with protein folding contributed to cotranslational targeting, we calculated the absolute contact order (ACO) scores for all mitochondrial proteins (Fig. 3.4D). ACO measures the total distance in primary sequence between residues that form native contacts in the folded protein, which correlates negatively with protein folding rates but positively with the peak force needed to unfold a protein[53,106]. We found that cotranslational TOM substrates displayed significantly higher ACO scores compared to post-translationally targeted mitochondrial proteins (Fig. 3.4D) or compared to a set of randomly sampled cytosolic proteins with the same size distribution (fig. S3.8D). Thus, in addition to the larger size, the presence of protein topology with higher folding and unfolding challenges correlated with the cotranslational import of these proteins. In further support of this notion, multidomain proteins, which are more resistant to unfolding due to interdomain interactions[107], dominated cotranslational TOM substrates (Fig. 3.4E and fig. S3.8E). Together, these observations suggest that the challenges in the folding/unfolding of large and topologically complex mitochondrial proteins necessitate a cotranslational mechanism for their import.

Finally, we assessed the requirement for cotranslational targeting using an *in vitro* import assay (Fig. 3.4F). NDUFS2, a cotranslational TOM substrate localized in the matrix, was translated in the rabbit reticulocyte lysate. Import into purified human mitochondria was carried out either cotranslationally, by adding mitochondria during translation, or post-translationally, by using puromycin to release the nascent chain from the ribosome before the addition of mitochondria. We observed a significantly higher import efficiency for NDUFS2 when import occurred cotranslationally than post-translationally (Fig. 3.4F), indicating that a cotranslational mode of targeting enhances the efficiency of its import.
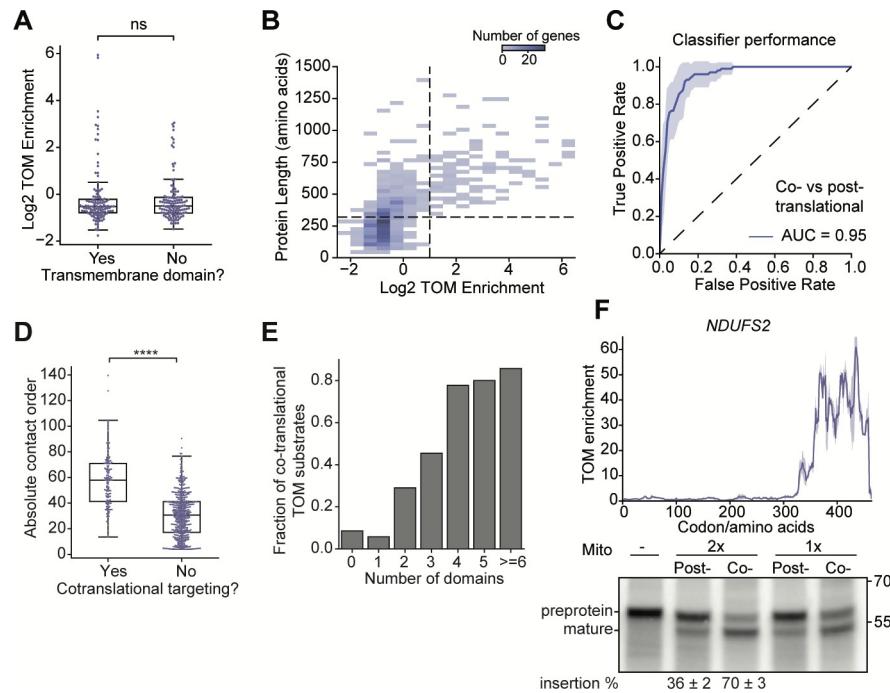
Fig. 3.4. Role of cotranslational mitochondrial protein targeting.

(**A**) Comparison of TOM enrichment between IMM proteins with and without TMDs. ns, not significant (p > 0.05). (**B**) Two-dimensional histogram of mitochondrial proteins binned by log2 TOM enrichment and protein length. (**C**) Receiver operator characteristic (ROC) curves from a supervised random-forest classifier that predicts whether a mitochondrial protein utilizes a co- or post-translational targeting pathway. The presence of an N-terminal MTS and protein length were used as the parameters to train the model. The solid line represents the mean performance, and the shaded area shows the range of data from 10-fold cross-validation. (**D**) Absolute contact order analysis of all mitochondrial proteins. ****, p<0.0001. (**E**) The fraction of cotranslational TOM substrates among mitochondrial proteins, grouped by the number of domains identified in each protein. (**F**) Upper panel, TOM interaction profile of NDUFS2, a mitochondrial matrix protein. The solid line shows the mean values, and the shaded area shows the range of data from two biological replicates. Lower panel, *In vitro* mitochondrial import of NDUFS2. ^{35}S-methionine labeled NDUFS2 was translated in rabbit reticulocyte lysate, and purified human mitochondria were added either immediately after translation initiation (co-) or after 40 min of translation followed by release of the nascent chain using puromycin (post-). Successful import is monitored by cleavage of the N-terminal MTS, which generates a product with lower molecular weight

(mature protein). The values below represent the quantification of percent insertion from two biological replicates, shown as mean ± SD.

3.4 Discussion

In this work, SeRP of the TOM complex bridges RNA localization and mitochondrial protein import, provides definitive evidence for cotranslational protein import into mitochondria, defines the timing and specificity of this process, and reveals its intimate link to cotranslational protein folding and unfolding. Our proteome-wide dataset also provides a rich resource to study the role of cotranslational targeting pathway in other important cellular processes, such as stress response.

Cotranslational protein targeting to the ER by signal recognition particle (SRP), a conserved pathway used by numerous proteins destined to the endomembrane system, has shaped many concepts about the roles and mechanism of the cotranslational mode of targeting. Unexpectedly, the results here show that cotranslational protein targeting to mitochondria is fundamentally different. First, the preponderance of integral membrane proteins among SRP substrates led to the notion that the cotranslational mechanism serves to prevent hydrophobic TMDs from aggregation in the cytosol[108], a model that was also proposed for cotranslational protein targeting to mitochondria in yeast [26]. However, no such correlation was found for mitochondrial proteins that are cotranslationally targeted to the TOM complex. Instead, cotranslational TOM substrates are characterized by a large size, the presence of multiple domains, and high ACO values, features that suggest they are slow to fold, prone to misfolding, and resistant to unfolding once the native structure is established. We therefore propose that cotranslational import prevents the stable folding of large mitochondrial proteins in the cytosol, which would otherwise pose a strong energetic barrier for their import and clog the TOM/TIM23 translocases.

Secondly, despite being predominantly used by proteins with an N-terminal MTS, cotranslational targeting does not initiate immediately upon the emergence of the MTS and instead, begins after 300-400 amino acids of the nascent mitochondrial protein have been

synthesized. This is distinct from SRP, which engages the nascent chain as soon as an ER targeting signal emerges from the ribosome and completes ER targeting before the translation of an additional 60 amino acids[109]. The late onset of mitochondrial targeting supports the hypothesis that cotranslational protein targeting to mitochondria would render translation elongation rate-limiting for import and must therefore be minimized[21]. Our results suggest the presence of a 'delay' mechanism to achieve this minimization, wherein the MTS is temporarily prevented from engaging the TOM complex until the emergence of a second molecular signal in the mature region of the nascent protein. Such a mechanism occurs in bacteria, where the chaperone TF binds early during the translation of many secretory proteins and delays their export by the SecB/A targeting machinery[3,110]. The 'delay' of mitochondrial protein import is potentially a conserved mechanism to optimize the balance between the speed and capacity of import machinery versus the prevention of irreversible folding in the cytosol.

Finally, our data indicate that the signal to initiate mitochondrial import is the exposure of a large, topologically complex protein domain during translation, implying an interplay between cotranslational protein folding and targeting. Structural models of cotranslational TOM substrates suggest that sufficient sequence information in the nascent protein has emerged at the onset of import to enable partial folding. These structures are potentially reversible and not prohibitory for import into mitochondria, because they lack the additional long-range interactions and interdomain contacts in the full-length protein and because the TOM/TIM23 machinery can import most single-domain proteins post-translationally. The interplay between folding and import is also supported by the finding of widespread cotranslational assembly of mitochondrial proteins, which resolved prior to the initiation of import. This observation suggests that nascent mitochondrial proteins expose unsatisfied contacts during translation, which make them susceptible to non-native interactions. In the simplest model, the co-co interaction prevents nascent mitochondrial proteins from targeting until the complete exposure of a large protein domain provides sufficient energetics for partial domain folding, which outcompetes co-co assembly and other inhibitory interactions. More elaborate mechanisms could involve a chaperone

analogous to TF, whose interaction with unsatisfied contacts on short nascent chains prevents their mitochondrial targeting. Structural alterations in the nascent chain upon the exposure of a large protein domain may lead to dissociation of this early-acting chaperone, thus allowing for the initiation of mitochondrial targeting. Our data place important constraints for the investigation of such targeting factors and mechanisms, providing a vital foundation for understanding the delivery of mitochondrial proteins.

3.5 Materials and Methods

Cell culture

HEK293-T cells were cultured in high glucose DMEM media supplemented with GlutaMAX™ and pyruvate (Gibco) + 10% fetal bovine serum (Gibco). Cells were grown at 37 °C and passaged regularly using Trypsin-EDTA for dissociation.

K562 cells were cultured in RPMI-1640 medium containing 25 mM HEPES, 2.0 g/L $NaHCO_3$, and 0.3 g/L L-glutamine supplemented with 10% Tet System Approved FBS, 100 units/mL penicillin, and 100 µg/mL streptomycin. A cell density between 0.25 - 1 million cells/mL was maintained. Cells were grown in 37 °C with 5% CO_2.

Antibodies

The following antibodies were used in this study: TOM40 (sc-365467, Santa Cruz), TOM22 (sc-101286, Santa Cruz), TOM70 (14528-1-AP, Proteintech), RPS6 (2317S, Cell Signaling), RPL10 (AP19053a, Abcepta), GAPDH (MA5-15738, Thermo Fisher), NACβ (ab203517, Abcam), NACα (PA5-116506, Thermo Fisher).

Cell line generation

TOM22-TwinStrep cell line

The sequence encoding a TEV cleavage site followed by Twin-Strep tag was inserted upstream of the stop codon of endogenous *TOMM22* via CRISPR/Cas9-mediated homology-directed repair (HDR) according to[111]. Briefly, the repair template was designed as a single-stranded DNA oligonucleotide (ssODN) with ~45 nt homology arms

at each side of insertion (Table S4) and purchased from IDT. Cells were transfected with pCas9-sgRNA plasmid (derived from PX459, Addgene) expressing the sgRNA targeting *TOMM22* (ATAACAATCTAGATCTTTCCAGG), and the ssODN repair template using Lipofectamine 3000 (Thermo Fisher). Cells were grown for 24 hrs and treated with puromycin for 48 hrs to select for successfully transfected clones. Single-cell clones were isolated by limiting dilution. Successful edits were verified by the sequencing of PCR (fwd GCACCTGAGTTGACCAACAGTT, rev CTCCCCACCAGGTTTAGATAGATC) products and western blotting.

NACα/β rapid depletion cell lines

Cell lines harboring the AtAFB2-miniIAA7 system were generated using the CRISPR/Cas-9 system following the protocol in [98]. Briefly, the auxin receptor F-box protein, AtAFB2, was first integrated into *TOMM22-TwinStrep* cells at the AAVS1 safe harbor locus via CRISPR/Cas9-mediated HDR. Cells were co-transfected with pCas9-sgAAVS1-1(129726, Addgene) and pSH-EFIRES-B-AtAFB2-mCherry (129718, Addgene) using Lipofectamine 3000 (Thermo Fisher). Cells were grown for 24 hrs and treated with puromycin and blasticidin to select for successfully transfected clones. Single-cell clones were isolated by limiting dilution, and successful edits were validated based on imaging of mCherry fluorescence.

To generate the NACα-AID cell line, the sequence of the degron, miniIAA7, was cloned from pSH-EFIRES-B-Seipin-miniIAA7-mEGFP (129719, Addgene). The sequence encoding GFP11-miniIAA7-GS linker was inserted at the N-terminus of *NACA* at its endogenous locus via CRISPR/Cas9-mediated HDR. The repair template was designed as a plasmid with ~700 nt homology arms at each side of the insertion (Table S4). AtAFB2 integrated cells were co-transfected with pCas9-sgRNA plasmid (derived from PX459, Addgene) expressing sgRNA targeting NACA (GCTTCGCCGGGCATTTCTGAAGG) and the repair template using Lipofectamine 3000 (Thermo Fisher). On the following day, cells were transfected with pcDNA5-GFP1-10 plasmid (a gift from Rebecca Voorhees) expressing GFP1-10. GFP$^+$ and mCherry$^+$ cells were selected by FACS after 24 hours.

Single clones were grown, and the successful edits were verified by the sequencing of PCR (fwd GCCCTCCTCGGAGTTTTTAAGAATA, rev AGGGACAGGAATTGCCCTTAA) products and western blotting.

The NACβ-AID cell line was made similarly as described above. A GS linker-miniIAA7-GFP11 sequence was inserted at residue 176 of endogenous *BTF3* via CRISPR/Cas9-mediated HDR. AtAFB2 integrated cells were co-transfected with pCas9-sgRNA plasmid (derived from PX459, Addgene) expressing sgRNA targeting BTF3 (TTTCCTTTCATAGCTGTGGATGG) and the repair template (Table S4) using Lipofectamine 3000 (Thermo Fisher). Single clones were selected by FACS, and successful edits were validated by the sequencing of PCR (fwd TGTGGGTTTTACCTGCACTCT, rev CCCTCTTCCCTGGGTAGTTTT) products and western blotting.

COQ3-repeat cell line

The sequence encoding the COQ3-repeat was integrated into the genome of *TOMM22-TwinStrep* cells via the PiggyBac transposon system[112]. The *COQ3-repeat* gene was generated by insertion of a tandem repeat of residues 94-137 in WT *COQ3*. It was then codon-changed to distinguish from WT *COQ3* (Table S4) and cloned into a PiggyBac vector (a gift from Shasha Chong), which co-expresses a puromycin resistance gene. Cells were co-transfected with PiggyBac-COQ3-repeat plasmid and SuperPiggyBac transposase plasmid using Lipofectamine 3000 (Thermo Fisher). Cells were grown for 24 hrs and treated with puromycin for 48 hrs to select for successfully transfected clones. After 48 hrs of recovery, cells were treated again with puromycin for 7 days to select for cells with successful gene integration.

Selective ribosome profiling of TOM complex

Cells were grown on 150 mm dishes to ~90% confluency. 8 dishes of cells were treated with 100 μg/ml cycloheximide for 2 min at 37 °C. The media was removed, and all subsequent steps were performed on ice using ice-cold solutions. Cells were detached by

pipetting 10 ml of 1x PBS supplemented with 100 µg/ml cycloheximide and 10 mM MgCl$_2$ per two dishes and centrifuged at 2000 ×g for 3 min at 4 °C. Pelleted cells were lysed in 7.2 mL Lysis Buffer (50 mM Tris pH 7.4, 100 mM NaCl, 10 mM Mg(OAc)$_2$, 1% digitonin, 100 µg/mL cycloheximide, 25 U/ml Turbo DNase (Invitrogen), 1x protease inhibitors (Complete EDTA-free, Roche)) and incubated on ice for 10 min. Cell lysate was triturated five times through a 26-G needle and clarified by centrifugation at 14, 000 rpm for 5 min at 4 °C. To test the effects of cycloheximide treatment, cells were treated with 100 µg/ml cycloheximide for 15 min at 37 °C before harvesting. Cells were then harvested and lysed as described above.

For DSP crosslinked samples, cells were grown on 150 mm dishes to ~90% confluency and treated with 100 µg/ml cycloheximide for 2 min at 37 °C. The media was removed, and cells were washed once using 1x PBS supplemented with 100 µg/ml cycloheximide and 10 mM MgCl$_2$. 10 mL crosslinking buffer (1x PBS, 100 µg/ml cycloheximide, 10 mM MgCl$_2$, 250 µM DSP) was added per dish and incubated for 5 min at room temperature. To quench the reaction, 50 mM Tris pH 7.4 was added and incubated for 5 min. All subsequent steps were performed on ice using ice-cold solutions. Crosslinking buffer was removed, and 900 µL Lysis Buffer was added to each dish. Cells were detached by scraping, and the cell lysate was collected and incubated on ice for 10 min. Subsequent steps were the same for crosslinked and uncrosslinked samples.

RNA concentration in the lysate was measured by Qubit HS RNA assay, and polysomes were digested with 150 U RNase1 (Ambion) / 40 µg RNA for 30 min at 4 °C with constant rotation. Digestion was stopped by 100 U/mL SUPERase*In RNase Inhibitor (Ambion) and chilling on ice. Monosomes were purified by centrifugation through Sucrose Cushion (1 M sucrose, 50 mM Tris pH 7.4, 100 mM NaCl, 10 mM Mg(OAc)$_2$, 0.1% digitonin, 100 µg/mL cycloheximide) in a TLA100.3 rotor at 100,000 rpm for 60 min at 4 °C. Pellets were washed once and resuspended in 1.2 mL Wash Buffer (50 mM Tris pH 7.4, 100 mM NaCl, 10 mM Mg(OAc)$_2$, 0.01% digitonin, 100 µg/mL cycloheximide). 10 µg of total RNA were removed from resuspended monosomes for ribosome profiling of the total translatome. 1.6 mL MagStrep "type3" XT beads (5% suspension, IBA Lifesciences, pre

washed for 3 times in Wash Buffer) was added to the remainder of the sample, and the suspension was rotated for 45 min at 4 °C. Beads were washed 4 × 5 min at 4 °C in Wash Buffer. During each wash, beads were transferred to a new low adhesion tube. After the fourth wash, beads and the total monosome sample were incubated with 400 μL TRIzol reagent (Thermo) for 10 min at room temperature. The supernatant was used for subsequent RNA extraction using the Direct-zol kit (Zymo) according to manufacturer's instructions.

Ribosome profiling libraries were prepared as described in [110] with an additional rRNA depletion step and sequenced on a Nextseq 2000. rRNA depletion was performed on the linker-ligated RNA footprints using the riboPOOL Ribo-Seq (human) kit (Galen) according to the manufacturer's instructions.

Data analysis

Processing of sequencing data

Raw sequencing data was processed as described in [91]. Briefly, 3' adaptor sequences were trimmed from sequencing reads with Cutadapt v4.1 [113] using the following command:

```
cutadapt --cores=0 -q20 -m24 -M42 --discard-untrimmed -O6 --no-indels -a adaptor_sequence -o
outfile.fastq.gz infile.fastq.gz 1> Cutadapt_report.txt
```

Unique molecular identifiers (UMIs, two random 5' nucleotides and five random 3' nucleotides) were trimmed from each read using a Julia script (Script1) from [91]. This generates an output fastq file containing the 7-nucleotide UMI information in the read name. The UMI-trimmed reads were mapped to the human ribosomal RNA sequences with Bowtie2 v2.4.5 [114] using the following command:

```
bowtie2 -p 32 -t -x rRNA_index -q infile.fastq.gz -p 16 --un outfile.fastq.gz -S /dev/null >
Bowtie2.report.txt
```

Reads that did not align to ribosomal RNA sequences were mapped to human reference genome (GRCh38p13 downloaded from NCBI) with STAR 2.7.10a [115] using the following command:

```
STAR --runThreadN 32 --genomeDir indexed_genome --readFilesIn infile.fastq.gz --
outFilterMultimapNmax 1 --outFilterType BySJout --alignIntronMin 5 --outFileNamePrefix Prefix --
outReadsUnmapped Fastx --outSAMtype BAM SortedByCoordinate --outSAMattributes All XS --
quantMode GeneCounts --twopassMode Basic
```

For each gene, the transcript with the longest coding sequence was selected, and ribosome density was assigned to the p-site of each read using a Julia script (Script2) from [91]. This generates an HDF5 file containing the 1-based indexing of nucleotide position within the CDS of each gene and the detected p-site reads at this position. Nucleotide reads at each codon were summed and used for all additional analyses via custom python scripts.

Single gene enrichment profiles

For each gene, the Reads Per Million (RPM)-normalized reads at each codon was computed for both the total and TOM-bound translatomes. The RPM-normalized reads were smoothed with a 15-residue sliding window, and a pseudocount of 0.05 RPM was added to both the total and TOM-bound translatomes for the enrichment calculation. TOM enrichment at each codon was calculated as the ratio of RPM-normalized reads in the TOM-bound translatome over that in the total translatome. The analysis of DiSP data [91] was performed similarly by calculating the ratio of RPM-normalized reads in the disome over that in the monosome.

Gene-level enrichment

The raw reads at each codon were summed over the entire coding sequence for each gene and normalized for gene size and sequencing depth to generate the Reads Per Kilobase per Million (RPKM)-normalized reads. This calculation was performed only for genes with greater than 50 raw reads in both the total and TOM-bound translatome datasets. Gene-level TOM enrichment was calculated as the ratio of RPKM-normalized reads from the TOM-bound translatome to that from the total translatome.

Peak detection

We defined the TOM binding peaks as the regions that met the following criteria: (1) TOM enrichment at each codon $> = 2.5$-fold for at least 7 consecutive codons; (2) The overlap of the peaks from two replicates $> = 7$ codons. The start of the first detected binding peak in each gene was defined as the onset of TOM binding. Only genes whose coverage is higher than 0.25 reads/codon in the total translatome or TOM-bound translatome in both replicates were used. For a few mitochondrial proteins, the longest coding sequence that was used previously in the sequencing reads alignment was not their major isoform annotated in MitoCarta 3.0 [116]. The onset of peaks on these proteins were manually curated to match their position on the major isoform.

Metagene profiles

Only genes whose coverage is higher than 0.25 reads/codon in the total translatome or TOM-bound translatome were used in all metagene analyses. For metagene analyses of the total and TOM-bound translatomes, the raw reads at each codon were first smoothed over a 15-residue sliding window, then divided by the average reads per codon of the respective gene to normalize for gene expression level. For metagene analyses of TOM enrichment and co-co interaction (disome/monosome enrichment), the enrichment scores at individual codons were calculated as described above. All genes that passed the filtering or subsets of genes (e.g. cotranslational TOM substrates) were aligned to the start codon or to the onset of TOM interaction as indicated. Metagene profiles were generated by calculating the position-wise arithmetic mean of expression-normalized reads or enrichment scores and the bootstrapped 95% CI.

Gene categorization

Mitochondrial genes and their mitochondrial subcompartments were annotated according to MitoCarta 3.0[116]. ER genes were annotated according to Uniprot[117].

Gene ontology analysis

Gene ontology analysis was performed on cotranslational TOM substrates using the Functional annotation clustering algorithm from DAVID Bioinformatics Resources[118] with

all detected mitochondrial proteins as the background. The analysis was performed with the default settings and 'Reactome pathway' was used as the annotation category. GO-terms were ranked by p-value.

Sequence feature analysis

WebLogo representation of the last 90 amino acids that have been translated at the onset of TOM engagement was performed using WebLogo3 (http://weblogo.threeplusone.com). The alpha-helix and beta-sheet propensity scores were calculated using NetSurfP-3.0 at each amino acid position[119]. The iMTS-L propensity score was calculated using iMLP at each amino acid position[120]. The charge and hydrophobicity of each amino acid was averaged over a 7 amino acids rolling window. All sequences were aligned to the onset of TOM interaction for metagene analysis. The position-wise arithmetic mean of charge, hydrophobicity, alpha-helix, beta-sheet and iMTS-L propensity scores from all sequences and bootstrapped 95% CI were calculated.

Mitochondrial targeting sequence (MTS) and domain predictions

The MTS of mitochondrial proteins was predicted by Mitofates[121] and TargetP 2.0[122] unless it has been experimentally verified. Proteins that were predicted to have an N-terminal MTS by either method were considered to be MTS-containing. TMDs were predicted by DeepTMHMM [123] unless it has been experimentally verified and deposited in Uniprot[117]. DomainMapper [124] was used to identify and annotate the protein domains in all mitochondrial proteins.

In silico protein structure prediction

The structure of COQ3-repeat was predicted using ColabFold[125], a Google Colab-based implementation of AlphaFold[126], using default settings. The structure alignment was performed using PyMOL.

Contact order calculation

The predicted structures of all mitochondrial proteins were downloaded from AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/,[127]). The ACO for each protein is calculated based on the structure, as previously described[53].

Logistic regression classifier

The features selected for classification were the length of the protein and the binary 'MTS?' indicator. The dataset was partitioned into training and test sets, with 20% of the data reserved for testing. A logistic regression classifier was used for model training. To evaluate the model's performance and generalizability, 10-fold stratified cross-validation was employed on the training data. In each fold, the model was trained on the training subset and predictive probabilities were obtained for the validation subset. These probabilities were used to compute the Receiver Operating Characteristic (ROC) curve, and the Area Under the Curve (AUC) was calculated as a metric of model performance.

Quantification and statistical analysis

All analyses were performed in python. Statistical significance was determined using independent-samples t-test. The p-values are indicated in figure legends.

Western blotting

For *in vivo* ribosome binding analysis, cells were lysed in Lysis Buffer (50 mM Tris pH 7.4, 100 mM NaCl, 10 mM Mg(OAc)$_2$, 1% digitonin, 100µg/mL cycloheximide, 25 U/ml Turbo DNase (Invitrogen), 1x protease inhibitors (Complete EDTA-free, Roche)), and the lysate was clarified by centrifugation at 14,000 rpm for 5 min at 4 °C. Ribosomes were pelleted by centrifugation through Sucrose Cushion (1 M sucrose, 50 mM Tris pH 7.4, 100 mM NaCl, 10 mM Mg(OAc)$_2$, 0.1% digitonin, 100 µg/mL cycloheximide) in a TLA100.3 rotor at 100,000 rpm for 60 min at 4 °C. Total lysate, supernatant, and resuspended pellet were analyzed by SDS-PAGE and immunoblotting.

For analysis of NACα/β depletion, indole-3-acetic acid sodium (IAA, Santa Cruz, sc-215171) was prepared as a 10 mg/ml stock solution in H$_2$O. Cells were treated with 100

µg/mL IAA or water for the indicated time, collected and lysed in WB lysis buffer (50 mM Tris pH 7.4, 100 mM NaCl, 10 mM Mg(OAc)$_2$, 1% NP-40, 1 mM DTT, 25 U/ml Turbo DNase (Invitrogen), 1x protease inhibitor (Complete EDTA-free, Roche)). The lysate was clarified by centrifugation at 14,000 rpm for 5 min at 4 °C and analyzed by SDS-PAGE and immunoblotting.

Mitochondrial isolation

Mitochondria were isolated from K562 cells according to a previously established protocol[128]. K562 cells were pelleted by centrifugation at 300g for 5 minutes. Cells were washed by resuspending in homogenization buffer containing 210 mM D-mannitol, 70 mM sucrose, 5 mM HEPES pH 7.4, 10 mM EDTA, 1 mM PMSF, and 2 mg/mL BSA and pelleting by centrifugation at 500g for 5 minutes. Pellets were resuspended in homogenization buffer and incubated on ice for 5 minutes. Cells were then lysed with 25 strokes in a glass Dounce homogenizer with a tight-fitting pestle. Lysed cells were centrifuged at 1300g for 5 minutes to remove nuclei and unbroken cells, then the supernatant was transferred to a clean tube. This step was repeated twice to obtain nuclei-free homogenate. Mitochondria were then pelleted by centrifugation at 11,000 g for 10 minutes. Mitochondria were washed twice by resuspending in isolation buffer containing 210 mM D-mannitol, 70 mM sucrose, 5 mM HEPES pH 7.4, 10 mM EDTA and pelleting by centrifugation at 11,000g for 10 minutes. The final mitochondrial pellet was resuspended with 10µl isolation buffer. To quantify mitochondrial samples, the protein concentration was measured using a Bradford assay.

***In vitro* mitochondrial import**

In vitro translations were carried out in rabbit reticulocyte lysate (RRL) according to[129]. Constructs for in *vitro* translation reactions were based on the pSP64 vector (Promega, USA). Templates for transcription were generated by PCR, with primers annealing to sequences upstream of the SP6 promoter and downstream of the stop codon. Transcription was performed at 37 °C for 1.5 hour with SP6 polymerase and a final concentration of 40 mM HEPES pH 7.6, 6 mM MgCl$_2$, 2 mM spermidine, 10 mM DTT, 1 x NTPs (0.5 mM

each), 0.5 mM Cap analog 7-methyl diguanosine triphosphate. The transcription reaction mixture was then directly used in a translation reaction. Each 50 µL translation contained 28.5 µL 1x T2 mix composed of the rabbit reticulocyte lysate, amino acids, and an energy mix, 10 µL transcription reaction, 50 µCi [^{35}S]-Methionine, 4 µM cold Methionine, 250 mM sucrose, 15 mM succinate, 0.1 µg/µL tRNA.

For cotranslational mitochondrial import, 5 µg of purified human mitochondria was added to a 7 µL translation reaction 1 min after initiation of translation at 32 °C, and the reaction was further incubated at 32 °C for 40 min. For post-translational import, the translation reaction was incubated at 32 °C for 40 min, followed by treatment with 1 mM puromycin for 5 min to release nascent chains. 5 µg of purified mitochondria was then added into a 7 µL translation reaction, and the mixture was further incubated at 32 °C for 40 min. The reactions were analyzed by SDS-PAGE and autoradiography.

Fluorescence microscopy imaging

NACα and NACβ rapid depletion cells were plated onto 35mm glass bottom dishes (MatTek) and treated with 100 µg/mL IAA or water for 4.5 hrs. Cells were then transfected with plasmids encoding mitochondrial matrix-targeted GFP1-10 and TRAP1-GFP11 with or without an MTS using Lipofectamine 3000 (Thermo Fisher). After 16 hrs, cells were imaged on a Leica Stellaris 8 FALCON laser scanning confocal microscope.

3.6 Acknowledgements

Data and materials availability: All sequencing data generated in this study are available at GEO under accession number GSE255657. Co-co interaction data is available at GEO under accession number GSE151959[91]. All datasets and predictions are provided as Supplementary Material. Custom python scripts used for data analysis are available upon request.

3.7 Supplementary figures
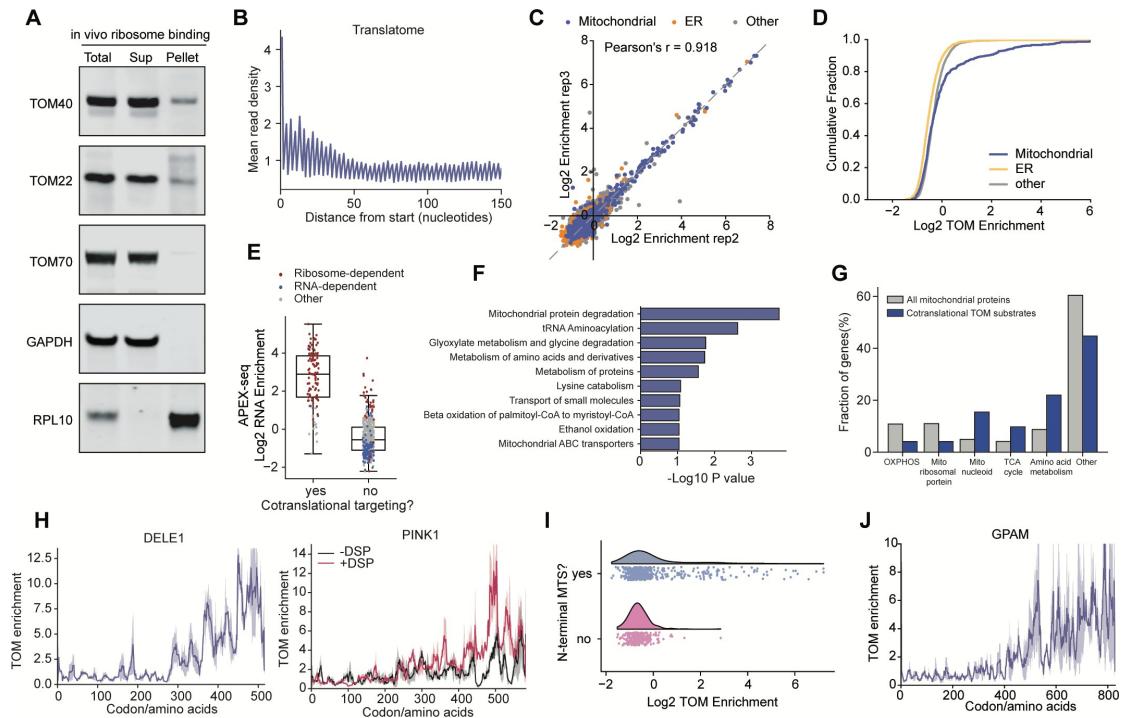
fig. S3.1. Selective ribosome profiling of the TOM complex.
(**A**) Western blots showing the co-sedimentation of TOM complex with ribosome in HEK293-T cells. (**B**) Average ribosome density of all genes in the total translatome data aligned to the start codon. The three-nucleotide periodicity reflects the translocation of ribosomes along the mRNA one codon at a time, indicative of data at near-codon resolution. (**C**) Reproducibility of gene-level TOM enrichment from two biological replicates. (**D**) Cumulative distribution of log2 TOM enrichment for genes categorized by subcellular localizations. (**E**) Comparison of the APEX-seq RNA enrichment at mitochondrial surface [23] between co- and post-translationally targeted proteins. (**F**) Gene ontology analysis of cotranslational TOM substrates. (**G**) Comparison of the fraction of genes in different functional classes between cotranslational TOM substrates and all detected mitochondrial genes. Genes are classified according to[23]. (**H**) TOM interaction profiles of *DELE1* and *PINK1*. A higher TOM enrichment on *PINK1* was observed in the presence of the crosslinker DSP. (**I**) Raincloud plot comparing the distribution of TOM enrichment on mitochondrial proteins with or without a predicted N-terminal MTS. (**J**) TOM interaction profile of *GPAM*, which encodes an OMM protein. In (H) and (J), solid lines show the mean values, and shaded areas show the range of data from two biological replicates.
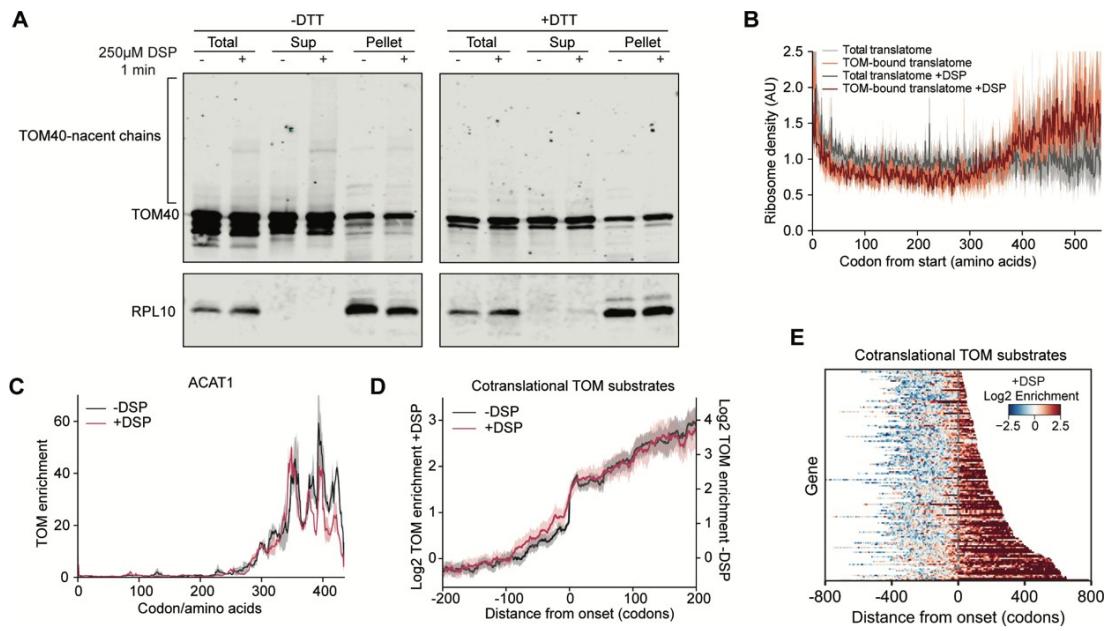
fig. S3.2. Selective ribosome profiling of the TOM complex with DSP crosslinking.
(**A**) Western blot analysis of TOM-nascent chain crosslinking using DSP. The collection of high molecular weight bands above TOM40 indicate crosslinking products, which were reversed by DTT (DSP contains a disulfide bond). (**B**) Metagene total translatome (*black* and *grey*) and TOM-bound translatome (*orange* and *brick*) profiles of mitochondrial proteins aligned to the start and stop codons, with and without DSP crosslinking. (**C**) Representative TOM interaction profiles of a mitochondrial gene with and without DSP crosslinking. Solid lines show the mean values, and shaded areas show the range of data from two biological replicates. (**D**) Metagene TOM enrichment profiles of all cotranslational TOM substrates aligned to the onset of TOM interaction, with and without DSP crosslinking. (**E**) Heatmap of log2 TOM enrichment at each codon for cotranslational TOM substrates after DSP crosslinking. Genes are aligned to the onset of TOM interaction detected without DSP crosslinking and sorted by the distance from the onset to the stop codon. In (B) and (D), solid lines show the mean values, and shaded areas show the 95% CI.
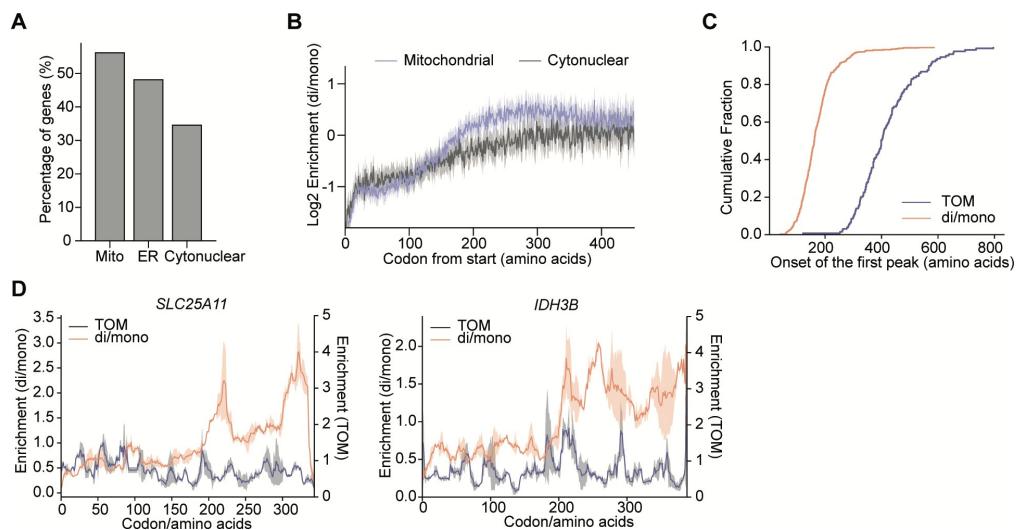
fig. S3.3. Analysis of disome selective profiling data to identify cotranslational protein assembly.

Disome selective profiling data are from[91]. (**A**) More than half of mitochondrial proteins exhibit co-co interactions. (**B**) Metagene co-co interaction profiles for mitochondrial (*blue*) and cytonuclear (*dark grey*) proteins aligned to the start codon. On average, nascent mitochondrial proteins initiate co-co interactions at a length of ~150 amino acids. Solid lines show the mean values, and shaded areas show the 95% CI. (**C**) Cumulative distribution of the onset of TOM interaction and co-co interaction. (**D**) Representative TOM interaction and co-co interaction profiles of post-translationally targeted mitochondrial proteins. Solid lines show the mean values, and shaded areas show the range of data from two biological replicates.
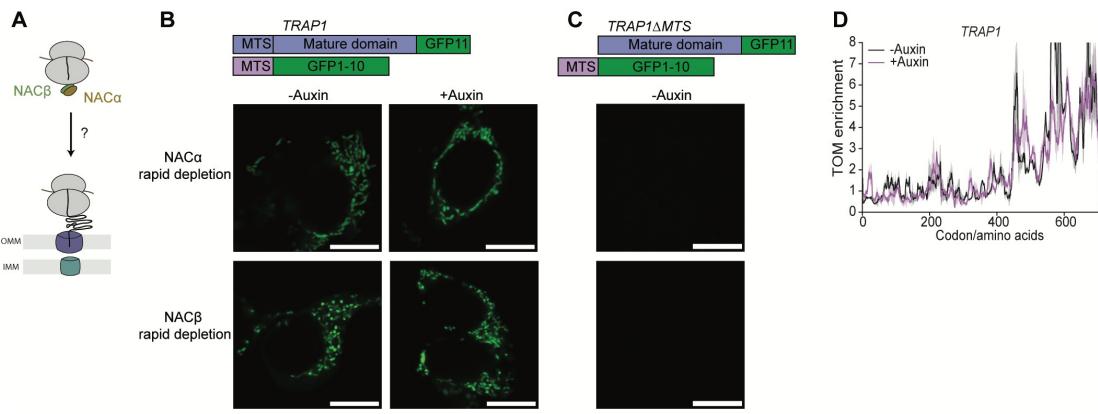
fig. S3.4. NAC is not required for the correct mitochondrial localization of TRAP1, a cotranslationally targeted protein.

(**A**) Schematic showing the potential role of NAC in cotranslational targeting of mitochondrial proteins. (**B**) Plasmids encoding mitochondrial matrix-targeted GFP1-10 and GFP11-tagged TRAP1 were co-transfected into NACα-AID or NACβ-AID cells 4.5 hr after auxin or H_2O (-auxin) addition. GFP fluorescence was imaged after 16 hr. Scale bars, 10 μm. (**C**) Same as in (B), except that the N-terminal MTS was deleted from TRAP1-GFP11. (**D**) Comparison of the TOM interaction profiles of *TRAP1* before and after auxin addition (4.5 hr) in NACβ-AID cells.
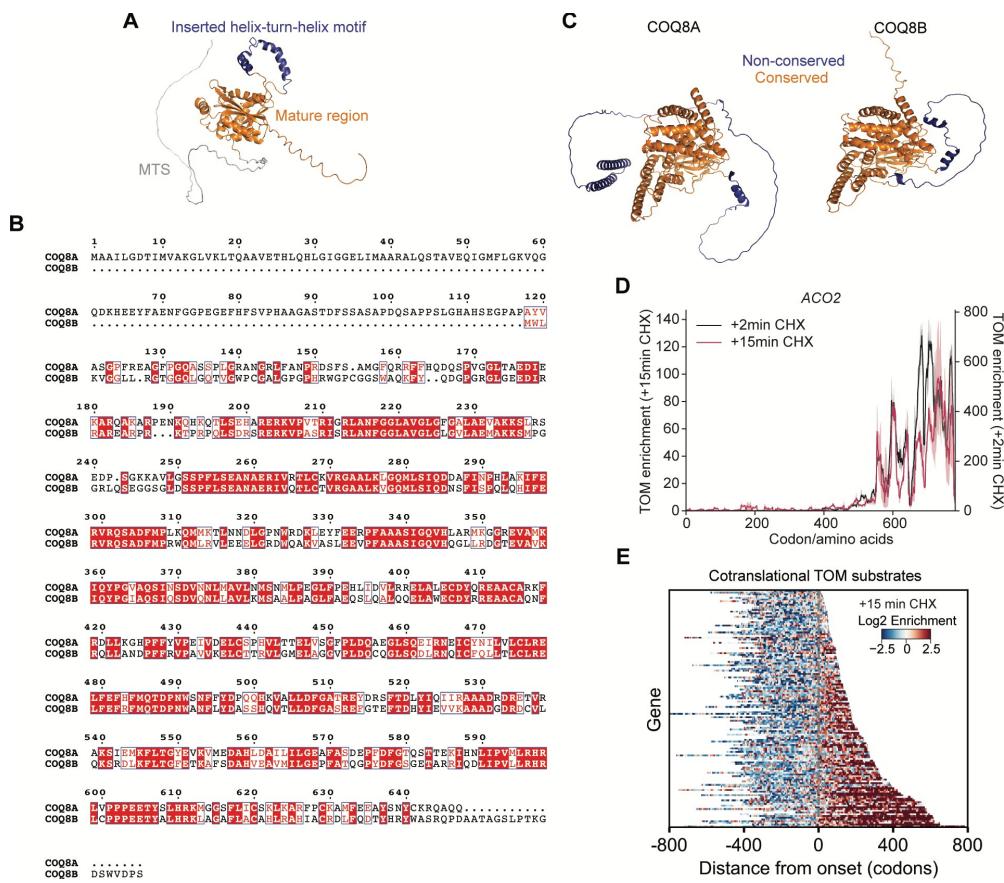
fig. S3.5. Cotranslational targeting to mitochondria initiates upon the emergence of specific features on the nascent protein.

(**A**) AlphaFold-predicted structure of COQ3-repeat, with MTS in *grey*, mature region in *orange* and the inserted helix-turn-helix motif in *blue*. The insertion does not affect the core structure. (**B**) Protein sequence alignment of COQ8A and COQ8B generated using Clustal Omega [130] and visualized using ESPRIPT 3.0[131]. (**C**) AlphaFold-predicted structure of COQ8A and COQ8B with non-conserved N-terminal regions in *blue* and conserved core regions in *orange.* COQ8A and COQ8B share similar core structures. The non-conserved regions are mostly unstructured and do not affect the core structures. (**D**) Representative TOM interaction profiles of a mitochondrial gene with 2 min or 15 min of CHX treatment. Solid lines show the mean values, and shaded areas show the range of data from two biological replicates. (**E**) Heatmap of log2 TOM enrichment at each codon for cotranslational TOM substrates after DSP crosslinking. Genes are aligned to the onset of TOM interaction detected under standard conditions and sorted by the distance from the onset to the stop codon.
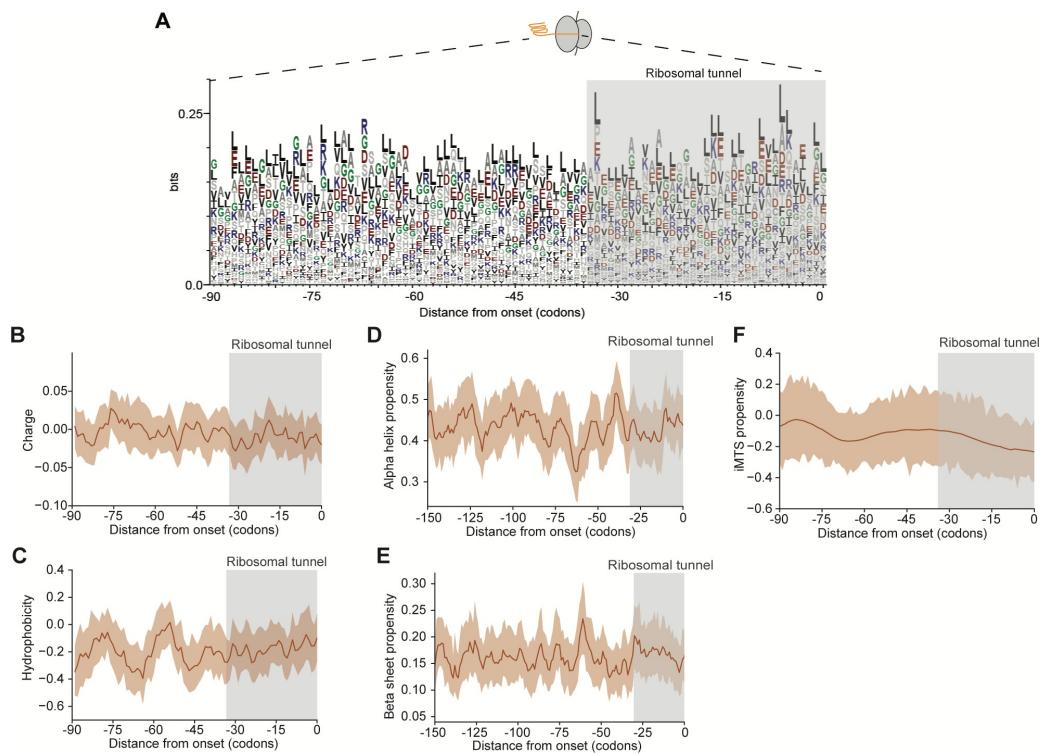
fig. S3.6. Cotranslational targeting to mitochondria is not dependent on the emergence of specific primary sequences or secondary structural features on the nascent chain.

(**A**) WebLogo representation of the sequences of C-terminal 90 amino acids that have been translated at the onset of TOM engagement. Residues to the left of the grey shaded area, which indicates the ~35 amino acids in the ribosome exit tunnel, are exposed at the onset of TOM interaction. (**B-F**) Averaged charge (B), hydrophobicity (C), α-helix (D), β-sheet (E) and internal MTS (iMTS) propensity (F) of exposed amino acids at the onset of TOM interactions. Solid lines show the mean values, and shaded areas show the 95% CI. Residues to the left of the grey shaded area, which indicates the ~35 amino acids in the ribosome exit tunnel, are exposed at the onset of TOM interaction.
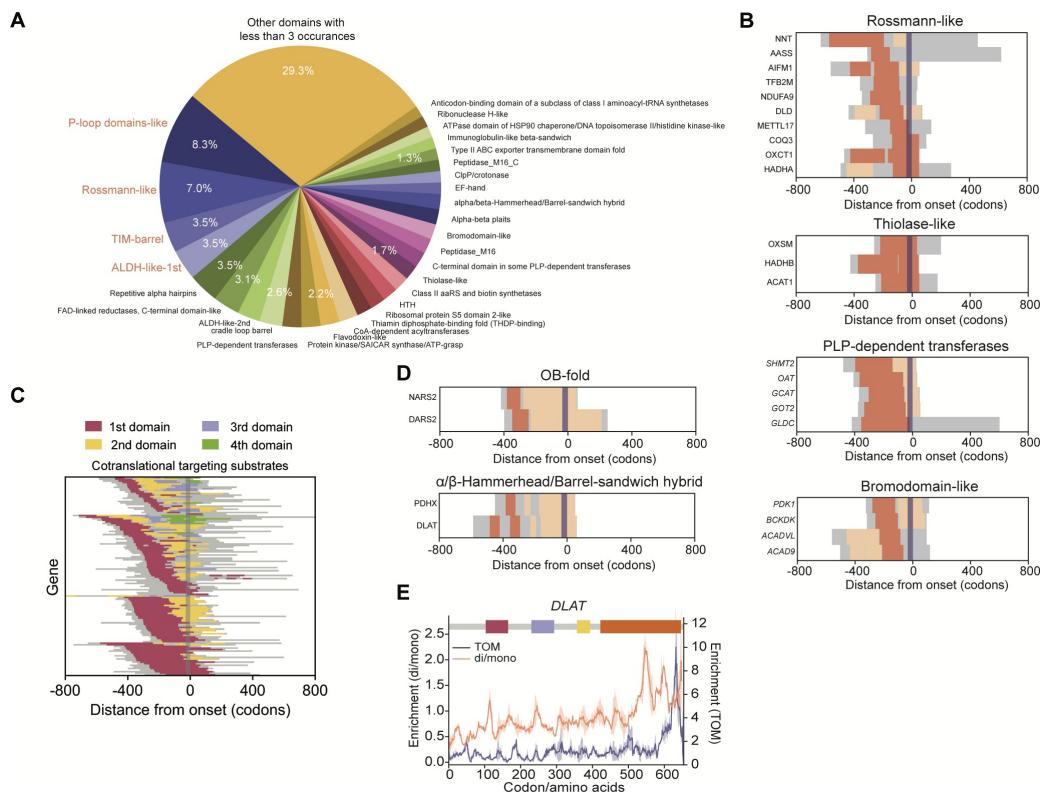
fig. S3.7. Exposure of small domains did not initiate cotranslational protein targeting to mitochondria.

(**A**) Frequencies of protein domains that have emerged from the ribosome at the onset of TOM engagement. (**B**) Proteins are grouped by shared domains and aligned to the onset of TOM interaction, with the shared domain in each group in *orange*, additional domains that have initiated translation at the onset of import in *yellow*, and the remainder of the protein in *grey*. Residues to the left of the purple shaded area, which indicates the ~35 amino acids in the ribosome exit tunnel, are exposed at the onset of TOM interaction. (**C**) Cotranslational TOM substrates are aligned to the onset of TOM interaction. Protein domains that have been translated at the onset are colored in order from the N- to the C-terminus. Genes are grouped based on the length of the first domain. When the first domain is small, cotranslational import did not initiate until a C-terminal domain or multiple domains have emerged. (**D**) As in (B), except that the shared N-terminal domains (*orange*) are small. Targeting initiated only when a large C-terminal domain is mostly emerged. (**E**) TOM interaction and co-co interaction profiles of *DLAT*. The scheme of the protein above depicts unstructured regions in grey and the individual folding

domains as solid bars in different colors. Solid lines show the mean values, and shaded areas show the range of data from two biological replicates.
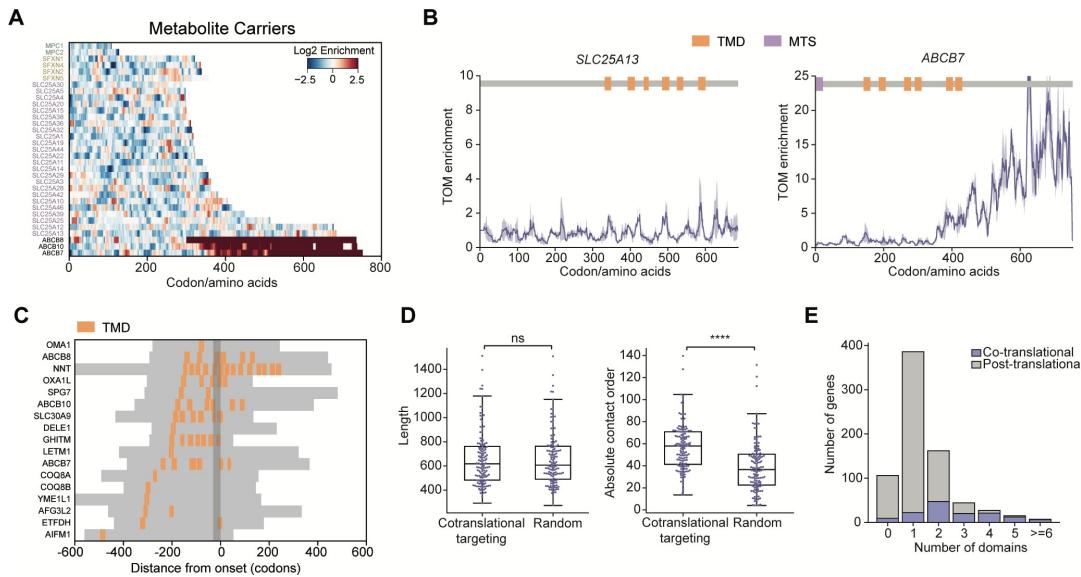
fig. S3.8. Cotranslational protein targeting to mitochondria does not correlate with the presence or number of TMDs, but with protein folding properties.

(**A**) Heatmap of log2 TOM enrichment at each codon of four families of metabolite carrier proteins (the mitochondrial pyruvate carrier (MPC), the sideroflexin proteins, the SLC25A family and the mitochondrial ABCB transporters). (**B**) TOM interaction profiles of two metabolite carrier proteins. The position of TMDs (*orange*) and MTS (*lavender*) were shown above. Solid lines show the mean values, and shaded areas show the range of data from two biological replicates. (**C**) Cotranslational TOM substrates containing predicted TMDs (*orange*) were aligned to the onset of TOM interaction. Residues to the left of the shaded area, which indicates the ~35 amino acids in the ribosome exit tunnel, are exposed at the onset of TOM interaction. (**D**) ACO values are compared between cotranslational TOM substrates and the same number of randomly sampled cytosolic proteins with the same length distribution. (**E**) Distribution of the numbers of co- versus post-translationally targeted proteins, grouped by the number of domains identified in each protein.

BIBLIOGRAPHY

1. Ingolia, N. T. Genome-Wide Analysis in Vivo of Resolution Using Ribosome Profiling. *Science (80-.).* **218**, 218–224 (2009).

2. Becker, A. H., Oh, E., Weissman, J. S., Kramer, G. & Bukau, B. Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nat. Protoc.* **8**, 2212–2239 (2013).

3. Oh, E. *et al.* Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308 (2011).

4. Saraogi, I. & Shan, S. ou. Co-translational protein targeting to the bacterial membrane. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1433–1441 (2014).

5. Cranford-Smith, T. & Huber, D. The way is the goal: How SecA transports proteins across the cytoplasmic membrane in bacteria. *FEMS Microbiol. Lett.* **365**, 1–16 (2018).

6. Walter, P., Ibrahimi, I. & Blobel, G. Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J. Cell Biol.* **91**, 545–550 (1981).

7. Schibich, D. *et al.* Global profiling of SRP interaction with nascent polypeptides. *Nature* **536**, 219–223 (2016).

8. Zhang, X. & Shan, S. Fidelity of Cotranslational Protein Targeting by the Signal Recognition Particle. *Annu. Rev. Biophys.* **43**, 381–408 (2014).

9. Schaffitzel, C. *et al.* Structure of the E. coli signal recognition particle bound to a translating ribosome. *Nature* **444**, 503–506 (2006).

10. Cabelli, R. J., Chen, L., Tai, P. C. & Oliver, D. B. SecA protein is required for secretory protein translocation into E. coli membrane vesicles. *Cell* **55**, 683–692 (1988).

11. Knyazev, D. G., Kuttner, R., Zimmermann, M., Sobakinskaya, E. & Pohl, P. Driving Forces of Translocation Through Bacterial Translocon SecYEG. *J. Membr. Biol.* **251**, 329–343 (2018).

12. Allen, W. J. *et al.* Two-way communication between SecY and SecA suggests a brownian ratchet mechanism for protein translocation. *Elife* **5**, 1–23 (2016).

13. Bauer, B. W., Shemesh, T., Chen, Y. & Rapoport, T. A. A 'push and slide' mechanism allows sequence-insensitive translocation of secretory proteins by the SecA ATPase. *Cell* **157**, 1416–1429 (2014).

14. Gupta, R., Toptygin, D. & Kaiser, C. M. The SecA motor generates mechanical force during protein translocation. *Nat. Commun.* **11**, 1–11 (2020).

15. Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P. & Wickner, W. The binding cascade of SecB to SecA to SecY E mediates preprotein targeting to the E. coli plasma membrane. *Cell* **63**, 269–279 (1990).

16. Weiss, J. B., Ray, P. H. & Bassford, P. J. Purified SecB protein of Escherichia coli retards folding and promotes membrane translocation of the maltose-binding protein in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8978–8982 (1988).

17. Castanié-Cornet, M. P., Bruel, N. & Genevaux, P. Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1442–1456 (2014).

18. De Geyter, J. *et al.* Trigger factor is a bona fide secretory pathway chaperone that interacts with SecB and the translocase. *EMBO Rep.* 1–17 (2020). doi:10.15252/embr.201949054

19. Pfanner, N., Warscheid, B. & Wiedemann, N. Mitochondrial proteins: from biogenesis to functional networks. *Nat. Rev. Mol. Cell Biol.* **20**, 267–284 (2019).

20. Becker, T., Song, J. & Pfanner, N. Versatility of Preprotein Transfer from the Cytosol to Mitochondria. *Trends Cell Biol.* **29**, 534–548 (2019).

21. Bykov, Y. S., Rapaport, D., Herrmann, J. M. & Schuldiner, M. Cytosolic Events in the Biogenesis of Mitochondrial Proteins. *Trends Biochem. Sci.* **45**, 650–667 (2020).

22. Cohen, B., Golani-Armon, A. & Arava, Y. S. Emerging implications for ribosomes in proximity to mitochondria. *Semin. Cell Dev. Biol.* (2023). doi:10.1016/j.semcdb.2023.01.003

23. Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490.e26 (2019).

24. Gadir, N., Haim-Vilmovsky, L., Kraut-Cohen, J. & Gerst, J. E. Localization of mRNAs coding for mitochondrial proteins in the yeast Saccharomyces cerevisiae. *RNA* **17**, 1551–1565 (2011).

25. Gold, V. A., Chroscicki, P., Bragoszewski, P. & Chacinska, A. Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Rep.* **18**, 1786–1800 (2017).

26. Williams, C. C., Jan, C. H. & Weissman, J. S. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science (80-.).* **346**, 748–751 (2014).

27. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. M. & Wickner, W. The purified E. coli integral membrane protein SecY E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62**, 649–657 (1990).

28. Schiebel, E., Driessen, A. J. M., Hartl, F.-U. & Wickner, W. ΔμH+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64**, 927–939 (1991).

29. Corey, R. A. *et al.* Specific cardiolipin–SecY interactions are required for proton-motive force stimulation of protein secretion. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7967–7972 (2018).

30. Rapoport, T. A., Li, L. & Park, E. Structural and mechanistic insights into protein translocation. *Annu. Rev. Cell Dev. Biol.* **33**, 369–390 (2017).

31. Tsukazaki, T. *et al.* Structure and function of a membrane component SecDF that enhances protein export. *Nature* **474**, 235–238 (2011).

32. Lee, H. C. & Bernstein, H. D. Trigger factor retards protein export in Escherichia coli. *J. Biol. Chem.* **277**, 43527–43535 (2002).

33. Ullers, R. S., Ang, D., Schwager, F., Georgopoulos, C. & Genevaux, P. Trigger factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. *Proc. Natl.*

Acad. Sci. U. S. A. **104**, 3101–3106 (2007).

34.   Huber, D. *et al.* SecA Interacts with Ribosomes in Order to Facilitate Posttranslational Translocation in Bacteria. *Mol. Cell* **41**, 343–353 (2011).

35.   Wang, S. *et al.* The molecular mechanism of cotranslational membrane protein recognition and targeting by SecA. *Nat. Struct. Mol. Biol.* **26**, 919–929 (2019).

36.   Deitermann, S., Sprie, G. S. & Koch, H. G. A dual function for SecA in the assembly of single spanning membrane proteins in Escherichia coli. *J. Biol. Chem.* **280**, 39077–39085 (2005).

37.   Neumann-Haefelin, C., Schäfer, U., Müller, M. & Koch, H. G. SRP-dependent co-translational targeting and SecA-dependent translocation analyzed as individual steps in the export of a bacterial protein. *EMBO J.* **19**, 6419–6426 (2000).

38.   Van Bloois, E., Haan, G. J., De Gier, J. W., Oudega, B. & Luirink, J. Distinct requirements for translocation of the N-tail and C-tail of the Escherichia coli inner membrane protein CyoA. *J. Biol. Chem.* **281**, 10002–10009 (2006).

39.   Van Der Laan, M., Nouwen, N. & Driessen, A. J. M. SecYEG Proteoliposomes Catalyze the Δψ-Dependent Membrane Insertion of FtsQ. *J. Biol. Chem.* **279**, 1659–1664 (2004).

40.   Saaf, A., Andersson, H., Gafvelin, G. & Heijnet, G. von. SecA-dependence of the translocation of a large periplasmic loop in the escherichia coli malf inner membrane protein is a function of sequence context. *Mol. Membr. Biol.* **12**, 209–215 (1995).

41.   Wang, S., Yang, C. I. & Shan, S. O. SecA mediates cotranslational targeting and translocation of an inner membrane protein. *J. Cell Biol.* **216**, 3639–3653 (2017).

42.   Rawat, S., Zhu, L., Lindner, E., Dalbey, R. E. & White, S. H. SecA Drives Transmembrane Insertion of RodZ, an Unusual Single-Span Membrane Protein. *J. Mol. Biol.* **427**, 1023–1037 (2015).

43.   Li, L. *et al.* Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **531**, 395–399 (2016).

44.   Park, E. *et al.* Structure of the SecY channel during initiation of protein translocation. *Nature* **506**, 102–106 (2014).

45.   Ma, C. *et al.* Structure of the substrate-engaged SecA-SecY protein translocation machine. *Nat. Commun.* **10**, 1–9 (2019).

46.   Duc, K. D., Batra, S. S., Bhattacharya, N., Cate, J. H. D. & Song, Y. S. Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res.* **47**, 4198–4210 (2019).

47.   Josefsson, L. G. & Randall, L. L. Different exported proteins in E. coli show differences in the temporal mode of processing in vivo. *Cell* **25**, 151–157 (1981).

48.   Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

49.   Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936–943 (2008).

50.   Bonardi, F. *et al.* Probing the SecYEG translocation pore size with preproteins conjugated

with sizable rigid spherical molecules. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7775–7780 (2011).

51. Randall, L. L. & Hardy, S. J. S. Correlation of competence for export with lack of tertiary structure of the mature species: A study in vivo of maltose-binding protein in E. coli. *Cell* **46**, 921–928 (1986).

52. Schierle, C. F. *et al.* The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the Escherichia coli periplasm via the signal recognition particle pathway. *J. Bacteriol.* **185**, 5706–5713 (2003).

53. Ivankov, D. N. *et al.* Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.* **12**, 2057–2062 (2003).

54. Kumamoto, C. A. & Francetic, O. Highly selective binding of nascent polypeptides by an Escherichia coli chaperone protein in vivo. *J. Bacteriol.* **175**, 2184–2188 (1993).

55. Randall, L. L. *et al.* Binding of SecB to ribosome-bound polypeptides has the same characteristics as binding to full-length, denatured proteins. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 802–807 (1997).

56. Breukink, E. *et al.* The C terminus of SecA is involved in both lipid binding and SecB binding. *J. Biol. Chem.* **270**, 7902–7907 (1995).

57. Baars, L. *et al.* Defining the role of the Escherichia coli chaperone SecB using comparative proteomics. *J. Biol. Chem.* **281**, 10024–10034 (2006).

58. Seinen, A. B., Spakman, D., Oijen, A. M. Van & Driessen, A. J. M. Cellular dynamics of the SecA ATPase at the single molecule level. *Sci. Rep.* 1–16 (2021). doi:10.1038/s41598-021-81081-2

59. Huber, D. *et al.* SecA Cotranslationally Interacts with Nascent Substrate Proteins In Vivo. *J. Bacteriol.* **199**, 1–14 (2017).

60. Wu, Z. C., De Keyzer, J., Kedrov, A. & Driessen, A. J. M. Competitive binding of the SecA ATPase and ribosomes to the SecYEG translocon. *J. Biol. Chem.* **287**, 7885–7895 (2012).

61. Catipovic, M. A. & Rapoport, T. A. Protease protection assays show polypeptide movement into the SecY channel by power strokes of the SecA ATPase. *EMBO Rep.* 1–8 (2020). doi:10.15252/embr.202050905

62. Allen, W. J., Watkins, D. W., Dillingham, M. S. & Collinson, I. Refined measurement of SecA-driven protein secretion reveals that translocation is indirectly coupled to ATP turnover. *Proc. Natl. Acad. Sci.* **117**, 31808–31816 (2020).

63. Dai, X. *et al.* Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nat. Microbiol.* **2**, 16231 (2017).

64. Corey, R. A. *et al.* ATP-induced asymmetric pre-protein folding as a driver of protein translocation through the sec machinery. *Elife* **8**, 1–25 (2019).

65. De Los Rios, P., Ben-Zvi, A., Slutsky, O., Azem, A. & Goloubinoff, P. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6166–6171 (2006).

66. Antonoaea, R., Fürst, M., Nishiyama, K. I. & Müller, M. The periplasmic chaperone PpiD

interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47**, 5649–5656 (2008).

67. Schäfer, U., Beck, K. & Müller, M. Skp, a molecular chaperone of Gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J. Biol. Chem.* **274**, 24567–24574 (1999).

68. Shimohata, N., Akiyama, Y. & Ito, K. Peculiar properties of DsbA in its export across the Escherichia coli cytoplasmic membrane. *J. Bacteriol.* **187**, 3997–4004 (2005).

69. Crane, J. M. & Randall, L. L. The Sec System: Protein Export in Escherichia coli. *EcoSal Plus* **7**, (2017).

70. Randall, L. L. Translocation of domains of nascent periplasmic proteins across the cytoplasmic membrane is independent of elongation. *Cell* **33**, 231–240 (1983).

71. Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. & Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**, 693–696 (1999).

72. Kramer, G. *et al.* L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419**, 171–174 (2002).

73. Tsirigotaki, A., De Geyter, J., Šoštarić, N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**, 21–36 (2017).

74. Chatzi, K. E. *et al.* Preprotein mature domains contain translocase targeting signals that are essential for secretion. *J. Cell Biol.* **216**, 1357–1369 (2017).

75. Steinberg, R., Knüpffer, L., Origi, A., Asti, R. & Koch, H. G. Co-translational protein targeting in bacteria. *FEMS Microbiol. Lett.* **365**, 1–15 (2018).

76. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).

77. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

78. Mohammad, F. & Buskirk, A. Protocol for Ribosome Profiling in Bacteria. *BIO-PROTOCOL* **9**, 139–148 (2019).

79. Mohammad, F., Green, R. & Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife* **8**, 1–25 (2019).

80. Busch, J. D., Fielden, L. F., Pfanner, N. & Wiedemann, N. Mitochondrial protein transport: Versatility of translocases and mechanisms. *Mol. Cell* **83**, 890–910 (2023).

81. Wiedemann, N. & Pfanner, N. Mitochondrial Machineries for Protein Import and Assembly. *Annu. Rev. Biochem.* **86**, 685–714 (2017).

82. Araiso, Y., Imai, K. & Endo, T. Role of the TOM Complex in Protein Import into Mitochondria: Structural Views. *Annu. Rev. Biochem.* **91**, 679–703 (2022).

83. Araiso, Y. *et al.* Structure of the mitochondrial import gate reveals distinct preprotein paths. *Nature* **575**, 395–401 (2019).

84. Tucker, K. & Park, E. Cryo-EM structure of the mitochondrial protein-import channel

TOM complex at near-atomic resolution. *Nat. Struct. Mol. Biol.* **26**, 1158–1166 (2019).

85. Lee-glover, L. & Shutt, T. Mitochondrial Quality Control Pathways Sense Mitochondrial Protein Import. *Trends Endocrinol. Metab.* **06**, 1–13 (2023).

86. Sekine, Y. *et al.* A mitochondrial iron-responsive pathway regulated by DELE1. *Mol. Cell* **83**, 2059-2076.e6 (2023).

87. Fessler, E., Krumwiede, L. & Jae, L. T. DELE1 tracks perturbed protein import and processing in human mitochondria. *Nat. Commun.* **13**, (2022).

88. Johnson, Z. L. *et al.* Structural basis of the acyl-transfer mechanism of human GPAT1. *Nat. Struct. Mol. Biol.* **30**, 22–30 (2023).

89. Fielden, L. F. *et al.* Central role of Tim17 in mitochondrial presequence protein translocation. *Nature* **621**, 627–634 (2023).

90. Zhou, X. *et al.* Molecular pathway of mitochondrial preprotein import through the TOM–TIM23 supercomplex. *Nat. Struct. Mol. Biol.* (2023). doi:10.1038/s41594-023-01103-7

91. Bertolini, M. *et al.* Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science (80-.).* **371**, 57–64 (2021).

92. Deuerling, E., Gamerdinger, M. & Kreft, S. G. Chaperone interactions at the ribosome. *Cold Spring Harb. Perspect. Biol.* **11**, 1–20 (2019).

93. M, G. *et al.* The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science (80-.).* **348**, 201–7 (2015).

94. Jomaa, A. *et al.* Mechanism of signal sequence handover from NAC to SRP on ribosomes during ER-protein targeting. *Science (80-.).* **375**, 839–844 (2022).

95. Wiedmann, B., Sakai, H., Davis, T. A. & Wiedmann, M. A protein complex required for signal-sequence-specific sorting and translocation. *Nature* **370**, 434–440 (1994).

96. Lesnik, C., Cohen, Y., Atir-Lande, A., Schuldiner, M. & Arava, Y. OM14 is a mitochondrial receptor for cytosolic ribosomes that supports co-translational import into mitochondria. *Nat. Commun.* **5**, (2014).

97. Gamerdinger, M. *et al.* Early Scanning of Nascent Polypeptides inside the Ribosomal Tunnel by NAC. *Mol. Cell* 1–11 (2019). doi:10.1016/j.molcel.2019.06.030

98. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat. Methods* **16**, 866–869 (2019).

99. Powers, T. & Walter, P. The nascent polypeptide-associated complex modulates interactions between the signal recognition particle and the ribosome. *Curr. Biol.* **6**, 331–338 (1996).

100. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).

101. Palmiter, R. D. Quantitation of parameters that determine the rate of ovalbumin synthesis. *Cell* **4**, 189–197 (1975).

102. Backes, S. *et al.* Tom70 enhances mitochondrial preprotein import efficiency by binding

to internal targeting sequences. *J. Cell Biol.* **217**, 1369–1382 (2018).

103. Bykov, Y. S. *et al.* Widespread use of unconventional targeting signals in mitochondrial ribosome proteins. *EMBO J.* **41**, 1–15 (2022).

104. Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. Chain length scaling of protein folding time. *Phys. Rev. Lett.* **77**, 5433–5436 (1996).

105. Garbuzynskiy, S. O., Ivankov, D. N., Bogatyreva, N. S. & Finkelstein, A. V. Golden triangle for folding rates of globular proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 147–150 (2013).

106. Mallik, S. & Kundu, S. Topology and Oligomerization of Mono- and Oligomeric Proteins Regulate Their Half-Lives in the Cell. *Structure* **26**, 869-878.e3 (2018).

107. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A. & Clarke, J. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* **8**, 319–330 (2007).

108. Akopian, D., Shen, K., Zhang, X. & Shan, S. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annu. Rev. Biochem.* **82**, 693–721 (2013).

109. Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science (80-.).* **346**, 1257521–1257521 (2014).

110. Zhu, Z., Wang, S. & Shan, S. Ribosome profiling reveals multiple roles of SecA in cotranslational protein export. *Nat. Commun.* **13**, 3393 (2022).

111. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

112. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1531–1536 (2011).

113. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

114. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

115. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

116. Rath, S. *et al.* MitoCarta3.0: An updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).

117. Bateman, A. *et al.* UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

118. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).

119. Høie, M. H. *et al.* NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res.* **50**, W510–W515 (2022).

120. Schneider, K., Zimmer, D., Nielsen, H., Herrmann, J. M. & Mühlhaus, T. IMLP, a predictor for internal matrix targeting-like sequences in mitochondrial proteins. *Biol. Chem.* **402**, 937–943 (2021).

121. Fukasawa, Y. *et al.* MitoFates: Improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteomics* **14**, 1113–1126 (2015).

122. Armenteros, J. J. A. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, 1–14 (2019).

123. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* 2022.04.08.487609 (2022).

124. Manriquez-Sandoval, E. & Fried, S. D. DomainMapper: Accurate domain structure annotation including those with non-contiguous topologies. *Protein Sci.* **31**, 1–13 (2022).

125. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

126. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

127. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

128. Richter-Dennerlein, R. *et al.* DNAJC19, a mitochondrial cochaperone associated with cardiomyopathy, forms a complex with prohibitins to regulate cardiolipin remodeling. *Cell Metab.* **20**, 158–171 (2014).

129. Sharma, A., Mariappan, M., Appathurai, S. & Hegde, R. S. In Vitro Dissection of Protein Translocation into the Mammalian Endoplasmic Reticulum. in *J Neurochem* (ed. Economou, A.) **619**, 339–363 (Humana Press, 2010).

130. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

131. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 (2014).