Dissertation

submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M.Sc. Kai Fenzl
born in Langen (Hessen), Germany
Oral examination: 19.05.2021

Disome selective profiling identifies nascent chain dimerization as a prevalent co-translational assembly mechanism in human cells

Referees:

Prof. Dr. Bernd Bukau

Prof. Dr. Dr. Georg Stöcklin

Contributions

Parts of this dissertation were performed in the context of a shared project together with Matilde Bertolini. All experiments performed by Matilde Bertolini are clearly indicated throughout this written thesis, if not stated otherwise, the data are the result of my own work. All data analysis scripts were developed in an iterative process that included contributions of Matilde Bertolini, myself and of Ilia Kats, who developed additional data analysis tools tailored for this project, which are also indicated as such. Frank Tippmann performed the analysis of crystal structures in context of co-translational assembly. Günter Kramer and Bernd Bukau supervised this PhD project.

Acknowledgment

The following page is dedicated to a great amount of people that I want to thank, many of which were essential for the success of my PhD adventure.

First of all, I want to thank Bernd Bukau for the opportunity to work in an environment that made me a better scientist, a better person and broadened my horizon. I was always fascinated by your enthusiasm for science and your support for young scientists.

Throughout my daily work I was constantly supported and supervised by Günter Kramer. This PhD would have lasted 10 years and would have contain only half the results (and mostly technical details) without his constant input of good ideas, suggestions and hundreds of hours of critical discussions.
I sincerely thank Matilde Bertolini, the best team mate that I could have wished for, who additionally became a great friend. My whole PhD experience would have been only half the fun and productivity without our team spirit. I believe that the trust and respect we have established between us is something unique, something that is so precious and fruitful that I will try each day to recreate this in the next step of my scientific career. Without our constant effort to maintain a productive team there would be no published co-co assembly paper during my stay in the Bukau lab. I know that I can be a pain in the neck and I really want to thank Günter and Mati for all the kindness and respect.

Of course, there are many other people from the Bukau lab that I have to thank sincerely. Ulrike Friedrich was the person who introduced me to ribosome profiling during a lab rotation in 2015. She evoked my fascination for this technique and gave me a kickstart in the beginning of my PhD by helping to establish the first mammalian data analysis pipeline. In this regard, I also want to sincerely thank Ilia Kats for making me realize multiple times that even after three years of handling deep sequencing data, I had no clue about data analysis.

One other important factor where I received constant help was the recurring amount of bureaucracy. I am convinced that I would have been on no conference and would have never received any salary without Jutta Rami. Big thanks to your constant help and patience with all of us. I also want to thank Martina Galvan and Rolf Lutz for the organization and coordination of all the HBIGS related bureaucracy, career support, additional trainings and HBIGS for partly funding my PhD project.

I am also truly thankful to all former and present members of the Bukau lab for their support, sharing knowledge and creating a great working atmosphere. Especially:
Aseem, Carla, Dorina, Eliana, Frank, Gamze, George, Igor, Ilgin, Jaro, Jessica, Jiri, Josef, Kristina, Lena, Oli, Regina, Shiran and Svenja.

I also want to a thank my Thesis Advisory Committee (TAC), consisting of Georg Stöcklin as my second referee and Claudio Joazeiro. You both were really supportive with your advices and all your criticism was really constructive. You made it possible that I enjoyed all of my TAC meetings. I also want to thank Mathias Mayer for the constant stream of critical questions in all our research seminars.

In the end, there are also people outside the ZMBH that had a major impact on my success. I want to thank my family and friends that believed in me and supported me throughout this journey. The last person that has to be mentioned by name and sincerely thanked is Elisa Kreibich, the best girlfriend a Kai can wish for. Without you at my side, I would have not managed to maintain a proper work-life balance, would have presumably died on a candy overdose, would have lived in a cold little apartment without plants nor love and would have missed a million wonderful moments.

Thank you all for your support! It will not be forgotten!

Table of Contents

1. Summary

1.1 English Summary

The formation of multi-protein complexes is a key feature of the cellular proteome in all kingdoms of life. The biogenesis of protein complexes *in vivo* is still poorly understood, but recent methodological advances now make it possible to reveal the underlying mechanisms. One milestone method, termed Selective Ribosome Profiling (SeRP), allowed to demonstrate the omnipresence of co-translational assembly in bacteria and yeast (Shieh *et al.* 2015, Shiber *et al.* 2018). This method provides codon-resolved information about heterodimer formation between already completed and folded proteins and their nascent partner subunits (termed co-post assembly). The fact that assembly can occur during translation raised the question whether co-translational assembly could also involve interaction between two nascent proteins translated by two neighboring ribosomes (termed co-co assembly). So far, indirect evidence suggested co-co assembly of only a few protein complexes. However, direct evidence that two ribosome-nascent chain complexes interact via their nascent chains is still scarce and we lack any information about the prevalence of this proposed process.

This dissertation focused on the investigation of the hypothesized co-co assembly mode. In collaboration with Matilde Bertolini (PhD student in the Bukau lab), we first developed an unbiased, proteome-wide screen based on ribosome profiling (Disome Selective Profiling, DiSP), to reveal the prevalence of co-co assembly in human cells. By applying DiSP to HEK293-T and U2OS cells, we identified hundreds of high confidence co-co assembling nascent proteins. Our proteome-wide data suggest that up to 30% of all annotated homomer subunits employ co-co assembly, most frequently induced by the formation of N-terminal coiled coils (mostly partially exposed at assembly onset) or interactions of well-known globular dimerization domains (that are generally fully exposed at assembly onset). We further show that co-co assembly of two human homodimeric candidates can be recapitulated in bacteria, in the absence of any eukaryote specific machinery. This suggests that assembly is solely facilitated by the intrinsic propensities of the nascent proteins to form quaternary structures. In addition, we validate the existence of co-co assembly also for endogenous *E. coli* proteins by DiSP.

The main outcome of this dissertation is the demonstration of co-co assembly as a mechanism mainly employed for homomer formation. Our experimental data indicate the high prevalence of co-co assembly in human cells to ensure productive protein complex biogenesis in the

crowded cytosol of cells. Initial findings suggest that co-co assembly is coordinated by a general transient slow-down of translation at the onset of assembly.

1.2 Deutsche Zusammenfassung

Die Mehrheit alle Proteine in Prokaryoten und Eukaryoten bilden Proteinkomplexe aus mehreren Proteinuntereinheiten. Die Proteinbiosynthese solcher Proteinkomplexe ist bisher nur sehr oberflächlich verstanden. Ein Meilenstein in der Untersuchung der Biogenese von Proteinkomplexen - war die Entwicklung und Anwendung der Methode Selective Ribosome Profiling (SeRP). Mit dieser Methode konnte bewiesen werden, dass die Bildung einiger heteromer Proteinkomplexe in Bakterien und in Hefe an die Translation gekoppelt ist (co-translationale Komplexbildungen). Die Assemblierung dieser Komplexe basiert auf der Interaktion einer vollständig synthetisierten und wahrscheinlich gefalteten Untereinheit mit einer naszierenden Untereinheit (co-post Assemblierung, Shieh *et al.* 2015, Shiber *et al.* 2018). Die Tatsache, dass Komplexbildung schon während der Translation direkt am Ribosom stattfinden kann, führte zu der Frage, ob und wie häufig Komplexbildung auch durch die Interaktion zweier naszierender Polypeptidketten stattfinden kann (co-co Komplexbildung). Zwar wurden bereits Einzelfälle beschrieben, allerdings fehlten direkte Beweise und Information darüber, wie häufig dieser Mechanismus die Komplexbildung unterstützt.

Der Focus dieser Dissertation lag auf der Untersuchung der co-co Komplexbildung. In Zusammenarbeit mit Matilde Bertolini (Doktorandin im Bukau Labor), wurde zuerst eine neue, Proteom-weite Detektionsmethode basierend auf Ribosome Profiling in menschlichen Zellen etabliert (Disome Selective Profiling, DiSP). Mit Hilfe von DiSP kann zum einen die Häufigkeit von co-co Komplexbildung in menschlichen Zellen bestimmt werden und zum anderen untersucht werden, wann während der Translation die Interaktion zweier naszierender Polypeptide erfolgt. Wir konnten hunderte Proteine in HEK293-T und U20S Zellen identifizierten, die klare Anzeichen für co-co Komplexbildung über ausschließlich naszierende Polypeptidkette aufzeigen. Unsere Daten zeigen, dass etwa 30% aller annotierten menschlichen Homo-Oligomere durch co-co Komplexbildung gebildet werden. In den meisten Fällen basiert die Assemblierung auf der Interaktion vollständig oder teilweise exponierter N-terminaler coiled coils, oder der Interaktion von bekannten globulären Dimerisierungsdomänen, die zum Zeitpunkt der Komplexbildung meist vollständig synthetisiert sind. Wir konnten darüber hinaus zeigen, dass zwei humane Homodimer-Kandidaten in *E. coli* co-co assemblieren können und sich damit unabhängig von

eukaryotischen Faktoren bilden. Dieses Ergebnis zeigt, dass die Komplexbildung nur von intrinsischen Eigenschaften der naszierenden Polypeptidkette abhängt.

Diese Dissertation beweist die Existenz und Prävalenz der co-co Komplexbildung und zeigt, dass dieser Mechanismus hauptsächlich zur Bildung von Homo-Oligomeren genutzt wird. Weitere Ergebnisse der Arbeit deuten darauf hin, dass die Komplexbildung im dicht gedrängten Zytosol menschlicher Zellen durch die kontrollierte transiente Verlangsamung der Translation zum Zeitpunkt der Komplexassemblierung koordiniert sein könnte.

2. Introduction

2.1 General protein biogenesis in eukaryotes

Proteins are vital components of a cell implicated in all major biological functions, including their own biosynthesis. The transcription of DNA into RNA in eukaryotes takes place inside the nucleus. The RNA polymerase transcribes DNA into a messenger RNA (mRNA) that contains genomic coding (exons) and non-coding (introns) sequences. Introns are removed inside the nucleus by a series of mRNA processing steps, known as splicing, to obtain a protein coding mRNA. In a final step, spliced mRNAs receive a methylated guanine at the 5' end and are polyadenylated at the 3' end, before they are exported from the nucleus through nuclear pore complexes into the cytoplasm. The exported mRNA nucleotide sequence is then translated into a linear string of amino acids by ribosomes, macromolecules consisting of more than 80 different proteins scaffolded by four ribosomal RNAs (rRNA). The ribosome is arranged in two subunits, the 40S small ribosomal subunit, and the 60S large ribosomal subunit. Translation is divided into three steps: initiation, elongation and termination. Initiation involves the 40S ribosomal subunit pre-loaded with a methionine coupled transfer RNA (tRNA), this complex scans along the mRNA until it reaches a start codon. From that point on, the large 60S ribosomal subunit joins in and the elongation starts with the stepwise movement of the translating ribosome by three nucleotides (one codon). During elongation, specific amino acid charged tRNAs bind to codons on the mRNA and thereby deliver amino acids to the ribosome. The ribosome harbors three tRNA binding sites, one aminoacyl- (A), peptidyl- (P) and exit-site (E). After selecting the right tRNA at the A-site, the ribosome mediates the formation of a peptide bond between the adjacent amino acids in A- and P-site to elongate the growing nascent peptide. This results in the gradual synthesis of a linear nascent polypeptide, which vectorially emerges from the ribosomal exit tunnel into the cytoplasm. The 80S ribosomal exit tunnel has a length of approx. 80 Å (Angstrom, 10^{-10} m) and can harbor around 30-40 amino acids of the nascent chain, depending on its compaction (Duc *et al.*, 2019). Once the 80S ribosome encounters a stop codon, termination occurs. Each step of translation is coordinated by initiation, elongation and release factors that are recruited to the translating ribosome.

In a human cell, around 3 million ribosomes (Duncan and Hershey 1983) actively translate more than 10 000 different kinds of mRNAs. The mRNA segment that is translated is called coding sequence (CDS), which is flanked by untranslated regions (UTRs) on both sides.

In order to obtain a functional protein, a linear polypeptide chain (primary structure) has to fold into a native protein with a defined, three-dimensional structure **(Fig. 1)**. The amino acids are connected by peptide bonds that are formed by the reaction of the amino group and the carboxyl group of neighboring amino acids in the peptidyl-transferase center (PTC) of the ribosome. The linear polypeptide exposes two different ends, one containing a free amino group (N-terminus) and the other a free carboxyl-group (C-terminus). The order of amino acids dictates the three-dimensional structure of the protein, including the secondary structure consisting of alpha-helixes, beta strands, and turns/loops mainly stabilized by hydrogen bounds of the amino acid backbones. The interaction of different amino acid residues leads to tertiary structure formation, which is mainly driven by the tendency to bury hydrophobic residues inside the protein structure. Most proteins form oligomers, the quaternary structure. This quaternary structure can be formed by two or more polypeptides of the same kind, leading to a homo-oligomer (homomer), or by different polypeptides, leading to the formation of a hetero-oligomer (heteromer).

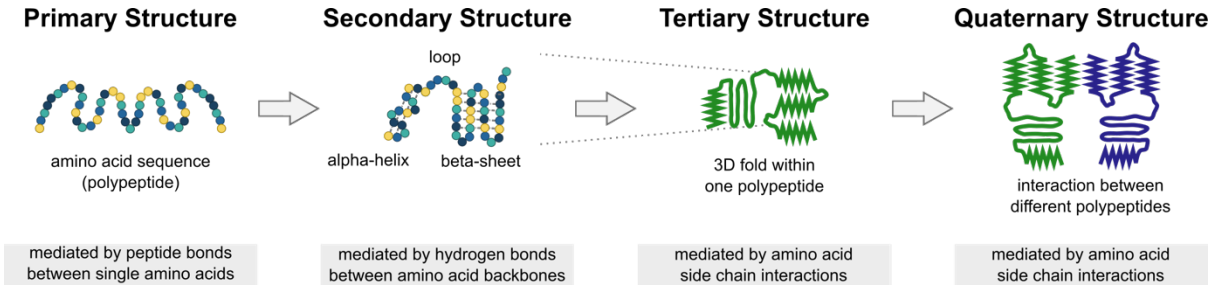

Fig. 1: The four states of protein folding. The classical view of protein folding is that a linear polypeptide forms first secondary structure element, which then interacts within one polypeptide with other secondary elements to form a 3D ternary structure. The final step is the assembly of multiple folded proteins into a protein complex (quaternary structure). Grey boxes indicate the essentially formed bonds to stabilize the respective structure.

The following chapters will focus on the co-translational events in eukaryotes, specifically in human cells, in the order of occurrence during translation to obtain a functional protein.

2.2 Co-translational maturation events during protein biogenesis in humans

Nascent polypeptides emerge vectorially (N- to C-terminus) from the ribosome and encounter the crowed cytoplasm, which encloses a high concentration of proteins and other macromolecules that have the potential to interact with the nascent protein. Such interactions can be helpful for co-translational folding, prevent further folding and keep the nascent polypeptide in an unfolded state for a certain time or lead to misfolding if the interactions are unspecific. The ribosome itself helps to prevent unspecific encounters and misfolding, by acting as a central hub to recruit a variety of co-translationally acting factors. These factors

assist various maturation processes of the nascent protein including: (1) the enzymatic modification of the nascent protein, (2) the targeting of specific ribosome-nascent chain complex (RNC) to the endoplasmic reticulum (ER) for membrane translocation, (3) the assisted folding and (4) the assembly of oligomeric complexes (**Fig. 2**) (Gloge *et al.* 2014, Kramer *et al.* 2019). These co-translational events are additionally controlled by the translating ribosome, the mRNA as well as intrinsic features of the nascent polypeptide, which will be discussed in the following chapters.

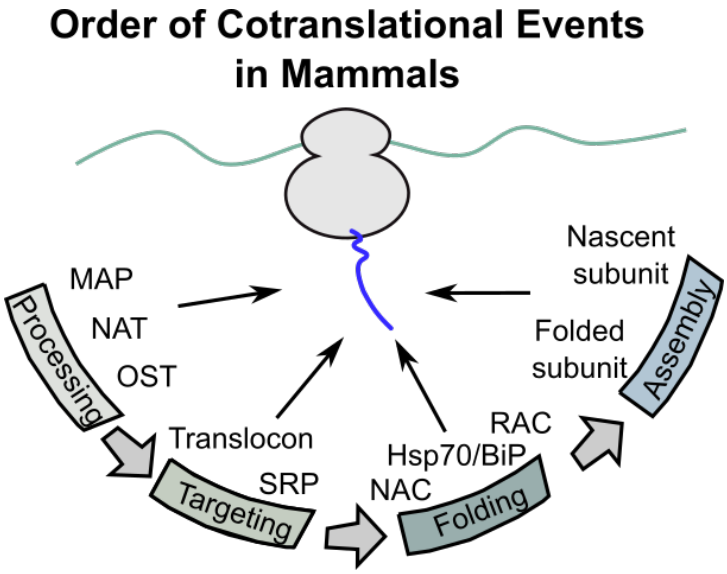

Fig. 2: Co-translational events during protein biosynthesis in mammals. Ribosomes translate an mRNA sequence (green line) into a linear polypeptide (blue line) composed of covalently connected amino acids in the cytosol. Upon emergence, the nascent polypeptide maturation can begin, which starts with the removal of the first methionine by methionine aminopeptidases (MAPs). Followed by acetylation of the most N-terminal amino acid by N-acetyltransferases (NATs). Protein translocation, facilitated by signal recognition particle (SRP) binding and transfer of the ribosome-nascent chain complex to the membrane-embedded translocon, which can be followed by another nascent chain processing step. The oligosaccharyl-transferase complex (OST) is able to couple oligosaccharides to translocated nascent proteins. Co-translational folding, can be supported by cytosolic chaperones like nascent-polypeptide-associated complex (NAC), ribosome-associated complex (RAC), and Hsp70s, including endoplasmic reticulum localized chaperones like the Hsp70 protein BiP (HSPA5). All bind the nascent chain to assist folding, keep a specific amino acid sequence unfolded or prevent interactions with other proteins. Co-translational assembly of functional oligomers can occur via assembly of ribosome-exposed oligomerization domains of proximal ribosomes (co-co assembly) on one mRNA or between co-localized mRNAs, or between a fully synthesized diffusing subunit and a nascent protein (co-post assembly).

2.2.1 Co-translational N-terminal processing in the cytosol

Ribosomes start translation at the start codon, mostly AUG, incorporating the amino acid methionine. Notably, the human proteome contains many proteins that have other amino acids than methionine at the most N-terminal end, which is accomplished by co-translational active methionine aminopeptidases (MAPs) (Kendall *et al.* 1992). Human cells encode two cytosolic

MAPs (MetAP1 and MetAP2), which cleave off the N-terminal methionine of nascent chains by hydrolysis. Both are selectively active on N-termini in which the methionine is coupled to a small and uncharged amino acid (Xiao *et al.* 2010). Therefore, at least two-thirds of the human proteome are estimated to be potential MAP substrates (Frottin *et al.* 2016). MAPs are likely the first interactors of nascent chains, followed by N-terminal acetyltransferases (NATs) catalyzing N-terminal acetylation (Kramer *et al.* 2019). In humans, seven different NAT complexes are described (Deng *et al.* 2020), which all have distinct amino acid recognition motifs. NatA, for example, acetylates non-methionine containing substrates, whereas NatB and NatC acetylate the N-terminal methionine. More than 80% of the human proteome is estimated to be N-terminally acetylated, supported by the broad range of recognition motifs (Arnesen *et al.* 2009, Deng *et al.* 2020). The role of N-acetylation is still not fully understood. Recent evidence implicates N-acetylation into the control of protein half-life, protein localization as well as protein-protein interactions (Nguyen *et al.* 2018, Ree *et al.* 2018, Friedrich *et al.* 2021).

2.2.2 Co-translational nascent protein folding and chaperone assistance

Protein folding frequently initiates co-translationally (Liutkute *et al.* 2020, Waudby *et al.* 2019) and it was recently estimated that more than one third of the *E. coli* proteome starts to fold co-translationally *in vivo* (Ciryam *et al.* 2013). The prevalence of co-translational folding may be similar or higher in eukaryotic cells. One model predicts that proteins that predominantly form local contacts, namely interactions between closely located amino acids in the primary sequence, fold co-translationally (Ciryam *et al.* 2013). Alpha-helical proteins are known to fold with a higher frequency co-translationally in comparison to proteins with more complicated mixed alpha-helical and beta-sheet structures (Ciryam *et al.* 2013). Co-translational folding of alpha-helical transmembrane domains (TMD) can already occur inside the ribosomal exit tunnel (Bañó-Polo et al. 2018) to prevent the exposure of these hydrophobic domains to the hydrophilic cytosol. While the formation of alpha-helices can already start deep inside the ribosomal exit tunnel, small tertiary structures or folding intermediates likely form only in the lower part of the tunnel near the ribosomal exit (Bhushan *et al.* 2010, Nilsson *et al.* 2015, Kosolapov and Deutsch 2009). The exit tunnel guides the folding, due to its narrow shape that restricts the folding space and directs interactions with the nascent chain. The exit tunnel is negatively charged, especially towards the distal end of the tunnel, which directly affects co-translational folding of charged nascent chains (Lu and Deutsch 2008). Emergence into the cytoplasm enlarges the available space and allows thereby the folding of bulky tertiary structures. Importantly, ribosome proximity often delays co-translational folding due to

transient interactions or repulsions between the nascent protein and the negatively charged ribosomal surface (Liutkute *et al.* 2020), suggesting a holdase activity of the ribosome itself (Liu *et al.* 2017, Farías-Rico *et al.* 2018). Ongoing translation increases the distance between the ribosome and N-terminal parts of the nascent chain and therefore lowers the impact of the ribosome on folding.

Co-translational folding processes often require the assistance of molecular chaperones, which guide nascent chains to the native tertiary structure and prevent non-specific interactions in the crowded cytosol (Gloge *et al.* 2014). Molecular chaperones are defined as "proteins that interact with, stabilize or help a non-native protein to acquire its native conformation, but are not present in the final functional structure" (Hartl *et al.* 2009). One major class of chaperones are Hsp70s, which reside in the cytosol, nucleus, ER and mitochondria. The human genome encodes for thirteen different Hsp70 members (Kampinga *et al.* 2009). A variety of specialized co-translational acting chaperones are known, which transiently bind directly to the vicinity of the ribosomal tunnel exit and/or the emerging nascent chain (**Fig. 3**, Hsieh *et al.* 2020, Kramer *et al.* 2019).

One of the early co-translational acting factors involved in nascent chain maturation is the heterodimeric nascent-chain associated complex (NAC), that binds to eukaryotic ribosomes in the proximity to the ribosomal tunnel exit to bind newly synthesized polypeptide chains as they emerge from the ribosome (Gamerdinger *et al.* 2019). NAC is stoichiometric to ribosomes in yeast and likely also in humans (del Alamo *et al*. 2011, Raue *et al*. 2007) and should be able to bind almost all ribosomes. Yeast NAC is among the first ribosome-associated maturation factors engaging all kinds of nascent proteins. NAC is essential in mammals but dispensable in yeast (Gamerdinger *et al.* 2019, Deng and Begring 1995). NAC contributes to substrate triaging by modulating SRP and thereby preventing mistargeting of cytosolic proteins to the translocon (Hsieh *et al.* 2020), but also acts as a chaperone involved in nascent chain folding and as a co-translational sorting factor by enhancing productive binding to mitochondria (Fünfschilling and Rospert 1999).

Another co-translational acting factor is the mammalian ribosome-associated complex (mRAC), which is a heterodimer consisting of the Hsp70 member Hsp70L1 (*HSPA14*) and the Hsp40 member MPP11 (*DNAJC2*). Mediated by MPP11, RAC directly binds to ribosomes to assist nascent chain folding (Otto *et al.* 2005). The co-translational activity of mRAC is not fully understood, but it is known that mRAC recruits HSPA1A to nascent chains, similar to the recruitment of the yeast specific Hsp70 chaperone Ssb by RAC, and stimulates the ATPase activity of Hsp70s (Jaiswal *et al.* 2011). Ssb binds a broad range of nascent chains, mostly hydrophobic and aromatic amino acid sequences, and thereby delays the onset of co-

translational folding of domains that will form the hydrophobic core (Döring *et al.* 2017). Recent evidence from the Bukau lab revealed the first nascent chain interactome of a mammalian cytosolic Hsp70 member, HSPA1A (unpublished data by Manuel Günnigmann), which has different nascent chain length requirements as compared to the Ssb in yeast. Other human Hsp70s are not known to directly bind to the ribosome, although they are described to bind to nascent chains. One co-translational active Hsp70 that engages nascent chains is the ER located HSPA5 (BiP, Grp78) that binds newly translocated proteins (Behnke *et al.* 2015). Like all Hsp70, it recognizes unfolded proteins by exposed hydrophobic amino acid sequences and thereby prevents such segments from misfolding or aggregation in the hydrophilic cytosol (Behnke *et al.* 2015).

Most ribosome-associated maturation factors are not yet fully characterized. One uncharacterized example is the mammalian protein Ebp1 that binds more than 50% of all cytosolic ribosomes, but the biological function is not yet understood (Wild *et al.* 2020, Kraushar *et al.* 2020). Considering the crowdedness at the ribosomal exit tunnel, it is clear that all these cytosolic ribosome-associated maturation factors must be coordinated and that the binding is not only dictated by binding to the ribosome itself. The exposed primary nascent chain structure, the folding state of the exposed segment and presumably the translation rate of the ribosome all contribute to the onset and duration of binding of these factors.

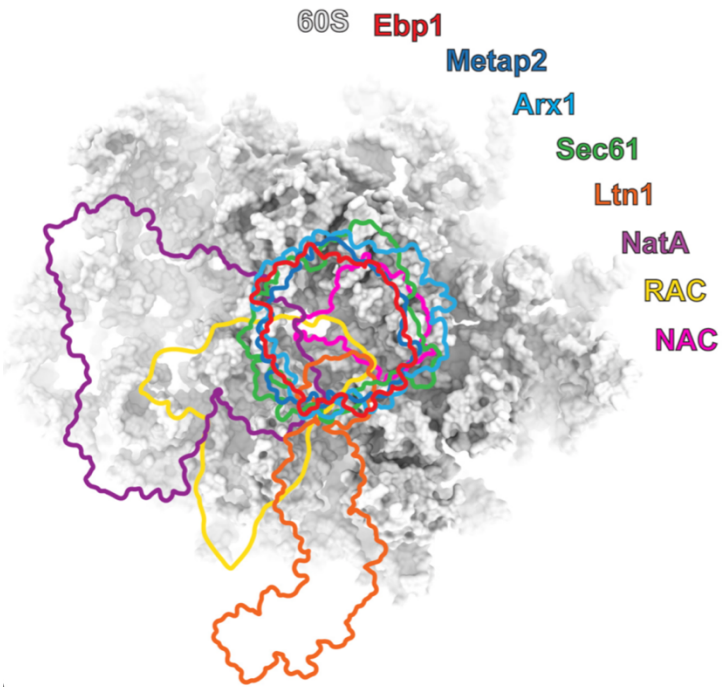

Fig. 3: Nascent chain binding factors that bind in the vicinity of the ribosomal tunnel exit and the emerging nascent chain. Grey background indicates the 60S surface from a mice ribosome. Overlapping structures of binding factors are indicated by different colors. (Adapted from Kraushar *et al.* 2020)

2.2.3 Coordination of co-translational nascent protein folding

It has become evident that the rate of translation can have a direct impact on nascent chain folding (Waudby *et al.* 2019, Jacobson and Clark 2016, Ciryam *et al.* 2013, Pechmann and Frydman 2013, O'Brien *et al.* 2012). The translation speed of the ribosome is not uniform and translation can slow-down or speed up along the coding sequence of a mRNA (Ingolia *et al.* 2009, Varenn *et al.* 1984). The average translation rate is estimated to be around five amino acids per second for eukaryotes. Folding of many fully synthesized, denatured proteins occurs within milliseconds to seconds *in vitro*, suggesting that the speed of translation is generally the rate-limiting step for folding (Kramer *et al.* 2019).

All amino acids, except methionine and tryptophan, are encoded by different synonymous codons. All organisms have a different set of frequently occurring codons in their protein coding sequences, which reflects the abundance of the respective tRNAs (Rocha 2004, Yu *et al.* 2015). It is well known that the more abundant, optimal codons are faster and more accurately translated, whereas non-optimal codons can induce translational pausing and cause more frequently mistranslation (Drummond and Wilke 2008, Pechmann and Frydman 2013). Optimal and non-optimal codons often form evolutionarily conserved clusters along the mRNA (Pechmann and Frydman 2013). Many reports revealed that changes in the translation rate caused by different synonymous codons can influence co-translational protein folding and maturation steps, indicating the co-evolution of folding and translation (Zhang *et al.* 2009, Zhang *et al.* 2011, The UniProt Consortium 2019, Jacobs *et al.* 2017). The slowing down of a ribosome can grant enough time for nascent chain maturation, local tertiary structure formation, or enable a time window for co-translational interactions. Therefore, non-optimal codons are necessary for the successful co-translational folding of protein domains (Zhang *et al.* 2009, Kramer *et al.* 2009, Wilson *et al.* 2012), but if the slow-down occurs at the wrong time it can also cause misfolding (Zhou *et al.* 2013, Kim *et al.* 2015, Yu *et al.* 2015, Buhr *et al.* 2016). Conversely, fast translation can also be beneficial as it limits the non-productive exposure of unfolded segments that require more C-terminally located amino acids to acquire a stable fold (Trovato and O'Brien 2017). Notably, optimal codons predominantly occur in beta-sheets, which are more sensitive to aggregation and require high translational fidelity (Lee *et al.* 2010, Pechmann and Frydman 2013). Taken together, the overall distribution of optimal and non-optimal codons along the coding sequence is an important feature that appears to affect the folding process profoundly.

Not only stretches of non-optimal codons are known to slow-down the rate of translation, but also mRNA secondary structures, which form a mechanical barrier to the ribosome (Kim *et al.* 2014, Chen *et al.* 2014). In addition, also interactions of the nascent chain with the ribosomal

exit tunnel or ribosome interaction of co-translationally acting factors such as SRP can lead to ribosome stalling (Wolin and Walter 1989).

Most nascent peptides experience transient pulling forces, for example by membrane integration (Ismail *et al*. 2012, Fujiwara *et al.* 2020) or co-translational folding (Goldman *et al.* 2015). Such pulling forces can increase the rate of peptide bond formation (Fritch *et al.* 2018, Nilsson *et al.* 2015), promote ribosomal frameshifting (Harrington *et al*. 2020) and even overcome translational pausing (Farías-Rico *et al.* 2018). So-called arrest peptides are special cases, in which a specific amino acid sequence interacts with the ribosomal exit tunnel, leading to ribosome stalling (Ito and Chiba 2013). Arrest peptide induced stalling can be resolved by co-translational folding near the exit tunnel, which generates a pulling force on the nascent chain (Farías-Rico *et al.* 2018). This principle of coupling folding and translation suggests that also other co-translational events may generate pulling forces. In the case of co-translational chaperone binding, it has been suggested that this can impact the translation elongation rate by pulling on the nascent chain (Rodriguez-Galan *et al.* 2021).

2.2.4 Co-translational protein translocation and maturation in the ER

Targeting of mRNAs to subcellular distinct compartments is a fundamental mechanism for localized protein synthesis, especially for proteins which require translocation into the ER. The Sec61 translocon facilitates the insertion of transmembrane proteins into the ER membrane or the translocation of soluble proteins across the membrane into the ER lumen. For human cells it is estimated that 27% of the cellular proteins contain TMDs (Almén *et al.* 2009) and approximately another 13% are secreted proteins (Uhlen *et al.* 2018). ER translocation of nascent proteins often requires the docking of the ribosome-nascent chain complex to the ER membrane. Co-translational targeting to the membrane is mainly initiated by the signal recognition particle (SRP), which recognizes hydrophobic sequences (transmembrane domains or signal sequences) within nascent chains and induces ribosomal pausing to allow the timely transport to the ER (Grudnik *et al.* 2009). This SRP induced translational pausing is released after the SRP–RNC complex docks, in a GTP-dependent interaction with the SRP receptor, to the Sec61 translocon (Miller *et al.* 1995). The nascent chain then enters the translocon and translation proceeds. Secretory and membrane proteins are not exclusively translocated by the SRP dependent pathway. SRP-independent targeting pathways to the ER were originally described as post-translational import mechanisms, since they mostly recognize C-terminally localized TMDs (Denic *et al.* 2013). However, a proximity-specific ribosome profiling study revealed that the SRP-independent targeting pathway (SND) can also function co-translationally (Aviram *et al.* 2016). This proximity suggest that the encoded

proteins are presumably not post-translationally inserted, and that additional targeting routes are active. It is now estimated that almost 60% of all secreted proteins in humans are co-translationally inserted into the ER (Jan *et al.* 2014). The guided entry of tail-anchored proteins (GET) pathway is another classical post-translational ER targeting process. However, a recent study in yeast identified two factors of this pathway, Get4 and Get5, as ribosome associated factors (Zhang *et al.* 2021). The binding site of both factors overlaps with the binding site of SRP and it seems that they capture tail-anchored proteins as soon as they are released from the ribosome and hand them over to other GET targeting components (Zhang *et al.* 2021).

The translocon itself allows, like the ribosomal exit tunnel, the formation of mostly alpha-helical secondary structures (Woolhead *et al.* 2004), but the translocon pore diameter does not allow fully folded proteins to be translocated (Mingarro *et al*. 2000), implying that post-translationally inserted proteins must be kept unfolded or require unfolding before passage.
The signal for ER-targeting of translocated proteins is a N-terminal hydrophobic region, called signal sequence, that is transported to the ER as the first step of translocation (Martoglio and Dobberstein 1998). As soon as the signal sequence is exposed into the ER lumen, it is co-translationally cleaved off by signal peptidases (Weihofen *et al.* 2002), which recognize and cleave a short sequence motif C-terminal of the signal sequence, and thereby release the nascent chain from the hydrophobic signal sequence. Further co-translational nascent chain modifications occur inside the ER. The oligosaccharyltransferase complex (OST), for example, is an integral component of the Sec61 translocon responsible for co-translational N-glycosylation of nascent proteins (Ruiz-Canada *et al.* 2009). OST transfers a core oligosaccharide consisting of glucose, mannose and N-acetylglucosamine to an asparagine residue in the polypeptide chain. The translocation of nascent chains into the ER is coupled to additional modifications which in parts can happen co-translationally. The most prominent and abundant modification is the formation of disulfide bonds, between cysteines of one nascent chain or of two nascent chains (Bergman and Kuehl 1979). The majority of ER proteins contain disulfide bonds, which stabilize the native protein. The oxidizing milieu of the ER allows two thiol groups of cysteine side chains to spontaneously form a disulfide bond, and protein disulfide isomerase can co-translationally rearrange these bonds within the ER (Bulleid and Freedman 1988). Once the polypeptide emerges from the translocon, co-translational folding continues, potentially assisted by ER specific chaperones (Harris *et al.* 2017). The ER-luminal Hsp70 protein BiP (HSPA5) is broadly involved in actively folding of nascent chains and can also prevent unspecific interactions with other proteins (Nguyen *et al.* 1991). Another, more substrate-specific ER chaperone is calnexin, which assists folding of

glycoproteins and ensures that only functional proteins enter the secretory pathway (Lamriben *et al.* 2016).

Another important organelle that requires the import of over 1000 nuclear-encoded proteins are mitochondria (Calvo and Mootha 2010). They have import receptors, which recognize signal sequences that often are N-terminally located and proteolytically removed after translocation, reminiscent to the ER transport (Fujiki and Vern 1993). Unlike ER translocation, protein import into mitochondria is considered to occur mostly post-translationally, even though multiple peptides have been shown to be co-translationally imported (Fujiki and Vern 1993). Recent publications showed that many ER proteins are translated in the proximity of mitochondria (Williams *et al.* 2014) and that some translating ribosomes are attached to the mitochondria outer membrane, indicating the existence of a not yet described co-translational import pathway (Gold *et al.* 2017).

2.3 Co-translational complex assembly

The majority of proteins function as multi-subunit oligomers. The building blocks for all protein complexes are either homomers, consisting of identical protein subunits, or heteromers, consisting of different protein subunits. *In vitro* assembly studies suggest that random collision of diffusing monomers can be sufficient for productive assembly (Friedman and Beychok 1979, Boulay *et al.* 1988, Jaenicke and Lilie 2000, Phillip and Schreiber 2013). However, many protein complexes are known to not efficiently form by mixing individually expressed subunits, hinting towards a coordinated co-translational assembly mechanism that begins before all subunits are fully synthesized (Table 1). It can be estimated that around 10^8 proteins are present inside a mammalian cell (Sims and Allbritton 2007), of which at least 55% are assumed to form oligomers (Lynch 2012). A growing cell continuously synthesizes thousands of different proteins per second, which have to faithfully assemble with the correct partner in the crowded cytoplasm. This dynamically changing environment makes the coordination of assembly based on random diffusion very challenging, especially because many subunits are not stable without their partner subunit and prone to aggregation. One solution to overcome the limitations of diffusion-based assembly is to coordinate complex formation with translation. Early co-translational assembly limits the time of exposure of unstable subunits and prevents undesired quaternary structure formations.

There are two possible modes of co-translational assembly. The first one is the directional or co-post assembly mode, which involves a directional interaction of a fully synthesized and

diffusing subunit (in its post-translational state) and a nascent subunit (in its co-translational state) (Shieh *et al.* 2015). The second possible mode involves two nascent chains, termed symmetrical or co-co assembly, which directly interact while being translated (**Fig. 4**, Schwarz and Beck 2019). Only co-co assembly is independent of diffusing full-length protein subunits. Co-post as well as co-co assembly can occur in theory via *trans* or *cis* assembly to form homo- as well as heteromeric complexes (**Fig. 4**). Co-post assembly *in cis* is a more predominant mechanism for prokaryotes that encode multiple subunits of one complex on polycistronic mRNAs in an operon structure. *Trans* co-post assembly is the predominant mechanism for colocalized mRNAs in eukaryotes. For co-co assembly, *cis* assembly involves that both ribosomes are located on one mRNA, whereas *trans* assembly involves two ribosomes translating different but colocalized mRNAs.

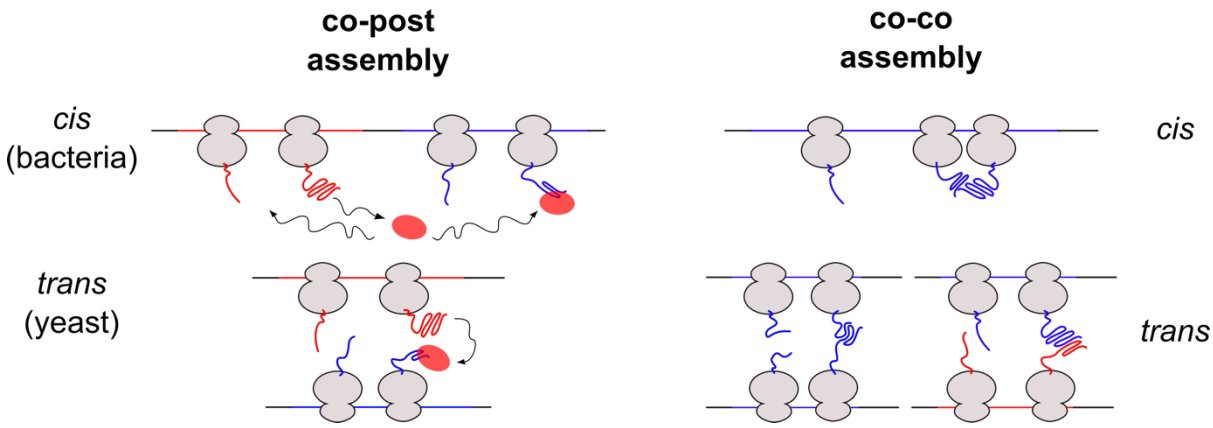

Fig. 4: Possible modes of co-translational protein assembly. Co-post as well as co-co assembly can occur in theory via *trans* or *cis* assembly to form homo- as well as heteromeric complexes. Co-post assembly involves the interaction of one fully-synthetized and presumably folded protein with one nascent subunit. The formation of a homomeric or heteromeric complex can occur via co-post assembly on one mRNA that encodes different subunits (*cis* assembly on a polycistronic mRNA, top left). Heteromer formation can also occur *in trans* between co-localized mRNAs (bottom left). Co-co assembly involves the interaction between two nascent subunits, which can occur on one mRNA to form homodimers (*cis* assembly, top right) or between different mRNAs to form homo- or heterodimers (*trans* assembly, bottom right).

The first indication of co-translational protein assembly was described in 1963, for the homo-tetrameric beta-galactosidase in *E. coli* (Zipser 1963). Zipser detected beta-galactosidase activity in the polysome fraction isolated via sucrose gradient ultracentrifugation. This result suggests that fully translated monomers form a functional tetramer that includes one nascent subunit (co-post assembly), or that the assembly process involves multiple nascent chains that form a tetramer before the subunits are fully synthesized (co-co assembly). Since this first indication, the list of complexes that are described to assemble co-translationally has steadily increased (Table 1). As described for post-translational assembly pathways, also co-

translational assembly often follows a specific pathway via formation of favorable assembly intermediates, for example that always fully synthesized subunit "A" interacts with the nascent chain of subunit "B" (Shiber *et al.* 2018, Shieh *et al.* 2015, Wells *et al.* 2015).

Co-translational assembly does not only take place in the cytosol but was additionally described to occur during translocation of transmembrane and secreted proteins, indicating that such an assembly mode is not exclusive for cytosolic proteins (Table 1, e.g.: Bergman and Kuehl 1979, Young *et al.* 1996, Zhang *et al.* 1999, Lu *et al.* 2001).

Table 1: Literature search for protein complexes that were suggested to employ co-translational assembly, in chronological order (if the detected complex assembly involved an mRNA instead of another protein subunit then gene names are written in lower case and italics)

Protein Subunits	Type	Organism	Source
β-Galactosidase	Homomer (Tetramer)	*E. coli*	Zipser & Perrin 1963
Immunoglobulin heavy and light chains	Heteromer	Mouse cells (MPC-11)	Bergman & Kuehl 1979
Myosin heavy chain	Homomer (Dimer)	Chicken cells (primary cultures of myoblasts)	Isaacs & Fulton 1987
Vimentin	Homomer (Dimer)	Chicken cells (primary cultures of myoblasts)	Isaacs 1989a
Titin	Heteromer	Chicken cells (primary cultures of myoblasts)	Isaacs 1989b
Tenascin	Homomer (Hexamer)	Human cells (U-138 MG)	Redick & Schwarzbauer 1995
Reovirus cell attachment protein σ1	Homomer (Trimer)	*in vitro*	Gilmore *et al.* 1996
SR receptor: α- and β-subunit	Heteromer	*in vitro*	Young *et al.* 1996
Tropomyosin	Homo- & Heteromer	Chicken cells (primary cultures of myoblasts)	L'Ecuyer *et al.* 1998
Photosystem II D1-D2 protein subunits	Heteromer	Spinach leaves	Zhang *et al.* 1999
NF-κB1: p105-p50	Heteromer	Hamster cells (CHO)	Lin *et al.* 2000
Voltage-gated K+ channel	Homomer (Tetramer)	*Xenopus* oocytes	Lu *et al.* 2001
p53	Homomer (Tetramer)	*in vitro*	Nicholls *et al.* 2002
IgE receptor: FcεRI-α, -β and -γ subunit	Heteromer	Human cells (HEK293) & hamster cells (CHO)	Fiebiger *et al.* 2005
Periferin	Homomer (Dimer)	Human cells (PC12 cells, HeLa)	Chang *et al.* 2006
Subunits of the SET1C/ COMPASS complex	Heteromer	*Saccharomyces cerevisiae*	Halbach *et al.* 2009
DrrA-DrrB	Heteromer	*Streptomyces peucetius* protein expressed in *E. coli*	Pradhan *et al.* 2009
Cdc2p-*rum1* and *cdc18*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Tea2p-*tip1*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011

Sty1p-*pyp2* and *cip2*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Rpt2p-*ubp6* and *rhp23*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Rpn12p-*ecm29*, *rpn1302* and *pn1301*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Atf1p-*pcr1*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Mnh1p-*mni1*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Arp6p-*alp5*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Arp9p-*snf21* and *snf22*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Arp42p-*snf21* and *snf22*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Arp8p-*ino80*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Arp2p-*arp8* and *arp9*	Heteromer	*Schizosaccharomyces pombe*	Duncan & Mata 2011
Major vault protein	Homomer	Insect cells (Sf9)	Mrazek *et al.* 2014
LuxA-LuxB	Heteromer	Vibrio harveyi protein expressed in *E. coli*	Shieh *et al.* 2015
hERG ion channel subunits	Heteromer	Human cells (iPSC-CMs, HEK293)	Liu *et al.* 2016
SAGA histone acetyltransferase subunits	Heteromer	*Saccharomyces cerevisiae*	Kassem *et al.* 2017
Fatty acid synthase	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Aminoacyl-tRNA Synthetase	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
N-acetyltransferase A	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
N-acetyltransferase B	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Anthranilate Synthase	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Carbamoyl phosphate synthetase	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Phosphofructokinase	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Translation Initiation factor eIF2	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Nascent chain associated complex	Heteromer	*Saccharomyces cerevisiae*	Shiber *et al.* 2018
Rpt1-Rpt2/ PSMC1-PSMC2	Heteromer	*Saccharomyces cerevisiae, Human cells (LNCaP cells)*	Panasenko *et al.* 2019
TAF6-TAF9	Heteromer	*Human cells (HeLa)*	Kamenova *et al.* 2019
TBP-TAF1	Heteromer	Human cells (HeLa)	Kamenova *et al.* 2019
ENY2-GANP	Heteromer	Human cells (HeLa)	Kamenova *et al.* 2019
TAF10-TAF8	Heteromer	Human cells (HeLa)	Kamenova *et al.* 2019
eIF3a-eIF3b	Heteromer	Human cells (HEK293-T)	Wagner *et al.* 2020

2.3.1 Detection of co-post assembly in prokaryotes and eukaryotes

One indication for co-post assembling complexes was the finding that single subunits were not stable when expressed in isolation but only upon co-expression of their partner subunit (Pradhan *et al.* 2009). This observation that one of two subunits is stable suggested a directionality of assembly where the stable subunit may protect the partner subunit from

misfolding or aggregation by co-translational engagement, acting like a dedicated chaperone. Another possible indication for the directionality of assembly is the localization of involved interaction domains. The nascent subunit must contain an N-terminal dimerization domain to allow co-translational engagement, while the position of the dimerization domain is not critical for the diffusing subunit (Shieh *et al.* 2015, Shiber *et al.* 2018, Kamenova *et al.* 2019, Lautier *et al.* 2020). In bacteria, the order of interactions often follows the architecture of the encoding operons. Here, the diffusing subunit is generally encoded upstream of the co-translationally engaged subunit (Shieh *et al.* 2015 and unpublished data from Josef Auburger).

The first milestone in directly detecting co-post assembly events with a resolution that allowed to determine the onset of the interaction was the use of Selective Ribosome Profiling (SeRP). This method is based on ribosome profiling (Ingolia *et al.* 2012), in combination with the affinity purification of a protein subunit of interest that engages a nascent chain (Becker *et al.* 2013). The interaction with the nascent protein allows to co-purify the associated ribosome and, thereby, the protected mRNA footprint. A detectable enrichment of sequenced ribosome protected footprints in the affinity purification compared to the total translatome allows to detect the start as well as the duration of nascent chain interactions with the affinity-purified protein (**Fig. 5**).

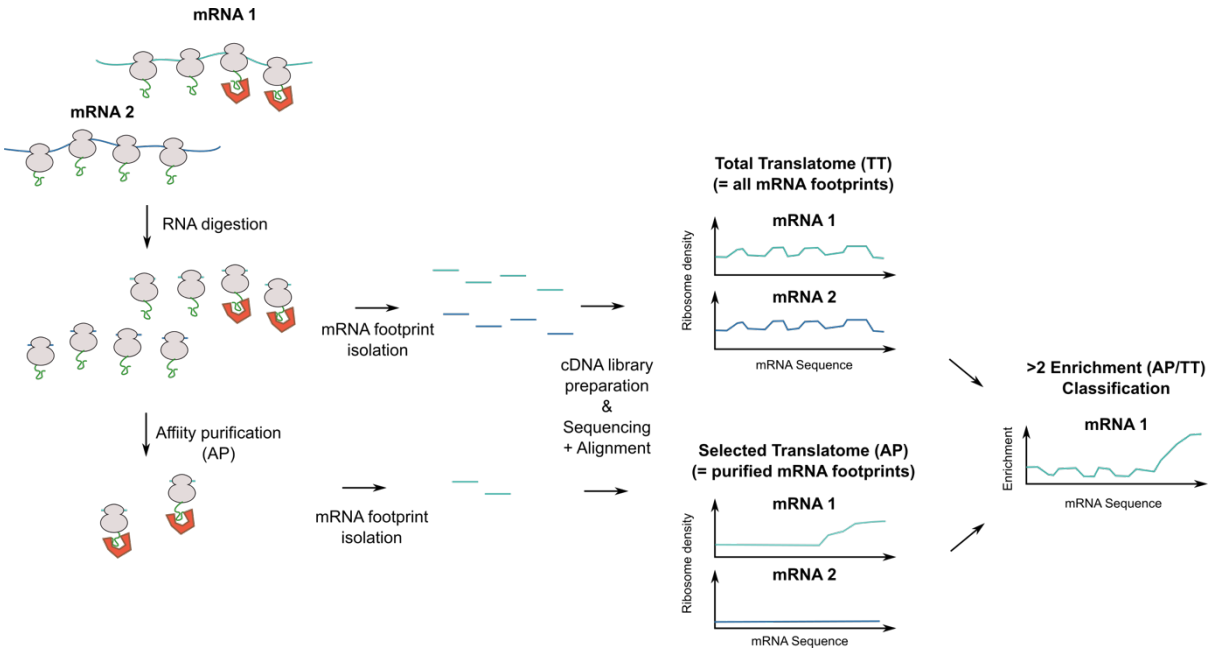

Fig. 5: Schematic overview of Selective Ribosome Profiling. Cells are lysed and the released mRNAs are digested with an unspecific RNase. This treatment destroys all mRNA linkages between ribosomes and generates single monosomes. A single ribosome protects a 30 nt mRNA footprint and all ribosome protected footprints can be recovered, sequenced and mapped to a reference genome. This resembles the total translatome of a cell. In addition, a separate translatome can be generated by

affinity purification of a selected ribosome-associated maturation factors (here shown as orange "U shape" bound to a nascent chain). This resembles the selected translatome or affinity purified translatome. The read distribution along a transcript in the total translatome reports on the local translation kinetics. Whereas the selected translatome reports on the association of the factor of interest with a subpopulation of translating ribosomes. The ratio of selected and total translatome reveals codon resolved nascent peptide interaction profiles of the factor of interest.

The first SeRP study analyzing assembly demonstrated the co-translational formation of the heterodimeric luciferase LuxA-LuxB from *Vibrio harveyi* (expressed in *E. coli*). In addition, the bacterial ribosome-associated chaperone Trigger Factor delays the co-translational engagement of LuxA with nascent LuxB, suggesting that chaperons can prevent premature, potentially unproductive interactions (Shieh *et al.* 2015). Shieh *et al.* 2015 further showed that the operon structure enhanced the efficiency of assembly, implying that the close proximity of genes on one mRNA facilitates co-translational interactions and correct assembly. This represents a common principle, as previous reports showed a reduction of functionally assembled complexes after splitting an operon in two parts (Pradhan *et al.* 2008). Furthermore, it was shown that the position of genes in an operon tends to be optimized for the order in which these protein subunits assemble into a complex (Wells *et al.* 2016). The expression levels clearly correlate of protein subunits encoded by the same operon and are thereby similarly co-expressed in contrast to subunits of a complex encoded on different mRNAs (Wang *et al.* 2005).

However, eukaryotes usually do not employ polycistronic mRNAs and proximal synthesis of subunits would therefore require the co-localization of different mRNAs. The co-localization of mRNAs encoding two co-translationally assembling human heteromer subunits, was demonstrated by single molecule fluorescence in situ hybridization (smFISH) experiments using two different fluorescence labels (Panasenko *et al.* 2019, Kamenova *et al.* 2019). How abundant such assembly coupled mRNA co-localizations occur in eukaryotes was so far not comprehensively investigated.

Most earlier studies focused only on co-translational assembly of selected candidates (Table 1). More recent studies started to explore the prevalence of co-post assembly in eukaryotes. The first low throughput study was conducted in *S. pombe* and published in 2011. This study used ribonucleoprotein co-immunoprecipitation followed by microarray analysis (RIP-Chip) and suggested that 12 out of 31 purified proteins led to the co-purification of mRNAs that encode for a known interaction subunit in a complex (Duncan and Mata 2011). The nascent chain dependent interaction was corroborated by puromycin controls that showed that the mRNA co-purification depended on nascent chains and active translation. A follow up study further indicated the biological relevance of this assembly mode, by providing evidence that

one fully synthesized subunit can regulate the phosphorylation of the nascent partner subunit, required to be biologically functional (Duncan and Mata 2014).

Another targeted low throughput study conducted in *S. cerevisiae* using SeRP detected 9 out of 12 selected complexes to co-translationally assemble (Shiber *et al.* 2018). The authors showed that the binding of the yeast specific ribosome-associated Hsp70 chaperone Ssb is coordinated with complex assembly, by correlating different SeRP data sets analyzing either assembly or co-translational chaperone binding. This result suggests that Ssb engages partially synthesized nascent chain segments until shortly before the partner subunit associates. This finding indicates that chaperones may facilitate assembly by ensuring timely folding prior to assembly, or by preventing unwanted, non-productive interactions. Therefore, it was speculated that chaperones that coordinate folding or assembly, should also prevent premature co-translational assembly (Shiber *et al.* 2018).

Co-post assembly was also proven in human cells. Kamenova *et al.* 2019 provided evidence that mammalian nuclear transcription complexes (TFIID, TREX-2 and SAGA), which are all composed of multiple subunits, employ co-translational assembly. They furthermore demonstrated that the position of the dimerization domain on both involved interaction partners dictates the mode of assembly. While N-terminal dimerization domains tend to co-co assemble, subunits enclosing a C-terminal dimerization domain employ co-post assembly. The study supported previous findings (Shiber *et al.* 2018) that indicated that the absence of the full-length assembling partner makes the nascent partner prone to aggregation or degradation. This indicates that protein synthesis and complex assembly are closely coupled to build multi-subunit complexes and suggests that co-translational assembly is a general necessity for faithful protein biogenesis.

2.3.2 Quality control of orphan subunits

Many cancer cells are aneuploid, which results in an imbalanced synthesis of individual subunits. Such imbalances cause an excess of unassembled “orphan” protein subunits (Chen *et al.* 2015, Dodgson *et al.* 2016), that are selectively eliminated to maintain protein homeostasis. The targeting process is supported by a conserved ubiquitin-conjugating enzyme called UBE2O, which recognizes basic and hydrophobic sequences on unassembled protein subunits to mediate ubiquitination (Yanagitani *et al.* 2017). It is still debated how many newly synthesized proteins are directly degraded by the proteasome (Kramer *et al.* 2019). Recent estimations suggest that approximately 10% - 15% of all newly synthesized proteins are often co-translationally polyubiquitinated in human cells and thereby rapidly degraded (Wang *et al.* 2013, McShane *et al.* 2016). Additional pathways that recognize unassembled

subunits exist, including the recently identified ubiquitin ligases UBR1 and HUWE1 involved in unassembled soluble protein degradation (USPD, Xu *et al.* 2016). The discovery of USPD indicates the importance of eliminating unassembled proteins to prevent the exposure of hydrophobic segments that can lead to protein aggregation. Such pathways for degradation of folded but unassembled orphan subunits are not only found for soluble but also membrane proteins in mammalian cells (Inglis *et al.* 2020). In yeast, orphan subunits diffuse more easily in the ER membrane than bulky protein complexes, and are thereby specifically recognized by the Asi complex, which directly binds transmembrane domains (TMD) and facilitates substrate ubiquitination (Natarajan *et al.* 2020). Another possible fate of unassembled subunits is demonstrated for the endoplasmic reticulum membrane protein complex (EMC), a post-translational and co-translational insertase (Volkmar and Christianson 2020). The cytosolic, unassembled subunit of EMC2, which usually assembles with EMC8, can be stabilized by binding to WNK1. WNK1 thereby prevents ubiquitination of EMC2 by competing with the binding of HUWE1, an E3 ligase, indicating a triaging mechanism selecting orphan proteins to be degraded or stabilized (Inglis *et al.* 2020, Xu *et al.* 2016).

2.3.3 Dimerization quality control

Recent work identified an additional surveillance mechanism for BTB domain containing complexes named Dimerization Quality Control (DQC). This mechanism recognizes unphysiological heterodimers of BTB domain containing proteins, which generally form homodimers (Mena *et al.* 2018). The mechanism appears to be kinetically controlled, supported by the rapid formation of functional homodimers. The dimerization buries short amino acid sequences that function as degrons, which are protein segments triggering protein ubiquitylation and proteasomal degradation. As aberrant BTB heterodimers assemble more slowly, they are recognized by DQC for degradation. The active E3 ligase complex SCF^{FBXL17} is the surveillance factor for BTB domain proteins (Mena *et al.* 2018). The authors speculate that more of such surveillance factors may exist that monitor the interaction status of other recurrent interaction modules, for example factors monitoring the productive formation of coiled coil domains (Mena *et al.* 2018).

2.3.4 Ribosome-associated quality control in humans

Ribosome pausing was implicated in a number of co-translational events, including nascent peptide folding, membrane targeting, and interactions with chaperones or other proteins (Collart and Weiss 2020). Ribosome slow-down and pausing can lead to the collision of two

ribosomes on one mRNA, which can activate multiple stress response pathways (Meydan and Guydosh 2020).

Classical ribosome profiling is based in the analysis of 30 nt footprints, which are protected by a single ribosome. Aligning all mRNA footprints to the respective full-length mRNA sequence allows precise determination of the ribosome positions and density along the detected mRNA (Ingolia *et al.* 2009). Recent studies started to investigate translation elongation pausing by isolating 60 nt footprints, which are protected by two ribosomes in close proximity that prevent RNase from cleaving in between them (Arpat *et al.* 2020, Zhao *et al.* 2020, Han *et al.* 2020). The isolation and analysis of 60 nt ribosome fragments indicate ribosome stalling events that are surprisingly not correlated with ribosome pausing detected by 30 nt fragments (Arpat *et al.* 2020, Zhao *et al.* 2020, Han *et al.* 2020). The frequency of ribosome collisions based on of 60 nt ribosome profiling was determined to involve approximately 10% of all translating ribosomes in mice (Arpat *et al.* 2020), while other studies suggest a more widespread distribution in other eukaryotic organisms (Zhao *et al.* 2020, Han *et al.* 2020). These collisions were mostly observed at specific sites, like on mRNA positions encoding either Pro-, Asp-, or Gly-containing dipeptide motifs as well as Lys- and Arg-rich polybasic motifs (Arpat *et al.* 2020, Han *et al.* 2020, Meydan and Guydosh 2020b), supporting previous findings that these sequences slow-down translating ribosomes. Such ribosome pausing events frequently occur in mRNA segments encoding unstructured regions of proteins, which often form a linker connecting two protein domains. This indicates that ribosome pausing contributes to the structure formation of the nascent chain and the folding of ribosome-exposed domains, before the next domain is emerging from the ribosomal tunnel exit (Apart *et al.* 2020).

Importantly, persisting ribosome pausing followed by ribosome collision triggers the ribosome-associated quality control (RQC) that ultimately causes nascent chain and mRNA degradation (Meydan and Guydosh 2020a). These surveillance mechanisms are mostly active on defective mRNAs. One frequent mRNA defect is the absence of a stop codon that enables ribosomes to translate parts of the mRNA poly(A) sequence. This leads to the so called "Non-Stop Decay" (NSD), caused by the translation of multiple AAA codons encoding for poly-lysin (Collart and Weiss 2020, Joazeiro 2019, Brandman and Hegde 2016). These positively charged lysin stretches interact with the negatively charged ribosome exit tunnel, resulting in translation slowdown, ribosome collision and the activation of the RQC (Collart and Weiss 2020). Ribosome exposed lysin residues serve as attachment sites for poly-ubiquitylation, which in the end leads to proteasomal degradation (Bengtson and Joazeiro 2010). Other events triggering ribosome collision are the translation of truncated mRNAs, which initiates the so called "No-Go Decay" (NGD). Here ribosomes reach the 3’-end of the truncated mRNA without

engaging a stop codon nor a poly(A) tail (Collart and Weiss 2020). Another mRNA feature causing collisions is the presence of a premature stop-codon leading to the activation of the "Non-sense Mediated Decay" (NMD). In all these cases, ribosome collision seems to be the trigger of ribosome disassembly, followed by mRNA and nascent chain degradation. Such collided ribosomes seem to form a unique interface that is recognized by the E3 ubiquitin ligase ZNF598 (Juszkiewicz *et al.* 2018). After binding ZNF598 ubiquitylates the small 40S ribosomal subunit, followed by ribosomal splitting to recycle ribosome subunits. Ribosomes with an empty A-site are splitted by the Pelota–Hbs1–ABCE1 complex, whereas ribosomes with an occupied A-site are splitted by activating signal cointegrator 1 complex (ASCC, Juszkiewicz *et al.* 2020, Phillips and Miller 2020, Shoemaker *et al.* 2010, Pisareva *et al.* 2011, Tsuboi *et al.* 2012). RACK1 acts as one abundant recruitment factor for ZNF598, mediating the recognition of ribosome collision (Ikeuchi *et al.* 2019, Sundaramoorthy *et al.* 2017). In addition, also ZNF598 independent pathways are known that detect ribosome collision. One is mediated by ribosomal bound ZAK that becomes phosphorylated by MAP2K after collision, and initiates a ribotoxic stress response that in severe cases can lead to apoptosis (Wu *et al.* 2020). Another factor is EDF1, which also identifies collision and thereby initiates initiation repression (Sinha *et al.* 2020).

In unstressed human cells the binding of ZNF598 to collided ribosomes further recruits GIGYF2 and 4EHP to stalled ribosomes. GIGYF2 and 4EHP are two cap-binding proteins, which block translational initiation upon binding. Both proteins are capable of repressing translation initiation also in the absence of ZNF598, indicating that additional factors may exist that similar to ZNF598 guide the recruitment of GIGYF2 and 4EHP to collided ribosomes to silence translation of faulty mRNAs (Hickey *et al.* 2020).

Once the ribosome is disassembled, the exposed tRNA bound to the 60S ribosomal subunit is recognized by NEMF, which recruits the E3 ubiquitin ligase Listerin to mark the nascent chain for degradation by the proteasome (Bengtson and Joazeiro 2010, Brandman *et al.* 2012, Shao *et al.* 2015). NEMF catalyzes the elongation of the nascent chain by adding alanine-threonine residues to the C-terminus (CAT tails) (Shen *et al.* 2015, Joazeiro 2019). This mechanism works independently of active translation and helps to push the nascent chain out of the ribosome exit tunnel to facilitate proteasomal degradation (Brandman *et al.* 2012, Defenouillère *et al.* 2013).

Recent data showed that similar RQC pathways can also function at the ER even in the absence of any mRNA defects (Brandman and Hegde 2016, Joazeiro 2019, Lakshminarayan *et al.* 2020, Trentini *et al.* 2020). Here, misfolded membrane proteins are subjected to co-translational degradation, depending on specific types of misfolding, induced by different

mutations (Lakshminarayan *et al.* 2020). In addition, another ER-associated quality control pathway exists, that employs UFMylation of collided ribosomes (Stephani *et al.* 2020).
Taken together, these data indicate that ribosomes are not only translation machineries but also important signaling molecules inside the cell. Furthermore, it becomes more and more clear that not a single stalled ribosome but rather the collision of two ribosomes triggers all downstream events. Of note, the recognition of collisions and the induction of one of the here described surveillance pathways critically depends on the context and the duration of ribosome stalling.

2.3.5 Indication for co-co assembly in prokaryotes and eukaryotes

It has been speculated that the proximity of neighboring nascent subunits of homomeric complexes on a polysome could facilitate early assembly mechanisms (Isaacs *et al.* 1989). Furthermore, the high prevalence of collided ribosomes (Arpat *et al.* 2020) implies that ribosomes may be very close to each other on one mRNA. This suggests that co-translational assembly could occur via the interaction of two proximal nascent subunits (co-co *assembly*). One of the first publications describing the possibility of co-co assembly studied the cytosolic intermediate filament vimentin (Isaacs 1989a), followed by a number of publications that also proposed co-co assembly of different proteins in various organisms, including the reovirus cell attachment protein, NfkB, p53, Rpn1-Rpn2, and TAF6-TAF9 (see Table 1). Among the described cases are not only dimeric complexes but also homo-trimeric and homo-hexameric complexes that were proposed to form during translation of one mRNA by interaction between neighboring ribosomes (Table 1, reovirus cell attachment protein σ1 and tenascin). In all cases N-terminal located interaction domains mediated the early assembly.
However, direct evidence that two ribosome-nascent chain complexes interact is still scarce and we lack any information on the molecular principles and prevalence of this proposed process. In theory, co-co assembly can involve the interaction of nascent chains from two neighboring ribosomes that translate the same (*cis*-assembly) or two different mRNAs (trans-assembly). For both variants it has already been speculated that translation pausing at specific mRNAs positions may facilitate the process (Isaacs and Fulton 1987, Isaacs *et al.* 1989a). Panasenko *et al.* 2019 speculate that the co-translational assembly of two subunits of the proteasome 19S complex may be triggered by ribosome pausing at a specific aspartate-proline codon pair. A *cis-assembly* mechanism for homo-oligomeric complexes would be independent of subunit-encoding mRNA co-localization. In addition, this homo-oligomer formation between neighboring ribosomes on one mRNA should allow to generate transcript-specific homomeric complexes. Such an assembly mechanism could be of great importance

for more complex organisms like humans, which encode many alternative splice variants and isoforms with different locations of one single protein. Co-co assembly *in cis* could suppress promiscuous interaction between closely related proteins.

3. Aims of this dissertation

The main aim of this dissertation was to study the prevalence and mechanistic principles of co-translational protein complex assembly in mammalian cells. Recent proof-of-principle studies based on Selective Ribosome Profiling from the Bukau lab demonstrated that co-translational complex assembly is a coordinated and prevalent mechanism in *E. coli* as well as *S. cerevisiae* (Shieh *et al.* 2015, Shiber *et al.* 2018). The specific assembly mechanism analyzed, termed co-post assembly, always included one fully synthesized protein engaging the nascent chain of another subunit. As the basis of this thesis, we hypothesized the existence of an alternative co-translational protein complex assembly mode mediated by two interacting nascent chains (co-co assembly). To investigate this hypothesis, it was necessary to established an experimental set-up, termed Disome Selective Profiling (DiSP), to investigate the prevalence of co-co assembly in human cell lines (developed in close collaboration with Matilde Bertolini). Since this led to the identification of a large number of potential co-co assembling complexes in U2OS and HEK293-T cells, we subsequently focused on the following questions:

1. Are the identified co-co assembling events mediated by nascent chain – nascent chain interactions?

2. What type of complexes and protein domains are involved in the identified co-co assembly proteome?

3. What are the basic requirements of co-co assembly, is this mechanism solely driven by intrinsic features of the nascent chains to form a quaternary structure or are dedicated assembly factors required?

4. Does co-co assembly between ribosomes on one mRNA ensure isoform specific homodimerization?

5. What is the distance between two co-co assembled ribosomes on one mRNA?

4. Results

4.1 Disome Selective Profiling (DiSP) detects a monosome to disome footprint density shift for hundreds of translated mRNAs in human cells

Recent findings have provided evidence that co-translational assembly occurs in the proteome of prokaryotes as well as eukaryotes, but up to now all identified proteins that engage in co-translational assembly had to be individually studied and validated. We aimed to analyze all potential proteins that dimerize co-translationally via exclusive nascent chain interactions (co-co assembly) in a proteome-wide screening approach employing ribosome profiling.

Ribosome profiling is based on the sequencing of ribosome protected mRNA footprints after cell lysis and RNase treatment. Translating ribosomes protect ca. 30 nt mRNA footprints from RNase digestion. Aligning these mRNA fragments to the human genome determines the position of all translating ribosomes on every mRNA in a cell. We rationalized that the interaction of two nascent chains will result in a ribosome pair (disome), which is resistant to RNase treatment. To detect co-co assembly by ribosome profiling, we had to physically separate fully RNase digested single ribosomes (monosomes) from nascent chain coupled disomes (**Fig. 6**). Two separate ribosome profiling experiments of both ribosome populations has thereby the potential to reveal, which mRNAs are translated by nascent chain coupled ribosomes.

The genome-aligned mRNA footprints of the monosome fraction should contain the majority of mRNA footprints from all translating ribosomes and generally resemble the total translatome. Following our rationale, the mRNA footprints from the disome fraction are expected to be enriched for mRNAs encoding for oligomeric proteins, which co-translationally assemble via nascent chain interactions. The ratio of disome over monosome reads for mRNA with co-co assembling ribosomes should indicate a footprint density shift between both ribosome populations. A disome enrichment is expected to indicate the start point of co-co assembly after at least a minimal length of a dimerization competent nascent peptide are exposed on translating ribosomes (**Fig. 6**). Furthermore, DiSP sequencing data allow then to determine the nascent chain sequences that are fully exposed on translating ribosomes at the start of the disome formation (30 codons are always assumed to be buried inside the ribosomal exit tunnel). Since this is a genome-wide unbiased approach, the sequencing data will directly report on the prevalence of co-co assembly in the cell and has the potential to reveal basic principles of this assembly mechanism.

The isolated disome fraction also contains collided ribosome pairs, which remain mRNA-coupled even after RNase treatment (Apart *et al.* 2020). However, such RNA coupled

ribosome pairs protect at least 60 nt long mRNA fragment, whereas fully RNA digested and exclusively nascent chain coupled disomes will protect two discrete 30 nt mRNA fragments. Therefore only 30 nt ribosome footprints of the mRNAs from both monosome and disome fractions are converted into a cDNA library for deep sequencing.

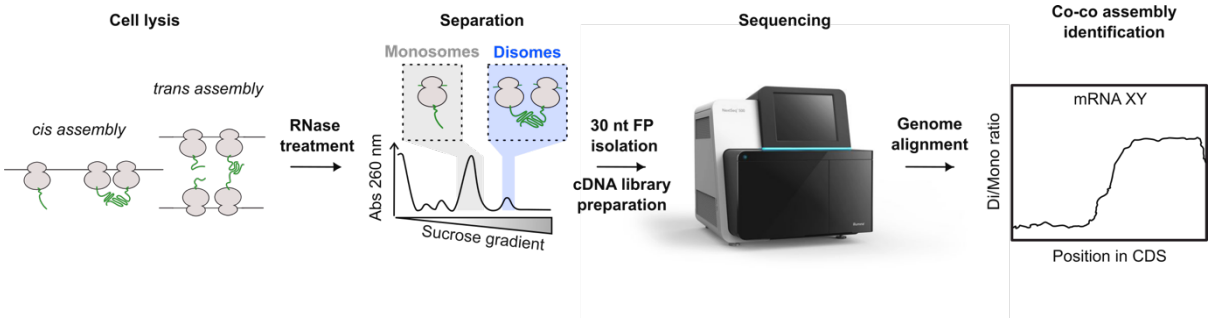

Fig. 6: Experimental set-up of Disome Selective Profiling (DiSP) to screen for nascent chain dependent co-translational assembly events. Human cells are lysed and RNase-treated, the resulting monosomes as well as disomes are separated by sucrose gradient ultracentrifugation. Only 30 nt ribosome footprints from both fractions are isolated, and converted into a cDNA library applicable for deep sequencing. Genome-aligned reads from the monosome and disome fraction are used to identify co-co assembly candidates based on a footprint density shift from the monosome to the disome state along individual mRNAs. This is visualized here by an expected disome over monosome enrichment profile that stays enriched till the end of translation.

The first DiSP results of the U2OS cancer cell line and the immortalized HEK293-T human embryonic kidney cell line, revealed in both cell lines more than hundred mRNAs that had a more than 2-fold disome over monosome enrichment (**Fig. 7A, blue dots**). A ≥ 2-fold disome over monosome footprint enrichment was used as first artificial threshold to identify potential assembly candidates and included lowly as well as highly expressed mRNAs in both cell lines (**Fig. 7A**). The small number of 2-fold disome enriched candidates suggested that a specific subfraction of all human mRNAs is translated by nascent chain connected ribosome pairs, supporting the assumption that only a subpopulation of proteins has the property to co-translationally assemble.

A metagene profile of the averaged monosome or disome footprint density distribution along all detected mRNA coding sequences (CDS) shows that early during translation, most ribosomes are in the monosome state (**Fig. 7B,** grey line). The absence of disomes in the beginning of the coding sequence is in agreement with our model of nascent chain mediated ribosomes dimerization (**Fig. 7B,** blue line). The disome enrichment starts near codons 200 and remains until the end of translation, in both the HEK293-T and U2OS data sets (**Fig. 7B**). This indicates that most nascent chain mediated interactions are stable till the end of translation and not only transiently maintained. It furthermore indicates that two nascent chain connected ribosomes terminate translation almost simultaneously, because no loss of disome

reads is detected before the stop codon (further explained in **4.3**). The same trend can be observed in an metagene enrichment plot by forming the disome over monosome ratio **(Fig. 7C).**

Both cell lines reveal highly similar lists of genes that display a 2-fold disome enrichment, demonstrating that disome formation is not a specific feature of the U2OS cancer cell line nor of the immortalized HEK293-T cell line **(Fig. 7D)**.

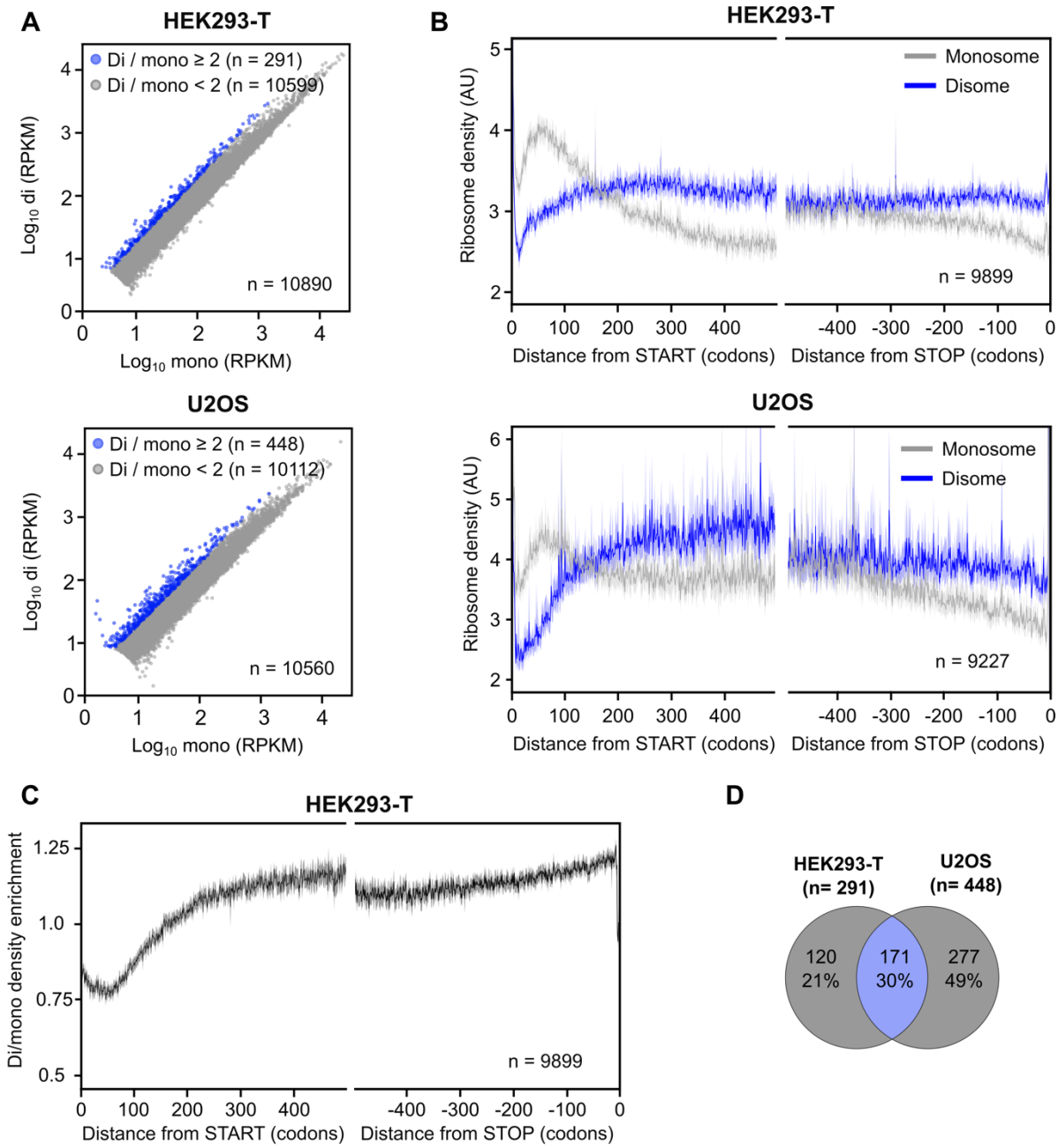

Fig. 7: Global analysis of disome footprint density shifts in two human cell lines. (A) Comparison of disome (di) and monosome (mono) footprint densities (RPKM = Reads Per Kilobase per Million mapped reads) of all detected mRNAs in HEK293-T cells (top) or U2OS cells (bottom), one replicate

shown of each data set. **(B)** Average footprint density along the coding sequence of all detected mRNAs (metagene profile) aligned to translation start (left) and stop (right) for HEK293-T (top) or U2OS cells (bottom), one replicate shown of each data set. **(C)** Disome over monosome ratio plot (metagene enrichment plot) of HEK293-T cells shown in (B). **(D)** Venn diagram indicating the overlap of the ≥ 2-fold disome enriched candidates shown in (A) of HEK293-T and U2OS cells.

Some of the detected disome enriched candidates are differently expressed in between both cell lines, which increases the variability in the respective candidate lists. Candidates that are expressed in both cell lines show highly correlated monosome as well as disome footprint density distributions along these translated mRNAs (**Fig. 8A**). Notably, a number of candidates reveal an almost complete shift of detected footprints from the monosome to the disome fraction, indicated by the loss of monosome reads after the onset of the disome enrichment (**Fig. 8A, grey lines**). This observation suggests that for these cases, the majority of ribosomes dimerize during translation, which implies a highly prevalent and efficient assembly process. Importantly, monosome footprint density loss was generally not detected for mRNAs lacking a detectable disome enrichment (**Fig. 8B, grey line**). Additionally, obtained total translatome data from HEK293-T cells indicate no footprint loss along the mRNAs of disome enriched candidates, suggesting that the observed monosome loss in our DiSP data is not related to ribosome disassembly, for example caused by ribosome quality control mechanisms (**Fig. 9**).

Among the top candidates of the 2-fold disome enriched candidates with a cytosolic localization (shown in **Fig. 8A**) were many homodimeric proteins. Most evident, we observed several intermediate filament proteins (e.g., all nuclear lamins as well as cytosolic vimentin), and other homodimeric proteins like dynactin motor proteins or other cytoskeleton proteins like myosin (**Fig. 8A**). This was a first validation of our findings, since nascent chain dimerization of cytoskeleton proteins like vimentin and myosin had been described (Isaacs and Fulton 1987, Isaacs 1989a). Importantly, the disome enrichment did not occur for all detected homomeric proteins, including homodimers that form via C-terminal dimerization domains or proteins that require post-translational modifications to dimerize (**Fig. 8B**). The depicted homodimers of the co-co and non-co-co class in Figure 3 reveal first common structural features. Supporting co-co assembly of our candidates, many of the 2-fold disome enriched homodimers employ a N-terminal homodimerization domain, often long alpha-helical coiled coils (for example all lamins, vimentin, dynactin subunit 1, mitotic spindle assembly checkpoint protein MAD1 and myosin-9). Furthermore, homodimers not included in our candidate list often enclose rather short dimerization domains that are near the C-terminus or scattered along the coding sequence, indicating that native folding and assembly can only occur post-translationally. Another interesting observation was that some profiles indicate

transient disome shifts, one example is the transient disome enrichment for Hsp90 between codon 200 and 400 (**Fig. 8B**, encoded by *HSP90AA1*). Such transient interactions often do not enrich disomes enough to rank these genes among the 2-fold enriched, but can report on real transient dimerization events (discussed in more detail in section **4.3** and **5.5**).
Some depicted candidates including glycerinaldehyd-3-phosphat-dehydrogenase show "gaps" within the monosome and disome footprint density along the mRNA (**Fig. 8B**, encoded by *GAPDH*). These gaps are produced by the employed data analysis tools, which exclude reads that cannot be uniquely aligned to the human genome and are found mostly in genes with highly repetitive regions or genes with annotated pseudogenes. Our analysis furthermore aligns all reads to the longest annotated coding sequence of one gene, and the expression of different isoforms will also impact the read distribution along open reading frames.
Thus, some single gene candidates like Lamin A/C (**Fig. 8A**, encoded by *LMNA*) show a reduction of reads in the last fraction of codons. For *LMNA,* we detected a drop of read density in the last 100 codons that is caused by the simultaneous expression of two prominent transcript variants, one long variant named lamin A and another shortened version called lamin C, which is lacking the last 91 codons. Therefore, only the reads of the last 91 codons of *LMNA* are unique for lamin A, whereas all other reads originate from the translation of lamin A and lamin C.

A prominent example that was previously suggested to co-translationally assemble by nascent chain dimerization is the tumor suppressor p53 (Nicholls *et al.* 2002). Surprisingly, our data did not show a disome shift for *TP53* (**Fig. 8B***, TP53)*. This finding agrees with the position of the annotated dimerization domain of p53 that is located at the far C-terminal end.

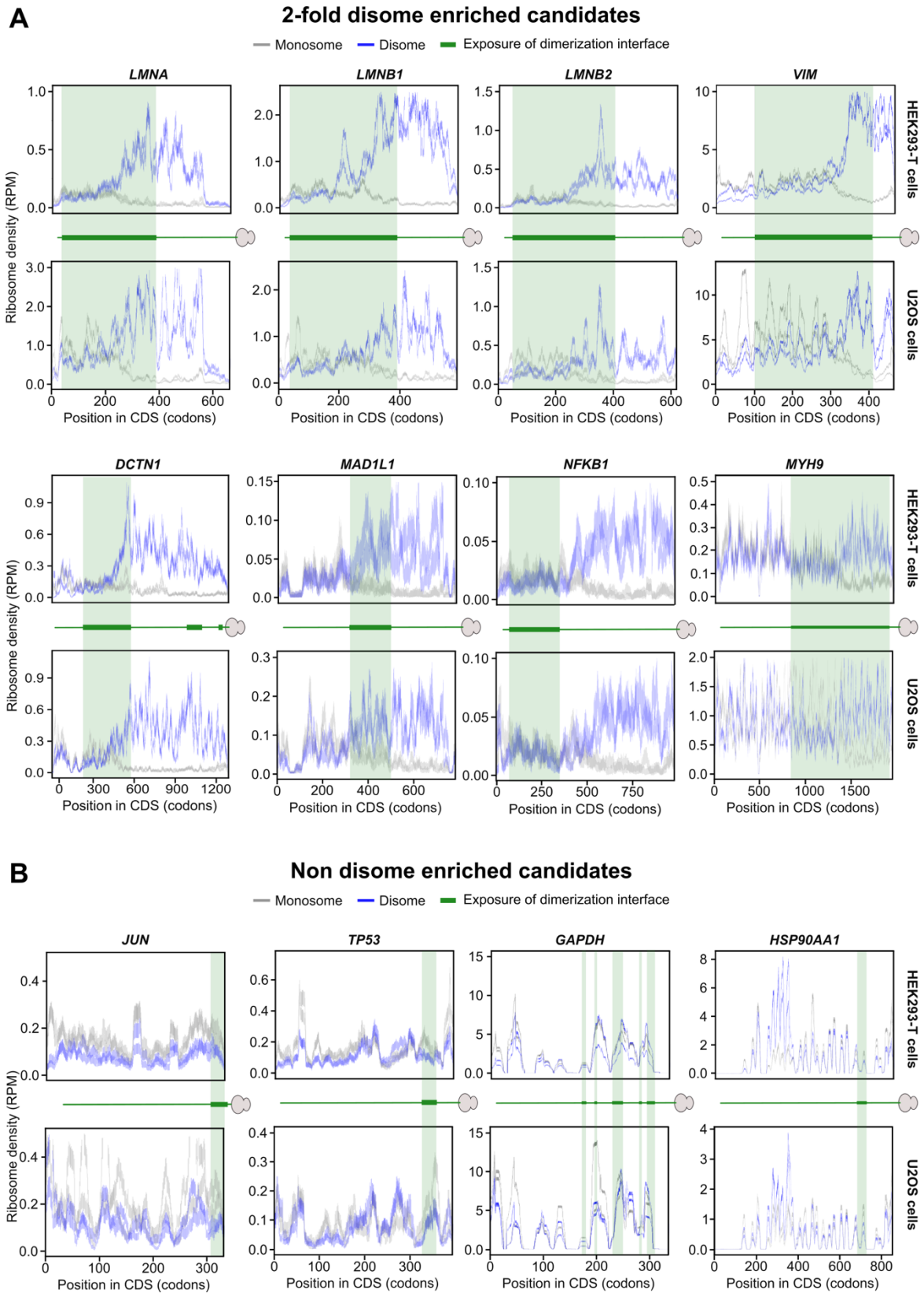

Fig. 8: DiSP profiles of selected cytosolic gene candidates. (A) Examples of prominent 2-fold disome enriched homodimers, detected in U2OS and HEK293-T cells. Green boxes indicate the exposure of annotated dimerization domains (boxes are shifted by 30 codons to account for the length of the ribosome exit tunnel, to guarantee full domain exposure). Two biological replicates are shown for each cell line. **(B)** Examples of prominent homomers without a detectable disome enrichment in U2OS and HEK293-T cells. Color code as in (A).

We also identified NfκB (encoded by *NFKB1*) as one candidate (**Fig. 8A**), in agreement with an earlier report (Lin *et al.* 2000). The translation products of the *NFKB1* mRNA are two protein isoforms, a shorter p50 and a full-length p105 isoform. The full-length p105 protein consists of an N-terminal RHD (Rel Homology Domain) required for dimerization, and a C-terminal domain that is responsible for the inhibition of NfκB. The p105 homodimer as well as a p50-p105 heterodimer is the inactivated state of NfκB. Active NfκB is generated by proteasomal degradation of the p105 C-terminus to form the active p50-p50 dimer, which functions as transcription factor (Lin *et al.* 2000).

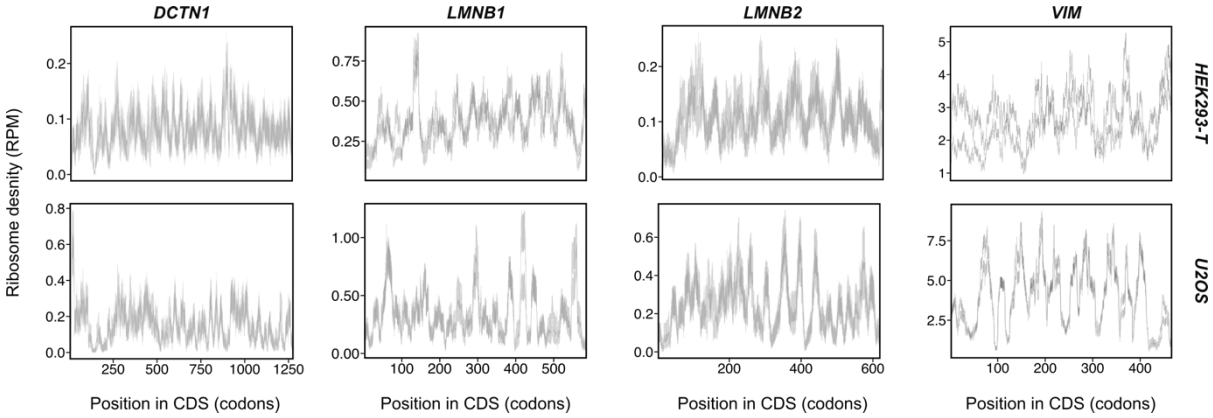

Fig. 9: Total translatome data show no footprint loss for disome enriched candidates. Examples of prominent disome enriched homodimers shown in Fig. 8. The ribosome density (RPM: reads per million) from total translatomes are shown obtained from U2OS and HEK293-T cells.

An initial analysis of the protein localization of all 2-fold disome enriched candidates revealed no clear enrichment of proteins that are located in one specific cellular localization in U2OS or HEK293-T cells, including translocated proteins (**Fig. 10A**). Among the enriched translocated proteins were many proteins with annotated transmembrane domains (TMD). Candidates with a single annotated TMD mostly revealed the onset of the monosome to disome footprint shift after full exposure of this domain, which indicates that TMDs may function as dimerization motif (**Fig. 10B**). This is in agreement with published data (Lemmon *et al.* 1992, Zhang *et al.* 1999). However, not all single TMD containing membrane proteins indicated a disome enrichment (**Fig. 10C,** a deeper analysis of TMDs is conducted in chapter **4.3.3**).

Additional analysis of the 2-fold disome enriched genes did not indicate a specific enrichment of homo- or heteromeric complexes over monomeric proteins, suggesting that this arbitrary chosen 2-fold enrichment selection criteria may not be specific enough to bioinformatically select co-co assembly candidates (**Fig. 10D**). The disome enrichment highly depends on the separation of the monosome and disome peak within a sucrose gradient (details in

Supplemental Material: Optimization of DiSP). The much bigger monosome peak always contaminates the much smaller disome fraction and thereby affects the 2-fold disome enrichment between experiments. The enrichment also depends on the sequencing depth of the monosome and disome sample especially for lowly expressed genes, which increase the number of wrongly assigned 2-fold enriched candidates. Therefore, we included additional control experiments (see chapter **4.2**) and improved the footprint shift detection tools in combination with more stringent selection criteria to define co-co assembly candidates (see chapter **4.3**).

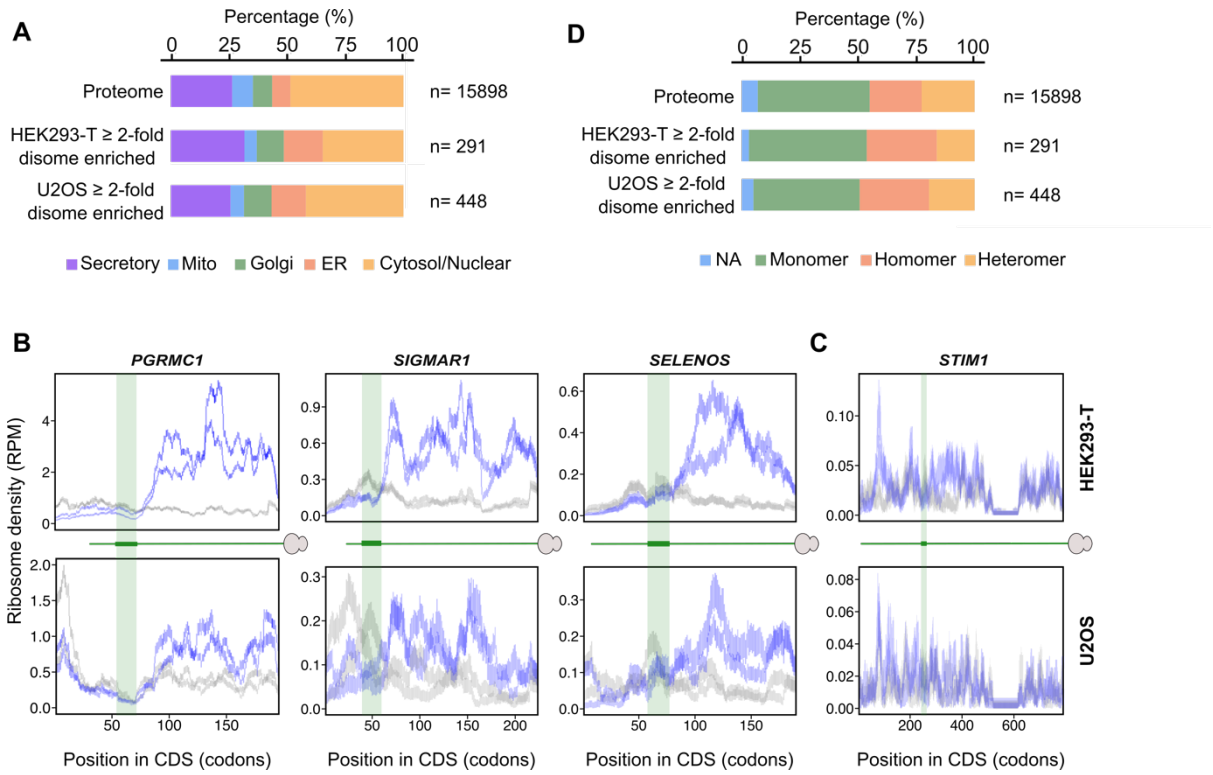

Fig. 10: DiSP identifies proteins from all cellular localizations. (A) Localization of all human proteins and disome enriched candidates in HEK293-T and U2OS cells (percentagewise). **(B)** Examples of disome enriched proteins with a single annotated transmembrane domain (TMD), detected in U2OS and HEK293-T cells. Green boxes indicate exposed annotated TMDs. All boxes are shifted by 30 codons to account for the length of the ribosome exit tunnel. Two biological replicates are shown for each cell line. **(C)** Example of a non-disome enriched protein with a single annotated N-terminally located TMD, detected in U2OS and HEK293-T cells. The gap of footprints at around 500 - 600 codons indicate a non-expressed exon. Colors as in (B). **(D)** Oligomeric state of all human proteins and disome enriched candidates in HEK293-T and U2OS cells (percentagewise).

4.2. The observed monosome to disome footprint density shift is mediated by nascent chain interactions

To challenge our model that the disome formation is mediated by nascent proteins, we explored whether the detected disome footprint density shifts remain when nascent chains are degraded upon treatment of lysates with increasing amounts of proteinase K (PK) (**Fig. 11A**). First, we tested the impact of proteinase digestion on the cellular proteome, as well as on ribosomes by employing Western blot analysis and sucrose gradient centrifugation. The amount of PK tested led to increasing degradation of full-length proteins such as the consecutively expressed Hsp70 (HSPA8), but also degraded surface exposed ribosomal proteins such as the 60S ribosomal protein L4 (RPL4, **Fig. 11B**). As expected, we detected that less surface exposed ribosomal proteins like the 60S ribosomal protein L32 were more resistant to PK digestion (RPL32, **Fig. 11B**). Importantly, polysome profiles using sucrose gradients suggested that even the highest PK condition did not detectably affect the ribosomal integrity (**Fig. 11C**).

The first PK titration DiSP data set was obtained from the two highest PK conditions (Very High PK= 1:200 and High PK = 1:2000 PK to total protein amount). As internal control, one non-PK treated sample was processed. The first sucrose gradients of RNase I digested and PK treated lysates revealed no change of the disome peak height between PK and non-PK treated samples (**Fig. 11D**). In several attempts to produce a replicate, it became clear that we observed in various times detectable aggregation of ribosomes, which seemed to be resolved by PK treatment (**Fig. 11E**, observed by an increase of monosome and disome peak heights in sucrose gradients with higher PK treatments). In a series of lysis buffer optimization experiments, we showed that the ribosome precipitation was lowest with high salt (500 mM KCl) lysis conditions, thus, the best prevention of ribosome aggregation during cell lysis, RNA digestion, PK treatment and ribosome isolation (**Fig. 11F**). Therefore, we performed another PK titration in a high salt lysis buffer and separated ribosomes on high salt gradients. This titration revealed also no loss of the disome peak height in sucrose gradients except for the highest tested PK concentration (**Fig. 11G**).

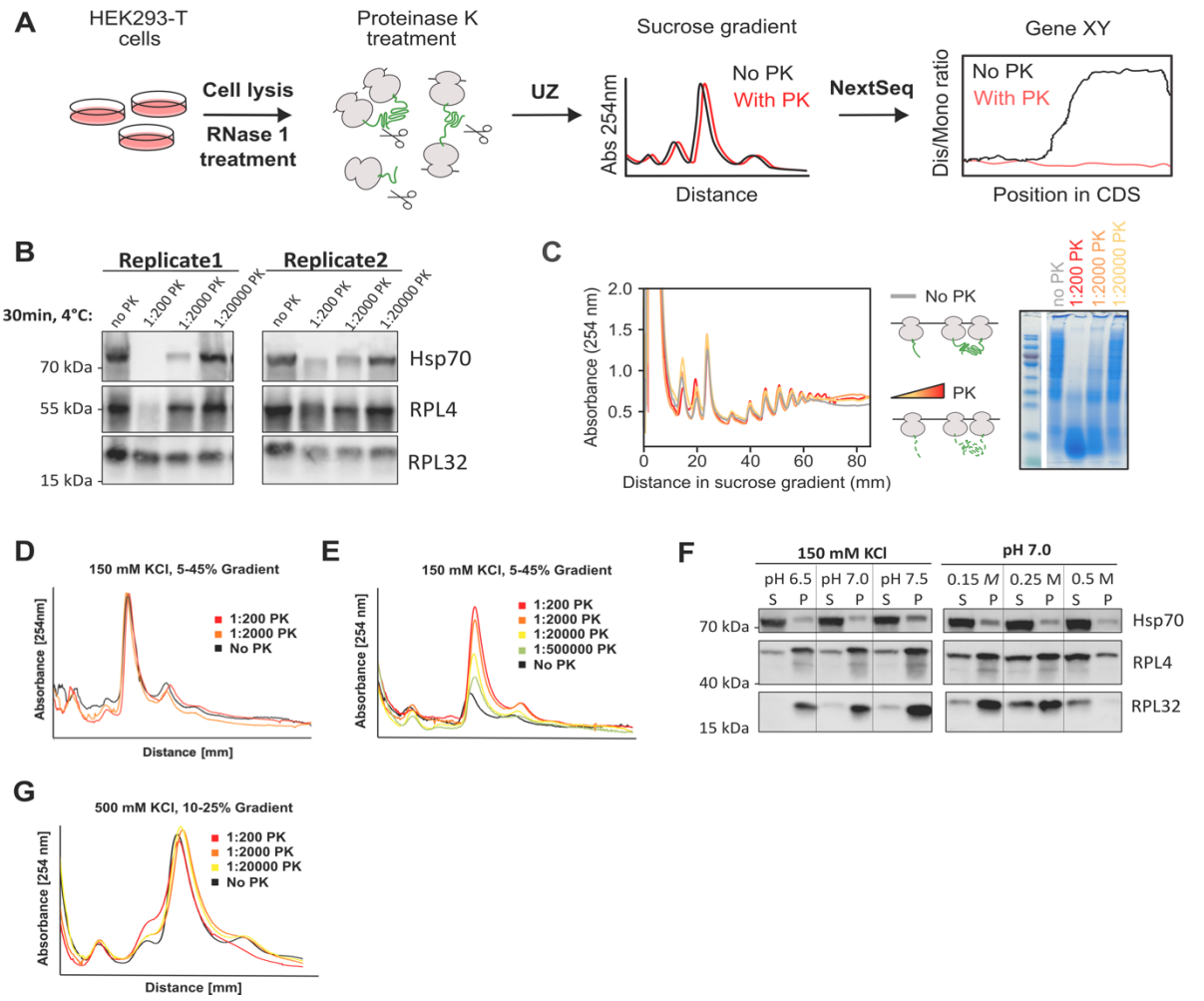

Fig. 11: Monosome to disome footprint shift is sensitive to proteinase K digestion. (A) Experimental set up of DiSP in combination with proteinase K (PK) treatment of RNase I digested lysates. **(B)** Immunostained SDS-PAGE with anti-HSP70 antibody (reflecting the digestion of a cytosolic full-length protein) and with two ribosomal protein specific antibodies (RPL4 is a ribosomal protein on the surface of the ribosome, whereas RPL32 is a buried ribosomal protein). Two replicates of PK titrations are shown. **(C)** Polysome profiles report on the ribosome integrity of lysates treated with the highest employed PK concentration (red line, 1:200) compared to non-treated control lysates (black line). PK digestion activity was proven by Coomassie staining of SDS-PAGE loaded with total lysates after treatment (right). **(D)** 5-45% (w/v) sucrose gradients of RNase I and PK digested lysates under low salt conditions. **(E)** 5-45% (w/v) sucrose gradients of RNase I and PK digested lysates under low salt conditions indicating a resolving effect of PK on aggregated ribosomes in low salt buffer. **(F)** Immunostained SDS-PAGE showing the effects of lysis buffer conditions (pH and salt concentrations was varied) on ribosomal aggregation. **(G)** 5-25% (w/v) sucrose gradients of RNase I and PK digested lysates under high salt conditions. From here on a total protein to PK concertation of 1:200 is termed "very high PK", 1:2000 "high PK", 1:6000 "mid PK", and 1:20000 "low PK" (see Material & Methods).

DiSP of PK digested lysates analyzed under physiological salt concentration of 150 mM KCl (from here on termed low salt) and 500 mM KCl high salt conditions revealed a stepwise loss of the previously observed 2-fold disome enrichment candidates with higher PK concentrations, except for the highest tested concentration of PK under high salt conditions (**Fig. 12A**, **B** for low salt and **Fig. 13A**, **C** for high salt conditions). Only the highest proteinase

K concentration with exposure to high salt appeared to affect the ribosome integrity, indicated by a merging of monosome and disome peaks in the sucrose gradient (**Fig. 11G**). This most extreme condition was therefore excluded from further analysis. Interestingly, the high salt data set in general revealed an overall increase of disome enriched candidates in comparison to the low salt data set, implying that high-salt is particularly well suited for DiSP (further discussed in **4.2.1**). Independent of the used salt conditions, increased PK concentrations generally led to a loss of 2-fold disome enriched candidates (**Fig. 12B** and **Fig. 13C**), a flattening of the disome enrichment at the metagene level (**Fig. 12C** and **Fig. 13B**) and in single gene examples (**Fig. 14A**). Whereas, single gene examples without a disome enrichment indicate no change with increasing PK concentrations (**Fig. 14B**).

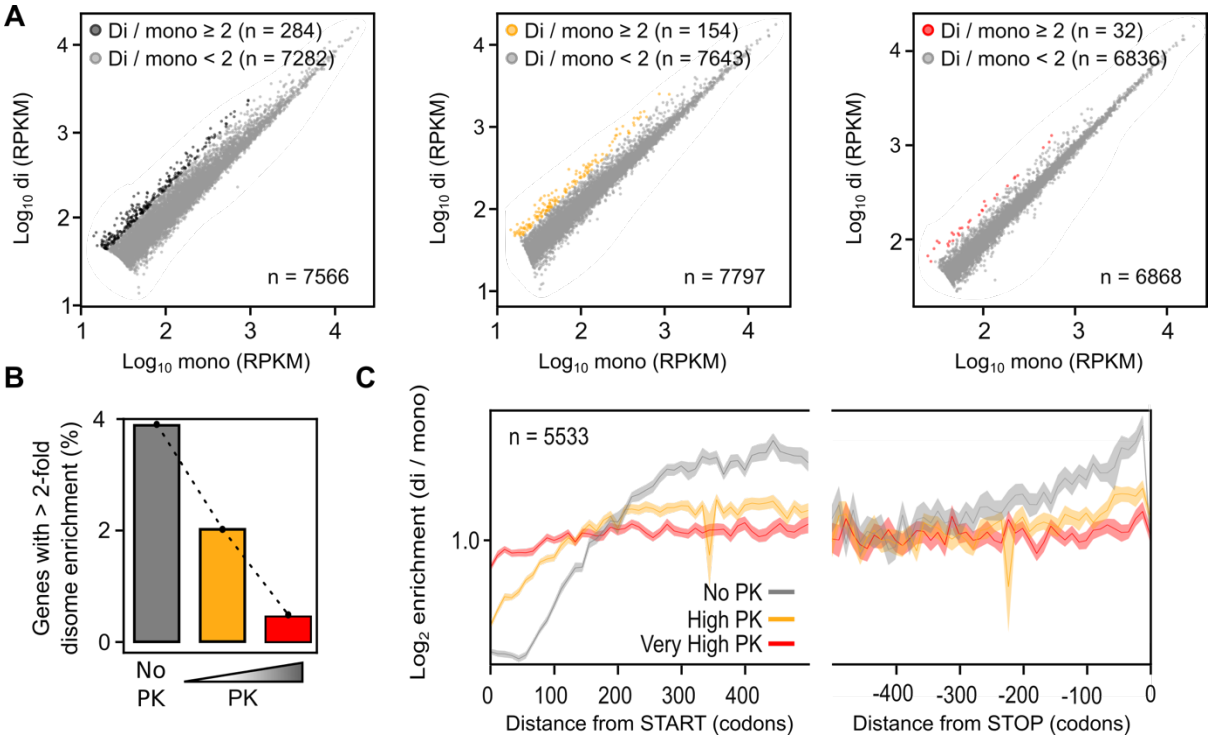

Fig. 12: Monosome to disome footprint shift is sensitive to proteinase K digestion (low salt data set). (A) Scatter plot of disome (di) and monosome (mono) footprint densities of all detected genes in HEK293-T cells without PK treatment (left), high PK treatment (middle), or very high PK treatment (right). **(B)** Bar plot of identified 2-fold enriched candidates depicted in (A). **(C)** Metagene disome enrichment profile aligned to translation start (left) and stop (right) for all PK conditions. For better visualization, all metagenes were averaged with a rolling averaging window of 10 codons.

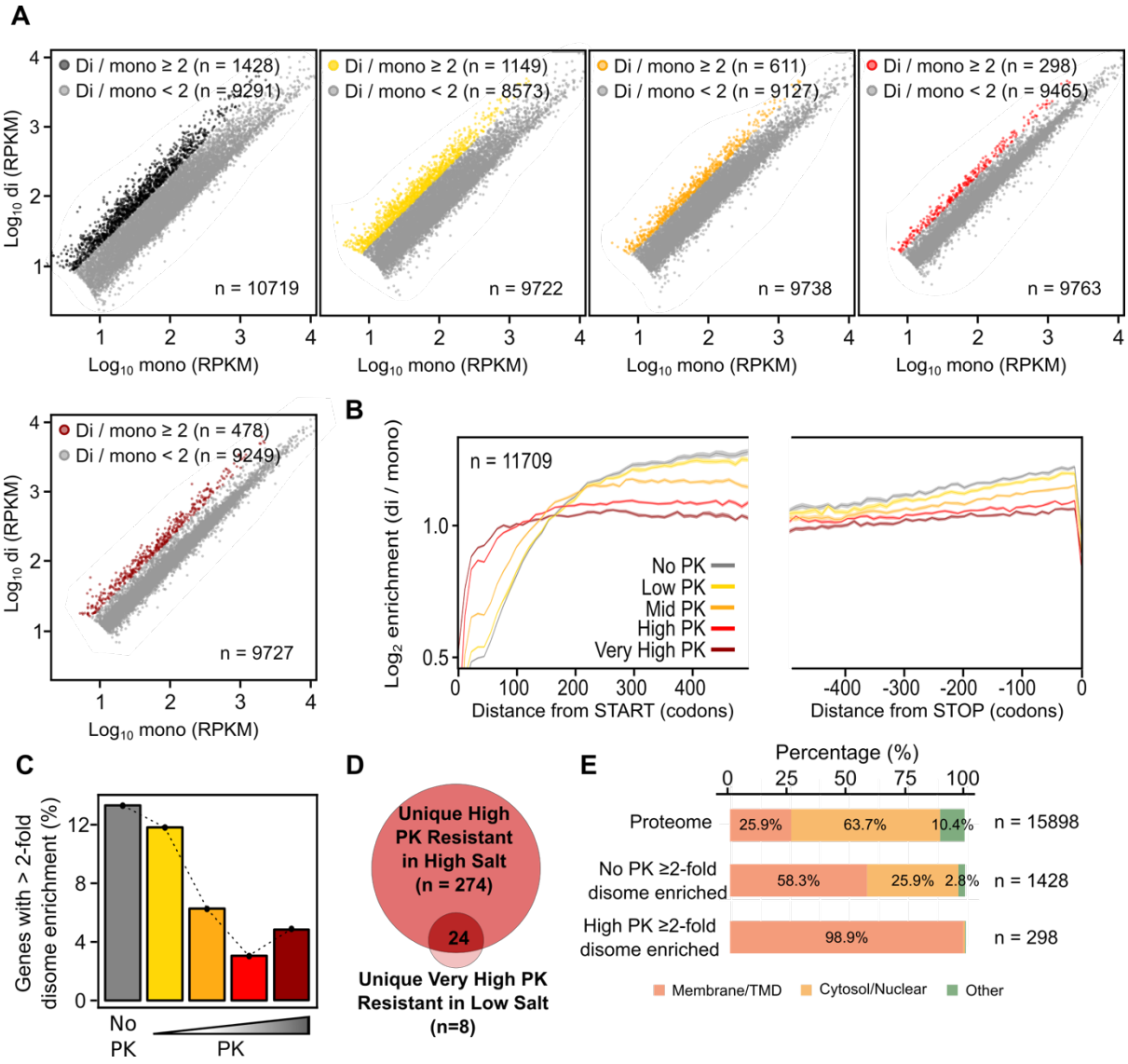

Fig. 13: Monosome to disome footprint shift is sensitive to proteinase K digestion (high salt data set). (A) Scatter plot of disome (di) and monosome (mono) footprint densities of all detected genes in HEK293-T cells for all tested PK treatments. From left to right: no PK, low PK, mid PK, high PK and very high PK treatment **(B)** Metagene disome enrichment profile aligned to translation start (left) and stop (right) for all PK conditions. For better visualization, all metagenes were averaged with a rolling averaging window of 10 codons. **(C)** Bar plot of identified 2-fold enriched candidates depicted in (A). **(D)** Venn diagram of all 2-fold enriched candidates detected under low salt conditions with very high PK or under high salt conditions with high PK (both classified as PK resistant). **(E)** Protein localization for the human proteome in comparison to the 2-fold disome enriched candidates in the no PK and high PK treated samples under high salt conditions.

Strikingly, some candidates did not lose the ≥ 2-fold disome enrichment upon PK treatment. The low salt condition revealed 32 "PK resistant" candidates, whereas the (much deeper sequenced) high salt condition revealed 298 "PK resistant" candidates (**Fig. 12A**, right panel and **Fig. 13A**, fourth panel). The PK resistant candidates identified in both conditions largely overlapped (**Fig. 13D**). The PK resistant class consisted almost exclusively of membrane spanning proteins (**Fig. 13E**). A simple explanation for the apparent PK resistance is that

membrane insertion and docking of ribosomes to the membrane-embedded translocon limits the accessibly of the nascent chains to PK, and thereby prevents nascent chain cleavage.

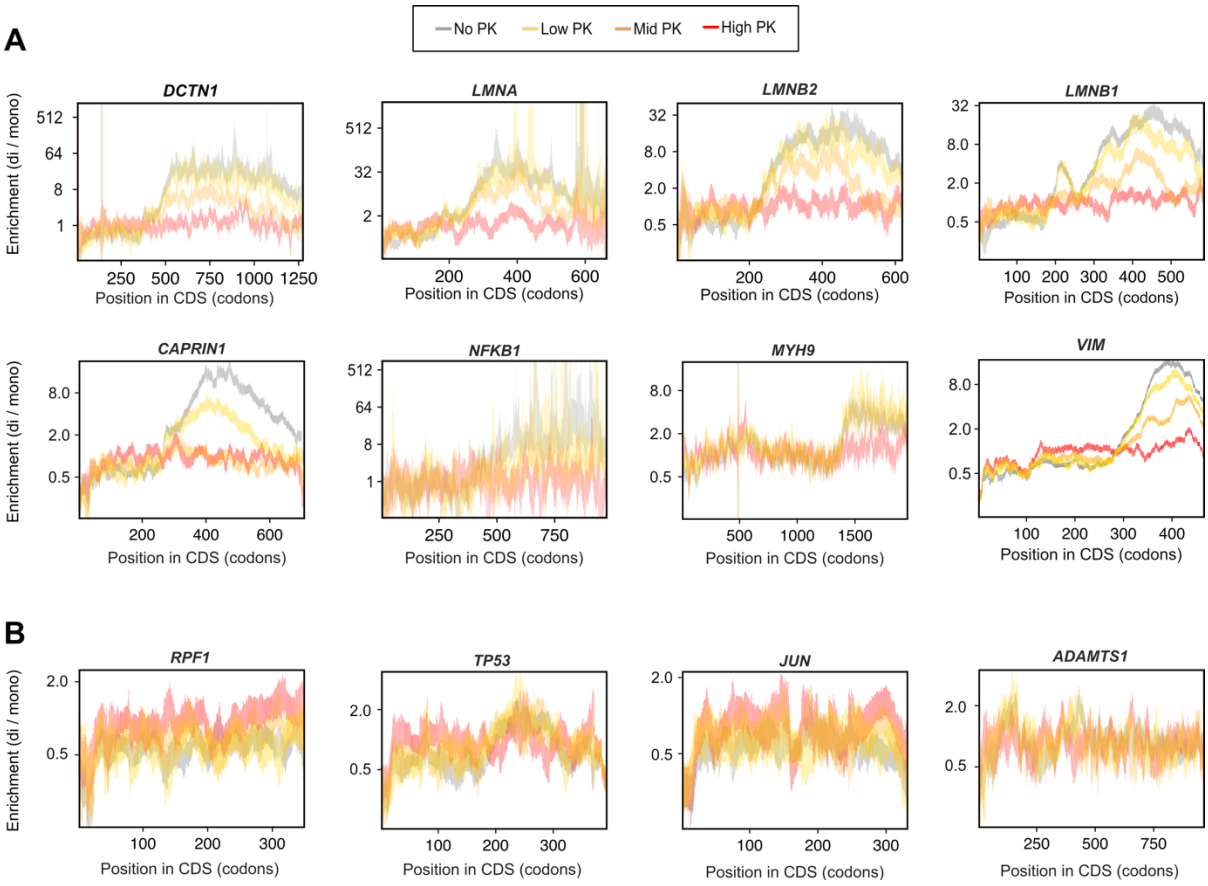

Fig. 14: Single disome over monosome enrichment plots reveal step wise footprint density loss. **(A)** Disome enrichment plots of prominent candidates that showed a loss of the disome over monosome footprint density shift with increasing PK concentrations (high salt DiSP data). **(B)** Disome enrichment plots of candidates without a disome enrichment indicate no change with increasing PK concentrations (high salt DiSP data).

Even though DiSP revealed a loss of 30 nt disome reads after PK digestion, the disome peak height in the sucrose gradients did not change to the same extent (**Fig. 11D**, **G**). This finding indicates that the disome population predominantly encloses other ribosomal pairs. Recent publications showed that collided ribosomes are inaccessible for RNase I and, therefore, generate 60 nt ribosomal footprints (Arpat *et al.* 2020, Han *et al.* 2020, Zhao *et al.* 2020). Such collided ribosomal pairs are not detected by DiSP, as they protect longer footprints. We, therefore, isolated and sequenced the 60 nt fragments from the high salt non-PK treated and high PK treated disome and monosome fractions (**Fig. 15A**). As expected, only the disome peak contained 60 nt fragments that aligned to the human genome (**Fig. 15B**). Metagene profiles revealed that 60 nt fragments mostly occur in the beginning and end of translation, in agreement with the fact that translation initiation and termination are the rate-limiting steps of translation (**Fig. 15C**). Since the metagene profiles indicate the position of the lagging

ribosome of a 60 nt ribosome pair, the footprint coverage ends slightly upstream of the stop codon (**Fig. 15C**). As expected, PK treatment did not alter the overall distribution of the 60 nt footprints along mRNAs. This finding also supports our assumption that the PK treatment did not affect ribosome integrity (**Fig. 15C**). The 60 nt footprint metagene profile further shows the opposite footprint distribution to the 30 nt footprint profiles. This finding hints towards a role of ribosome collision in the initiation phase of co-translational assembly (discussed in greater detail in chapter **4.3.4**).

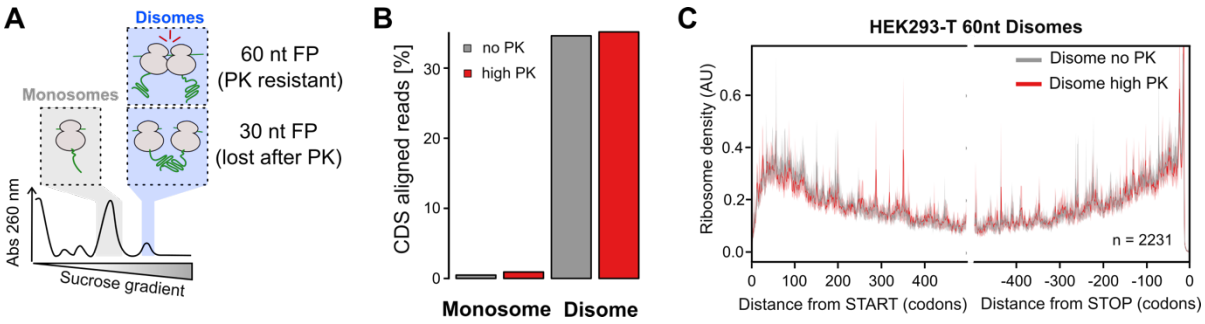

Fig. 15: 60 nt footprints in the disome fraction do not respond to PK treatment. (A) Schematic depiction of a sucrose gradient, showing that the disome peak contains not only nascent chain coupled ribosomal pairs but also collided and incompletely RNase digested ribosome pairs. **(B)** Bar plot showing the percentage of aligned 60 nt footprints isolated from the monosome and the disome fraction of untreated and proteinase K treated samples. **(C)** Metagene profiles of 60 nt disome reads aligned to the start (left) and stop codon (right) of untreated and proteinase K treated samples. The footprint assignment is to the A-site of the lagging ribosome of a collided ribosome pair.

Overall, the PK titration coupled to DiSP analyzing 30 nt footprints indicates that the disome footprint shifts are caused by protein-based interactions. These experiments however do not exclude the possibility that other proteins than nascent chains may contribute to or even cause disome shifts, for example factors that directly bind and couple two ribosomes. Therefore, another series of control experiments was conducted that included the release of nascent chains by puromycin. This set of experiments required a series of optimizations and modifications of the general DiSP protocol to ensure that puromycin treatment does not induce ribosomal disassembly after nascent chain release (see Material & Methods; all puromycin experiments and data analysis was performed by Matilde Bertolini). In brief, it was necessary to omit cycloheximide during ribosome isolation because it prevents nascent chain release by puromycin. Additionally, these control experiments required high salt conditions (500 mM KCl) to allow the efficient release of nascent chains (Blobel and Sabatini 1971). The efficient release was tested by immunostaining of puromycylated nascent chains, which were only detected in the supernatant (ribosome released nascent chain) and not in the ribosome bound fraction (**Fig. 16A**, right). Finally, all steps had to be performed at 0°C to prevent ribosome

disassembly after nascent peptide release by puromycin (**Fig. 16A**, left, polysome structure is unaffected after puromycin treatment). These necessary adaptations of the protocol caused a global flattening of the disome to monosome enrichment (**Fig. 16B**) and a general loss of the 2-fold disome enriched candidates (**Fig. 16C**). As detected in the PK control data set, the disome shift of ribosomes detected for co-co assembly candidates was reduced upon puromycin treatment, whereas all other footprint density distribution of non-co-co assembly candidates was unaffected **(Fig. 16D**, more details about this data set can be found in the PhD thesis of Matilde Bertolini). This data further supports our model that the observed disome footprint shift is caused by nascent chain interactions and does not depend on other factors that directly couple two ribosomes.

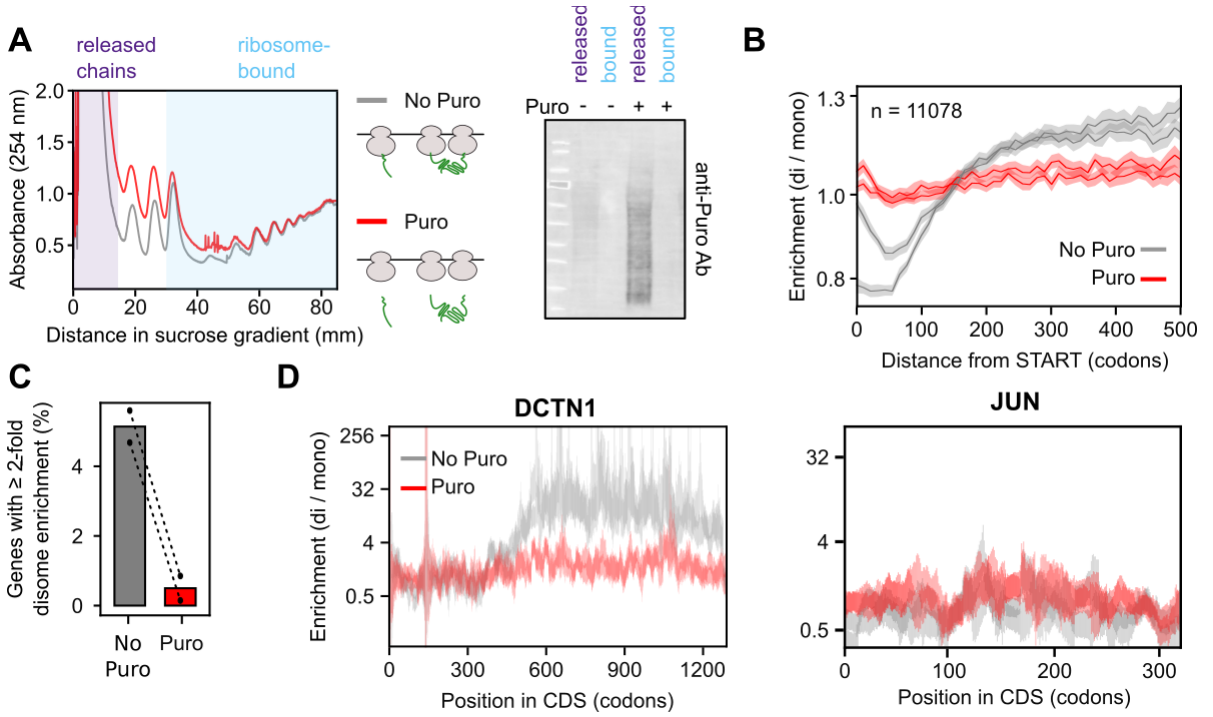

Fig. 16: Nascent chain release via puromycin treatment reduces monosome to disome footprint shifts. (A) Polysome profiles indicating that the ribosome integrity is not affected by puromycin treatment (red line) compared to non-treated control lysates (grey line). Puromycin related nascent chain release was detected by immunostaining (right). **(B)** Metagene profiles are aligned to translation start sites for puromycin treated (red line) and control lysates (grey line). For better visualization, all metagenes were averaged with a rolling averaging window of 10 codons. **(C)** Bar plot of detected 2-fold enriched candidates, with and without puromycin treatment. **(D)** Disome enrichment of *DCTN1*, one of the 2-fold disome enriched candidates (left). Disome enrichment of JUN, that is not identified as co-co assembly candidate (right). Data presented in this figure were generated and analyzed by Matilde Bertolini.

4.2.1 High salt conditions generate a more pronounced disome enrichment

We prepared one final DiSP data set of HEK293-T cells lysed in high salt buffer and loaded on high salt sucrose gradients to compare it to the previous DiSP data performed under physiological salt conditions (150 mM KCl) in HEK293-T cells. The two high salt replicates showed again the expected read distribution in metagene profiles for disome as well as monosome reads (**Fig. 17A**). This new set of replicates revealed around one thousand 2-fold disome enriched candidates out of 10 000 detected transcripts (**Fig. 17B**). The comparison of monosome and disome reads between both high salt replicates revealed a highly correlated read distribution (**Fig. 17C**). Only few candidates differed between both replicates, including only lowly expressed genes for which random fluctuations in read counts most likely affected the calculated enrichment values (**Fig. 17C**, red dots). A similar trend was also obtained by comparing the disome over monosome ratios (di/mono enrichments) from both high salt replicates, indicating that the variation of 2-fold disome enriched candidates between both replicates is not caused by fundamental differences in the determined disome over monosome enrichments nor by a distinct class of candidates (**Fig. 17D**, left).

The reproducibility of DiSP under high- and low salt conditions was supported by a clear correlation of the disome over monosome enrichments obtained in HEK293-T cells (**Fig. 17D**, right). Interestingly, the scatter plot indicates a slightly tilted data point cloud towards the high salt conditions. The generally higher enrichment scores obtained under high salt conditions are presumably caused by greater separation of monosome and disome fractions upon sucrose gradient centrifugation, indicated by sharper monosome and disome peaks of high salt samples (data not shown) that lead to higher enrichment scores (see Supplemental Material: Optimization of DiSP). A sigmoidal fitting tool, developed by Ilia Kats (described in chapter **4.3**), was employed to determine and correlate the footprint density shift onset of all candidates with a disome shift in the low salt and high salt HEK293-T data sets. The determined onsets between low- and high salt data sets highly correlate (**Fig. 17E**). In summary, these results suggests that the use of high salt conditions improves the overall disome enrichment but does not affect the overall shape of detected single candidate disome enrichment profiles.

These untreated high salt data sets were therefore used for all downstream analysis of co-co assembly by comparing it to both high salt PK and puromycin control data sets (see **4.3**).

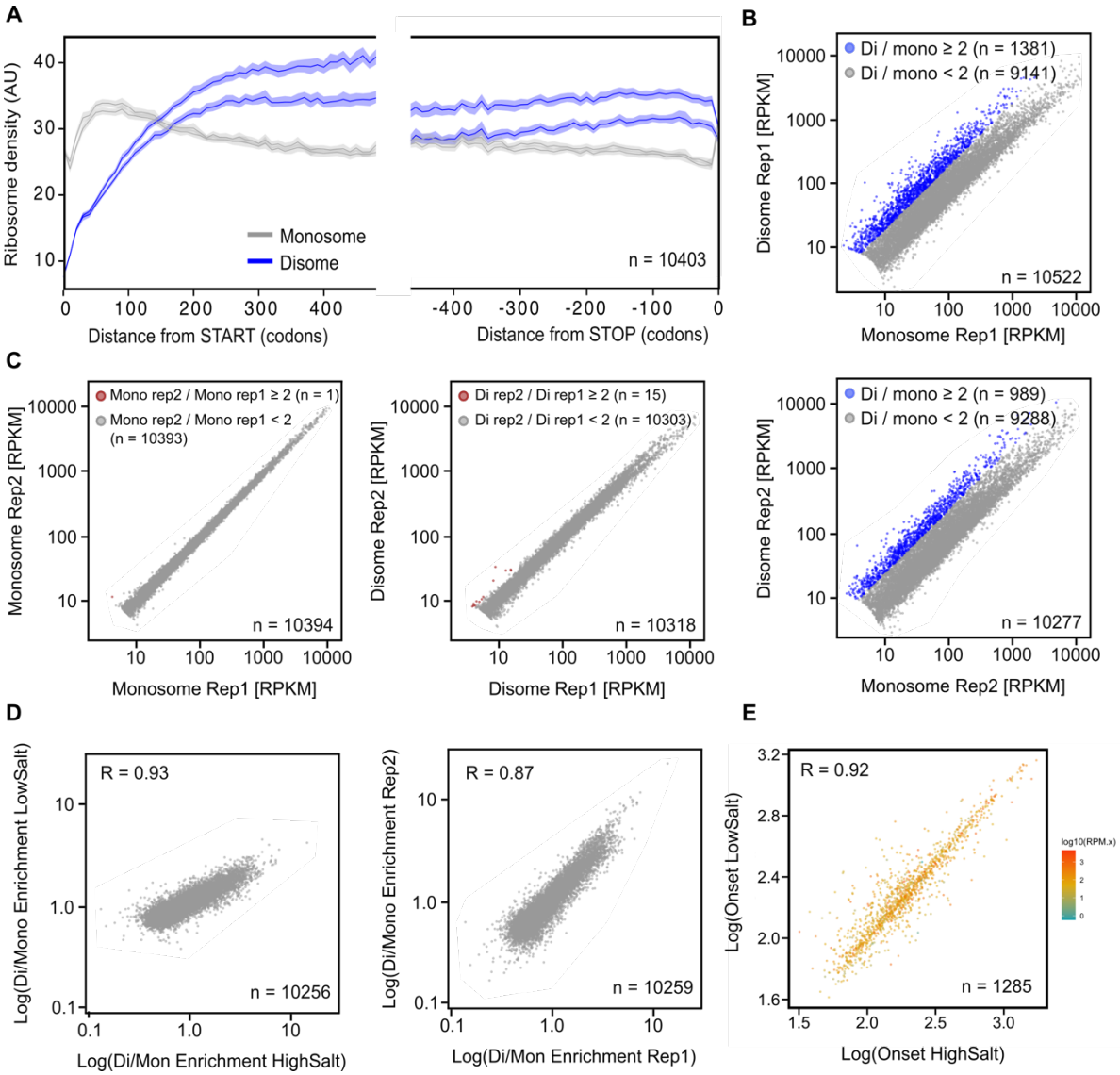

Fig. 17: DiSP performed using high salt and low salt buffers highly correlate. (A) Metagene profiles aligned to translation start (left) and stop (right) for both high salt replicates. For better visualization, all metagenes were averaged with a rolling averaging window of 10 codons. **(B)** Scatter plots of monosome versus disome reads in both high salt replicates. Blue dots indicate all 2-fold disome enriched candidates. **(C)** Scatter plot comparing monosome reads in both high salt replicates (left, Pearson correlation = 0.996) and disome reads in both high salt replicates (right, Pearson correlation = 0.971). **(D)** Scatter plots comparing the disome over monosome enrichments between both high salt replicates (left, Pearson correlation = 0.93) or between the low and high salt conditions (right, Pearson correlation = 0.87). **(E)** Scatter plot comparing the determined onsets of the disome shift between low and high salt conditions of candidates with a sigmoidal footprint detection in both data sets (Pearson correlation = 0.92).

4.3 How prevalent is co-co assembly and what type of complexes and protein domains are involved?

Calculating enrichments of ribosome densities in disome over monosome datasets have the intrinsic limitation that short disome shift periods are poorly detected. To overcome this limitation and to also reveal when during translation disome shift occur, an alternative analysis tool had to be developed. This tool employed the newly obtained high salt data sets of HEK293-T cells and also considered the data of both control DiSP experiments (one set of PK titration experiments and two DiSP-puromycin replicates) to reveal a high confidence co-co assembly candidate list. The tool is based on a sigmoid fitting procedure that identifies all single gene enrichment profiles, which display a disome footprint density shift (**Fig. 18A left**; the tool was developed by Ilia Kats, see Material & Methods). This tool not only detected stable co-co assembly events that remained until the end of translation but also identified transient disome shifts, suggested by a double sigmoidal disome enrichment profile (**Fig. 18A, B**). A double sigmoidal shape can be caused by transient interactions, but also if the interacting ribosome terminate at different time points (**Fig. 18B**). In this case the delayed ribosome would shift back to the monosome fraction before it reaches the end of the coding sequence (**Fig. 18B**). Such asynchronous termination of nascent chain coupled ribosomes could occur for two neighboring ribosomes on one mRNA (*cis* assembly) as well as ribosomes on different mRNAs (*trans* assembly) (**Fig. 18B**).

All single gene enrichment profiles that indicated no footprint density shift were classified as non-co-co assembling candidates. This does not exclude the possibility that other co-translational assembly events occur like co-post assembly, which cannot be detected by DiSP (**Fig. 18B**). This unbiased bioinformatics selection regime allowed to identify and rank potential co-co assembly candidates. To be considered as a high confidence co-co assembly candidate the following selection criteria had to be fulfilled: (i) Detection of a single or double sigmoidal shape within the disome to monosome enrichment profile. (ii) The sigmoidal enrichment must be significantly lost upon treatment with puromycin and proteinase K. (iv) The mature protein must localize to either the cytoplasm or the nucleus.

All proteins that did not meet criteria (i) were directly classified as non-co-co assembly proteins, whereas all that met criteria (i) but not (ii - iv) were classified as low confidence co-co assembly candidates (**Fig. 18A**). We decided to categorize translocated proteins as low confidence because we aimed for a high confident class, which should contain only candidates of high certainty. For membrane proteins we could not formally exclude the possibility that the interaction with membrane components of the translocation machinery may cause a migration into the disome fraction during sucrose gradient centrifugation. Importantly, all downstream

analyses were performed on both, high- and low confidence candidates, membrane proteins were analyzed separately (see result section **4.3.3**).

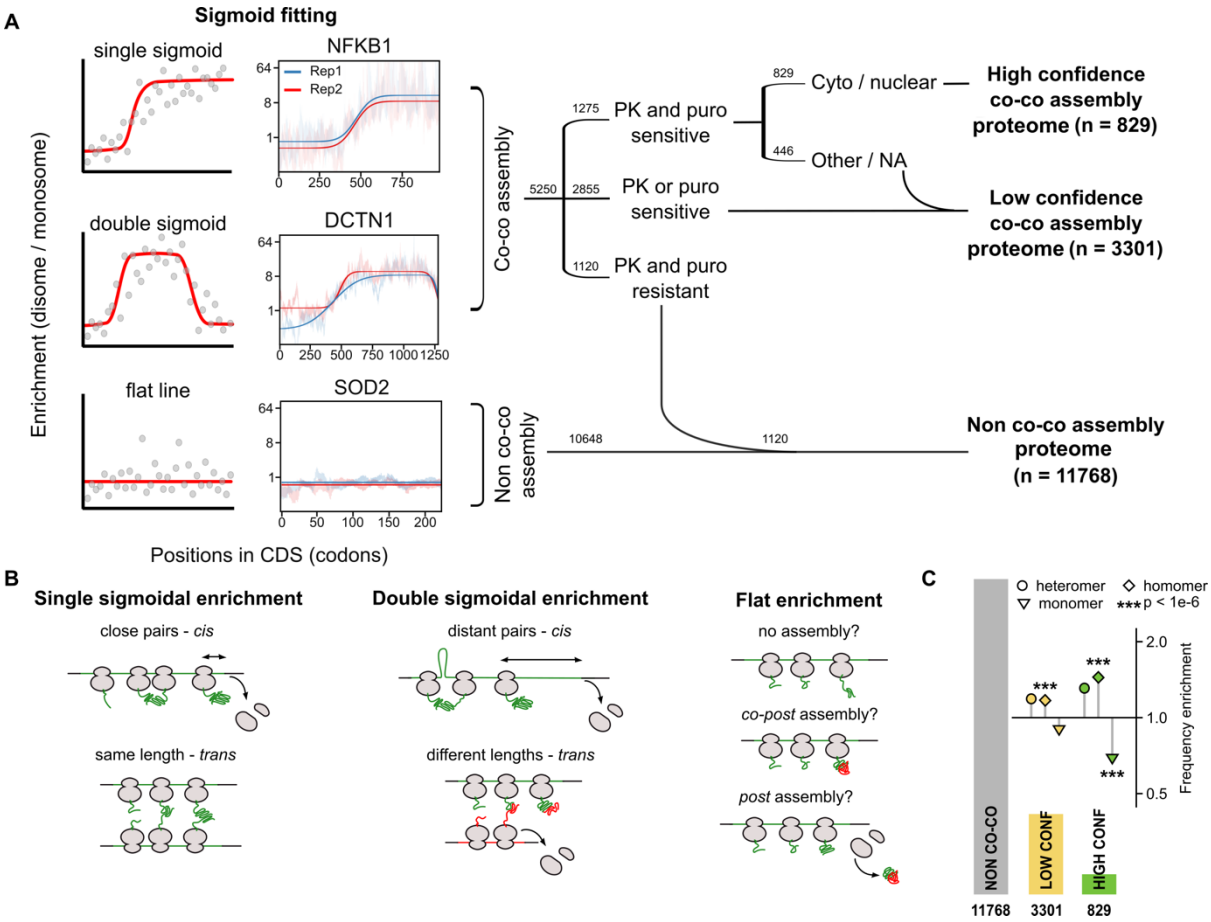

Fig. 18: High confidence co-co assembly candidates involve many homomeric complexes. (A) Schematic representation of disome over monosome enrichment read distributions (grey dots in the background) and the corresponding fitted curves (red solid lines) for each possible model (single, double sigmoid or flat line). All single and double sigmoidal candidates are further filtered based on their sensitivity to PK and puromycin treatments and their annotated subcellular localization to define a high- and low confidence class. All disome over monosome enrichment read distributions without detectable sigmoidal fits were classified as the non-co-co assembly class. **(B)** Schematic depiction of mechanisms causing single and double sigmoidal disome over monosome enrichment read distributions as well as mechanisms associated to assembly than can cause flat enrichment profiles. **(C)** The bar plot shows the numbers of annotated homomer, heteromer and monomer subunits detected in each class defined in (A). The enrichment of homomers, heteromers and monomers in the high and low confidence class is shown as odds ratio (enrichment of annotated complex subunits in each class is normalized by the frequency of these subunits in the proteome). The high confidence class only includes cytosolic or nuclear localized proteins, while the low confidence class encloses proteins of all subcellular localization, therefore both were compared to different backgrounds (see Material & Methods).

The sigmoidal fitting tool classified many genes as sigmoidal, even though the disome over monosome enrichment was almost invisible. This mild sigmoidal shift was mostly caused by a steady increase of the disome fraction alone, combined with an almost non-observable

decrease of monosome reads. The PK and puromycin sensitivity criteria removed most of these weak sigmoidal candidates, as this phenomenon was insensitive to both treatments.
The list of high- and low-confidence co-co assembly candidates was significantly depleted for annotated monomeric proteins and enriched for complex forming subunits (**Fig. 18C**). The indicated enrichments are calculated by dividing the frequency of homomers or heteromers in each class by the frequency of homomers or heteromers in the respective background proteome. The high confidence class includes only cytosolic or nuclear localized proteins, therefore the frequency is calculated in regard to the cytosolic or nuclear human proteome. While the low confidence class encloses proteins of all subcellular localization and is therefore compared to the entire human proteome.
The high and low confidence lists were only significantly enriched for homomeric proteins. Heteromers indicate also a visible but non-significant enrichment in both classes. This is caused by the employed significance test, that is corrected for expression biases (developed by Ilia Kats, see Material & Methods). Additionally, the low confidence list suffered from poor oligomeric annotations of the included membrane-associated proteins, based on the employed UniProtKB/Swiss-Pro database (The UniProt Consortium 2019).

4.3.1 What nascent chain segments are exposed at assembly onset?

By having established a high confidence list and the corresponding onset of disome enrichment, we explored what part of the nascent chain is exposed before the onset of assembly. The determined onsets for all high confidence list candidates revealed that most co-co assembly interactions started before 50% of the mRNA sequence was translated (**Fig. 19A**). Consistently, we identified that homodimerization interfaces are enriched in the N-terminal halves of high confidence candidates, based on publicly available crystal structures (**Fig. 19B** left, analysis of crystal structures was performed by Frank Tippmann). This seems to be a specific feature of homomers in the high confidence class, because the same analysis performed on all available crystal structures of human homomers does not show such an enrichment (**Fig. 19B**, right). This is in agreement with published data that in general, homodimerization interfaces in the human proteome are enriched towards the C-terminus (Natan *et al.* 2018).
Residues involved in subunit interactions were mostly found to be exposed at assembly onset for homomers, whereas this enrichment was much weaker for heteromeric proteins of our high confidence list (**Fig. 19C**). Further indicating that co-co assembly involves mostly homomers.

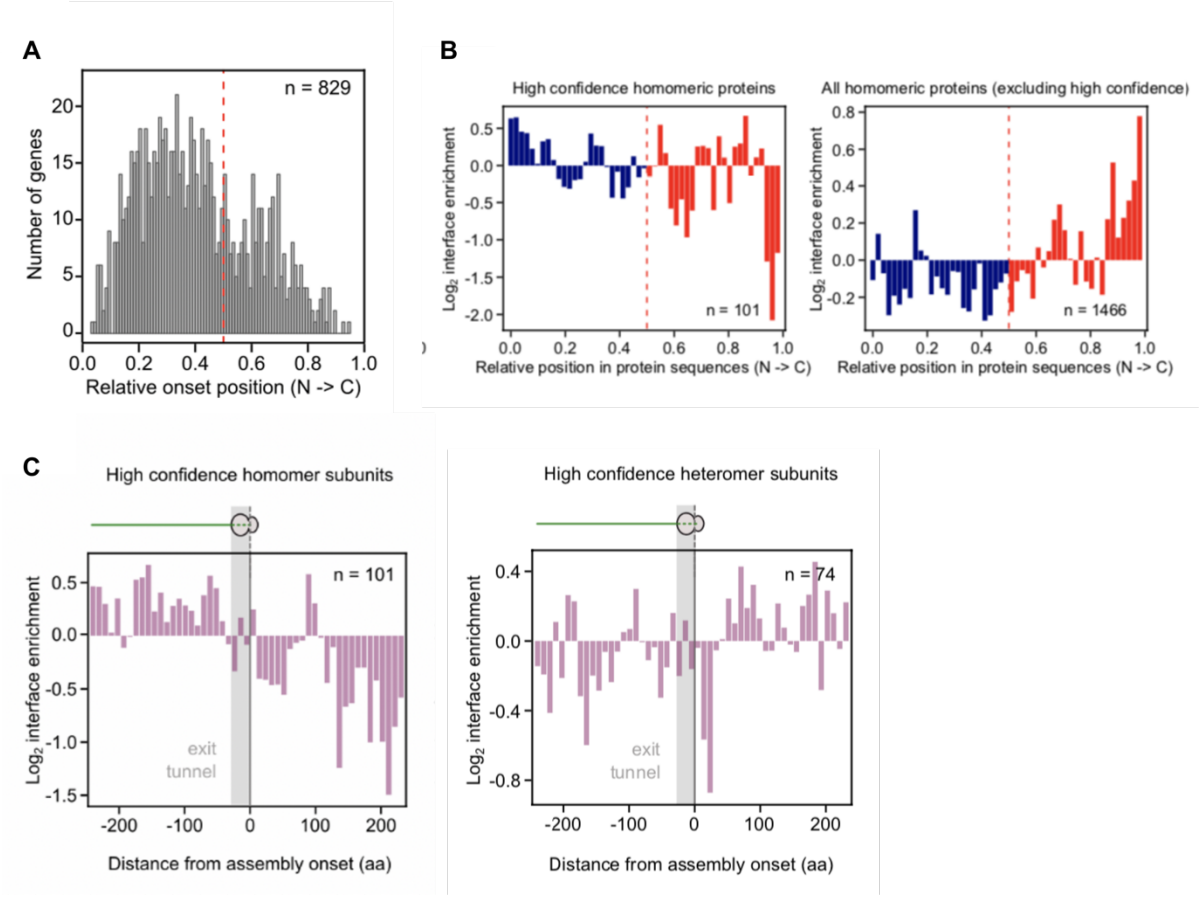

Fig. 19: The onset of co-co assembly occurs together with the exposure of N-terminal interaction domains (A) Histogram of the relative onset positions of high confidence co-co assembly candidates. The red dashed line indicates the middle of the coding region. **(B)** Relative enrichments of segments forming the complex subunit interface for high confidence homomeric complexes (left) or including all homomeric complexes in the human proteome excluding high confidence candidates (right). Interface positions were determined from crystal structures. All genes are normalized to the same length. Blue and red bars left and right of the vertical dashed line indicate interface enrichment in the N-terminal and C-terminal halves of proteins, respectively. **(C)** Distribution of residues of the inter-subunit interface of homomers (left) or heteromers (right) of the high confidence list determined from available crystal structures. The interface enrichment is shown along the proteins' primary sequence aligned to the assembly onset. The data analysis shown in (B) and (C) were performed by Frank Tippmann.

In total, seven significantly enriched homodimerization interfaces were identified in our high confidence list (**Fig. 20A**). Overall, 193 out of 829 high confidence candidates revealed an annotated coiled coil domain that was partially or fully exposed before the onset of co-co assembly (**Fig. 20B**, left). The median length of exposed coiled coils was 111 residues, and strongly varied between different candidates (**Fig. 20B**, right). This finding is in agreement with the extensive heterogeneity of coiled coil domains, and the highly variable interactions strengths that depends on involved residues and their arrangement according to the required heptameric repeat to stabilize a coiled coil. We classified seven additional domains that are generally positioned N-terminally to coiled coil domains (colored orange in **Fig. 20A**), and

therefore, are exposed already at the onset of co-co assembly, but presumably do not contribute to oligomerization.

Another dimerization motif, which is only partially exposed at the onset of co-co assembly are BAR domains (**Fig. 20C**). These alpha-helical domains consist of three (classical BAR) or five (F-BAR) antiparallel alpha-helices, and are known to mediate membrane curvature. For all identified candidates we found that at least the most N-terminal alpha-helix (helix1) that interacts with its partner helix1' in an antiparallel fashion, was exposed at the onset of co-co assembly (**Fig. 20C**)**.**

Our DiSP data also revealed globular dimerization domains, which were always fully exposed at the onset of co-co assembly. This is in agreement with the requirement of co-translational domain folding before assembly**.** The most abundant globular dimerization domains were BTB domains found in 36 candidates of the high confidence list, followed by 12 SCAN domain candidates and 4 Rel-like homology domains (RHD) (**Fig. 20E**). Visual inspection of single candidate enrichment profiles suggested that the bioinformatically determined onsets are not always precise, in particular when profiles showed long shallow increases of enrichment values. A too early determined onset, by approximately 10-30 codons, was often determined for the lower half of the BTB candidates shown in the heatmap (**Fig. 20E**, left), indicating that really all BTB domains are fully exposed before onset. All three motifs are N-terminally located domains known to mediate mostly homo- but also hetero-oligomerizations. Interestingly, all identified dimerization domains are conserved and highly redundant in the human proteome. Our finding suggests that many proteins employ these five dimerization domains to facilitate co-co assembly.

In addition, we identified two less characterized motifs leading to co-co assembly. The first are STI1 repeats, mostly found in ubiquilin proteins (**Fig. 20D**). This domain mediates homo- and heterodimerization of ubiquilin 1 and ubiquilin 2, two high confidence candidates that fully expose the second STI1 repeat (STI1 2) at the onset of assembly (**Fig. 20D**). The second motif is GBD/FH3, a regulatory element in Diaphanous-related formins, which is involved in nucleation and remodeling of the actin cytoskeleton. The FH3 domain has been implicated in dimerization of the mouse homologue of human *DIAPH1* (Rose *et al.* 2005b). We found six human formins among our high confidence proteins and in all cases the FH3 domain was exposed at assembly onset, suggesting that formins may co-translationally assemble via the FH3 domain (**Fig. 20D**).

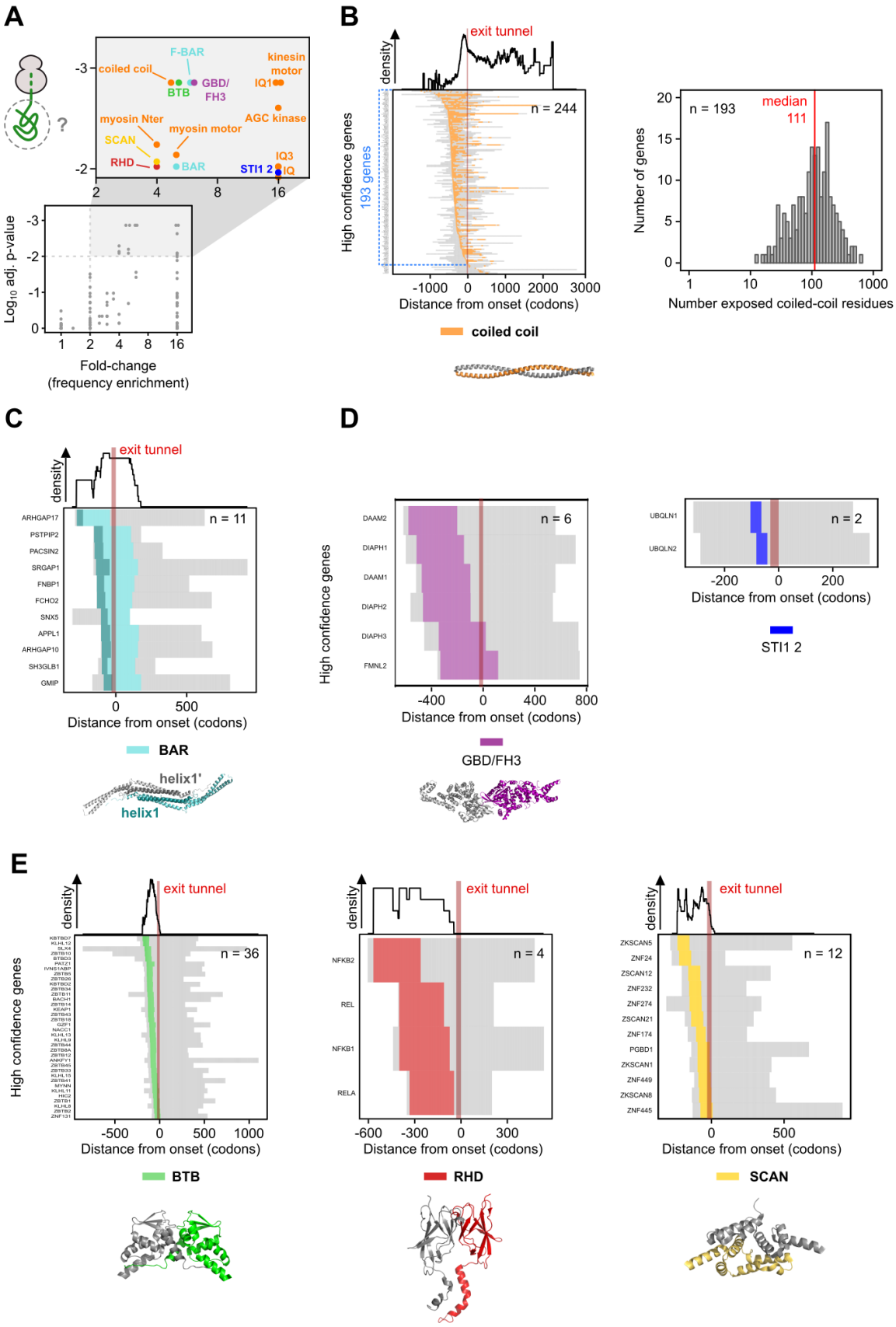

Fig. 20: Protein domain classes conferring co-co assembly. (A) Analysis of protein domains exposed at assembly onset. The frequency of each domain in the high confidence class is compared to their general frequency in the human proteome (see Material & Methods). **(B)** Heatmap of all coiled

coil domain encoding mRNAs in the high confidence class (left). Everything that is left from the red bar (bar indicates the ribosome exit tunnel) are exposed nascent chain segments at assembly onset. All candidates that expose at least parts of a coiled coil domain at the onset of assembly are highlighted with a blue dotted line (n = 193). Density plot on top indicates the distribution of coiled coil residues relative to the onset of co-co assembly. Indicated structure is from PDB: 1D7M. The number of ribosome-exposed coiled coil residues at assembly onset are visualized as histogram (right). Median of exposed residues is shown by a red line. **(C)** Heatmap of all BAR domain encoding mRNAs in the high confidence class (residues forming helix1 of the BAR domain are colored dark cyan in the heatmap and in the indicated structure, PDB: 3Q0K). Corresponding domain density profiles are shown on top. **(D)** Heatmap indicating the STI1 2 and GBD/FH3 domain positions at the assembly onset of high confidence candidates. Shown structure is from PDB: 1z2c. **(E)** Heatmap of completely exposed domains: BTB (left), RHD (middle) and SCAN (right). Corresponding domain density profiles shown on top. Representative structures are PDB: 1BUO, 1K3Z, 3LHR.

4.3.2 What fraction of monosomes shift to the disome state after onset of co-co assembly?

As described for the first DiSP data set, we do not only observe increased disome reads at the assembly onset, but in most cases also a simultaneous monosome depletion. We reasoned that the monosome depletion can report on the efficiency of co-co assembly, as it indicates the fraction of translating ribosomes that shift to the disome state and are engaged in co-co assembly. For a global analysis, it was important to exclude that the change of ribosome footprint densities is not caused by different translation kinetics that occurs in the 3' end of the coding sequence. Therefore, we determined the reduction of footprints from a total translatome before and after initiation of co-co assembly and normalized it to the reduction of footprint from the isolated monosomes (see Material & Methods). A metagene profile with total translatome data including all high confidence candidates showed that the footprint distribution does not change significantly after the onset of co-co assembly. The steady continuation of translation after onset in a total translatome suggest that the translation rate is unaltered and ribosomes do not disassemble or slow-down after onset of assembly (**Fig. 21A**, left). Whereas a metagene profile of all high confidence candidates in the monosome sample indicates a clear loss of footprints after co-co onset, which indicates the footprint shift to the disome state (**Fig. 21A**, left). The comparison of monosomes to total translatome footprints before and after onset allowed to calculate the percentage of ribosomes engaged in co-co assembly. The median monosome footprint reduction was around 40%, but exceeded for several candidates 90%, indicating that in many cases the majority of nascent chains assembled co-translationally (**Fig. 21A**, right). The same trend was detected for all low confidence list candidates (**Fig. 21B**). The three proteins featuring the highest efficiency (≥ 90% depletion), are *TPR*, *EEA1* and *CLIP1*. All contain an extremely long coiled coil homodimerization domains (between 1000 and 1500 coiled coil involved residues, compared to a median coiled coil length of 66 residues in the cellular proteome). This suggests a high stability of the co-co assembled

disomes, which may explain the strong monosome depletion (**Fig. 21C**). Total translatome data indicate no footprint loss along the mRNAs of these candidates, suggesting that the observed monosome loss in our DiSP data is caused by a significant footprint shift to the disome fraction and not caused by a footprint loss due to ribosome disassembly (**Fig. 21D**).

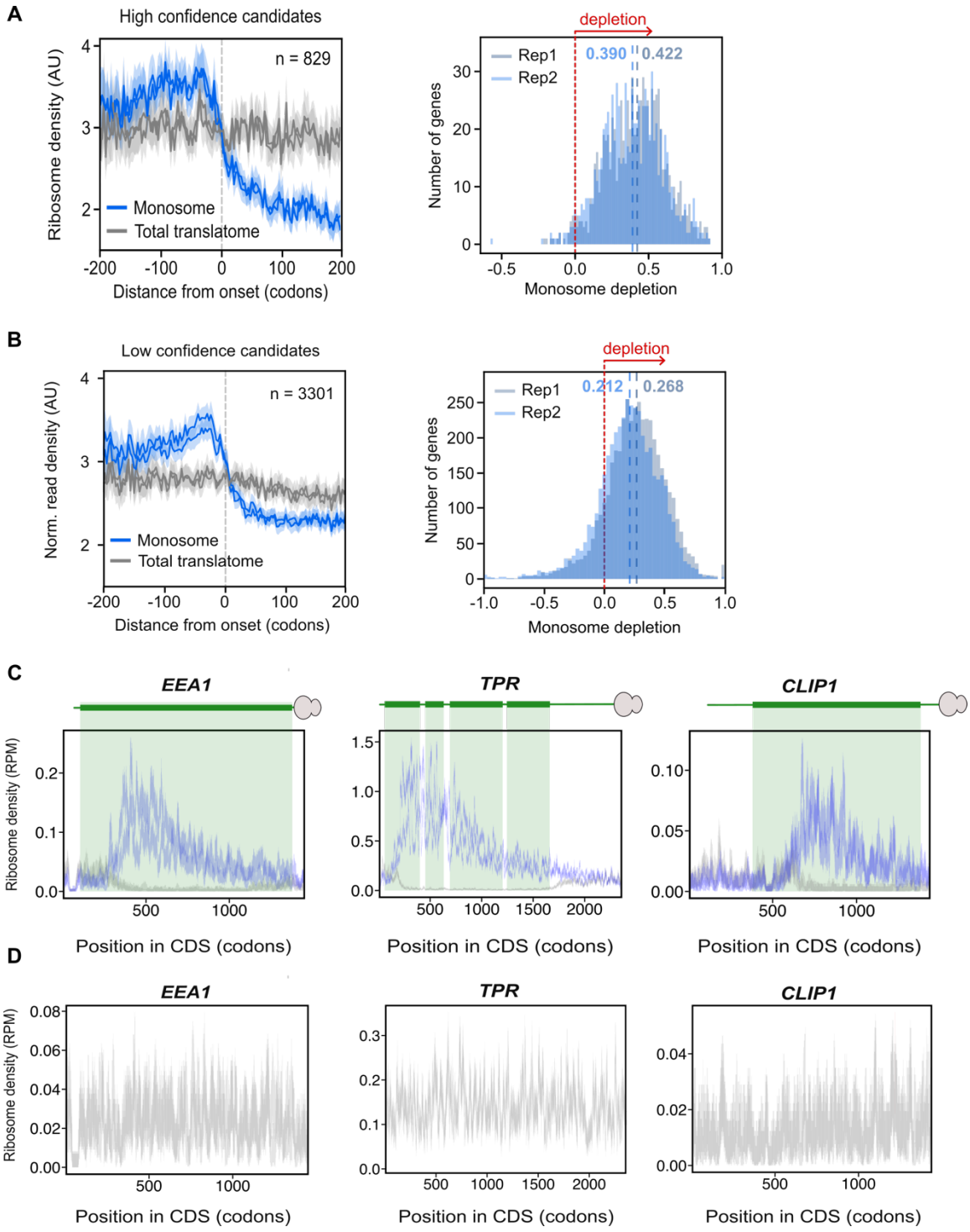

Fig. 21: Monosome depletion reports on the efficacy of co-co assembly. (A) Metagene profiles of the monosome fraction and a total translatome including all high confidence candidates aligned to co-

co assembly onset (left). Histogram of monosome depletion of all high confidence candidates. Monosome depletion is quantified for each gene separately by analyzing the fraction of remaining footprints downstream compared to upstream of assembly onset, each normalized by the total translatome (right). Median monosome depletion in two replicates are shown by blue dashed lines. **(B)** Same as in (A) but for the low confidence class. **(C)** Ribosome density plots for single gene candidates with the highest monosome depletion (>90% monosome depletion after onset of assembly). Green boxes indicate the exposure of the annotated coiled coil domains. Two biological replicates are shown. **(D)** Ribosome density plots from total translatome data of HEK293-T cells for candidates shown in (C).

4.3.3 Co-co assembly of transmembrane containing proteins

The first analysis of membrane proteins was restricted to candidates that have been excluded from the high confidence list because of the presence of a transmembrane domain, yet fulfilled all other criteria of our ranking. A heatmap of all these candidates revealed that the onset of co-co assembly mostly occurs after exposure of the first TMD (**Fig. 22A**). This onset after only one exposed TMD suggests that co-co assembly involves interactions of two peptides, each providing one TMD within the endoplasmic reticulum membrane (**Fig. 22A**). We measured the distance between the onset of assembly and the first amino acid of the nearest upstream TMD domain. This revealed a median of 61 translated amino acids, implying that the onset of assembly occurs in most cases almost immediately after full TMD exposure from the ribosome (assuming an average ribosome exit tunnel length of 30 amino acids and an average TMD length of 21-24 residues in humans) (**Fig. 22B**). The identified co-co TMD candidates showed no strong preference for the TMD membrane orientation (**Fig. 22C**).

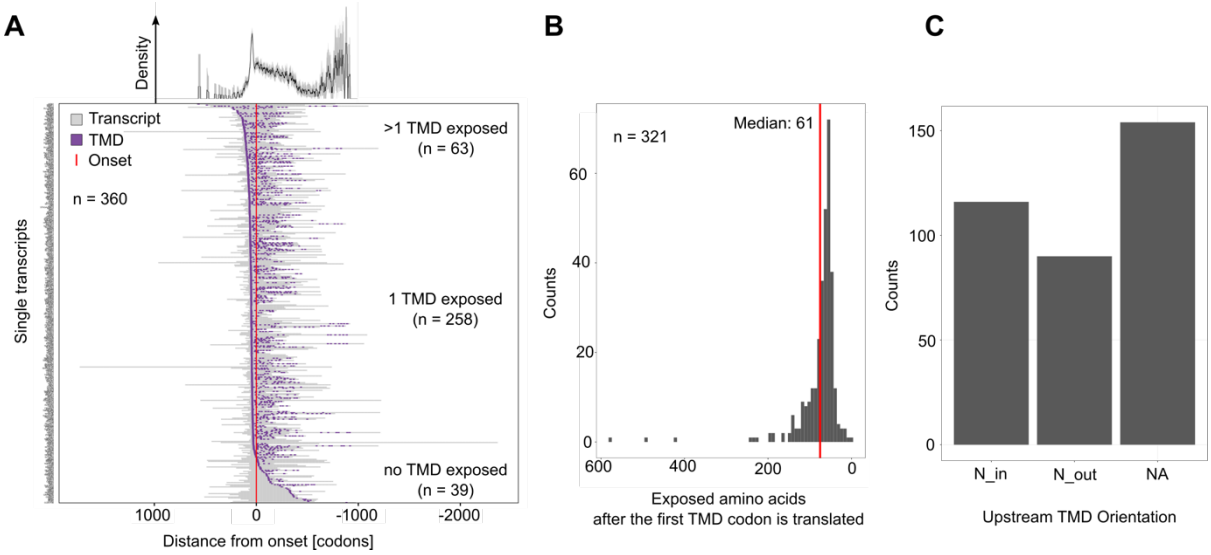

Fig. 22: Characterization of TMD containing proteins in the low confidence list. (A) Heatmap of all transmembrane domain encoding mRNAs aligned to the onset of co-co assembly (red line). Included are all genes that were classified as low confidence candidates by fulfilling criteria (i) to (iii) of our ranking system (described in **Fig. 18**). Density plot on top indicates the distribution of TMD residues along the coding sequence of all included genes. **(B)** Distribution of the number of residues exposed before co-co assembly onset, counted from the start of the TMD to the C-terminal residue positioned in

the ribosomal P-site. Included are all candidates with an exposed TMD before co-co assembly onset. Median of exposed residues is indicated by a red line. **(C)** Bar plot of the first exposed TMD orientations before the onset of co-co assembly. N_{in}: N-terminus of the proteins faces towards the cytosolic side; N_{out}: N-terminus of the proteins faces towards the ER lumen.

Considering that the TMD drives the dimerization, we next asked what feature of TMDs may trigger co-co assembly. The majority of co-co assembling membrane proteins revealed a clear onset after exposure of the first TMD (258 out of 360 candidates, **Fig. 22A**). Other candidates exposed multiple exposed TMDs before onset. Because these multiple TMDs are mostly very close along the peptide structure, it is often not possible to define, which TMD leads to the onset of co-co assembly. Therefore, candidates with multiple TMDs before onset were excluded and only single exposed TMDs before onset were classified as "co-co (mediating) TMDs". To define a "non-co-co TMD" class, we first selected all TMD containing proteins that were classified as non-co-co assembling proteins and assigned onsets along the coding sequence (based on a tool developed by Ilia Kats, which randomly assigns an assembly onset based on the onset distribution of all co-co assembly candidates with a TMD, see Material & Methods). Afterwards, single TMDs exposed before the assigned onset were chosen as the "non-co-co TMD" control group. The mean hydrophobicity scores were calculated for both groups of TMDs (**Fig. 23A**). The co-co TMD class had a slightly lower average hydrophobicity score, presumably caused by the higher frequency of glycine residues within the co-co assembly transmembrane domain (**Fig. 23B**). However, not only the amino acid composition but also the order of amino acids within a TMD is important. We searched for a potential motif enriched in co-co TMDs that may confer co-co assembly. One motif known to promote and stabilize helix–helix interactions of membrane proteins is the GxxxG motif (x represents any hydrophobic amino acid) (Kleiger *et al.* 2002, Kleiger and Eisenberg 2002). The higher frequency of glycine in the middle of all co-co TMDs is in agreement with the hypothesis that these TMDs mediate dimerization inside the ER membrane.

The group of Gunnar van Heijne developed a prediction tool that calculates the apparent free energy difference ΔG_{app}, for insertion of a putative TMD sequence into the ER membrane by the Sec61 translocon (Hessa *et al.* 2007). Their calculation tool was employed to calculate the ΔG_{app} of the "co-co TMD" and "non-co-co TMD" class (http://dgpred.cbr.su.se). We observed a mean negative ΔG_{app} value for the non-co-co TMD class, which indicates that the sequence is predicted to be recognized as a transmembrane helix by the Sec61 translocon and integrated into the membrane (**Fig. 23C**). The "co-co TMD" class has a mean positive ΔG_{app} value of membrane insertion, which suggests that the TMD segment cannot be efficiently inserted without a stabilizing interaction with surrounding helices (**Fig. 23C**). This observation indicates that co-co assembling TMDs may require the interaction with another

TMD to stably integrate into the membrane bilayer. Taken together, these findings support the assumption that the onset of co-co assembly is coordinated by the exposed TMD.

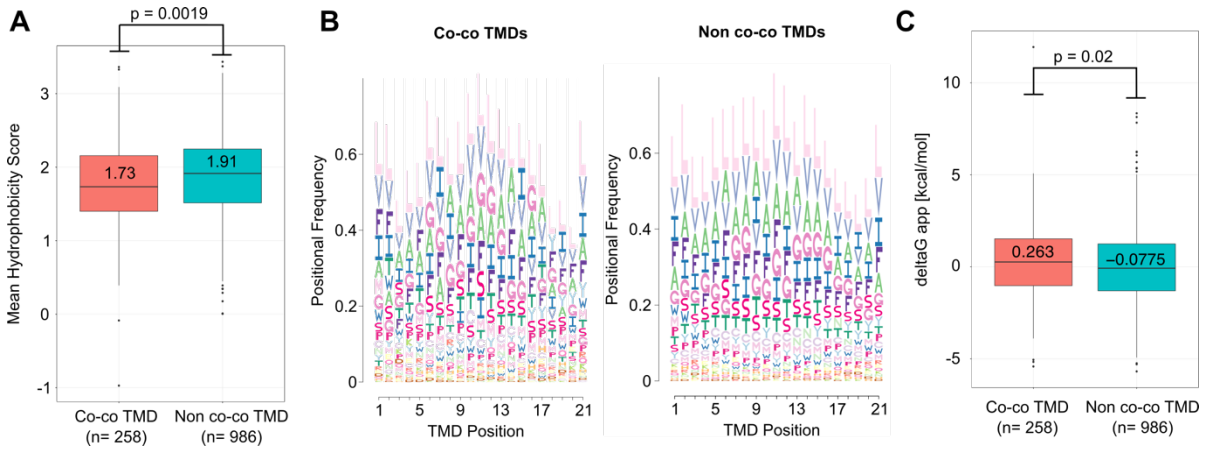

Fig. 23: Co-co assembling TMDs may require the interaction with another TMD to stably integrate into the membrane bilayer. (A) Mean Kyte-Doolittle hydrophobicity plot for "co-co TMDs" and "non-co-co TMDs". **(B)** Average amino acids frequency plot for each position along a 21 long TMD from both TMD class. **(C)** Apparent free energy difference ΔG_{app} calculations for each TMD domain of the "co-co TMD" and "non-co-co TMD" class. The statistical significance in (A) and (C) was calculated with a Welch's t-test (see Material & Methods for further explanations).

Another analysis of membrane proteins was focused on candidates that were classified as "exclusive PK resistant", meaning that the sigmoidal footprint shift did not significantly disappear after PK treatment but after puromycin treatment. This group enclosed many short proteins (around 300 amino acids) with multi-spanning transmembrane domains, indicating that membrane association may render disomes resistant against PK treatment (**Fig. 24**).

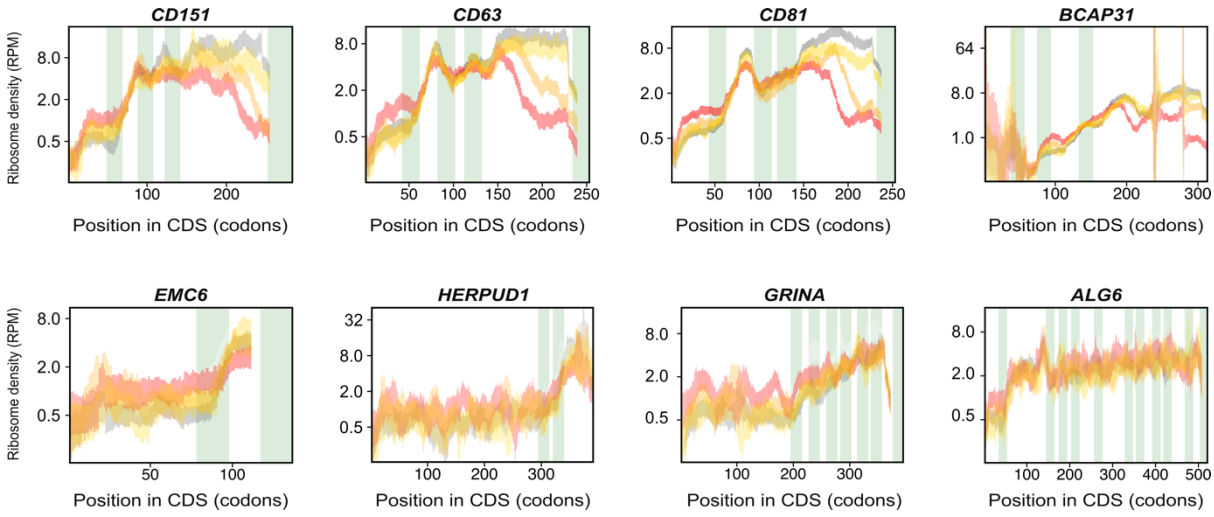

Fig. 24: Many multi-spanning membrane proteins are classified as PK resistant. Single gene ribosome density profiles for multiple PK resistant classified candidates. Green boxes indicate the exposure of a transmembrane domain. All boxes are shifted by 30 codons to account for the length of

the ribosome exit tunnel. This shift leads to the sudden end of the shown single gene profiles before the end of the plot.

4.3.4 Is co-co assembly coordinated by ribosome stalling?

By isolating and analyzing 60 nt ribosome protected footprints, several recent publications revealed that ribosome collisions are a widespread phenomenon in eukaryotes (Arpat *et al.* 2020, Han *et al.* 2020, Zhao *et al.* 2020). In addition, it was speculated that ribosomal stalling could be a mechanism to coordinate co-translational assembly (Panasenko *et al.* 2019). In order to explore if co-co assembly onset positions may coincide with frequent ribosome stalling, we analyzed 60 nt fragments from high salt HEK293-T disome fractions. We observed an accumulation of collided ribosomal pairs yielding 60 nt footprints before the onset of co-co assembly (**Fig. 25A**, left). The 60 nt stalling peak occurs approx. 35 nt before the onset of co-co assembly, which is caused by the alignment of the lagging ribosome of each ribosome pair, suggesting that the leading ribosomes is positioned at the co-co assembly onset (**Fig. 25A**, left). Total translatome metagenes of 30 nt data showed that this 60 nt footprint enrichment is not caused by an overall increased ribosome density around the onset of co-co assembly. It is surprising that the occurrence of 60 nt stalling peaks does not co-occur with a simultaneous increase of 30 nt stalling peaks, because we would expect that one ribosome has first to be stalled before a collided state is formed. This finding is in agreement with previous studies, which showed that translation slow-down detected by classical ribosome profiling (30 nt) and by disome profiling (60 nt) are not correlated with each other (Arpat *et al.* 2020, Han *et al.* 2020, Zhao *et al.* 2020). To proof the significance of the observed 60 nt stalling before assembly onset we randomly assigned co-co assembly onsets for all candidates and observed that the 60 nt footprint enrichment is lost (termed "scrambled onset", **Fig. 25A**, right).

The enrichment of stalled ribosomes before co-co onset agrees with our model for co-co assembly and with published data, that suggest coordination of translational pausing and co-translational assembly of specific complexes (Panasenko *et al.* 2019). For both, *cis* and *trans* assembly, pausing of the leading ribosome could cause ribosome collision and could facilitate co-translational assembly by bringing nascent subunits into close proximity or creating a time frame to allow assembly to occur (**Fig. 25B**).

Different mechanisms causing translation slow-down are feasible. Possibilities are secondary structures in the mRNA, as well as slowly translated codons or codon pairs as well as nascent chain segments that interact with the ribosomal tunnel to slow-down translation (Drummond and Wilke 2008, Pechmann and Frydman 2013, Kim et al. 2014, Sharma *et al.* 2019, Arpat et al. 2020). The pause of the leading ribosome could be resolved by the collision itself or by a pulling force on the nascent subunit, caused by the physical interaction of nascent chains or

co-translational folding (**Fig. 25B**, described in more detail in the Discussion and Outlook section). More detailed single gene candidate analysis of a deeper sequenced 60 nt DiSP data set can be found in the PhD thesis of Matilde Bertolini. In brief, our data suggest that lowly expressed genes indicate only a single 60 nt stalling peak before co-co assembly onset, whereas highly expressed genes show multiple 60 nt stalling peaks before assembly onset. Interestingly, after the onset of co-co assembly almost no 60 nt peaks are observed, indicating that no stalling events are happening further downstream along the mRNA.

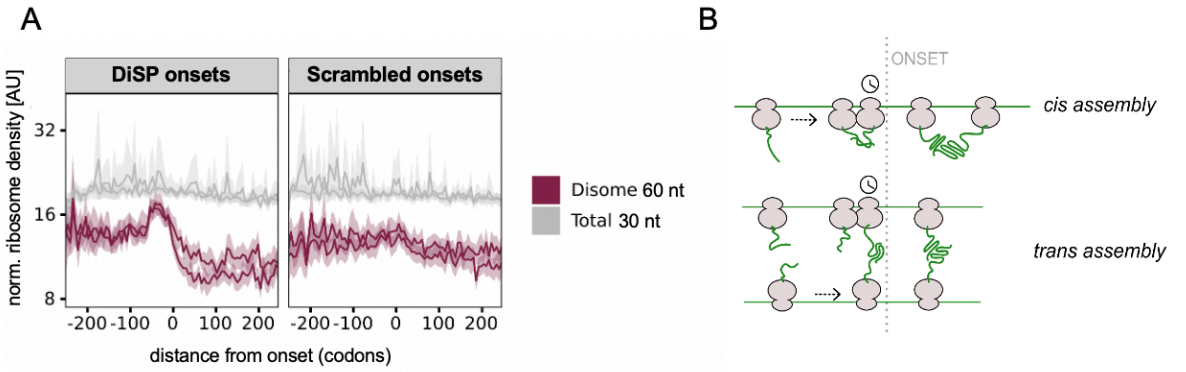

Fig. 25: Co-co assembly and ribosome stalling may be coordinated. (A) Metagene profiles of 60 nt disome reads (purple, indicating ribosome stalling/collision) and 30 nt total translatome reads (grey, indicating all translating single ribosomes) aligned to co-co assembly onset of the high and low confidence list. The enrichment of 60 nt reads before onset (left) is lost by assigning random onset positions (right). **(B)** Schematic depiction of possible modes that lead to ribosome collision (60 nt data were generated and sequenced by myself, cartoon and metagenes were made by Matilde Bertolini).

4.4 Attempt to employ SeRP experiments for measuring the distance of two co-co assembling ribosomes translating the same mRNA

Our findings that numerous homodimers utilize co-co assembly and in addition ribosome collision and co-co assembly seem to be coordinated hint together towards a *cis*-assembly mechanism between neighboring ribosomes on one mRNA. This led to the question how distant two nascent chain-connected translating ribosomes are on one mRNA. The first indirect approach to determine the distance was based on the footprint density loss towards the end of translation (**Fig. 26A**, top). The rationale was that the disomes convert into monosomes as soon as the leading ribosome terminates. The remaining ribosome will continue to translate but the footprints will end up in the monosome data set. *Cis*-assembly of two ribosomes in the closest possible way would lead to a loss of disomes 30 nt before the stop codon. We observed varying lengths of monosome shifts upstream of the stop codon for single candidates (**Fig. 8A**, *LMNB1*, *DCTN1*, *MAD1L1*) and a shift of about 65 nucleotides in a metagene profile including all homodimeric high confident candidates aligned to the stop codon (**Fig. 26A**, bottom). Ribosomes forming disomes are therefore not permanently in a

collided state after the onset of co-co assembly. This is in agreement with the 60 nt data, that showed a reduced occurrence of collided ribosomes after the onset of assembly.

To determine the distance between two co-co assembled ribosomes *in cis*, we chose to employ SeRP with two homodimers of our high confidence co-co assembly list, lamin C (*LMNA*) and dynactin subunit 1 (*DCTN1*). The rational of this experiment is based on the assumption that if the leading ribosome terminates, it will release its nascent chain, allowing us to affinity purify the nascent chain coupled ribosome via the freely accessible C-terminal tag. The enrichment plot (affinity purified ribosomes/ total translatome) should directly report on the distance between the last two ribosomes for these two homodimers (**Fig. 26B**). In principle SeRP could also detect the last step of co-co assembly *in trans* between ribosomes on two mRNAs, if the two ribosomes are not fully synchronized till the end of translation.

Using a CRISPR approach, we generated a heterozygote *LMNA-GFP11-TwinStrep* HEK293-T cell line as well as a homozygous *DCTN1-GFP11-TwinStrep* cell line (**Fig. 26C**, see Material & Methods). Initial pull-down experiments with Strep-Tactin coated magnetic beads showed that both tagged proteins were immunoprecipitated from each cell lysates (**Fig. 26D, E**). These two cell lines were separately used to perform SeRP. However, SeRP for C-terminally tagged dynactin subunit 1 or lamin C did not reveal a detectable AP/total translatome enrichment toward the end of translation (**Fig. 26F**). Such a flat AP/total translatome enrichment plot suggest that the short-lived last step of co-co assembly is not detectable by SeRP. One possible explanation is that the complete nascent chain release of the leading ribosome may not happen before the lagging ribosome also reaches the stop codon and terminates translation. Thereby no ribosome-nascent chain complex coupled to a fully synthesized partner subunit would exist, which was the target of the affinity purification.

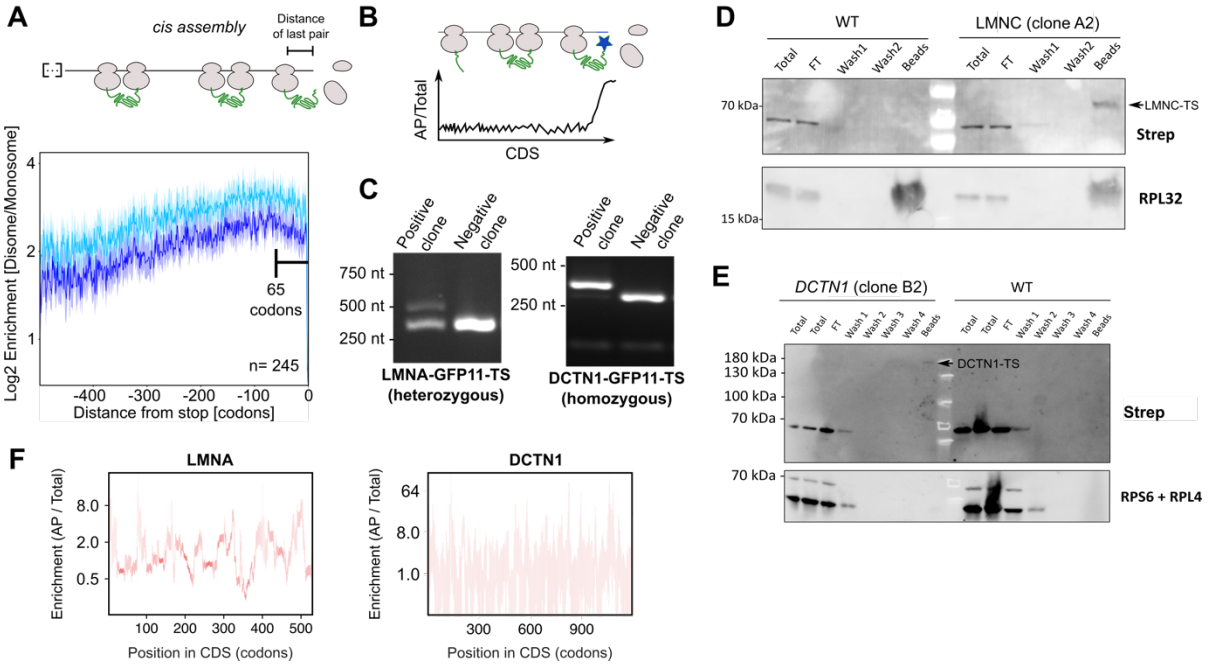

Fig. 26: A SeRP-based attempt to detect the distance between two ribosomes for *LMNA* or *DCTN1*. (A) Homodimers formed by co-co assembly *in cis* should lose the disome state after the leading ribosome reaches the stop codon and disassembles. Therefore, a disome footprint density loss towards the end of translation is expected. The distance to the stop codon should resemble the distance of two ribosomes engaged in co-co assembly on one mRNA (top). Metagene profile of all high confidence co-co assembling homodimers aligned to the stop codon (two replicates are shown). Highlighted is the distance between the stop codon and the first occurrence of a disome footprint density loss (bottom). **(B)** Schematic representation of expected result of the SeRP experiment with endogenous C-terminally tagged proteins. The enrichment of affinity purification (AP) to total translatome should indicate when one ribosome terminates and releases the fully synthesized, tagged peptide and allow the co-purification of the following ribosome. The distance between the enrichment onset and the stop codon should indicate the distance between both nascent chain coupled ribosomes. **(C)** PCR validation of single edited cells via CRISPR-Cas9. One heterozygote clone next to a negative clone for *LMNC-GFP11-TS* (left) is shown; as well as one homozygous clone next to a negative clone for *DCTN1-GFP11-TS* (right). **(D + E)** Western blots of fractions taken during SeRP. Both show that the AP with the MagStrep “type3” XT beads in the respective cell lysate isolated the expected full-length C-terminally tagged protein. **(F)** AP/total enrichment profiles indicate no enrichment of footprints towards the end of translation for *LMNC-GFP-TS* nor *DCTN1-GFP11-TS* (the plotted confidence intervals are so faint caused by the low amounts of reads in the affinity purification translatome, see Material & Methods).

4.5 Is it possible to recapitulate co-co assembly of human proteins in *E. coli*?

One important question to understand the mechanism of co-co assembly is whether the specific assembling properties of nascent chains suffice for disome formation. For single validation experiments, we planned to employ DiSP of *E. coli* cells over-expressing a human co-co candidate. Heterologous expression allows to determine if co-co assembly is an intrinsic feature of the nascent chains, outside the eukaryotic folding environment. It also allows studying the assembly of mutants of otherwise essential proteins.

The first human co-co assembly candidate for validation tests in *E. coli* was selected based on three criteria. The protein of interest had to (i) be an exclusive homodimer that does not require any other interacting factors for reaching the correctly folded and assembled native state, (ii) show a clear disome enrichment after exposure of a characterized dimerization domain and (iii) be included in the high confidence list. One protein class that passed all selection criteria were all four type V intermediate filaments in humans (lamin A, C, B1 and B2), which are the major constituents of the nuclear lamina. All lamins are high confidence co-co assembly candidates and structurally similar, consisting of three main domains: an unstructured N-terminal head domain, a central alpha-helical rod domain and a C-terminal and globular tail domain (**Fig. 27A**). Lamin homodimer formation in the cytosol is mediated by the 300 amino acid long rod domain, which forms a coiled coil connecting two monomers (Ruan *et al.* 2012, Makarov *et al.* 2019). In the nucleus, multiple lamin homodimers then polymerize to filaments by head-to-tail assembly, which form the nuclear lamina meshwork (**Fig. 27A**). Lamin B1 and B2 derive from separate genes, whereas alternative splicing of the *LMNA* gene produces two isoforms: lamin A and C that only differ in the C-terminal tail domain. Although both isoforms share exactly the same N-terminal dimerization domain, they form exclusively isoform-specific homodimers *in vivo* (Kolb *et al.* 2011). So far it was not understood how isoform specific lamins are formed *in vivo*. A yeast two-hybrid screen indicated that human lamin proteins can form homo- as well as heterodimers in different combinations, caused by the assembly of the highly conserved rod dimerization domain (Ye and Worman 1995). We speculated that co-co assembly could provide a simple mechanism to achieve isoform-specific homomer assemblies, by allowing two nascent chains to assemble in a zipper like mechanism during ongoing translation (**Fig. 27B**). Therefore, lamin C was chosen for the first reconstitution of co-co assembly in *E. coli* (Rosetta was chosen as *E. coli* strain, Table 7).

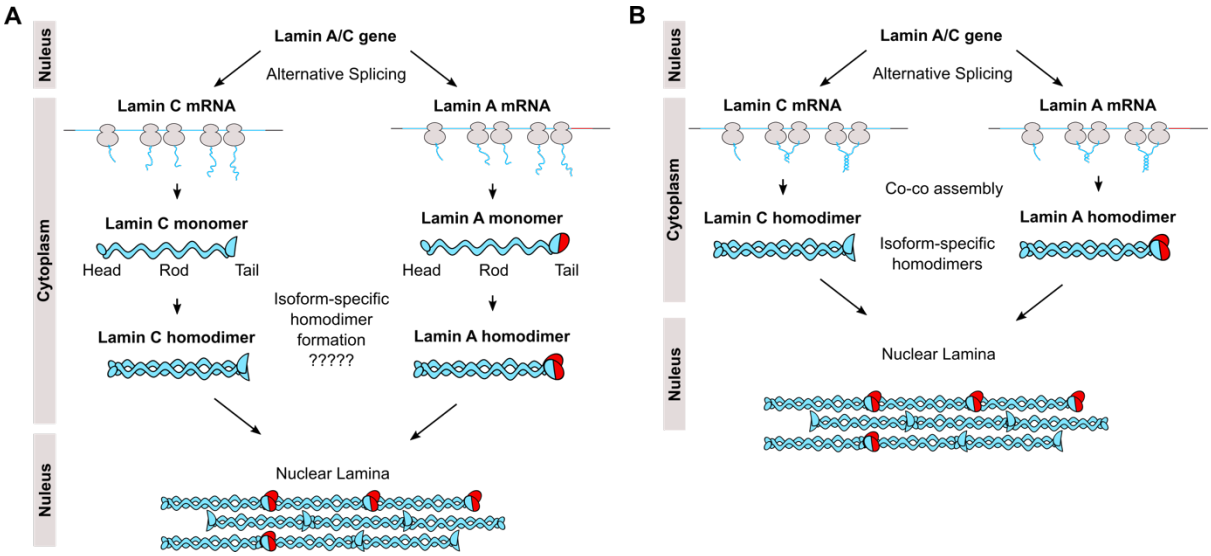

Fig. 27: Model of isoform specific lamin homodimer formation. (A) Current view of lamin assembly *in vivo*. The *LMNA* gene encodes two lamin mRNAs generated by alternative splicing. Both mRNAs are translated into lamin monomers, which form post-translational isoform specific homodimers via the rod domain in the cytosol (question marks indicate the unknown mechanisms to obtain isoform specific homomers). These lamin homodimers assemble head to tail into filaments inside the nucleus and generate the nuclear lamina by antiparallel filament layers. **(B)** Proposed model for isoform specific lamin formation via co-co assembly between neighboring ribosomes on each lamin mRNA. This simplified assembly mechanism for homodimer formation would make any further regulation obsolete, and allows the downstream nuclear lamina assembly.

Importantly, the formation of human lamin C dimers in *E. coli* was already described in combination with structural studies that showed functional human lamin polymers (Moir *et al.* 1991). We performed DiSP of *E. coli* synthesizing human lamin C upon IPTG induction from a plasmid. The lamin C overexpression generated an RNase resistant disome peak in sucrose gradients, suggesting that nascent lamin C can dimerize and connect two ribosomes (**Fig. 28A**).

The used expression plasmid introduced an N-terminal bipartite tetra-cysteine motif known as split-FlAsH tag, which forms a tetra-cysteine motif only when two split-FlAsH tags come in close proximity (<7 Å, Luedtke *et al.* 2007). This split-FlAsH tag system was previously used to study protein-protein interactions, by employing the chemical compound FlAsH-EDT2 (4,5-bis(1,3,2-dithiarsolan-2-yl fluorescein)), termed from here on FlAsH, which has a high affinity for tetra-cysteine motifs and becomes fluorescent upon binding (**Fig. 28A**). The detection of protein dimerization with FlAsH is known as bipartite tetra cysteine display (Luedtke *et al.* 2007). We postulated that the formation of a native lamin coiled coil homodimer will bring cysteine pairs in close enough proximity to allow FlAsH to bind and become fluorescent. The used lamin construct does not include the unstructured 90 nt long head domain sequence, to ensure close proximity of both split-FlAsH tags in the final dimer.

In agreement with our assumption, we detected FlAsH fluorescence in the disome peak suggesting lamin co-co assembly (**Fig. 28B**). Surprisingly, a clear fluorescent peak was also detectable in the monosome fraction (**Fig. 28B**).

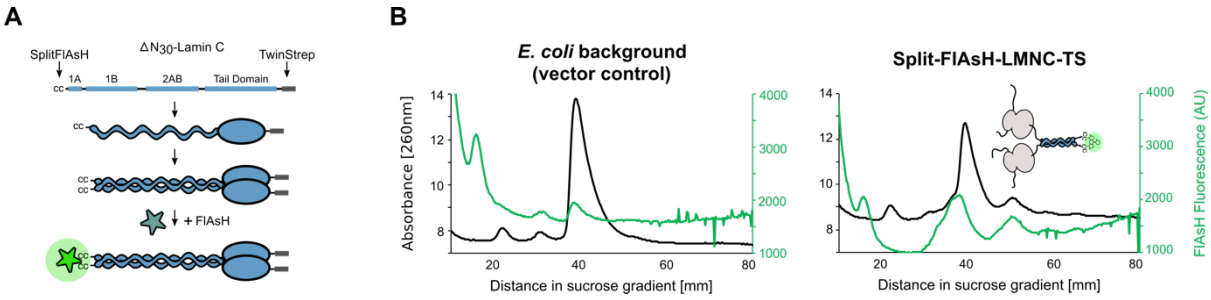

Fig. 28: Disome formation is caused by expression of human lamin C in *E. coli*. (A) Structural features of the employed lamin C construct. Lamin dimerization allows FlAsH to bind to two split-FlAsH tags in close proximity, which generates a fluorescent signal. **(B)** Sucrose gradients of *E. coli* lysates from cells transformed with empty control vector (left) or a plasmid encoding for human lamin C, with a N-terminal split-FlAsH tag and a C-terminal TwinStrep tag lacking the 30 amino acid unstructured N-terminal head domain (right). Black lines indicate the UV trace at 260 nm, green lines the measured FlAsH fluorescence trace (emission: 474 nm, excitation: 525 nm).

Control experiments performed in our lab by Jiri Koubek and Niklas Stockburger revealed that not only a tetra-cysteine motif is able to bind and activate FlAsH fluorescence. The presence of DTT in the lysis buffer or sucrose gradient buffer is also able to activate FlAsH fluorescence in the absence of split-FlAsH tag dimerization, implying that the presence of a single split-FlAsH tag and DTT is enough to generate a clear fluorescence signal. The structure of DTT could explain this background fluorescence (**Fig. 29A**). In the absence of DTT, a significant decrease of FlAsH fluorescence for the monosome as well as disome fraction was detectable (**Fig. 29B**).

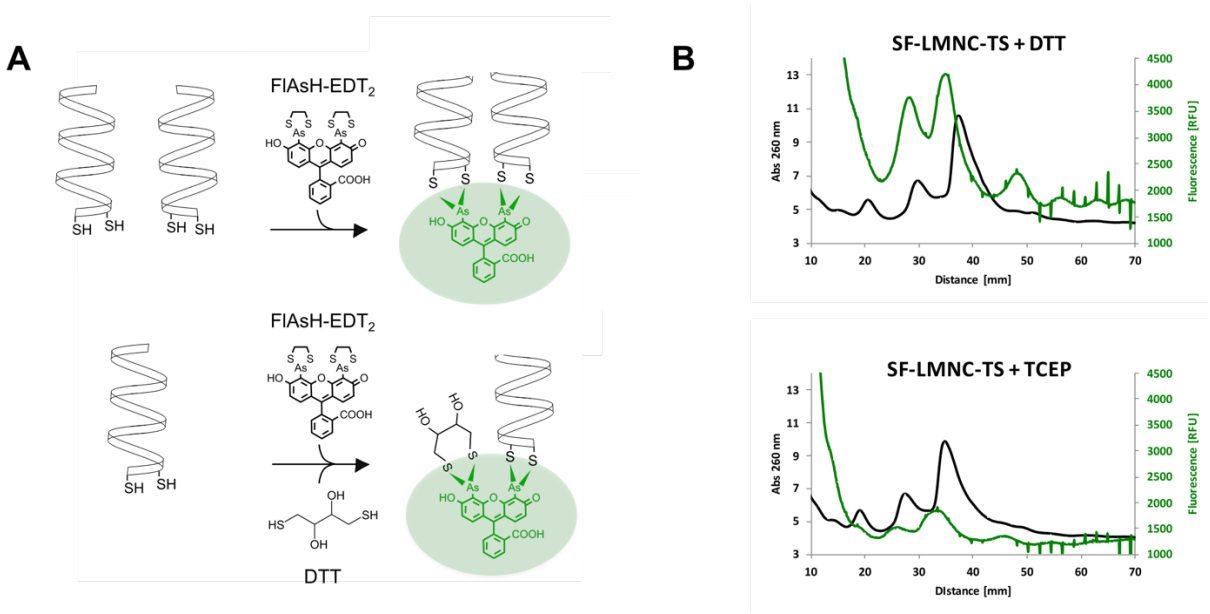

Fig. 29: FlAsH background fluorescence activated by DTT. (A) Principle of FlAsH-EDT2 fluorescence activation in the presence of two split FlAsH-tags, depicted as two SH groups in an alpha helix (top, Luedtke *et al.* 2007). Model of FlAsH-EDT2 fluorescence activation in the presence a single split FlAsH-tag and DTT (bottom). **(B)** Sucrose gradients of lamin C overexpressing *E. coli* lysed in the presence of DTT (left, standard lysis buffer) or TCEP (right).

The monosome and disome isolation was repeated in the absence of FlAsH to ensure that the presence of the dye does not change the assembly behavior of lamin C. DiSP revealed that the observed disome peak encloses ribosomes translating *LMNA* (**Fig. 30A**). About 60% of all ribosomes were translating the plasmid encoded lamin C after the IPTG induction (**Fig. 30A**). The single gene profile of lamin C revealed that the minimal length of nascent lamin C mediating the disome shift in *E. coli* was very similar to that of the endogenously expressed lamin C in mammalian cells (**Fig. 30B**). The conserved onset in a fundamentally different folding environment highlights the robustness of co-co assembly and reveals that co-co assembly is not strictly dependent on eukaryotic specific assembly factors, translation regulation or mRNA subcellular localization. Additionally, this DiSP experiment in another host revealed the specificity of the detected footprint density shift for lamin C in humans, because lamin co-co assembled in the absence of any other human nuclear envelop or intermediate filament protein. This result further indicates that the footprint shift for lamin C in humans is caused by specific homomer formation.

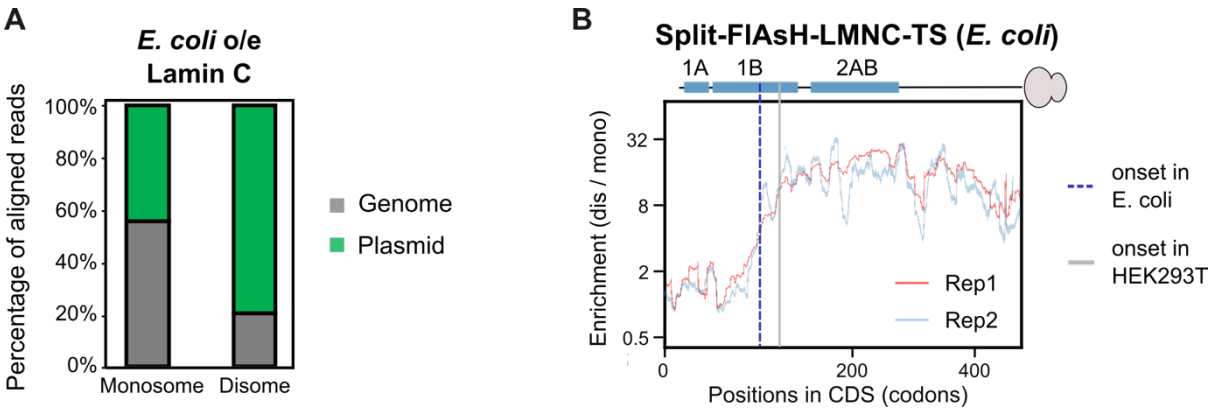

Fig. 30: Disome footprint density shift is detectable for human lamin C in *E. coli*. (A) Bar plot showing percentages of aligned reads to the *E. coli* genome or plasmid encoded lamin sequence from the isolated monosome and disome fraction. **(B)** Disome over monosome enrichment profile of plasmid-encoded lamin C expressed in *E. coli.* The ribosome-exposed coiled coil interfaces are indicated by blue bars on top. Two vertical dotted lines indicate the determined onset in *E. coli* or human HEK293-T DiSP data.

The absence of an observable disome peak in *E. coli* wild type cells (Rosetta strain, Table 7) suggests that *E. coli* does not employ co-co assembly (**Fig. 28A**). However, we performed DiSP also with wild type cells and detected over 600 different 2-fold disome enriched candidates (**Fig. 31A**). Important to note is that no *E. coli* protein was specifically enriched in the disome fraction of lamin over-expressing cells compared to the wild type background strain (**Fig. 31B**). This indicates that the artificial presence of lamin C does not shift specifically *E. coli* proteins to the disome fraction.

The investigation of co-co assembly in *E. coli* is now continued by Jaro Schmitt (PhD student in the Bukau lab). In brief, we observed comparable results to the human DiSP data set and identified similar monosome to disome footprint distributions in metagene profiles as well as for single cytosolic co-co assembly candidates (**Fig. 31C, D**). As in humans, we also detected TMD containing inner membrane proteins, suggesting that co-co assembly may be a major route for IMP folding and assembly (**Fig. 31E**). The high abundance of membrane proteins is not in agreement with findings from Josef Auburger, a previous PhD student in the Bukau lab, who performed the first DiSP experiments in *E. coli*. The major difference is the presence of detergents in the lysis buffer of this study (see Material & Methods). This means that, by omitting detergents, all membrane associated ribosomes were presumably lost in the lysate clearance step or during sucrose gradient fractionation of the previous DiSP experiments.

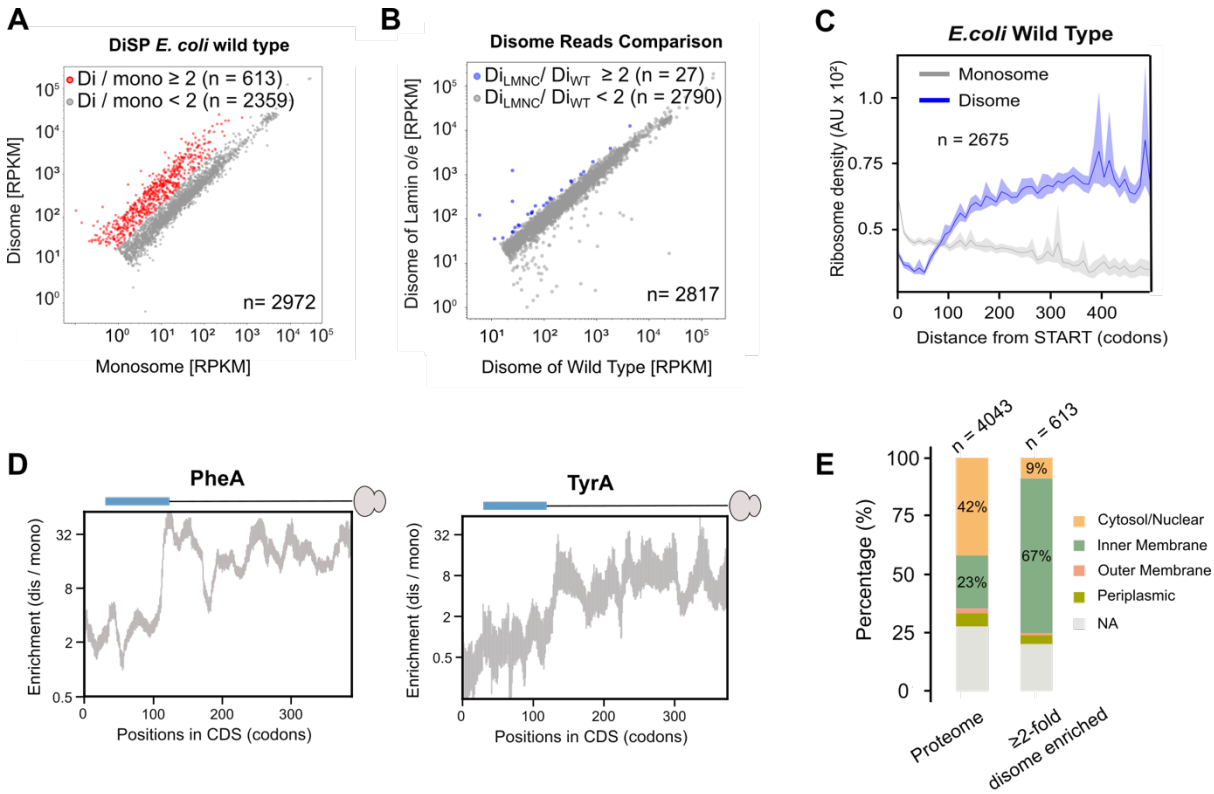

Fig. 31: Disome footprint density shifts are also detectable in the *E. coli* proteome. (A) Comparison of disome and monosome footprint densities of all detected genes in *E. coli* wild type cells. Genes with a more than 2-fold disome over monosome enrichment are highlighted in red. **(B)** Comparison of all disome reads from the *E. coli* wild type cells and all disome reads from the *E. coli* cells overexpressing human lamin C. Genes with a more than 2-fold enrichment in the lamin expressing cells are highlighted in blue. **(C)** Average footprint density along the coding sequence of all detected genes (metagene) aligned to translation start in *E. coli* wild type cells. **(D)** Single gene enrichment plot examples of two prominent disome enriched homodimers of the *E. coli* proteome. The blue box indicates the exposure of the annotated chorismate mutase domain in both proteins, which is known to facilitate homodimerization for PheA. **(E)** Stacked bar plot showing the percentage of annotated localizations of the *E. coli* proteome and the 2-fold disome enriched candidates.

To prove that the observed co-co assembly properties are not lamin specific, we performed DiSP of *E. coli* overexpressing another human homodimer. Dynactin subunit 1 (isoform p150^{glued} encoded by *DCTN1*) was chosen, which also dimerizes via a coiled coil and is included in our high confidence list. Similar to the expression of *LMNA*, the plasmid-based expression of*DCTN1* (including a N-terminal split-FlAsH tag and C-terminal TwinStrep tag) generated a disome peak that was enriched with ribosomes translating *DCTN1* (**Fig. 32A-C**). The monosome and disome isolation was performed in the absence of FlAsH to ensure that the presence of the dye does not change the assembly behavior. The disome over monosome enrichment profile of *DCTN1* revealed that the assembly onset was similar to one detected in human cells and involved the exposure of the coiled coil domain (**Fig. 32D**). In the presence of FlAsH, we again observed a fluorescence signal in the monosome as well as disome fraction, which was caused by the presence of DTT in the lysis buffer (**Fig. 32A**). Together with the previous observation, this data suggested that the experimental set up using FlAsH

and DTT in the lysis buffer created an unspecific FlAsH signal that is not caused by binding of a split-FlAsH, but by binding a single cysteine pair and DTT.

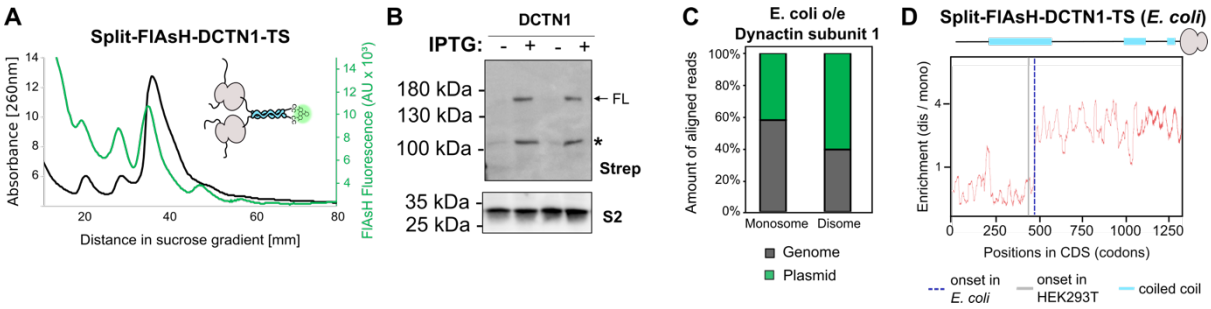

Fig. 32: Human dynactin subunit 1 can co-co assemble in *E. coli*. **(A)** Sucrose gradients of *E. coli* cells transformed with a plasmid encoding human *DCTN1* with a N-terminal splitFlAsH tag and C-terminal TwinStrep tag. Black lines indicate the UV trace at 260 nm, green light the measured FlAsH fluorescence trace (emission: 474 nm, excitation: 525 nm). **(B)** Western blot showing the overexpression of *DCTN1*. The anti-Strep antibody detects the C-terminal TwinStrep tag of dynactin, the arrow indicates the full-length protein, and the asterisk indicates presumably an N-terminal degradation product. The anti-S2 antibody detects the small ribosomal subunit S2 protein, here used as loading control. **(C)** Percentage of aligned reads to the *E. coli* genome vs plasmid encoded *DCTN1* from the isolated monosome and disome fraction. **(D)** Disome over monosome enrichment profile of plasmid-encoded *DCTN1* construct shown in (A) expressed in *E. coli*. The ribosome-exposed coiled coil interfaces which are involved in the dimerization are indicated by blue bars on top. The vertical lines indicate the onset determined in humans (grey solid line) or in *E. coli* (blue dotted line).

Our DiSP sequencing results revealed that more than 60% of all ribosomes translate the plasmid encoded lamin sequence, indicating the presence of large amounts of full-length as well as nascent lamin proteins inside the cytosol of *E. coli* (**Fig. 30A**). This high amount of interaction partners could increase the probability of co-post assembly employing orphan and interaction competent lamin monomers. To test this possibility, we performed SeRP by pulling C-terminally TwinStrep tagged lamin C from *E. coli* lysates (**Fig. 33A**). In principle, this experiment could also reveal the distance of the last two translating ribosomes (**Fig. 33A** ,as explained in **4.4**). The conducted SeRP experiment indicated that the pull-down of lamin C from the lysate leads to a decrease of the disome as well as monosome fluorescence observed in sucrose gradients (**Fig. 33B, C**). This result suggests that full-length lamin could interact with nascent chains of monosomes as well as disomes. Western blot analysis revealed that the affinity purification of full-length lamin co-purified ribosomal proteins (co-AP), hinting towards the interaction of a full-length lamin with nascent lamin (**Fig. 33D**). The sequencing results of SeRP employing a TwinStrep tag pull-down demonstrated no clearly enriched *E. coli* candidates, indicating that fully synthesized lamin C does not unspecifically interact with distinct nascent proteins of the *E. coli* proteome *in vivo* or *post*-lysis (**Fig. 33E**). However, the single gene enrichment profile of lamin C indicated only a very weak C-terminal AP/Total translatome enrichment, suggesting that co-post assembly may be possible in this

artificial over-expression situation, but is not a major path of lamin assembly (**Fig. 33F**, left). The last weak peak before the stop-codon, indicates a distance of approximately 50 codons, which could report on the distance between the last two coupled ribosomes (**Fig. 33F**, left). This observation is in agreement with the previous determined distance based on a metagene profile of all homodimers of the high confidence list (see **Fig. 26A**).

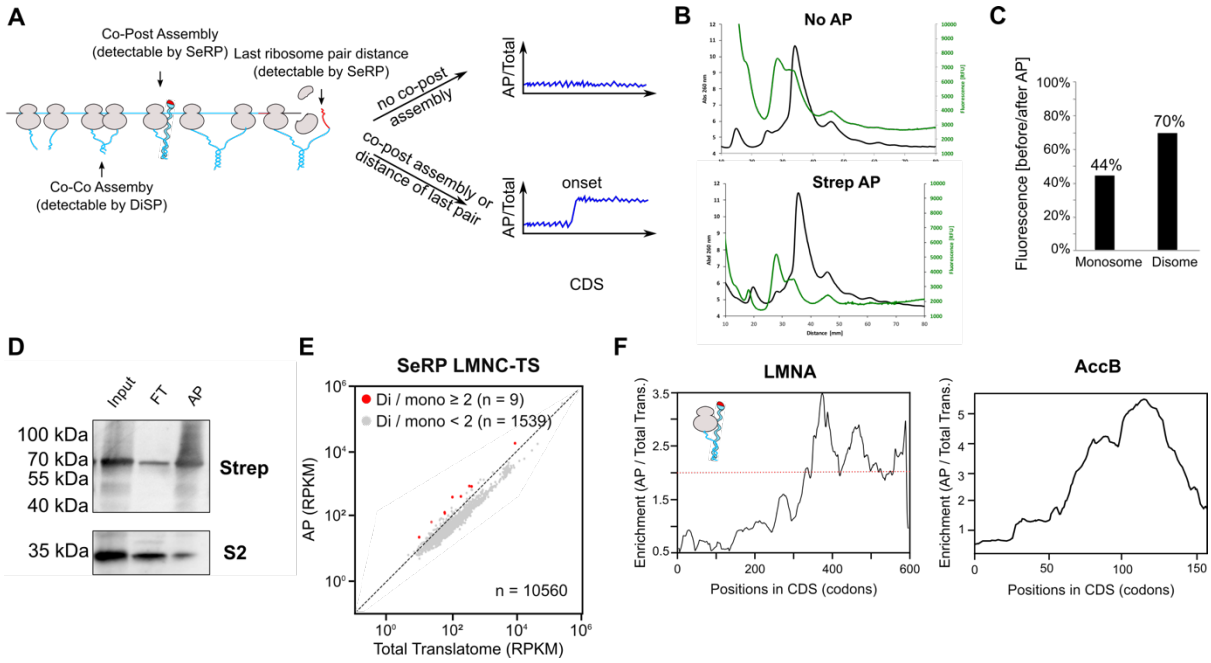

Fig. 33: SeRP of human lamin C in *E. coli* suggest that co-post assembly of lamin C is not frequently occurring. (A) Model of co-post assembly that is detectable by SeRP and the expected lamin enrichment profiles. **(B)** Sucrose gradients of lysates from *E. coli* cells overexpressing pET3a-FLAG-SF-LMNC-TS before lamin C affinity purification (top) or after affinity purification (bottom). Black lines indicate the UV trace at 260 nm, green lines the measured FlAsH fluorescence trace (emission: 474 nm, excitation: 525 nm). **(C)** Quantification of the fluorescence signal before and after affinity purification observed in (B). **(D)** Immunostaining before and after affinity purification of C-terminally tagged lamin C from *E. coli* lysates. The anti-Strep antibody detects full-length lamin C, the anti-S2 antibody detects the small ribosomal subunit S2 protein, here used to indicate the co-purification of ribosomes with the lamin pull-down. **(E)** Comparison of affinity purified (AP) and total translatome footprint densities of all detected genes in *E. coli*. **(F)** AP over total enrichment profiles of lamin C (left) and the strongest enriched *E. coli* candidate AccB (right).

The strongest enriched *E. coli* candidate in the lamin C pull-down was an acetyl-CoA:carbon-dioxide ligase (*AccB*) (**Fig. 33F**, right). This protein is active in a dimeric state, by C-terminal domain dimerization, but its biotinylation prevents dimerization and keeps the protein in a monomeric state (Chapman-Smith *et al.* 1997). The co-AP of this protein in the Strep-Tag affinity purification presumably indicates that this protein is co-translationally biotinylated and does not reports on the interaction with TwinStrep tagged lamin C.

4.5.1 Identification of the minimal structural requirements of co-co assembly

The lamin rod domain can be subdivided in three coiled coil segments, called 1A, 1B, and 2AB, which are interconnected by linkers (**Fig. 28A**). The onset of co-co assembly for lamin C occurred always within coil 1B, in human and *E. coli* cells (**Fig. 30B**). This shows that 1A and parts of 1B suffice to initiate assembly. We aimed to explore what parts of the coiled coil domain are minimally required for co-co assembly and whether coil 1A, as most N-terminal segment, is crucial for assembly initiation.

We first tested individually the dimerization capabilities of all three coiled coil segments of lamin C. The employed dimerization assay is based on the fusion of a protein of interest to the N-terminal DNA binding domain (λN) of the phage lambda repressor domain (λcI). This transcription repressor can only bind to the specific DNA operator sequence in the dimeric state. An *E. coli* strain encoding *lacZ* under control of the λ promoter was used to identify the dimerization propensity of λN fusion proteins by a simple beta-galactosidase assay (Hu *et al.* 1990, see Material & Methods, Table 7). The dimerization propensity of all tested lamin C rod segments (coil 1A, coil 1B, coil 1A-1B, coil 2AB and the unstructured head-domain with coil 1A-1B-2AB) repressed *lacZ* expression to the same extent as the wild type lambda repressor (λcI), indicating that each coil segment can homodimerize (**Fig. 34A**). The monomeric λN fraction alone served as negative control. This experiment revealed that all coiled coil segments mediate homodimerization, but did not report if the dimerization occurs co-translationally.

We started to N-terminally truncate lamin C by deleting the 40 amino acid long coil 1A and performed DiSP. This N-terminal deletion indicated a clear disome formation in sucrose gradients (**Fig. 34B**) and DiSP revealed the expected footprint density shift (**Fig. 34C**). The determined co-co assembly onset of the lamin 1A deletion was shifted by 42 codons downstream in comparison to the 40 amino acid longer full-length lamin C (**Fig. 34C**). This result proves that co-co assembly of lamin C is not strictly dependent on the most N-terminally located coil 1A domain, but initiates after enough of the coiled coil is exposed to mediate stable dimerization. These findings agree with a model of a zipper-like co-co assembly mechanism for coiled coil domains.

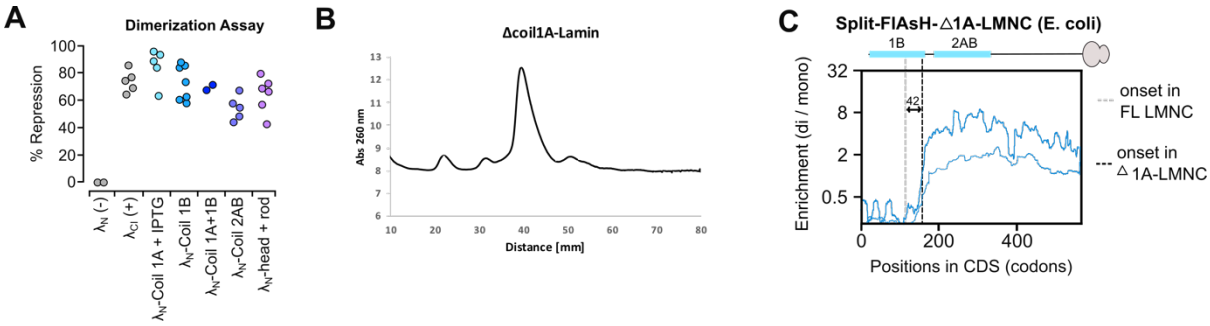

Fig. 34: Lamin C coil 1A is not required for co-co assembly. (A) Dimerization assay with different λ_N-lamin C rod sub-domain fusion constructs. Dimerization propensity of each tested construct was measured based on the repression of beta-galactosidase expression in *E. coli* (assay is based on Hu *et al.* 1990, Material & Methods). The isolated lambda-N domain alone (λ_N), cannot dimerize, and was used as negative control, whereas the dimeric full-length lambda repressor (λ_{CI}) was used as positive control. The λ_N-coil 1A construct required 20 μM IPTG (Material & Methods) **(B)** Sucrose gradients of *E. coli* cells transformed with a plasmid encoding human Δcoil1A-lamin C, with a N-terminal split-FlAsH tag and C-terminal TwinStrep tag. The black line indicates the UV trace at 260 nm. **(C)** Disome over monosome enrichment profile of plasmid-encoded Δcoil1A lamin C construct expressed in *E. coli* (two replicates are shown). The ribosome-exposed coiled coil interfaces are indicated by blue bars on top. Onset difference between full-length (FL) lamin and the coil1A deletion is indicated with dotted lines.

We generated a plasmid that encoded the isolated coil 1B sequence upstream of the full-length mCherry gene with a C-terminal TwinStrep tag, to first test if coil 1B is able to facilitate co-co assembly. In addition, we tested if the equally long coil 2AB segment of lamin C can also mediate co-co assembly by exchange of coil 1B with the coil 2AB sequence.

Simultaneously we focused on the importance of coiled coil quaternary structure formation during co-co assembly by generating a mutated coil 1B construct. The coiled coil amino acids sequence is formed by multiple heptameric repeats ((***abcdefg***)$_n$), in which hydrophobic residues are preferentially found in position ***a*** and ***d***, which are critical for the coiled coil quaternary structure formation. The swop of the ***a*** and ***e*** positions within all heptameric repeats of coil 1B was predicted to retain the helical structure, the overall amino acid composition and the hydrophobicity of the nascent chain, but eliminate the proficiency of coil 1B to form a coiled coil (**Fig. 36A**).

The coil 1B wild type construct showed a clear disome fluorescence, whereas the mutated coil 1B* construct as well as the coil 2AB construct showed almost no disome fluorescence but increased monosome fluorescence (**Fig. 35A, B**). The increase of monosome fluorescence agrees with the assumption that more split-FlAsH tagged nascent chains are present in the monosome fraction that, by reacting with DTT, display background fluorescence (see result section above).

Western blot detection of lysates from *E. coli* over-expressing one of the three constructs indicated equal steady state levels after IPTG induction (**Fig. 35C, D**). All three constructs showed additional, smaller protein bands detected by the Strep-tag antibody, indicating N-terminal partial degradation. Importantly, a dimeric state that was resistant to SDS-PAGE

sample preparation was only detected for the coil 1B construct and the full-length lamin C construct (**Fig. 35C**), in agreement with the prominent disome peak in the sucrose gradient profiles (**Fig. 35A**).

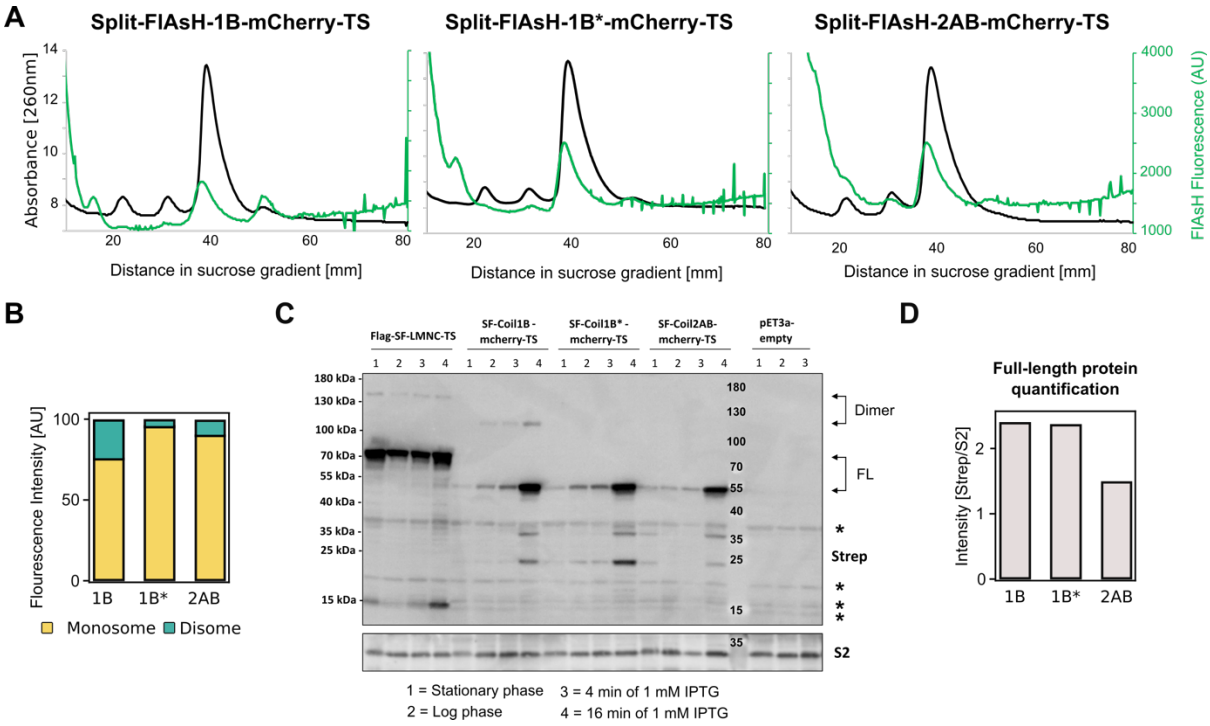

Fig. 35: Coil 1B forms SDS-PAGE resistant dimers. (A) Sucrose gradients of *E. coli* cells expressing coil1B (left), coil1B* (middle) or coil2AB (right) encoded on plasmids. Black lines indicate the UV trace at 260 nm, green light the FlAsH fluorescence (emission: 474 nm, excitation: 525 nm). **(B)** Quantification of the fluorescence signal of the monosome and disome peaks shown in (A). **(C)** Western blot detection of full-length lamin C or coil1B-, coil1B*- or coil2AB-constructs expressed in *E. coli*. Dimeric and full-length (FL) bands highlighted by arrows and asterisks indicate background signal that was also detected in control lysates. The ribosomal protein S2 was used as loading control. **(D)** Quantification of full-length constructs normalized to the ribosomal S2 protein.

The DiSP sequencing results revealed the expected onset of co-co assembly for the coil 1B construct (**Fig. 36A**), whereas no detectable monosome to disome footprint density shift was detectable for the mutated coil 1B* construct (**Fig. 36A**). These DiSP data verify that nascent coil 1B, N-terminally fused to mCherry, efficiently mediates co-co assembly, whereas the coil 1B* is incapable of forming co-co assembly dimers and does not cause co-translational disome formation in *E. coli*. These results support that DiSP detects *in vivo* interactions between nascent chains, which drive protein oligomer formation. Additionally, this set of experiment provides evidence that co-co assembly is not caused by hydrophobic interactions of nascent chains, since the physical properties are almost identical for all tested constructs.

The sequencing results for the coil 2AB construct revealed a detectable onset of assembly after the translation of approximately 100 codons, similar to coil 1B, yet the disome enrichment

was reduced (**Fig. 36B**). The reduced enrichment agrees with the reduced disome peak height obtained in the sucrose gradient (**Fig. 35A**). This difference of co-co assembly between the equally long 1B and 2AB coiled coils may be explained by the reduced stability of 2AB. The coil 2AB segment can be further subdivided in coil 2A and 2B (Ruan *et al.* 2012), which is indicated as discontinuous coiled coil by *in silico* predictions (**Fig. 36B**).

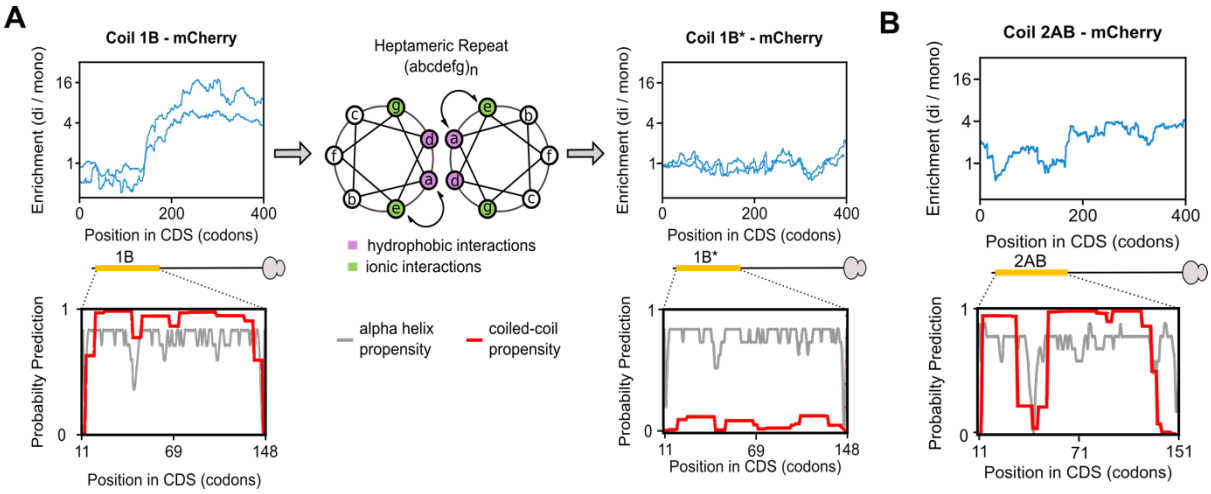

Fig. 36: Lamin co-co assembly does not rely on eukaryote-specific factors and depends on intrinsic nascent chain features. (A) Disome over monosome enrichment profiles of *LMNA* encoding lamin coil 1B (left) or the a/e swapped version of coil 1B* (right) fused N-terminally to mCherry and expressed in *E. coli* (n = 2). The ribosome-exposed coiled coil interfaces are indicated by yellow bars. Helical wheel projection shows residue arrangements (a-g) of the heptameric repeat (middle). Coiled coil (red) and alpha-helical (grey) probability predictions are shown for both wild type and mutant 1B (insets). **(B)** Same as in (A) for the coil 2AB construct.

4.6 Does co-co assembly ensure specific homodimerization of protein isoforms with identical oligomerization domains?

The bacterial experiments have not only proven that co-co assembly can happen outside the eukaryotic folding environment, but have also demonstrated the specificity of co-co assembly for the lamin C isoform alone. Lamin A and C are two isoforms, which originate from the same gene but are translated from two alternatively spliced transcripts. Both isoforms exclusively form homodimers *in vivo* and do not form heterodimers, even though they share exactly the same N-terminal dimerization domain (Kolb *et al.* 2011). We speculated that co-co assembly could provide a simple mechanism to achieve isoform-specific homomer assemblies (**Fig. 27B**). In principle, two possible modes of co-co assembly could lead to isoform specific lamin homodimer formation *in vivo*. If co-co assembly happens *in trans* it would require co-localized mRNAs of the same kind, presumably mediated by mRNA localization motifs in the UTRs, which differ for both isoforms. However, if co-co assembly happens *in cis* on one mRNA no further regulation would be necessary, in agreement with our findings in bacteria that co-

co assembly is driven only by the close vicinity of nascent N-terminal dimerization interfaces (**Fig. 37A**).

To explore this hypothesis in humans, we employed a heterozygous HEK293-T cell line that encodes for a C-terminally TwinStrep-tagged lamin C previously used for SeRP (see result section **4.4**). The tagged and untagged lamin C alleles in our heterozygous human cell line enclose the same UTRs. We surmised that in these cells, *trans* assembly should generate mixed oligomers enclosing tagged and untagged lamin C, while *cis* assembly would exclusively allow isoform specific homomer formation. The affinity purification of the tagged lamin C protein never co-purified untagged lamin C, which strongly supports the *cis* assembly model (**Fig. 37B**). We conclude that co-co assembly *in cis* facilitates isoform-specific lamin dimerization in human cells.

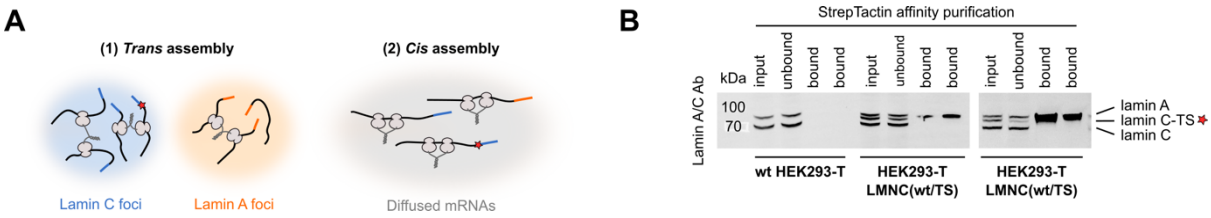

Fig. 37: Co-co assembly *in cis* ensures specific lamin homodimerization in human cells. (A) Schematic representation of the two possible assembly modes of lamins (*trans* vs *cis* assembly). The red star represents the integrated TwinStrep tag, and the blue/yellow bars the lamin A or C specific 3' UTRs. **(B)** Western Blot detection of TwinStrep tag affinity purifications from wild type or heterozygous *LMNC*(wt/TS) HEK293-T cells. Two technical replicates of the lamin C pull-down in HEK293-T LMNC(wt/TS) cells are shown (the here depicted lamin pull-down experiments were performed by Matilde Bertolini).

4.7 Can DiSP be modified to estimate the number of heteromeric co-co assembly candidates?

One limitation of DiSP is the missing information on the identity of the two nascent chain coupled footprints, which is lost upon RNase digestion. In other words, DiSP cannot reveal which footprints are protected by the same pair of nascent chain coupled ribosomes and, thus, cannot directly reveal whether the disome enrichment is due to formation of one exclusive homodimer or additional heterodimers.

Our DiSP approach identified subunits of heteromeric complexes enclosed in both, the low and high confidence lists, indicating that co-co assembly may be employed for heteromer formation. The identified heteromers included candidates that were already described to presumably employ co-co assembly to form heterodimers, like for TAF6 and TAF9, or for subunits of the proteasome 19S regulators particle (**Fig. 38 A-D**, see Discussion chapter **5.1**).

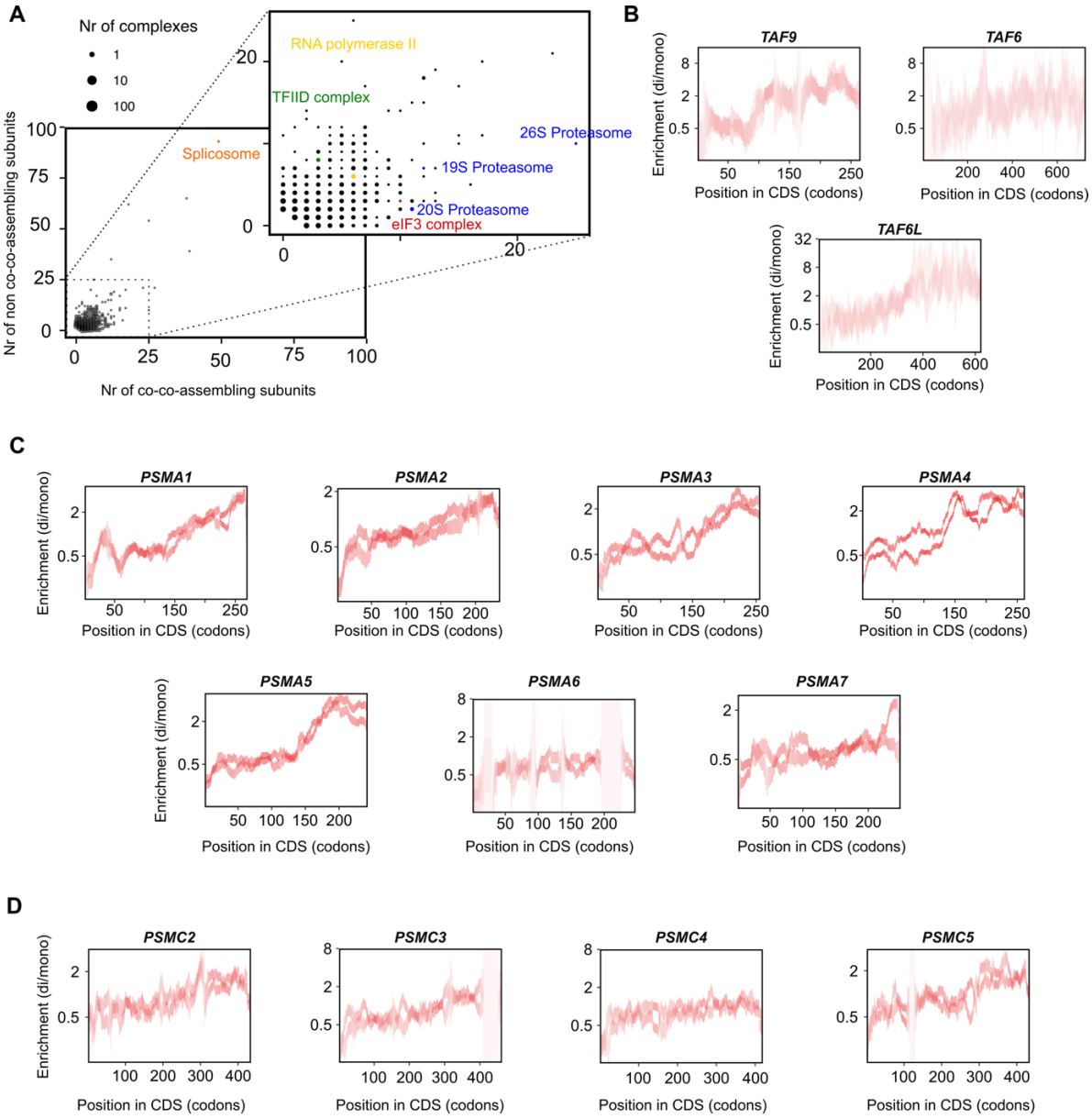

Fig. 38: Potential co-co heteromeric candidates. (A) Scatter plot of subunits with or without a disome shift from all heteromeric complexes annotated in the CORUM database. Specific single complexes are highlighted, e.g., the spliceosome (35 of 142 subunits co-co assembled), RNA polymerase II (6 of 17 subunits co-co assembled), TFIID complex (3 of 12 subunits co-co assembled), 26S proteasome (21 of 35 subunits co-co assembled), and eiF3 complex (8 of 13 subunits co-co assembled). **(B-D)** Single gene disome over monosome enrichment plots for different heteromeric candidates. **(B)** Depicted are the three TFIID subunits that are classified as co-co candidates. TAF6 and TAF9 were described to co-co assemble (Kamenova *et al.* 2019). **(C)** Depicted are all alpha subunits that form a heptameric ring of the proteasome. All are classified to have a sigmoidal shift except of PSMA6, but only PSMA4 and PSMA5 are classified as high confidence co-co assembly candidates. **(D)** Depicted are some subunits of the 19S proteasomal subunit, PSMC2/3/5 show a sigmoidal enrichment, but only PSMC5 is classified as high confidence candidates. PSMC4 shows no disome enrichment.

Aiming for a method that can reveal the identity of interacting nascent chains, we performed DiSP experiments that implemented an RNA ligation step to re-connect two ribosome footprints of one disome pair. A ligation of two 30 nt footprints would generate a chimeric RNA

fragment of around 60 nt. Sequencing this chimer footprint would directly reveal the existence of heteromeric dimers that form via co-co assembly. In addition, 60 nt re-ligated footprints enclosing fragments of identical transcript could reveal the distance of ribosomes on one mRNA, and if the sequences are partially identical, prove the existence of *in trans* assembly for homomer assembly.

To fuse footprints, monosomes and disomes were isolated following standard DiSP protocols with minor modifications. One modification was to reduce the stringency of the RNase I digest, to enhance the footprint ligation efficiency. This partial footprint digestion should generate longer footprint extensions and may enhance the efficiency of ligation. The chosen ligase was RtcB from *E. coli,* which can directly ligate RNase I digested RNA fragments (**Fig. 39,** Tanaka and Shuman 2011). The limited RNase I digest was in addition beneficial to reduce the digestion of rRNA, which will also be ligated by RtcB. The monosome and disome samples were diluted prior to ligation to favor footprint ligation of ribosomal pairs over ligations of footprints of randomly colliding ribosomes in solution (based on established protocols used for Hi-C, van Berkum *et al.* 2010). The monosome fraction was used as a background control for inter-molecular footprint ligations. Sequencing of the 60 nt fragments revealed up to 95% rRNA fragments, which are more prevalent among the 60 nt fragments than among the 30 nt fragments. The remaining 5% of the 60 nt footprints indeed contained ligated RNA fragments. However, the vast majority (> 85%) of the ligation products included rRNA-rRNA or rRNA-mRNA fusion products. The remaining 15% revealed some true mRNA – mRNA fusions, however we did not succeed to identify specific chimeric ligation products that were significantly enriched in the disome over the ligated monosome fraction (data not shown). The project was stopped at this point, as it was not possible to identify nascent chain connected heteromers or the distance between neighboring ribosomes on one mRNA. However, other experimental methods can be employed to further explore this mechanism on a global scale (see Discussion and Outlook).

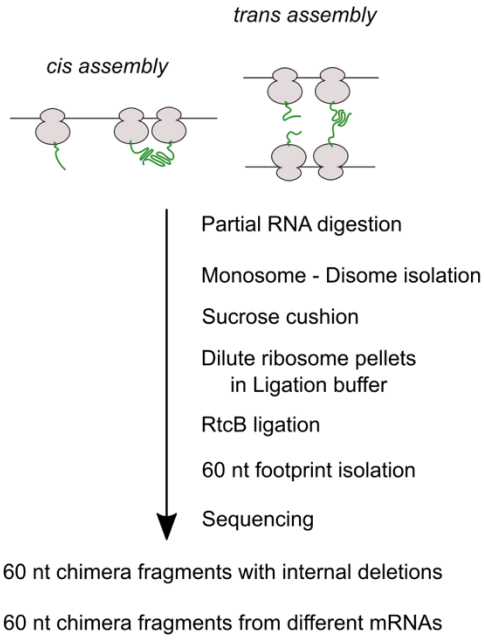

Fig. 39: Experimental set up to re-ligate disome footprints.

5. Discussion

Our DiSP data reveal co-co assembly as a widespread mechanism employed by human cells to efficiently generate in particular homomeric protein complexes. Our DiSP data allow to determine, whether and when two nascent subunits interact *in vivo*. The investigation of co-co assembly by DiSP is the first comprehensive study that presents direct proof of productive nascent chain dimerization. The nascent chain dependency of co-co assembly candidates was corroborated by the loss of the footprint density shift after nascent chain degradation via proteinase K treatment or puromycin-mediated nascent chain release. For all candidates that showed a significant footprint density shift we could additionally determine the efficiency of co-co assembly, based on the loss of monosome reads after assembly. Finally, our data allowed to explore the nascent chain segments that mediate co-co assembly. We find that more than 30% of all annotated homodimers are formed via co-co assembly.

5.1 Comparison of DiSP results with suggested co-co assembly candidates

Several cytoskeletal proteins had been proposed before to employ co-co or co-post assembly to assemble into filaments (reviewed in Fulton and L'Ecuyer 1993). The following list of proteins were proposed to co-co assemble and are expressed in our tested human cell lines: myosin (Isaacs and Fulton 1987), vimentin (Isaacs 1989a), NF-kB1 (Lin *et al.* 2000), p53 (Nicholls *et al.* 2002), peripherin (Chang *et al.* 2006), hERG potassium channel (Liu *et al.* 2016), Rpt1-Rpt2 (Panasenko *et al.* 2019), and the TAF6-TAF9 complex (Kamenova *et al.* 2019).

The assembly of vimentin was proposed to co-co assemble by Isaacs *et al.* 1989: "[...] it is possible that formation of the double-stranded coiled coil could begin while two nascent vimentin chains are still attached to ribosomes, generating a mature amino-terminal end of the coiled coil during translation". Our DiSP data supports this model and demonstrates that nascent vimentin dimerizes once the annotated coiled coil domain is partially exposed (**Fig. 8A**). DiSP provides additional insights and further evidence on the existence of co-co assembly including many more intermediate filaments, for example all nuclear type V intermediate filaments known as lamins (**Fig. 8A**).

Another important example is the co-co assembly for NfκB, which according to our DiSP data occurs after full exposure of the annotated dimerization domain (**Fig. 8A**). This onset of assembly fully agrees with the previously proposed co-translational assembly mode including two nascent NfκB chains on one mRNA (Lin *et al.* 2000). The NFKB1 mRNA encodes for two

proteins, the shorter p50 and the full-length protein p105. The C-terminus of p105 encloses the iκb domain, which keeps NfκB in an inactive state. The active protein consists of only a p50 homodimer, after the C-terminal half is cleaved by proteolysis. One model of NfkB1 assembly and maturation suggests co-translational nascent chain dimerization and C-terminal cleavage by the proteasome of only one C-terminal half, therefore generating an inactive p50-p105 heterodimer from one mRNA (Lin *et al.* 2000). Our data support the co-translational dimerization but do not provide evidence for co-translational nascent chain cleavage, as we do not observe a loss of disome reads after the onset of co-co assembly. According to our data both nascent chains stay intact and in the dimeric state till the end of translation. The evidence provided in the Lin *et al.* 2000 paper does not explain how the proteasome co-translationally discriminates between two identical and almost equally long assembled NfkB nascent chains. Their model is also challenged by other reports that claim to observe a signal-induced post-translational proteasome-dependent p105 proteolysis (Heissmeyer *et al.* 2001). It is however possible that both, the experimental conditions and the employed cell lines, impact the detection of co-translational versus post-translational p50-p105 maturation.

Another complex that has been described to co-co assemble in human HeLa cells was the TAF6 and TAF9 heterodimer, which are part of the transcription factor IID (TFIID, Kamenova *et al.* 2019). Both TAF6 and TAF9 proteins are ranked as low confidence co-co assembly candidates in our DiSP data. The disome enrichment for TAF6 spans about 500 codons while for TAF9 the disome enrichment only spans about 160 codons (**Fig. 38B**). Assuming that TAF6 and TAF9 heterodimerize via co-co assembly *in trans*, we would expect that the length of the disome enrichment along each mRNA should be similar, assuming that ribosomes on both mRNAs translate with a comparable speed. Interestingly, we observed that the TAF6 homologue TAF6L showed a more compatible disome enrichment for the heterodimer formation with TAF9, indicating that this homologue may assemble with TAF9 in HEK293-T cells (**Fig. 38B**). It is to note that the co-purifications and the microscopy results provided in the Kamenova *et al.* 2019 paper are also fully compatible with a bi-directional co-post assembly mechanism, as described for three protein yeast complexes studied by Shiber *et al.* 2018.

Another class suggested to form a heterodimer via co-co assembly was *RPT1-RPT2* in yeast or *PSMC1-PSMC2* in humans, which are subunits of the proteasome 19S regulatory particle (Panasenko *et al.* 2019). Three of these 19S subunits are in our low confidence list and one is in the high confidence list (**Fig. 38**). Of note, we detected even stronger disome shifts for almost all 20S subunits, 11 out of 14 have a sigmoidal footprint shift (*PSMA1-PSMA7* and *PSMB1-PSMB7,* **Fig. 38C, D**). This indicates that the proteasome is a strong candidate for

co-co heterodimer formation, but further experimental evidence is required to prove that the disome shifts are caused by heterodimer formation, and if so which subunits co-translationally pair with each other.

No clear sigmoidal footprint shift was detectable for any hERG potassium channel protein subunit (*KCNH1- KCNH7*) in both tested cell lines, which is mainly due to the extremely low expression levels. Liu *et al.* 2016 used transient overexpression to study the assembly of two isoforms arising from the *KCNH2* gene. We could also not support the suggested co-co assembly of peripherin (Chang *et al.* 2006), which again is due to the extremely low expression levels in both tested cell lines.

Another candidate that has been suggested to co-co dimerize is p53, which also showed no footprint density shift in our DiSP data (**Fig. 8B**). The assembly mechanism of p53 was described to involve co-translational homodimerization of two nascent p53 chains between ribosomes translating one mRNA, followed by a post-translational dimerization of two homodimers to form the final homotetramer (Nicholls *et al.* 2002). It is possible that we failed to identify p53 in our screen caused by the rather C-terminal position of the dimerization domain (**Fig. 8B**). The dimerization could therefore happen during the end of translation after almost full exposure of the nascent peptide. The disome formation depends on a stable interaction that must survives the monosome-disome separation. Such a late and short-lived co-co assembly event at the end of translation would be in full agreement with the previously published data and the absence of a detectable disome enrichment in our DiSP data.

5.2 Localization and abundance of co-co dimerization domains

DiSP data from human cells revealed that mostly homomeric subunits are significantly enriched (**Fig. 18C**). Domains that mediate homodimerization are known to be enriched rather at the C-terminal halves in archaea, bacteria and eukaryota (Natan *et al.* 2018). This evolutionarily conserved C-terminal localization of homodimerization domains was proposed to ensure that folding occurs before assembly, to prevent misfolding and aggregation of nascent chains (Natan *et al.* 2018). Our own analyses could recapitulate this C-terminal enrichment, based on available crystal structures of all human homodimers (**Fig. 19B**, right). However, we observed a clear N-terminal enrichment of dimerization domains among our high confidence co-co assembly candidates (**Fig. 19B**, left). This N-terminal enrichment is caused by a certain set of N-terminal homodimerization domains, in particular coiled coil, BTB, SCAN, RHD, and BAR domains, which suggests that these protein domains are particularly suited to facilitate co-co assembly (**Fig. 20**).

The annotation of the human proteome reveals that the total amount of dimerization domains is rather limited. Coiled coils as well as BTB domains are two prevalent, conserved and reoccurring dimerization domains employed by many proteins, which all comprise a tendency to form unintended mixed heteromeric interactions. In humans 177 different BTB domain containing proteins are annotated (according to the SMART database, Letunic *et al.* 2018). This protein-protein interaction domain is mostly N-terminally located and mediates dimerization of especially transcription factors and potassium ion channels. The amount of coiled coil containing proteins is even higher and currently includes more than 2000 different coiled coil proteins that are annotated in the human proteome (2274 according to UniProtKB, 2400 according to Rose *et al.* 2005a). Coiled coils are employed by a variety of different protein classes to mediate dimerization, especially transcription factors, molecular motors and intermediate filaments.

The problem of forming unintended, mixed heterodimers becomes even more evident if one considers that the current human gene annotation includes on average seven transcript isoforms for each protein coding gene based on alternative splicing, of which many are not forming heteromers. Many of these splice variants have alternative subcellular locations (Uhlén *et al.* 2015). Therefore, it seems reasonable to assume that co-co assembly serves an additional function in higher eukaryotes, which is the prevention of unintended heteromer formation This would create selective pressure for early co-co assembly to avoid homodimer formation based on random diffusion (**Fig. 27B**).

Interestingly, cells have simultaneously evolved a quality control system to recognize and eliminates misassembled heteromers. This so-called Dimerization Quality Control (DQC) surveys BTB domain assemblies, but it was speculated that such a surveillance mechanism could also exist to monitor the composition of other complexes, including coiled coils (Mena *et al.* 2018).

Long coiled coils as well as BTB domains have evolved rather late in evolution. BTB domains are so far not found in bacteria (Stogios *et al.* 2005). Coiled coil domains longer than 250 amino acids are almost completely absent in the proteome of bacteria, but prevalent in archaea and eukaryotes (Rose *et al.* 2005a). Most long coiled coil proteins in eukaryotes are parts of motor proteins involved in transport or function as filamentous stabilators of the cytoskeleton or nucleus. In general, coiled coil domains are more abundant in the human proteome (roughly 10% of the proteome contains coiled coils) in comparison to the *E. coli* proteome (only 3% of the proteome contains coiled coils). This finding indicates that co-co assembly became presumably more relevant and abundant later in evolution, in organisms

with a higher degree of complexity including an increase of cell volume and compartmentalization. It is known that most known ordered protein assembly pathways tend to be evolutionarily conserved (Marsh *et al.* 2013), which suggests that our identified co-co assembling protein domains should co-co assemble in all three domains of life.

5.3 Co-co assembly depends on intrinsic features of nascent chains

We employed PK and puromycin treatments coupled to DiSP to test whether the detected footprint density shift is lost by destroying nascent chain interactions, which was the case for most detected candidates (**Fig. 12 - 15**). This set of experiments does not exclude that a "nascent chain bridging factor" is involved in co-co assembly. This need of an involved eukaryotic specific binding factors was disproved by the reconstruction of lamin and dynactin co-co assembly in *E. coli* (**Fig. 30 and Fig. 32**). We could further show that the disome formation of lamin in *E. coli* depends on the quaternary structure formation between two nascent chains (**Fig. 36**). This was shown by swapping the first and fifth amino acid along all heptameric repeats of the coiled coil 1B domain. These swap-mutations did not affect the hydrophobicity nor the ability to form α-helices, but dramatically reduced the calculated ability to form a coiled coil (**Fig. 36**). Taken together, co-co assembly appears to be independent of assembly factors and may be generally driven by the dimerization-propensity of N-terminal nascent chains segments. Such nascent chains segments that drive co-co assembly are therefore expected to assemble quickly and may even assist simultaneously co-translational folding.

5.4 Nascent chain folding and assembly are intertwined processes

Our DiSP data revealed two modes of co-translational assembly, on the one hand the onset of assembly after full exposure of a dimerization domain, suggesting folding precedes assembly (e.g. BTB, RHD or SCAN domains), and on the other hand the start of assembly after only partial domain exposure (coiled coil or BAR domains) (**Fig. 20**). The second mode of assembly indicates that folding and assembly are not distinct, consecutive processes but occur simultaneously and are functionally intertwined. For co-co dimerization involving a coiled coil we hypothesized that the mechanism of assembly works in a zipper-like manner. Indicating that two alpha-helical ends start to interact and gradually form the coiled coil resembling a zipper while translation continuous. This would indicate that secondary and tertiary structure formation occur simultaneously. In the context of a collaboration, the Tans lab proofed by single molecular optical tweezer experiments that the productive coiled coil

formation of lamin depends on simultaneous folding and assembly. Uncoupling assembly and folding lead to the formation of a monomeric compacted fold that was unable to dimerize, implying folding and assembly must be integrated to prevent early nascent chain misfolding (unpublished data, experiments were performed by Jaro Schmitt and Florian Wruck). These *in vitro* measurements via optical tweezers are the first demonstration of an direct interplay between assembly and folding. The results highlight the importance of co-co assembly to ensure correct coiled coil formation and prevent nascent chain aggregation. Alpha-helices are supposed to fold quickly, often during synthesis. The formation of coiled coils may predominantly occur via co-co assembly and are thereby generally implicated in productive homomer biogenesis, considering that alpha-helix interactions are the most frequent and most efficient co-co assembly domains (**Fig. 21C**). This demonstrates that co-co assembly is an integral part of the folding process and suggest a physiological relevance of co-co assembly. It is known that the translation speed variations can influence co-translational protein folding and maturation steps, which indicates the co-evolution of folding and translation (Zhang et al. 2009, Zhang et al. 2011, Jacobs et al. 2017). Our DiSP data suggest that a similar dependency and co-evolution could exists between co-translational folding and assembly.

Nascent chain aggregation is often described as a thread for newly synthesized proteins. The here identified mostly alpha-helical co-co dimerization domains resemble a "closed interface", that dimerize and have mostly no tendency to form higher multimers, which could trigger the formation of larger aggregates (as shown for β-strand interactions that tend to self-associate into higher assemblies). Of note, DiSP is restricted to the comparison of monosomes and disomes, and is therefore limited to the detection of dimerization. It is feasible to also perform TriSP, trisome selective ribosome profiling, to identify also trimerization, or even higher oligomerization by isolating higher sucrose fractions. The isolation of such higher sucrose fractions after RNase treatment could also include ribosomes that are coupled by co-translational aggregated nascent chains. Therefore, additional experimental controls would be required to prove that the footprint density shift towards higher sucrose fractions is caused by productive multi-subunit formation and not by nascent chain aggregation.

The yeast specific chaperone Ssb binds to nascent chains before the onset of co-translational assembly *in vivo* (Shiber *et al.* 2018). Along the same line, it is tempting to speculate that also in human cells folding as well as assembly of native co-co dimers is supported by co-translationally acting chaperones and maturation factors. Presumably, co-co assembly does not compete with chaperone binding, but is rather coordinated with folding and chaperone interactions. Our data showing assembly of lamin in bacteria do not exclude that in human cells, chaperones may bind to prevent early premature assembly, or that additional

mechanisms are involved to ensure correct complex formation (for example by equal subunit expression or mRNA co-localization). In addition, our DiSP data also do not exclude the co-existence of co-co and co-post assembly for the same complex.

5.5 DiSP reveals transient assembly interactions

Our initial focus was set on stable co-co dimerization events with a constant disome enrichment that remained almost to the end of translation. However, also transient co-translational protein dimerization could form *in vivo*. Our applied fitting model allowed the robust detection of single and double sigmoidal enrichments, including transient peaks of disome enrichment (**Fig. 18A**). The majority of detected double sigmoidal candidates did not indicate a transient disome enrichment in the middle of the coding sequence but indicated a steep footprint loss, right before the end of translation. This was interpreted as indication of *asynchronous termination of* the leading and lagging ribosomes of a nascent chain connected pair. The leading ribosome will terminate and thereby the lagging ribosome shifts to the monosome fraction (**Fig. 18B**).

However, transient peaks in the middle of the coding sequence occurred rarely, which could report on meaningful biological events. Such a transient short-lived disome enrichment could be caused by nascent chain interactions that are recognized and resolved by chaperones. This would prevent co-translational aggregation or could be intended and required as a transient scaffold to assist productive co-translational folding. It is also possible that transient dimerization occurs naturally, but are resolved by other co-translational nascent chain modifications that resolves the dimeric and favor the monomeric state. Such transient nascent chain dimerization that are actively resolved should be detected by an antiparallel density shifts back from the disome to the monosome state. Such a footprint density shift from disomes back to the monosome was seen for HSP90AA1 (**Fig. 8B)** or TRP (**Fig. 21C)**. This may suggest that co-translational modifications or chaperone binding dissolved the dimerization.

Another explanation of the detection of only transient disome enrichments could be caused by an active ribosome quality control mechanism, which detects the disome state and thereby disassembles ribosome. However, single candidates that showed a disome footprint loss and no simultaneous increase of monosome reads showed no general footprint loss in the corresponding total translatome data (e.g. EEA1 or CLIP1 in **Fig. 21C**). This disqualifies an active disassembly mechanism for our detected co-co assembly candidates. Therefore, a second explanation is that such a disome read loss without a gain of monosome reads, is caused by the formation of higher oligomeric assemblies. Such a formation of higher

oligomeric states would shift reads from the disome fraction to the trisome or quatrosome fraction, which are not detectable in the current DiSP approach. This indicates that also higher oligomerization could be detectable by preforming Trisome selective profiling.

5.6 Are co-co assembling disomes potential substrates of ribosome quality control?

Several ribosome quality control mechanisms are described that detect and eliminate collided ribosomes (see Introduction **2.3.4**). These mechanisms are known to initiate ribosome disassembly followed by mRNA and nascent chain degradation. We detected that ribosome collision occurs more frequently before the onset of nascent chain dimerization, which indicates that translation slow-down or stalling does occur before nascent chain dimerization events. However, most of our detected co-co assembly candidates indicate a stable disome enrichment till the end of translation and not a loss of reads after ribosome collisions occurred, which suggests that RQC is not predominantly degrading co-co assembled nascent chains. It is known that certain fractions of newly synthesized proteins are rapidly degraded caused by cotranslational ubiquitination. In human cells it was proposed that about 12%–15% of all nascent polypeptides are co-translationally polyubiquitinated resulting in proteasomal targeting (Wang *et al*. 2013). Therefore, it is possible that some of the collided ribosomes that we observed before onset of assembly are targeted by RQC, without affecting the bulk disome enrichment (**Fig. 25A**). In general, further experimental evidence is required to understand how collided ribosomes escape RQC. On possible answer is that this is kinetically controlled and that transient ribosome collisions escape RQC, as long as the duration of the collided state is short-lived. In this regard it is important to understand how the collided state can be released before RQC is activated. Folding near the ribosome exit tunnel can generate a pulling force on the nascent chain C-terminus, which can release ribosome stalling and facilitate translation elongation (Fritch *et al.* 2018, Farías-Rico *et al.* 2018). Similar to the coupling of translation and folding, a recently published study suggests that co-translational assembly can also generate a pulling force (Fujiwara *et al*. 2020). Therefore, it is tempting to speculate that the collision of two ribosomes could occur before assembly to ensure close proximity of two nascent chains, and the collided state is then released by nascent chain assembly. A failure of nascent chain assembly would extend the lifespan of the collided state and eventually activate RQC.

5.7 Disease relevance of co-co assembly

Misfolded proteins are often toxic and are implicated in different human diseases. We provide initial evidence that co-co assembly ensures correct complex assembly, especially for the formation of isoform specific lamin homodimers (**Fig. 27B**). The assembly mechanism for lamins was so far not understood and it was speculated that mRNA localization or heteromer degradation sustain the exclusive lamin homomer state. The here performed experiments indicate that lamin homomer formation is presumably not driven by active mRNA localization. Considering that protein synthesis is an energetically costly process, it seems unlikely that lamin degradation is the general mechanism to ensure the exclusive existence of isoform specific homodimers *in vivo*. Out proposed model of co-translational subunit dimerization provides a simple and elegant mechanism to regulate quaternary structure acquisition in the crowded cytoplasm for homodimers.

Our new model of lamin homomer formation based on co-co assembly could be important to understand different lamin related diseases, known as laminopathies. The best described laminopathies are caused by mutations of the *LMNA* gene include the Emery-Dreifuss muscular dystrophy or the Hutchinson–Gilford progeria syndrome (Ho and Lammerding 2012). The most common class of *LMNA*-linked diseases are mediated by deficient lamin A processing, which includes farnesylation of a C-terminal cysteine (Schreiber and Kennedy 2013, de Leeuw *et al.* 2018). However, mutations occurring within the lamin rod dimerization domain have been identified in a growing number of human diseases associated with skeletal and cardiac muscle weaknesses, including the Charcot-Marie-Tooth disease, Severe fetal akinesia or Congential-type muscular dystrophy (Ho and Lammerding 2012, Rose *et al.* 2005a). These disease-causing mutations in the central rod domain are affecting both lamin A and lamin C. All of these mutations lead to an altered residues composition of the lamin rod domain, implying an essential structural and/or functional role of these residues. It is possible that such mutations cause disease by impairing co-co assembly and result in deficient oligomeric states of lamin.

A number of our co-co assembly candidates are also relevant cancer drivers, our high confidence list includes for example TPR (proto-oncogene) or PRKAR1A (tumor-suppressor). One main hallmark of cancer cells is the resistance to apoptosis, which leads to higher survival rates. Nuclear factor kappa B (NFκB), regulates cell survival by inducing the expression of many apoptosis suppressor genes (Li *et al.* 2016, Karin and Lin 2002). Formation of NfkB p50 dimers resembles the active state and its formation must be suppressed in healthy cells. Therefore, the NfkB gene encodes simultaneously for the inhibitor of p50, names p105. Such

activator-inhibitor pairs, including toxin-antitoxin pairs in bacteria, are presumably highly dependent on efficient complex formation to ensure that the potentially harmful oligomeric state is not unintentional produced and released into the cytosol.

5.8 Co-co assembly of membrane and secreted candidates

Many translocated proteins indicated a clear disome enrichment, which suggested that co-co assembly could happen inside a membrane bilayer or inside the ER lumen for secreted proteins during co-translational translocation (**Fig. 22**). Since most secreted extracellular matrix proteins form oligomers (Redick and Schwarzbauer 1995), it is plausible that a number of translocated protein complexes co-co assemble. The frequency of disome enriched candidates for translocated proteins was comparable to the frequency of co-co assembly of cytosolic proteins (**Fig. 10A**). The disome enrichment often initiated upon complete exposure of the first N-terminal transmembrane domain (**Fig. 22A**). Co-co assembly between two TMD domains inside a membrane could stabilize less hydrophobic or amphipathic transmembrane helices(**Fig. 23**). For co-co assembly *in cis*, we further speculate that a cluster of translocons inside a planar ER membrane may facilitate co-co assembly by local proximity and directed release of interacting TMDs into the membrane bilayer. However, further control experiments are required to exclude that some of the disome enriched candidates is caused by interactions of a single TMD with bulky components of the translocation machinery, that may shift a monosome into the disome fraction.

Our DiSP data revealed an enrichment of glycine residues in the TMD sequences that are enriched before onset of assembly (**Fig. 23B**). The GxxxG motif is a commonly described transmembrane dimerization domain (Lemmon *et al.* 1992), supporting the model that the observed disome shifts are reporting on *in vivo* dimerization events. A number of publications indicate that one simple sequence motif like the GxxxG motif cannot be the main mediator of TMD dimerization, and that other factors are involved (reviewed in Li *et al*. 2012). It remains to be experimentally explored how this whole process of transmembrane dimerization is controlled (see Outlook).

Along this line, it was also already reported that transmembrane proteins employ co-translational assembly, mediated by TMD dimerization inside the ER membrane (Lemmon *et al.* 1992). Co-post assembly was suggested for D1 and D2 membrane proteins (Zhang *et al.* 1999), evidenced by the co-purification of full-length D2 proteins with D1 ribosome-nascent chain complexes. Co-translational assembly was also suggested for IgE receptors, which depend on specific stoichiometries (Fiebiger *et al*. 2005). These receptors only efficiently form

fully functional complexes in the presence of the known partner subunits. In the absence of the required partner subunits improperly assembled receptors were detected that were subjected to the ER quality control system and immediately directed toward degradation (Fiebiger *et al.* 2005).

Indications for co-translational assembly are also published for the secreted extracellular matrix protein tenascin (Redick and Schwarzbauer 1995), which is exclusively found as a hexamer while precursors, like monomers, dimers, or trimers are undetectable. This lack of assembly intermediates suggests that nascent tenascin polypeptides associate prior to completion of translation to form hexamers co-translationally. The multimeric tenascin complex is stabilized by amino-terminal inter-subunit disulfide bonds, which also form co-translationally, in agreement with many reports that show disulfide bonds can be formed inter- and intra-molecular during co-translational translocation (Bergman and Kuehl 1979). The tenascin gene encodes for at least two different transcripts via alternatively splicing (Erickson 1993). Both tenascin isoforms have never been reported to mix into heteromeric hexamers (Taylor *et al.* 1989), which is reminiscent for the here presented results for lamins, which also have multiple isoforms that exclusively form homomers. As for lamins, tenascin contains amino-terminal coiled coil structure proposed to mediate assembly (Spring *et al.* 1989).

6. Outlook

The here presented DiSP data describe an unrecognized level of coordination of protein synthesis with protein assembly and open up many exciting questions. One interesting aspect will be the analysis of the evolutionary conservation of co-co assembly, especially since we have already identified co-co assembly candidates from bacteria (PhD project from Josef Auburger and Jaro Schmitt) as well as yeast (unpublished work from Dorina Merker). A comparison of three species specific co-co assembly candidates could reveal conserved protein orthologs, which presumably lead to the co-evolution of specific homodimerization domains. Our finding that a human protein co-co assembles in *E. coli* indicates that the folding environment is not absolutely critical to induce nascent chain assembly and that the process is mediated by intrinsic features of nascent chains. Therefore, it would be intriguing to see that our identified human co-co assembling domains, or smaller folds within these domains, are conserved. The experimental DiSP set up and the data analysis tools can in principle be applied to any organism, allowing even broader comparisons. In addition, cross-correlations of co-co assembly onsets and position resolved nascent interactomes of difference chaperones will presumably indicate a coordination of assembly with chaperone binding, as seen for co-post assembly in yeast (Shiber *et al*. 2018). Such correlations will further reveal the coordination and interplay of different ribosome-associated maturation factors with co-translational assembly. However, not only data set comparisons but also a series of different experimental approaches will be required to further reveal the detailed mechanisms that drives co-co assembly (listed below).

(i) It will be exciting to further explore co-co assembly within membranes and during translocation of our identified ER localized co-co assembly candidates. However, further control experiments will be needed to understand the driving forces of this process. We identified multi-transmembrane candidates with an onset of assembly after the exposure of the second or third TMD. Therefore, it would be interesting to investigate what features drive assembly and support this by mutations and TMD swop experiments. It will also be worth studying if a single TMD alone can mediate co-co assembly. Alternative scenarios may be that specific N-terminal peptide sequences or a nucleotide sequence in the UTRs guide the assembly path.

Considering that the oligomeric state of most membrane proteins is only poorly annotated, it is necessary to first proof that our disome enrichment after exposure of the first TMD is really connected to a cotranslational dimerization event. Therefore, we planned a first set of

experiments to determine the oligomeric state for single TMD candidates by adding an N-terminal splitFlAsH tag to single ER membrane proteins with a clear co-co assembly onset after TMD exposure. N_{in} topology proteins could be used to visualize by fluorescence microscopy that the dimerization is happening inside the ER membrane in living cells (Griffin *et al.* 1998).

(ii) We speculate that the arrangement of neighboring ribosomes on a polysome connected to nearly located translocons should align nascent chains in a planar area, this parallel nascent peptide orientation should facilitate co-co assembly. The hypothesis that the polysome structure could facilitate co-co assembly should also be true for cytosolic translated mRNAs. The three-dimensional orientation of ribosomes translating a single mRNA has been described to either avoid proximal nascent proteins from interactions (Brandt *et al.* 2010) or to facilitate assembly (Mrazek *et al.* 2014). These two opposing findings indicate presumably two different modes of polysome conformation. For monomeric proteins and all non-co-co assembly candidates it would be beneficial if the distance between exit tunnels is maximized. Instead, for co-co assembly candidate it would be beneficial to have two exit tunnels in close proximity. This opens up an exciting direction for further structural studies to visualize co-co assembly in the context of a polysome for example using *in vivo* cryo-EM tomography. Constructs generated for this thesis could be used as first starting points to analyze the connection between co-co assembly and polysome structure. We showed that plasmid-encoded lamin mRNAs can occupy more than 60% of all translating ribosomes in *E. coli* after 15 min of IPTG induction. This could be used to study *in vivo* the polysome structure of wild type lamin coil 1B, which co-co assembles, in comparison to the mutated coil 1B construct, which does not co-co assemble. This specific comparison in the artificial *E. coli* system could be used as starting point to understand if co-co assembly can alter the polysome structure or not. If this is not the case, it would be still interesting to see if co-co assembled ribosomes are pairwise oriented along a single mRNA or not, based on the pictures obtained via cryo-EM tomography.

(iii) The analysis of 60 nt disome footprints revealed ribosome pausing before the onset of co-co assembly. This hints towards a coordination between translation elongation and assembly in mammalian cells. Different reports highlighted that collided ribosomes occur during the translation of specific codon pairs, which leads to stalling (Arpat *et al.* 2020, Zhao *et al.* 2020, Han *et al.* 2020). Along this line it was proposed that ribosome pausing and co-translational interactions are coordinated (Panasenko *et al*. 2019). We speculate that nascent peptide folding and assembly could relieve ribosome pausing. Accordingly, data generated in

collaboration with the group of Sander Tans show that co-co assembly between two nascent peptides can be analyzed using optical tweezer and that such a dimerization event can produce a pulling force (in line with results from Fujiwara *et al.* 2020). Further data collection and analysis will be required to corroborate such a model of nascent chain assembly and ribosome stalling relief. Additional cross correlations between co-co assembly onsets and 60 nt stalling peaks could also help to identify regulating mRNA or nascent peptide motifs of this assembly mechanism.

(iv) One outstanding goal is to understand how prevalent heterodimers are formed by co-co assembly. The here presented proteome-wide approach based on disome footprint re-ligations failed, however it is still possible to overcome this limitation of DiSP. A method called "split-pool recognition of interactions by tag extension" (SPRITE) could be employed to resolve the two footprints of one disome pair (Quinodoz *et al.* 2018). Instead of ligating two footprints, this method would require dilutions of the disome sample to separate single disome pairs in multi-well plates followed by barcoding of both footprints in one well. An iterative process of sample pooling, splitting and barcoding is repeated several times, and thereby generates a unique string of barcode for all footprint pairs that always end up in the same well. This method was originally developed to cluster higher-order chromatin structures, but has the potential to be also used for RNA fragments. This method could allow us to answer if disomes are formed between different mRNAs encoding for heteromers. Such identified candidates could then be validated by smFISH approaches *in vivo* to further cooperate the existence of spatially defined co-localized mRNA transcripts that allow the formation of higher order assemblies.

(v) In general, the here presented proteome-wide data may be corroborated by single candidate studies and detailed mechanistic analysis. One possible direction is to further study the co-co assembly mechanism of lamins and other highly disease relevant candidates like NfkB, TPR (proto-oncogene) or PRKAR1A (tumor-suppressor). For these candidates it would be of high interest to study the impact of naturally occurring mutations within the dimerization domains. For lamins it is known that several missense mutations within the rod dimerization domain can cause distinct diseases that are associated with skeletal and cardiac muscle weaknesses (reviewed in Ho and Lammerding 2012). Therefore, they could have an impact on co-co assembly, and could interfere with quaternary structure formation of lamin

7. Material & Methods

Table 2: Instruments used for this study

Instrument	Type / Model	Company
Agarose gel chamber and trays	/	ZMBH workshop
Balances	PG603-S, PB1502-S	Mettler-Toledo International Inc.
Bioanalyzer	2100 Bioanalyzer Instruments	Agilent Technologies
Biological safety cabinet	Safe 2020 1.2 (class II laminar flow cabinet)	Thermo Fisher Scientific
Blue Light LED Transilluminator	UVT-22 BE-LED	Herolab
Filtration System	All-Glass Filter (90 mm)	EMD Millipore
Fractionator	Model 2110 Fraction collector	Bio-Rad
French Pressure Cell Press	FLA-078	SLM AMINCO
Gel Image System	GenoSmart gel documentation system	VWR
Gel System	XCell SureLock Mini-Cell Electrophoresis System	Invitrogen
Gel System	Criterion Vertical Electrophoresis Cell	Bio-Rad
Image Analyzer	ImageQuant LAS 4000	GE Healthcare
Incubator	ISF1-X, Climo-Shaker / MIR-254, MIR-154-PE	Kuhner / Panasonic
Incubator (with CO_2)	HERAcell 150i	Thermo Fisher Scientific
Magnetic stirrer	MR Hei-Mix L	Heidolph
Microscope	ECLIPSE TS100	Nikon
Mini-centrifuge	C1202	Labnet
Mixer Mill	MM400, with 10 ml jars and 12 mm balls	Retsch
Nanodrop	Nanodrop 2000 UV-VIS Spectrophotometer	Thermo Fisher Scientific
Overhead roller	Intelli mixer	Neolab
pH-Meter	FE20, pH-electrode LE438	Metter-Toledo International Inc.
Photometer	NovaSpec UV/VIS Spectrophotometer	GE Healthcare
Pipetboy	Pipetboy2	Integra
Plate Reader	FLUOstar Omega, SPECTROstar Nano	BMG LABTECH
Power Supply	Consort EV231 / Power PAC 300	Sigma-Aldrich / Bio-Rad
Qubit	Qubit 3.0 Fluorometer	Thermo Fisher Scientific
Sequencer	NextSeq550 / HiSeq2000	Illumina
Shaker	Model 3015	GFL
Sucrose Gradient Detector	TRIAX FC2	BioComp
Sucrose Gradient Station	GRADIENT Station ip	BioComp
Tabletop Centrifuges	5424, 5424R	Eppendorf AG
Thermocycler	T personal / FlexCycler2	Biometra / Analytikjena
Thermomixer	Comfort / TS pro	Eppendorf / CelMedia
Turbo Blotter	Trans-Blot Turbo Transfer System	Bio-Rad
Ultracentrifuges	Sorvall Discovery 90SE / M120SE	Thermo Fisher Scientific
UV source	INTENSILIGHT C-HGFI	Nikon

UV table	CHROMA 43	Vetter
Vacuum pump	Vacuum pump unit	Vacuubrand
Vortex	Genie 2	Scientific Industries
Water Bath	1012	GFL
Western blot apparatus midi/maxi	/	ZMBH workshop

Table 3: Consumables and equipment used for this study

Consumables & Equipment	Company
1.5 ml + 2 ml test tube	Sarstedt
1.5 ml DNA LoBind Safe-Lock tubes	Eppendorf
1.5 ml non-stick RNase-free tube	Ambion
12% ExpressPlus PAGE Gels (12 and 15 well)	GeneScript
15 ml + 50 ml centrifuge tubes	Sarstedt
26-G needle	Sarstedt
6/12/24/96-well plates	Greiner
Cell scraper	TPP
Cellulose acetate filters, 0.2 µm	Sartorius
Criterion TGX Precast Gels	Bio-Rad
Cryo-Vials S	Greiner
Cuvettes	Sarstedt
Erlenmeyer flasks	Schott
Filter Tips (10, 20, 200 and 1000 µl)	Neptune Scientific
Freezing Containment (Mr. Frosty)	Thermo Fisher Scientific
Gel breaker tubes	IST Engineering
Glass bottles with cap	Schott
Instant blue solution	Expedeon
Latex Gloves	Semperguard
Magnetic separation stand (MagneSphere)	Promega
measuring cylinder	EcoLab
Micro loop	Biosigma
Neubauer counting chamber	Brand
Nitril Gloves	Ansell
Parafilm	Bremis
PCR tubes	Sarstedt
Penicillin-Streptomycin mix	Gibco
Petri dishes	Greiner
PVDF membrane, 0.2 µm	Carl Roth
Rotor (SW40 / TLA120S)	Beckman Coulter
Spin-X-cellulose acetate columns	Sigma-Aldrich
Sterile filters, 0.2 µm	Sarstedt
SW40 centrifugation tubes	Seton
Syringe filters, 0.22 µm	Sarstedt
Syringe, 5, 10 and 50 ml	Becton Dickinson BD
T25 + T75 flask	Greiner
TBE-Urea polyacrylamide gel, 8, 10 or 15%	Invitrogen
Tips (10, 20, 200 and 1000 µl)	Steinbrenner
Whatman Paper, 3 mm	GE Healthcare Life sciences

Table 4: Commercial kits used for this study

Kit	Company
Agilent High Sensitivity DNA kit	Agilent Technologies
Agilent Small RNA kit	Agilent Technologies
GenElute Gel Extraction Kit	Sigma-Aldrich
GenElute HP Plasmid Miniprep Kit	Sigma-Aldrich
GenElute PCR Clean-Up	Sigma-Aldrich
Mix&Go Kit	Zymo Research
NEXTflex small RNA-Seq Kit v3	PerkinElmer
NextSeq 500/550 High Output Kit v2/2.5 (75 / 150 cycles)	Illumina
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific
QIAprep Spin Miniprep Kit	QIAGEN
QIAquick gel extraction	QIAGEN
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific
SMARTer smRNA-Seq Kit for Illumina	TaKaRa
SuperScript III first-strand synthesis kit	Thermo Fisher Scientific

Table 5: Commercial Enzymes/Proteins used for this study

Enzyme/Protein	Company
Benzonase	Millipore
CircLigase ssDNA Ligase (c = 100 U/µl)	Epicentre
DNase	Lab stock
Dnase1 (25 U/ml)	Roche
HF Phusion Polymerase (c = 2 U/µl)	New England BioLabs
Lysozyme 10 mg/ml	Sigma-Aldrich
MNase	Lab stock
Murine RNase Inhibitor (c = 40,000 U/ml)	New England BioLabs
Opti-Taq DNA polymerase	EURx Ltd
OptiTaq polymerase	Roboklon
Phusion polymerase	Thermo Fischer Scientific
Proteinase K from Tritirachium album	Sigma-Aldrich
rAPID alkaline phosphatase	Roche
Restriction enzymes	New England BioLabs/ Thermo Fisher Scientific
RNase I (c = 100 U/µl)	Ambion
RNase-free DNase I (c = 10 U/µl)	Roche
RtcB	New England BioLabs
Superscript III Reverse Transcriptase (c = 200 U/µl)	Invitrogen
T4 DNA Ligase	New England BioLabs
T4 DNA Ligase	Lab stock
T4 DNA Ligase	Thermo Fischer Scientific
T4 Polynucleotide Kinase (c = 10 U/µl)	New England BioLabs
T4 RNA Ligase 2, truncated (c = 100 U/µl)	New England BioLabs
Taq polymerase	Lab stock
TrueCut Cas9 protein v2	Invitrogen

Table 6: Antibodies used for this study

Antibody Target	Used dilution	Host	Provider	Specificity
anti-mouse-AP	1:20 000	Horse	Vector Laboratories	/
anti-rabbit-AP	1:20 000	Goat	Vector Laboratories	/
anti-sheep-AP	1:20 000	Rabbit	Vector Laboratories	/
HSP70/Hsc70	1:2000	Mouse	Santa Cruz Biotechnology	/
Lamin E1	1:3000	Mouse	Santa Cruz Biotechnology	sc-376248
Puro	1:1000	Mouse	Millipore	MABE343
RPL32 (human)	1:500	Rabbit	Human protein atlas	HPA047501
RPL4 / RPL6 (human)	1:5000	Rabbit	Lab collection	/
S2 (bacteria)	1:2500	Sheep	Lab collection	/
Strep-Tactin-AP	1:1000	/	IBA	

Table 7: Cells used or generated for this study

Cells	Source	Genotype
Rosetta *E. coli* strain	Novagene	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR)
FI8202 *E. coli* strain	Lab Stock	[ΔntrBCfadAB101::Tn10 laqIq lacL8/λ202]
XL1-Blue E. coli cells	Lab Stock	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacIqZΔM15* Tn*10* (Tetr)]
BL21 (DE3) E. coli cells	Lab Stock	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*
U2OS cells	ATCC	/
HEK293-T cells	DSMZ	/
HEK293-T cells	This study	Heterozygous *LMNA*(wt/gfp11-TS), clone A2
HEK293-T cells	This study	Heterozygous *DCTN1*(wt/gfp11-TS), clone B7
HEK293-T cells	This study	Homozygous *DCNT1*(wt/gfp11-TS), clone B2

Table 8: Software used for this study

Software	Version	Company/Reference
2100 Expert Software	B.02.08	Agilent
Anaconda Navigator	1.9.7	Anaconda Inc.
bcl2fastq converter	2.20	Illumina
Bowtie	1.2.2	Langmead *et al.*, 2009
Bowtie2	2.2.9	Langmead and Salzberg, 2012
cutadapt	1.13	Martin, 2011
FastQC	0.11.5	Simon Andrews
Fiji with ImageJ	2.0.0	Wayne Rasband
IGV	2.3.91	Robinson *et al.*, 2011
ImageQuant LAS 4000 Control Software	/	GE Healthcare
Inkscape	0.92	Free Software Foundation, Inc.
Julia	1.1	Bezanson *et al.,* 2012
Python 2	2.7.15	Python Software Foundation
Python 3	3.6.6	Python Software Foundation
R	3.6.1	The R Foundation
RStudio	1.2.5019	RStudio, Inc.
Samtools	1.5	Li *et al.*, 2009
Serial Cloner	2.6.1	Franck Perez
STAR	2.7.1a	Dobin *et al.,* 2013
Sypder	3.3.4	Pierre Raybout
TopHat	2.1.0	Kim *et al.,* 2013
Triax Flow Cell Software	1.50A	BioComp
Word, Excel, Power Point for Mac	16.39	Microsoft

Table 9: Chemicals used for this study

Chemicals	Company
2-Propanol (>99.8)	Sigma-Aldrich
20% SDS (sodium dodecyl sulfate) solution	Ambion
Acetone (99.5%)	Bernd Kraft
Acid Phenol:Chloroform (5:1, pH 4.5, >99%)	Ambion
Agarose	Sigma-Aldrich
Biotin	Carl Roth
Bromphenol blue	Carl Roth
Calcium chloride	Carl Roth
Chloroform	Merck
DEPC	Carl Roth
DMSO (dimethyl sulfoxide)	Sigma-Aldrich
dNTP mix	Bioline
DTT	Sigma-Aldrich
EGTA (ethylene glycol tetra acetic acid)	AppliChem
Ethanol	AppliChem
Ethanol (>99.8%)	Sigma-Aldrich
Ethidium bromide (1% stock solution)	Applichem
Ethylendiamine-tetraacetic acid (>99%)	Carl Roth
FlAsH-EDT2	Lab collection
Glycerol (99.0 – 101.0%)	Sigma-Aldrich
HEPES	AppliChem
Isopropanol	Sigma-Aldrich
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Carl Roth
Magnesium chloride (>98.5%)	Carl Roth
Magnesium sulphate heptahydrate (>99%)	Carl Roth
Methanol (99.8%)	Sigma-Aldrich
Monosodium phosphate ($NaH_2PO_{4)}$	AppliChem
NP40	Sigma-Aldrich
o-Nitrophenyl-β-D-galactopyranoside (ONPG)	Sigma-Aldrich
PMSF	Applichem
Polyethylene glycol 6000	Carl Roth
PopCulture Reagent	Sigma-Aldrich
Potassium chloride	Carl Roth
Potassium chloride (>99.5%)	Carl Roth
Potassium hydroxide	Carl Roth
SDS-Pellets (>99%)	Carl Roth
Skim milk powder	Fluka
Sodium chloride (>99.8%)	Sigma-Aldrich
Sodium hydroxide (98%)	Carl Roth
Sodium phosphate dibasic dihydrate ($Na_2HPO_{4)}$	AppliChem
Sucrose	Sigma-Aldrich
TCA	Applichem
TCEP bond breaker	Thermo Fisher Scientific
Tris	Carl Roth
Tris ultrapure	Applichem
Triton X-100	Sigma-Aldrich
Tween-20	Carl Roth
β-mercaptoethanol	Carl Roth

Table 10: Plasmids used and generated for this study

Plasmid	Features	Source	Comments
pET3a_SF-LMNC-TS	ColE1, AmpR, Inserted ORF: ATG-SplitFlasHTag-LMNC_31-572,-TwinStrepTag	This Study	Frameshift before the TwinStrep Tag Employed for DiSP in *E. coli*
pET3a_FLAG-SF-LMNC-TS	ColE1, AmpR, Inserted ORF: ATG-FLAGTAG-SplitFlasHTag-LMNC_31-572-TwinStrepTag	This Study	Constructed by Jiri Koubek based on pET3a_SF-LMNC-TS, frameshift corrected Used for SeRP in *E. coli*
pET3a_SF-deltaCoil1A-LMNC-TS	ColE1, AmpR, Inserted ORF: ATG-SplitFlasHTag-deltaCoil1A_LMNC_71-572-TwinStrepTag	This Study	generated based on pET3a_SF-LMNC-TS Used for DiSP in *E. coli*
pENTR221_DCTN1	pUC ori, KanR, DCTN1 cDNA sequence encoding for p150glued	This Study	Ordered via DKFZ Gateway ORF collection
pET3d_SF-DCTN1-TS	ColE1, AmpR, Inserted ORF: ATG-SplitFlasHTag-DCTN1-TwinStrepTag	This Study	Employed for DiSP in *E. coli*
pET3a_SF-Coil1B-mCherry-TS	ColE1, AmpR, Inserted ORF: ATG-SplitFlasHTag-LMNC-Coil1B-WT_mCherry-TwinStrepTag	This Study	Employed for DiSP in *E. coli*
pET3a_SF-Coil2AB-mCherry-TS	ColE1, AmpR, Inserted ORF: ATG-SplitFlasHTag-LMNC-Coil2AB-WT_mCherry-TwinStrepTag	This Study	Employed for DiSP in *E. coli*
pET3a_SF-Coil1B*-mCherry-TS	ColE1, AmpR, Inserted ORF: ATG-SplitFlasHTag-LMNC-Coil1B-a<->e-swap mutant-mCherry-TwinStrepTag	This Study	Ordered via General Biosystems Inc. Employed for DiSP in *E. coli*
pJH391-lambda-coil1A-TS	AmpR, first 132 nt of λ *cI indI* fused with LMNC coil1A-TwinStrep tag, driven by the lacUV5 promoter	This Study	Used for Dimerization Assay
pJH391-lambda-coil1B-TS	AmpR, first 132 nt of λ *cI indI* fused with LMNC coil1B-TwinStrep tag, driven by the lacUV5 promoter	This Study	Used for Dimerization Assay
pJH391-lambda-coil2AB-TS	AmpR, first 132 nt of λ *cI indI* fused with LMNC coil2AB-TwinStrep tag, driven by the lacUV5 promoter	This Study	Used for Dimerization Assay
pJH391-lambda-coil1A-1B-2AB-TS	AmpR, first 132 nt of λ *cI indI* fused with LMNC coil1A-1B-2AB-TwinStrep tag, driven by the lacUV5 promoter	This Study	Used for Dimerization Assay
pJH391-lambda-Head-coil1A-1B-2AB-TS	AmpR, first 132 nt of λ *cI indI* fused with LMNC Head_domain-coil1A-1B-2AB-TwinStrep tag, driven by the lacUV5 promoter	This Study	Used for Dimerization Assay
pJH370	AmpR, fusion between λ *cI indI* and GCN4(751-843), driven by the lacUV5 promoter	Hu *et al.* 1990	Used for Dimerization Assay
pFG157	AmpR, full-length λ cI indI	Hu *et al.* 1990	Used for Dimerization Assay
pKH101	AmpR, first 132 nt of λ cI indI	Hu *et al.* 1990	Used for Dimerization Assay
pET3a	ColE1, AmpR	Lab collection	/
pET3d	ColE1, AmpR	Lab collection	/

pJH391-TS	AmpR, first 132 nt of λ *cI indI* fused to TwinStrep Tag, driven by the *lacUV5* promoter	This Study	Constructed by Anja Schubert based on pJH391
pcDNA3.1-GFP(1-10)	f1 ori, SV40 ori, AmpR, Neo/KanR, Inserted ORF: GFP(1-10)	This Study	Ordered via Addgene Used for CRISPR-Cas9
pcDNA3.1	f1 ori, SV40 ori, AmpR, Neo/KanR	Lab collection	Ordered via Addgene

Table 11: DNA markers used for this study

Size standards	Company
10 bp DNA Ladder	Invitrogen
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific
PageRuler Prestained Protein Ladder	Thermo Fisher Scientific
Ultra Low Range DNA Ladder	Thermo Fisher Scientific

Table 12: Solutions used for this study

Solutions	Company
10x ECF solution	GE Healthcare Life Sciences
5x HF or GC Phusion buffer	Thermo Fischer Scientific
Bradford Solution	Bio-Rad
DMEM (high glucose, GlutaMAX, pyruvate)	Gibco
Heat-inactivated FCS	Gibco
GlycoBlue (c = 15 mg/ml)	Ambion
Lipofectamine 2000	Invitrogen
Lipofectamine Cas9 Plus reagent	Invitrogen
Lipofectamine CRISPRMAX reagent	Invitrogen
Novex TBE-Urea sample buffer (2x)	Thermo Fisher Scientific
OptiMEM	Gibco
PBS (10X), pH 7.4	Gibco
SYBR Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific
Trypsin-EDTA 0.05% solution	Gibco

Table 13: Affinity matrices used for this study

Affinity matrix	Company
Strep-Tactin Sepharose (50% slurry)	IBA
Streptavidin magnetic beads	New England BioLabs
MagStrep "type3" XT beads (5% suspension)	IBA

Table 14: Inhibitors used for this study

Inhibitor	Company
protease inhibitor (complete EDTA free).	Roche

Table 15: Antibiotics used for this study

Antibiotics	Used Concentration	Company
Ampicillin (dissolved in water)	100 µg/ml	Carl Roth
Chloramphenicol (dissolved in isopropanol)	25 µg/ml	Sigma-Aldrich
Cycloheximide (dissolved in ethanol)	100 µg/ml	BioChemica
Kanamycin (dissolved in water)	50 µg/ml	Carl Roth
Penicillin-Streptomycin (10,000 U/mL)	100 U/ml	Gibco
Puromycin Dihydrochloride (dissolved in water)	2 mM	Gibco

Table 16: DNA Primers used for this study

ID	Descriptive Name	Sequence (5' to 3')
MB 36	pCA528-T7_prom-fwd	TAATACGACTCACTATAGGG
MB 37	pCA528-term-rev	GGCTTTGTTAGCAGCCGGATC
MB 44	GFP11_inside fw	TCATGAGTATGTAAATGCTGCTGGGA
MB 46	GFP11_inside_rev	TCCCAGCAGCATTTACATACTCATGA
MB 75	LMNA_inside_rev	CTCCTTGGCTACTGAGTCAA
MB 76	LMNA_inside_fw	GATGAGGAGGGCAAGTTTGT
MB 78	T7 fw	TAATACGACTCACTATAGGGAGA
MB 108	LMNA template for Gibson rev	TAGCCTCTCCTCCTCGCCC
MB 132	LMNC_gRNA5_valid_fwd	GTAGACATGCTGTACAACCCTTCC
MB 133	LMNC_gRNA5_valid_rev	GGTATAGGGAGGAGAGAGAAGAAAGG
MB 134	DCTN1_gRNA11_valid_fwd	GATGACACAGTCTACATGGGCAAAG
MB 135	DCTN1_gRNA11_valid_rev	GAGGGAGCAGTTGAACAACAAATTATG
MB 143	NdeI-flash-d30LMNC_fw	ggccggcatATGGCGGGAAGTTGCTGCGGCGGCGAGAAGGAGGACCTGCAGGAG
MB 144	BamHI-TwinStrep_rev	CCGGCCGGATCCTTAGGCGCCTTTTTCGAACTGC
MB 158	pJH391_SalI_LMN_ATG_fwd	GATGCGGAGAGATGGGTGTCGACAATGGAGACCCCGTCCCAG
MB 159	pJH391_SalI_LMN_Coil1A_fwd	GATGCGGAGAGATGGGTGTCGACAGACCTGCAGGAGCTCAATG
MB 160	pJH391_BamHI_LMN_Coil1A_rev	ACTGCGGGTGAGACCAGGATCCGACCACCTCTTCAGACTCG
MB 161	pJH391_SalI_LMN_Coil1B_fwd	GATGCGGAGAGATGGGTGTCGACATACGAGGCCGAGCTCGGG
MB 162	pJH391_BamHI_LMN_Coil1B_rev	ACTGCGGGTGAGACCAGGATCCGGTCTCACGCAGCTCCTC
MB 163	pJH391_SalI_LMN_Coil2AB_fwd	GATGCGGAGAGATGGGTGTCGACAGATGCGCTGCAGGAACTG
MB 164	pJH391_BamHI_LMN_Coil2AB_rev	ACTGCGGGTGAGACCAGGATCCCTCGCCCTCCAAGAGCTTG
MB 184	Split_FLAsH_rev	GCCGCCGCAGCAACTTCCCGCCAT
MB 189	LMNA_after_Coil1A_fwd	AGCCGCGAGGTGTCCGGCATCAAG
MB 197	DCTN1_seq_fwd	CAAGAGCTGGAAGTTGTGAG
MB 198	DCTN1_seq_fwd2	CTACAGATATTGCCCTCCTG
MB 205	His_GA_Strep_fwd	CACCACCACCACCACCACGGCGCCTGGAGCCACCCGCAGTTCGAGAAAGGTGGA
MB 206	Strep_SpeI_pet3d_rev	GCTTTGTTAGCAGCCGGATCCACTAGTTTACTTAAGGGCGCCTTTTTCGAA

MB 209	pET3d_NcoI_SF_DCTN1_fwd	AAGAAGGAGATATACCATGGGAAGTTGCTGCGGCGCACAG AGCAAGAGGCACGTGTAC
MB 210	His_GA_DCTN1_x_stop_rev	GTGGTGGTGGTGGTGGTGGGCGCCGGAGATGAGGCGACT GTGAAG
MB 212	SpeI-Coil1B-fwd	AGA-ACTAGT-TACGAGGCCGAGCTCG
MB 213	XhoI-Coil1B-rev	ATA-CTCGAG-GGTCTCACGCAGCTCC
MB 214	SpeI-Coil2AB-fwd	ATA-ACTAGT-GATGCGCTGCAGGAACTG
MB 215	XhoI-Coil2AB-rev	ATA-CTCGAG-CTCGCCCTCCAAGAG
MB 216	mCHerry_inside_fwd	CCAGACCGCCAAGCTGAAGG
MB 217	mCHerry_inside_rev	CCTTCAGCTTGGCGGTCTGG
#1229	pJH391-fwd	GGCAGGGATGTTCTCACCTGAGC
#1230	pJH391-rev	GGACGATATCTCAGCGGTGGCAGC

Table 17: gRNAs and donor templates used for this study for CRISPR insertions purchased from IDT (underlined sequence represents the homology arms; gRNA5 + T5 were used for *LMNA* and gRNA11 + T11 for *DCTN1* endogenous tagging)

gRNA/ template	Target region	Sequence (5' to 3')
gRNA5	*LMNA* exon 10, cut exactly before STOP codon	GTGAGTGGTAGCCGCCGCTG (PAM: AGG)
gRNA11	*DCTN1* exon 10, cut exactly before STOP codon	AGCAGGGGAAAGGAGTGCTT (PAM: AGG)
T5	Donor template for insertion into *LMNA*	CTGCTCCATCACCACCACGTGAGTGGTAGCCGCCG CAGCGCCAGAGACCACATGGTGCTGCATGAGTACG TAAACGCCGCCGGGATTACAGGTAGCGCCTGGTCT CACCCCCAGTTCGAGAAAGGAGGTAGCGCCTGGA GCCACCCCCAGTTCGAAAAATGAGGCCGAGCCTGC ACTGGGGCCACCCAGCCAGG
T11	Donor template for insertion into *DCTN1*	AGCAGCTGCACCAGCTTCACAGTCGCCTCATCTCC CAGCGCCAGAGACCACATGGTGCTGCATGAGTACG TAAACGCCGCCGGGATTACAGGTAGCGCCTGGTCT CACCCCCAGTTCGAGAAAGGAGGTAGCGCCTGGA GCCACCCCCAGTTCGAAAAA TAAGCACTCCTTTCCCCTGCTGTCCCCTTCGACCCT CAG

Table 18: Self-designed biotinylated depletion oligos for human rRNA fragments (ordered from siTOOLS)

Oligos for 30 nt RNA footprints	Oligos for 30 nt circ. DNA	Oligos for 60 nt RNA footprints
ACCGGCTATCCGAGGCCAAC	GTTGGCCTCGGATAGCCGGT	CCTGGTTGATCCTGCCAGTA
GACCGGCTATCCGAGGCCAA	TTGGCCTCGGATAGCCGGTC	TTGATCCTGCCAGTAGCATA
CGGCTATCCGAGGCCAACCG	CGGTTGGCCTCGGATAGCCG	AAGATTAAGCCATGCATGTC
CCGGCTATCCGAGGCCAACC	GGTTGGCCTCGGATAGCCGG	CGACTCTTAGCGGTGGATCA
CGGGCGCTTGGCGCCAGAAG	CTTCTGGCGCCAAGCGCCCG	CTTAGCGGTGGATCACTCGG
CCGGGCGCTTGGCGCCAGAA	TTCTGGCGCCAAGCGCCCGG	TGAAGAACGCAGCTAGCTGC
CAGACAGGCGTAGCCCCGGG	CCCGGGGCTACGCCTGTCTG	TTGCCCTCGGCCGATCGAAA
GACGCTCAGACAGGCGTAGC	GCTACGCCTGTCTGAGCGTC	CTCGGCCGATCGAAAGGGAG
CGACGCTCAGACAGGCGTAG	CTACGCCTGTCTGAGCGTCG	CGAATCCGGAGTGGCGGAGA
GCGACGCTCAGACAGGCGTA	TACGCCTGTCTGAGCGTCGC	ATACCACTACTCTGATCGTT
AGCGACGCTCAGACAGGCGT	ACGCCTGTCTGAGCGTCGCT	TTTTCACTGACCCGGTGAGG
GACAGGCGTAGCCCCGGGAG	CTCCCGGGGCTACGCCTGTC	GTATAGTGGTGAGTATCCCC
GCCGGGCGCTTGGCGCCAGA	TCTGGCGCCAAGCGCCCGGC	GTGGTGAGTATCCCCGCCTG
CCTCGATCAGAAGGACTTGG	CCAAGTCCTTCTGATCGAGG	GGGTTCGATTCCCCGACGGG
GCCTCGATCAGAAGGACTTG	CAAGTCCTTCTGATCGAGGC	CCTTCCCGTGGATCGCCCCA
TGCGATCGGCCCGAGGTTAT	ATAACCTCGGGCCGATCGCA	TCCCGTGGATCGCCCCAGCT
CGATCGGCCCGAGGTTATCT	AGATAACCTCGGGCCGATCG	ATTGATCATCGACACTTCGA
GCGATCGGCCCGAGGTTATC	GATAACCTCGGGCCGATCGC	TCATCGACACTTCGAACGCA
GGGCCGGTGGTGCGCCCTCG	CGAGGGCGCACCACCGGCCC	TTCCTCCCGGGGCTACGCCT
CGGGCCGGTGGTGCGCCCTC	GAGGGCGCACCACCGGCCCG	CCTACTTGGATAACTGTGGT
GACGGCGCGACCCGCCCGGG	CCCGGGCGGGTCGCGCCGTC	AGAGCTAATACATGCCGACG
ACCGGGTCAGTGAAAAAACG	CGTTTTTTCACTGACCCGGT	ACATGCCGACGGGCGCTGAC
ACTCCGCACCGGACCCCGGT	ACCGGGGTCCGGTGCGGAGT	GAGGTGGGATCCCGAGGCCT
ACAGGCGTAGCCCCGGGAGG	CCTCCCGGGGCTACGCCTGT	CCTGGTTGATCCTGCCAGTA
ACAGGCGTAGCCCCGGGAGA	TCTCCCGGGGCTACGCCTGT	TTGATCCTGCCAGTAGCATA
CGACGGCGCGACCCGCCCGG	CCGGGCGGGTCGCGCCGTCG	AAGATTAAGCCATGCATGTC
AGGACTTGGGCCCCCCACGA	TCGTGGGGGGCCCAAGTCCT	CGACTCTTAGCGGTGGATCA
CCGGGTCAGTGAAAAAACGA	TCGTTTTTTCACTGACCCGG	CTTAGCGGTGGATCACTCGG
CGGGTCGACTCCGTGTACAT	ATGTACACGGAGTCGACCCG	TGAAGAACGCAGCTAGCTGC
AGGCCTCGGGATCCCACCTC	GAGGTGGGATCCCGAGGCCT	TTGCCCTCGGCCGATCGAAA

Table 19: Genome-aligned reads of all employed data sets for this thesis (low salt refers to 150 mM KCl and high salt to 500 mM KCl in the lysis- and sucrose gradient buffer. All data sets labeled with *"E. coli"*, refer to a Rosetta *E. coli* background, Table 7. The stable HEK293-T cell lines (here labeled with clone A2 and B2) employed for SeRP can be found in Table 7)

Experiment	Replicates	Sample	Aligned reads (million)
DiSP U2OS low salt*	Rep1	mono	16.5
		dis	12.6
	Rep2	mono	19.2
		dis	23.5
DiSP HEK293-T low salt (PK1, re-sequenced)**	Rep1	mono	22.5
	Rep2	dis	21.3
DiSP HEK293-T high salt	Rep1	mono	18.3
		dis	16.9
	Rep2	mono	20.5
		dis	20.0
DiSP HEK293-T high salt + proteinase K (PK9)	Rep1	mono1 (noPK)	22.5
		mono2 (VeryHighPK)	10.8
		mono3 (HighPK)	11.7
		mono4 (MidPK)	11.5
		mono5 (LowPK)	10.3
		dis1(noPK)	20.6

		dis2 (VeryHighPK)	11.1
		dis3 (HighPK)	8.59
		dis4 (midPK)	9.2
		dis5 (LowPK)	11.0
DiSP HEK293-T high salt + puromycin	Rep1	mono-puro	11.3
		dis-puro	10.8
		mono+puro	9.7
		dis+puro	10.3
	Rep2	mono-puro	12.3
		dis-puro	11.3
		mono+puro	9.35
		dis+puro	15.9
DiSP HEK293-T low salt + proteinase K (PK1)**	Rep1	mono1 (noPK)	3.9
		mono2 (VeryHighPK)	3.4
		mono 3 (HighPK)	6.2
		dis1(noPK)	4.0
		dis2 (VeryHighPK)	2.6
		dis3 (HighPK)	2.6
DiSP *E. coli*	Rep1	mono	11.2
		dis	9.3
DiSP *E. coli* + *SF-LMNA-TS*	Rep1	mono	4.7
		dis	6.4
	Rep2	mono	8.9
		dis	7.5
DiSP *E. coli* + *SF-DCTN1-TS*	Rep1	mono	9.3
		dis	5.0
DiSP *E. coli* + *SF-Δcoil1A-LMNA-TS*	Rep1	mono	9.6
		dis	10.8
SeRP *E. coli* + *Flag-SF-LMNA-TS*	Rep1	Total Translatome	6.8
		AP	5.2
SeRP HEK293-T + Lamin (clone A2***)	Rep1	Total Translatome	18.7
		AP	23.8
SeRP HEK293-T + *DCTN1* (clone B2)	Rep1	Total Translatome	3.7
		AP	0.6
60 nt DiSP HEK293-T +/- PK	Rep1 (noPK)	mono	0.2
		dis	1.2
	Rep1 (highPK)	mono	0.4
		dis	1.4
60 nt DiSP HEK293-T	Rep1	dis	3.44
	Rep2	dis	2.44

* libraries were prepared according to Döring et al. 2017 to be sequenced on a HiSeq2000
** libraries were prepared with the TAKARA SMARTer smRNA-Seq Kit for Illumina
*** libraries were prepared with the NEXTflex small RNA-Seq Kit v3

7.1 Cell culture work

The following steps were performed in a laminar flow cabinet (SAFE2020, Thermo Fisher Scientific, class II). For the laminar flow cabinet all used materials were cleaned rigorously with 70% (v/v) ethanol.

Thawing of cells - A T25 flask was filled with 5 ml high glucose DMEM media containing GlutaMAX and pyruvate (Gibco) supplemented with 10% (v/v) heat inactivated FCS (Gibco), 100 units/ml penicillin and 100 µg/ml streptomycin (Gibco) and pre-heated at 37°C in a humidified incubator with 5% CO2 (HERAcell 150i). One Cryo-S vial (Greiner Bio-one) of cells was removed from the liquid nitrogen tank and transported on dry ice to the cell culture. The Cryo-S vial was placed in a pre-heated 37°C water bath for approximately 1 min and removed before the whole sample was thawed. The vial was disinfected with 70°C ethanol and the complete content was added to the pre-warmed media inside a T25 flask. The flask was slowly shaken and then placed in the 37°C incubator for 24 hours. On the next day the growth medium was replaced with fresh media to remove residual DMSO. From that day on cells were subcultured according to the cell line specific instructions. Cells were at least three times subcultured before first experiments were started.

Cell culture subculturing - U2OS cells (ATCC) and HEK293-T cells (DSMZ, Table 7) were cultivated in high glucose DMEM media containing GlutaMAX™ and pyruvate (Gibco) supplemented with 10% (v/v) heat inactivated FCS (Gibco), 100 units/ml penicillin and 100 µg/ml streptomycin (Gibco) in T25 or T75 culture flasks (CellStar, Greiner Bio-one). Cells were passaged regularly (each second or third day, according to the cell line specific instructions) through trypsinization (Gibco) and grown in a humidified incubator with 5% CO2 at 37°C (HERAcell 150i). To do so the growth media was aspirated and attached cells quickly washed with 1x PBS (Gibco) and detached by Trypsin-EDTA 0.05% solution (Gibco) for 2-3 min (HEK293-T cells at room temperature, U2OS cells at 37°C). Trypsinization was stopped by adding at least five times higher volume of DMEM media with supplements to the flask (T25= 1 ml Trypsin + 5 ml DMEM with supplements, T75= 3 ml Trypsin + 15 ml DMEM with supplements).

Cell counting - 10 µl of diluted or undiluted cell suspension was transferred to a Neubauer counting chamber (Ansell). Cells were counted under a Nikon Eclipse TS100 microscope with 200x magnification. Each time all cells in four big 16-corner squares were counted and the average amount was used for the cell concentration calculation.

Seeding - Cells were seeded 18-24 hours before lysis in 15 cm^2 dishes (3.5 million U2OS or 6 million HEK293-T cells) to reach 70-90% confluency at the time of harvesting. A single dish of cells seeded in this way is enough for performing one DiSP experiment. The number of 15 cm^2 dishes for SeRP varied between 8 - 15 dishes for one experiment.

Stable cell line generation - Three HEK293-T knock-in cell lines were generated via CRISPR-Cas9 homology-directed repair: one homozygous cell line expressing *DCTN1-GFP11-TS* (clone B2), one heterozygous cell line expressing *DCTN1-GFP11-TS* (clone B7), and one heterozygous cell line expressing *LMNC-GFP11-TS* (clone A2, Table 7). The sequence encoding for GFP11 followed by a TwinStrep tag was provided as single-stranded donor oligonucleotide (ssODN) with 35 nt homology arms at each side (purchased from IDT, Table 17). The guide RNA was designed according to the Dharmacon online design tool (http://dharmacon. horizondiscovery.com/gene-editing/crispr-cas9/crispr-design-tool/). The insertion of the GFP11-TwinStrep tag sequence deleted the protospacer adjacent motif (PAM sequence) and thereby avoided re-cleavage by Cas9 at the same position.

The gRNA5 was used to generate the in-frame knock-in of *GFP11-TS* into *LMNA* (behind exon 10, therefore only affecting lamin C and not lamin A) and gRNA11 for the knock-in of *GFP11-TS* into *DCTN1* (in the last exon 31, affecting multiple isoforms including the longest p150 isoform, Table 17). The GFP11-TS tag was inserted in both cases directly before the stop codon. The TwinStrep tag sequence contained a short linker (GGGSA) between two Strep tags (WSHPQFEK). The GFP11 was included to allow to enrich positive clones via FACS by simultaneous transient co-expression of GFP1-10. A C-terminally GFP11 tagged protein can complement non-fluorescent GFP1-10 to generate fluorescence GFP (Cabantous *et al.* 2005). The TrueGuide Synthetic gRNA reagents (Invitrogen) were used according to the provider instructions. HEK293-T cells were first transfected with purified Ribo-Nucleo-Proteins (TrueCut Cas9 Protein v2, Invitrogen) including a two-piece annealed guide RNA (gRNA, composed of a crRNA and a tracrRNA) followed by the transfection of a single-stranded donor oligonucleotide (ssODN) for homology directed repair. Therefore, a poly-L-lysine coated 24-well plates (Greiner) was prepared and 60 000 HEK293-T cells were seeded per well on one day before transfection. On the next day, per planed transfection a 20 µl mixture was prepares and incubated for 5 min at room temperature, consisting of 7.5 pmol TrueCut Cas9 protein v2, 7.5 pmol crRNA:tracrRNA duplex and 1:10 v/v Lipofectamine Cas9 Plus Reagent (in Opti-MEM medium). Then 5.5 pmol of ssODN template and 1.5 µl Lipofectamine CRISPRMAX reagent diluted in 25 µl OptiMEM was added to the transfection mixture and added to one well with attached cells. After 24 hours of incubation, cells were trypsinized and transferred to poly-L-lysine coated 6-well (Greiner) with fresh DMEM supplemented with 10% (v/v) FBS. On the following day, cells were transfected with 1.5 µg of pcDNA3.1-GFP1-10 plasmid, 4.5 µl Invitrogen Lipofectamine 2000 Reagent in OptiMEM (180 µl transfection mix per well). After 24 hours, single cells were FACS-sorted at the ZMBH Flow Cytometry and FACS facility into 96 well plates according to the GFP signal to enrich for positively edited clones. Single clones were grown until 80% confluency and the edit was validated by genome extraction, targeted PCR amplification and sequencing (see next section).

Genome extraction - The media of confluent wells of a 96 well plate was aspirated and cells detached with 100 µl Trypsin-EDTA 0.05% solution (Gibco) for 5 min at room temperature. Then 200 µl DMEM with 10% (v/v) FCS was added per well and all cells were resuspended by pipette and transferred to a 1.5 ml reaction tube. Remaining cells were left in the well with fresh media. All tubes were centrifuged for 2 min at 1000 xg, media removed and washed with 500 µl 1x PBS. The washed cell pellet was resuspended in 50 µl lysis buffer (10mM Tris 7, 10 mM EDTA pH 8, 0.5% (w/v) SDS, 10 mM NaCl, 1 mg/ml Proteinase K) and incubated for 3 hours at 60°C at 550 rpm (cell pellets can also be frozen in liquid nitrogen and stored at - 20°C). Each sample was then mixed with 6.5 µl 5 M NaCl (f.c. 0.15 M) and 150 µl pure ethanol and precipitated for at least 1 hour or overnight at -80°C. On the next day all samples were centrifuged for 30 min at 20000 xg at 4°C and the pellet washed with 150 µl 70% (v/v) ethanol. Then again centrifuged for 5 min and the supernatant discarded and the DNA pellet tried for 5 min at 55°C (on a thermoblock). The dried pellet was resuspended in 20 µl 10 mM Tris pH 8 and 1:5 diluted for Nanodrop measurements. Genome extractions were diluted to 1-10 ng/µl with water and 1 µl was then used for the PCR validation. The PCR amplification was done with home-made Phusion polymerase in 5x GC Buffer (Thermo Fischer Scientific) supplemented with 1.25 M Betaine and 1.5% (v/v) DMSO. The used primers for determining the correct edit of LMNC were MB132 + MB133, and for *DCTN1* MB134 + MB135 with the following PCR program (see Table 16):

Initial denaturing:	98°C	30 sec
Cycles (39x):	98°C	10 sec
	60°C	30 sec
	72°C	30 sec
Final elongation:	72°C	5 min

The PCR reaction was analyzed on 1% (w/v) agarose gels and stained with EtBr. The tag insertion should lead to a 120 nt band shift of the final PCR product (no insertion: approx. 350 nt, positive homozygous clone: approx. 420bp, positive heterozygous clone: band at approx. 350bp and 420bp). In addition, GFP11 specific primers (MB44 + MB46, Table 16) were used for additional PCR amplification reactions, to check for the correct tag insertion and orientation.

Cryo stocks - A confluent 75 cm^2 tissue culture flask (CellStar, Greiner Bio-one) was used to fill up three Cryo-S vials (Greiner Bio-one). The growth media was aspirated and cells were detached by trypsinsation and stopped by adding DMEM with supplements (see *Cell culture subculturing*). The cell suspension was transferred into a 15 ml Falcon tube (Sarstedt) and centrifuged for 3 min at 100 xg. The cell pellet was washed in 1x PBS and the centrifugation step was repeated. The whole cell pellet was resuspended in 3 ml DMEM supplemented with 10% (v/v) FCS and 10% (v/v) DMSO. All three Cryo-S vials were filled with 1 ml of the cell suspension. The properly closed samples were transferred into a freezing container filled with 250 ml isopropanol (Mr. Frosty, Thermo Fisher Scientific). The whole container was incubated at -80°C overnight and on the next day all vials were transferred into the liquid nitrogen tank.

Mycoplasma check - The MYCOPLASMACHECK service of GATC Biotech was used to test our cultures for contaminations based on a mycoplasma-specific PCR assay. This assay includes primer pairs to detect the following mycoplasma species: *M. arginini, M. fermentans, M. orale, M. hyorhinis, M. hominis, M. genitalium, M. salivarium, M. synoviae, M. pirum, M. gallisepticum, M. pneumoniae, M. yeatsii, Spiroplasma citri* and *Acholeplasma laidlawii.*

7.2 Disome Selective Profiling (DiSP)

Cell lysis for U2OS cells - A 15 cm^2 dish with 70-80% confluent U2OS cells were taken from the 37°C incubator immediately before harvesting. After removing the growth media by inversion, all subsequent steps were performed on ice, using ice-cold and RNase-free solutions and tools in a 4°C room. Dishes on ice were first washed by gently pouring 10 ml of 1x PBS supplemented with 100 μg/ml cycloheximide and 10 mM $MgCl_2$ and the whole dish slightly tilted from one to the other side. Then the PBS solution was poured out of the dish and remaining solution removed by tapping the whole dish several times on a stack of tissue papers. Back on ice 100 μl of 5x concentrated lysis buffer (250 mM HEPES pH 7.0, 50 mM $MgCl_2$, 1 M KCl, 5% (v/v) NP40, 50 mM DTT, 500 μg/ml cycloheximide, 25 U/ml recombinant Dnase1 (Roche) and protease inhibitor (complete EDTA free, Roche)) was added to each dish and cells were scraped from each dish. The final cell lysate of one plate (around 500 μl after lysis, leading to a 1x lysis buffer concentration) was transferred to a 1.5 ml non-stick RNase-free tube (Ambion) and incubated for 15 min on ice. Cell lysates were triturated five times through a 26-G needle and cleared by centrifugation for 5 min at 20,000 ×g at 4°C.

Cell lysis for HEK293-T cells - Two 15 cm^2 dishes with 70-80% confluent HEK293-T cells were taken from the 37°C incubator immediately before harvesting. After removing the growth media by inversion, all subsequent steps were performed on ice, using ice-cold and RNase-free solutions and tools in a 4°C cold room. HEK293-T cells were detached from the dish by pipetting 10 ml of 1x PBS supplemented with 100 μg/ml cycloheximide and 10 mM $MgCl_2$ on top and slowly resuspending the cells with a 10 ml pipette. The cell suspension was collected in 50 ml falcon tubes and pelleted for 3 min at 2000 ×g, 4°C. The cell pellet derived from one dish was resuspended in 200 μl 1x low-salt (50 mM HEPES pH 7.0, 10 mM $MgCl_2$, 150 mM KCl, 1% (v/v) NP40, 10 mM DTT, 100 μg/ml cycloheximide, 25 U/ml recombinant Dnase1 (Roche) and protease inhibitor (complete EDTA free, Roche). Given the requirement for high salt concentrations in the puromycin DiSP experiment *(see below)*, a high-salt lysis buffer

containing 500 mM KCl was employed for all DiSP experiments of HEK293-T cells to allow comparison of the main and control datasets. In all cased lysates were incubated for 15 min on ice and then triturated five times through a 26-G needle and cleared by centrifugation for 5 min at 20000 ×g at 4°C.

RNase I digestion - RNA concentration in the cleared lysate was determined by Qubit HS RNA assay with 1:100 dilutions in water. Lysates were digested with 150U RNase I (Ambion) per 40 μg RNA for 30 min at 4°C and 500 rpm on a thermomixer.

MNase digestion - RNA concentration in the cleared lysate was determined by Qubit HS RNA assay with 1:100 dilutions in water. Lysates were mixed with $CaCl_2$ to reach a final concentration of 5 mM and digested with various amounts between 10 - 500 U MNase per 40 μg RNA were added (self-purified MNase according to Becker *et al.* 2013). Digestions were performed at 4°C for 30 min and 500 rpm on a thermomixer. The digestions were stopped by adding EGTA to a final concentration of 6 mM.

Sucrose gradient formation - The Gradient Station (BioComp) was used in combination with SW40 centrifugation tubes (Seton). The high and low sucrose solutions were prepared in sucrose buffer (50 mM HEPES pH 7.0, 5 mM $MgCl_2$, 150 mM KCl or 500 mM KCl, 100 μg/ml cycloheximide, EDTA Free protease inhibitor tablet Roche) and filtered through a 0.22 μm filter. Short caps were used to seal the tubes and linear gradients were formed with the following custom mixing program for a 5% - 45% gradient: M#1: 09 sec/ 83°/ 30 rpm M#2: 09 sec/ 83°/ 0 rpm M#3: 01 sec/ 86°/ 40 rpm M#4: 7 min/ 90°/ 0 rpm, sequence: 12121212121234. For the mixing of 10% - 25% gradients the following mixing program was used: 2:19 min /81.5°/14 rpm. Gradients were stored at 4°C for at least 1 hour before use.
For measuring the linearity of formed gradients 150 μl Trypan blue was added to the heavy sucrose solution before mixing. The A_{260} profile indicates then the distribution of Trypan blue along the gradient and reports on the linearity of the formed gradient.

Monosome-disome separation - Gradients were stored at 4°C for at least 1 hour before use. Up to 300 µg total RNA was loaded per gradient, 5-45% (w/v) gradients were centrifuged for 3.5 hours and 10-25% (w/v) gradients for 3 hours at 35,000 rpm, 4°C (SW40-rotor, Sorvall Discovery 100SE Ultracentrifuge) to allow maximum separation of monosome and disome peaks (the fastest acceleration and breaking times are recommended). Samples should be loaded slowly on top of the gradient to avoid any mixing of the sample with the top layer of the gradient, the resolving power of the gradient is greater the narrower the initial sample zone is. After centrifugation, absorbance profiles at 254 nm were recorded using the Piston Gradient Fractionator (BioComp) and gradients were fractionated in 40 fractions of 300 µl that were immediately frozen in liquid nitrogen. Fractions corresponding to monosome and disome peaks were pooled separately and subjected to acid phenol RNA extraction (see below). Note that 5 to 8 fractions between the monosome and disome peaks were usually excluded to minimize contamination between the two samples.

Polysome profiling - Samples meant for polysome profiling were kept permanently on ice after cell lysis, and handled like samples for *monosome-disome separation* except that no RNase treatment was performed. Lysates were then directly loaded on a 5-45% sucrose gradient.

DiSP with Proteinase K treatment - As described in Bertolini *et al.* 2021: A PK stock concentration (10 mg lyophilized Proteinase K from Tritirachium album (Sigma-Aldrich) mixed with 1 ml ice cold PK storage buffer (50 mM Tris pH 7.5, 5 mM $CaCl_2$, 40% (v/v) glycerol)) was aliquoted, snap-frozen in liquid nitrogen and stored at -80°C. For PK treatments one aliquot

was thawed and immediately used. All steps were carried out on ice, using pre-cooled ice-cold solutions and tools. DiSP with PK treatment was performed as described above using HEK293-T cells with some modifications. Briefly, cells were harvested and resuspended in 1x high salt lysis buffer without protease inhibitors (composition as for DiSP, see above). Protein concentration in the cleared lysate was determined by Bradford assay (Bio-Rad Protein Assay) followed by a standard RNA digestion with RNase I (see above).
Next, lysates were supplemented with different PK concentrations and incubated for additional 30 min at 10 rpm on a rotation wheel at 4°C. According to the protein content in the lysate, PK was titrated as follows:

No PK	= PK storage buffer was added instead of PK
Low PK	= 1:20000 PK to total protein amount
Mid PK	= 1:6000 PK to total protein amount
High PK	= 1:2000 PK to total protein amount
Very High PK	= 1:200 PK to total protein amount

Samples were loaded on a sucrose gradient to separate monosomes from disomes (see above) and ribosome-protected RNA footprints were processed as described below.
RNase I digestion was omitted in control samples to verify polysome integrity after PK digestion by polysome profiling. Total lysates were also analyzed on SDS PAGE to visualize the degree of protein degradation upon different PK treatments (described below).

DiSP with puromycin treatment - As described in Bertolini *et al.* 2021: Conditions suited to release nascent chains with puromycin without dissociating ribosomes from mRNAs were adapted from (Blobel and Sabatini 1971). Cycloheximide had to be omitted from all solutions because it is incompatible with puromycin activity. All steps were carried out working on ice with ice-cold solutions and tools. HEK293-T cells were seeded on poly-L-lysine coated 15 cm^2 dishes and lysed on dish as follows: cells were rinsed with ice-cold 1x PBS supplemented with 10 mM $MgCl_2$ and lysed by scraping in 100 µl 5x concentrated standard lysis buffer lacking cycloheximide. Next, cleared lysates (roughly 500 µl / dish after scraping) were mixed with 3M KCl solution to obtain a final concentration of 500 mM KCl. Puromycin samples were supplemented with 2 mM puromycin (Gibco) and control samples with the same volume of 1x lysis buffer. RNase I was less active at 0°C compared to 4°C, therefore, all RNA digestions were performed with 750U RNase I per 40 µg total RNA in an ice-bath for 25 min with occasional shaking. After incubation, lysates were cross-linked using 0.5% formaldehyde (Pierce 16% Formaldehyde (w/v), Methanol-free) and incubated for 30 additional minutes in an ice-bath. Samples were loaded as described above on a sucrose gradient to separate monosomes from disomes and ribosome-protected RNA footprints were processed as described below.
RNase I digestion was omitted in control samples to verify polysome integrity after puromycin treatment by polysome profiling. In these cases, sucrose fractions corresponding to the supernatant (containing released nascent proteins) and to polysomes (containing ribosome-bound nascent proteins) were collected. Proteins were precipitated with Trichloroacetic acid (TCA) and separated by SDS PAGE. Puromycilated nascent proteins were detected by Western blot using anti-puromycin antibody (Table 6).

DiSP of E. coli cells - As described in Bertolini *et al.* 2021: Rosetta cells (Novagene, Table 7) were always freshly transformed and selected on LB agar plates with the required antibiotics (Table 15) before each experiment. Colonies were picked for overnight cultures in EZ Rich Defined Medium (Neidhardt *et al.* 1974), which were used on the next day to inoculate 200 ml EZ-RDM to an initial OD_{600} of 0.05. Cells were grown at 37°C in 1L baffled Erlenmeyer flasks with shaking at 120 rpm. Following procedures were performed as described in *(Shieh et al.*

2015, Becker et al. 2013) with minor modifications. Briefly, cells were harvested during log phase (OD_{600} = 0.5-0.6), if not otherwise stated, cells were induced for 16 min with 1 mM IPTG, isolated by fast-filtration and flash-frozen in liquid nitrogen. Frozen cell pellets were lysed by mixer milling (2 min, 30 Hz, Retsch) together with 500 µl frozen lysis buffer (50 mM HEPES pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 0.4% (v/v) Triton X-100, 0.1% (v/v) NP-40, 1 mM chloramphenicol, protease inhibitor tablets (Roche), DNase I (Roche), and 1 mM TCEP or 1 mM DTT). The RNA concentration of the lysate was determined by Nanodrop and then digested with 150U MNase per 40 µg RNA (MNase was produced by myself according to Becker *et al.* 2013) at 25°C and 650 rpm on a thermo mixer. Digestion was stopped by placing samples in ice and adding a final concentration of 6 mM EGTA. Lysates were loaded on pre-cooled 5-45% (w/v) sucrose gradients (sucrose dissolved in 50 mM HEPES pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, protease inhibitor tablets (Roche), and 1 mM TCEP or 1 mM DTT), and centrifuged for 3.5 h at 35,000 rpm, 4°C. Fractions corresponding to monosomes and disomes were isolated and ribosome-protected RNA footprints were processed as described below.

FlAsH measurements - For all experiments with simultaneous FlAsH detection during the sucrose gradient fractionation 3 µM FlAsH-EDT_2 was added directly to the cell lysate before RNase digestion. The excitation maximum of FlAsH is at 508 nm and the emission maximum at 528 nm. The detection of FlAsH fluorescence was done with the TRIAX detector (FC-2 dual wavelength flow cell, BioComp). Therefore, a 474 nm/50 nm filter was used for emission and a 535 nm/50 nm filter to detect the fluorescence (the 50 nm indicates the filter bandwidth of the used filters in the TRIAX, so this is the +/- range of light that can still pass the indicated light range).

Acid phenol RNA extraction - 700 µl of pooled sucrose fractions were mixed with 40 µl 20% (w/v) SDS (Ambion) and quickly vortexed. 750 µl acid phenol:chloroform (5:1, pH 4.5, Ambion, pre-heated to 65°C) were added and incubated for 5 min at 65°C on a thermomixer (Eppendorf) at 1400 rpm. All samples were incubated 5 min on ice and centrifuged at 20000 xg for 2 min. The top aqueous phase was transferred to a fresh tube and mixed additionally with 700 µl non-heated acid phenol. After 5 min of incubation at room temperature with occasional vortexing all samples were centrifuged for 2 min at 20000 xg. Again, the top aqueous phase was transferred to a fresh tube and immediately mixed with 600 µl chloroform. After several rounds of pulse vortexing all samples were centrifuged for 1 min at 20000 xg and the top aqueous phase was transferred to a fresh tube. The remaining 550 µl of aqueous phase was mixed with 62 µl 3 M NaOAc pH 5.5 (1/9 equivalence volume), 2 µl Glycoblue (Ambion) and 700 µl isopropanol (1 equivalence volume) and incubated for 1 hour or overnight at 80°C. Frozen samples were centrifuged for at least 30 min at 20000 xg and 4°C and pellets were washed with 750 µl ice-cold 70% (v/v) ethanol. Followed by a quick centrifugation for 5 min at 20000 xg (4°C) and all remaining liquids were aspirated by pipette. To ensure that no ethanol remains all pellets were dried for 2-5 min at 55°C. Dried pellets were then resuspended in 20 µl 10 mM Tris-HCl pH 7.0.

Footprint isolation - Phenol extracted RNA, dissolved in 10 mM Tris pH 7, was mixed with 2x Novex™ TBE-Urea Sample Buffer (Invitrogen) and denatured for 2 min at 80°C. 1-20 µg of sample was loaded per pocket on to 15% TBE-Urea polyacrylamide gels (Invitrogen) and run for 70 min at 200V in 1x TBE buffer (Ambion, one gel pocket was loaded with only loading dye between different samples). An in-house DNA ladder consisting of 25, 35, 64, 71, 104 nt long poly-N or poly-T nucleotides was employed for all footprint isolations. Gels were stained for 5-20 min in 1x TBE buffer containing 0.01%(v/v) SYBR gold (Invitrogen) at room temperature with slight agitation. Desired gel pieces were extracted on a blue light table with

a scalpel (30 nt footprints = excised range from 25-35 nt, 60 nt footprints = excised range from approximately 55 - 90 nt).
The gel pieces were placed into 0.5 ml gel breaker tubes (IST Engineering), nested into a 1.5 ml tube and centrifuged for 3 min at 20000 xg. To each sample 500 µl 10 mM Tris-HCl pH 7.0 was added and tubes were incubated at 70°C for 15 min in a thermomixer at 1400 rpm (Eppendorf). Gel pieces were removed by using a Spin-X cellulose acetate column (Sigmar) and the flow through was transferred to a new tube. RNA was precipitated by adding 55 µl 3 M NaOAc pH 5.5, 2 µl Glycoblue and 550µl isopropanol. After mixing, tubes were frozen at 80°C for at least 1 hour or overnight. Frozen samples were centrifuged for 30 min at 20000 xg and 4°C and pellets were washed with ice-cold 70% (v/v) ethanol and resuspended in 15 µl of 10 mM Tris-HCl pH 7.0.

Library preparation - Ribosome profiling libraries of all U2OS samples were prepared as described in great detail in (Döring *et al.* 2017) and sequenced on a HiSeq 2000 (Illumina) at the DKFZ Core Facility for Sequencing. All other libraries were sequenced on a NextSeq550 sequencer in our lab. The HEK293-T low salt proteinase K samples were prepared with the SMARTer smRNA-Seq Kit (TaKaRa) according to the providers manual. The SeRP data for LMNC-TS were prepared with the NEXTflex small RNA-Seq Kit v3 according to the providers manual (Table 19). For all other HEK293-T samples ribosome profiling libraries were prepared as described in great detail in (Galmozzi *et al.* 2019, MGlincy *et al.* 2017).

Custom rRNA depletion - As described in Bertolini *et al.* 2021: The most prevalent human rRNA fragments were removed from samples meant to be deeply sequenced by employing custom biotinylated reverse complement DNA oligonucleotides (ordered and designed in collaboration with siTOOLs Biotech, Table 18), followed by a pull-down via magnetic Streptavidin beads (New England BioLabs). In general, rRNA depletions were performed on the adaptor-ligated RNA footprints. To maximize efficiency, an additional depletion step was optionally performed on the circularized DNA using a reverse-complement pool of biotinylated oligos (Table 18). In brief, 5 µl ligated RNA or circularized cDNA was mixed with a 4-fold molar excess of the respective rRNA depletion oligo pool and DEPC water to a final volume of 25 µl. Then 2x wash/binding buffer (40 mM Tris pH7, 1 M NaCl, 2 mM EDTA, 0.1% (v/v) Tween 20 supplemented with 2 µl murine RNase inhibitor for depletions at the RNA stage) was added to a final volume of 50 µl. Nucleic acids were denatured in a thermocycler for 90 sec at 99°C and hybridization was performed by decreasing the temperature by 0.1°C per second to 37°C, followed by a 15 min incubation at 37°C. For each reaction, a 2-fold excess Streptavidin Magnetic Beads (New England BioLabs) was calculated based on the beads binding capacity and the amounts of used biotinylated oligos. Beads were washed three times with 750 µl 1x wash/binding buffer before usage and resuspend in 10 µl 1x wash/binding buffer. Beads were then added to the hybridized RNA/DNA-oligo mix and incubated for 15 min at room temperature (with occasional mixing). Biotinylated oligos hybridized to target rRNA were then magnetized and removed from the sample. The remaining nucleic acids were precipitated as described in the paragraphs before.
The same procedure was done to remove the most prevalent rRNA fragments from E. coli samples, with biotinylated oligos described in Becker *et al.* 2013.

7.3 Selective Ribosome Profiling (SeRP)

Total translatome / Sucrose cushion - To enriched ribosomes for total translatome analysis 100 µl lysates were carefully loaded on 800 µl sucrose cushion (25% (w/v) sucrose, 50 mM HEPES pH 8, 150 mM KCl, 5 mM $MgCl_2$, 100 µg/ml cycloheximide, and protease inhibitor (complete EDTA free, Roche), filter through a 0.22 µm filter) in TL120-rotor tubes. Samples were centrifuged for 1h at 100k rpm and 4°C, the supernatant was immediately removed after

the centrifugation and the ribosome pellet overlaid with 100 µl of 1x lysis buffer. Solid pellets were first resuspended by slight agitation on a shaking table in a 4°C cold room for 30 min - 1 hour and then fully resuspended by pipette (careful resuspension takes time for bigger ribosome pellets). Resuspended samples were flash frozen in liquid nitrogen and stored at -80°C until subjected to acid phenol RNA extraction.

Affinity purification of ribosome-nascent chain complexes - Two kinds of beads were used to enrich ribosome-nascent chain complexes. For SeRP of *E. coli* Strep-Tactin Sepharose beads (IBA) were used, whereas for SeRP of human cells MagStrep "type3" XT Beads (IBA) were used (Table 13). For human SeRP (cell lines are indicated in Table 7), eight 70-80% confluent 15 cm^2 dishes were lysed and for the affinity purification of *LMNC-TS* 300 µl and for the affinity purification of *DCTN1-TS* 150 µl of bead slurry were used. For SeRP of Rosetta *E. coli* cells expressing Flag-SF-LMNC-TS from a plasmid 200 µl of bead slurry was used to pull from a lysate of 50 ml of *E. coli* culture (harvested as described above at OD_{600} = 0.5-0.6). In all cases, the indicated amount of bead slurry was transferred into 1.5 ml non-sticky tube and washed at least three times with wash buffer (50 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 150 mM KCl, 0.1% (v/v) Tween 20, 10% (v/v) glycerol, 1 mM DTT, 100 µg/ml cycloheximide or 1 mM chloramphenicol). Sepharose beads were centrifuged for 1 min at 600 xg, whereas magnetic beads were separated for 1 min at a magnetic stand. The washed beads were added simultaneously with 150U RNase I per 40 µg RNA to lysates from human cells and incubated for 30 min at 4°C on an overhead shaker (program C1 + 10 rpm). In the case of *E. coli* SeRP, lysates were mixed with $CaCl_2$ to reach a final concentration of 5 mM and were then digested with 150 U MNase per 40 µg RNA for 6 min at 25°C and 650 rpm on a thermomixer. The digestion was stopped by adding EGTA to a final concentration of 6 mM. The washed beads were then added and the affinity purification was performed for 45 min at room temperature on an overhead shaker (program C1 + 20 rpm).
The unbound solution (= called flow trough/FT) was kept on ice and the beads were washed 4 times with 1 ml 1x wash buffer. Each time the sample was inverted for 30 sec by hand and then transferred to a fresh non-sticky tube, the wash solution was kept on ice. Before taking samples for Western blot analysis, all flow through and wash solutions were centrifuged for 5 min at 20000 xg and 4°C to get rid of remaining beads. From the washed beads 1/10 was transferred for Western blot analysis. All remaining beads were flash frozen in liquid nitrogen after aspiration of all upper liquid and stored at -80°C for hot acid phenol extraction.

7.4 Standard molecular biology methods

PCR - The enzymatic amplification of specific DNA fragments for the integration into a plasmid was done with the Phusion polymerase (Thermo Fischer Scientific). All reaction components were mixed in PCR tubes (Sarstedt) on ice, then shortly mixed with a Vortex (Vortex Genie2, Scientific Industries) and centrifuged for 2 seconds (Mini-centrifuge, Sunlab). All samples were quickly transfer to a thermocycler (T Personal Thermocycler, Biometra) preheated to the denaturation temperature (98°C). The general PCR reaction mixture contained the following components in a 25 µl reaction volume:

50 – 100 ng	DNA Template
10 pmol	Forward primer
10 pmol	Reverse primer
10 mM	dNTP mix
1x	HF Phusion buffer (Thermo Fischer Scientific)
1U	Phusion polymerase (Thermo Fischer Scientific)

PCR program:

Initial denaturation:	98°C	2 min
Cycle (30x):	98°C	10 sec
	50-65°C	30 sec*
	72°C	0.5-2 min**
Final elongation:	72°C	5 min

* The annealing temperatures were chosen according to the primer T_m
** Elongation rate was estimated with 1 kbp per minute

DNA separation - All PCR reactions and plasmid digestions were analyzed on a 1% (w/v) agarose gel. The agarose powder (Sigma-Aldrich) was added to 1x TAE buffer (20 mM Tris, 1 mM EDTA, 20 mM acetic acid) and the whole mixture was heated up in a microwave for approximately 1 min at 900 W. In general, 1-0.5 μg/ml of Ethidium bromide (1% stock solution, Applichem) was added before the liquid gel was poured into a chamber to solidify with the appropriate comb. All sample for the DNA separation were mixed beforehand in a 1:6 ratio with 6x sample buffer (30% (v/v) glycerol, 1 mM EDTA, 0.25% (w/v) Bromphenol blue, 0.25% (w/v) xylene cyanol). The DNA was separated at 120 V in 1x TAE buffer for 30-45 min. The Gene ruler 1kb DNA ladder (Thermo Fisher Scientific) was used as marker. The separated DNA fragments were visualized with UV-light on an UV-table within the gel documentation system from Vilber Lourmat. The CHROMA 43 UV table (Vetter) was used to cut out single lanes with a scalpel for gel extractions.

Gel extraction - The GenElute™ Gel Extraction Kit (Sigma-Aldrich) was used according to the provider protocol to extract single bands from a 1% (w/v) agarose gel. The purified product was used directly for enzymatic restrictions or was stored at -20°C.

PCR clean up - The GenElute™ PCR Clean-Up (Sigma-Aldrich) was used according to the provider protocol to purify DNA from remaining salts and proteins from the PCR reaction mixture. The purified product was used directly for enzymatic restrictions, yeast transformations or was stored at -20°C.

Plasmid preparation - The GenElute HP Plasmid Miniprep Kit (Sigma-Aldrich) was used according to the provider protocol to isolate plasmids from a 4 ml *E. coli* overnight culture. The isolated plasmids were stored at -20°C.

DNA concentration determination - The NanoDrop 2000 (Thermo Fisher Scientific) was used to measure the absorbance of all DNA samples at 260 nm and to calculate the final nucleotide concentration (A_{260} of 1.0 = 50 μg/ml). Additionally, also the 260nm/280nm (DNA to protein ratio) and 260nm/230nm absorbance values (DNA to organic solvent/chaotropic salts ratio) were monitored to identify impurities.

Enzymatic restriction - Restriction enzymes were used in different buffer systems to generate sticky ends for the ligation of two specific DNA fragments (all enzymes and buffers were purchased from Thermo Fischer scientific or New England BioLabs). For double digests the appropriate buffer was chosen according to the Thermo Fischer Scientific or New England BioLabs double digest tables. All restrictions reactions contained at least 1000 ng of target DNA with 5-10 U of each restriction enzyme. Each digest was incubated for at least 1.5 h at 37°C and was afterwards inactivated for 20 min at 80°C.

Dephosphorylation - Single digested plasmids were dephosphorylated to avoid unspecific self-ligation. For this purpose, the restricted sample was incubated for 30 min at 37°C with 1 U

rAPID alkaline phosphatase (Roche) within 1x rAPID alkaline phosphatase buffer (Roche). The whole reaction was inactivated at 75°C for 2 min. All dephosphorylated samples were directly used for ligations or were stored at -20°C.

Ligation - The T4 DNA Ligase (Thermo Fischer Scientific) was used to connect the phosphate-sugar backbone of two restricted DNA fragments. In general, the ligation was set up with a 3 to 1 molar ratio of insert to vector with at least 100 ng of the respective vector. The reaction was done in 20 µl volume with 5 U Ligase in 1x Ligation buffer (50 mM Tris pH7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg/ml BSA) for 2 hours at room temperature or overnight at 4°C. All ligation mixtures were inactivated for 10 min at 65°C.

Competent E. coli cells - Competent *E. coli* cells (Table 7) *were prepared with the* Mix&Go Kit from Zymo Research. A dense overnight culture (grown at 37°C) was used to inoculation 50 ml LB medium (in a 250 ml flask) to OD_{600} = 0.05. This cell suspension was incubated at 30°C and 150 rpm for approximately 2.5 hours until the OD_{600} value reached 0.6. The whole cell suspension was transferred in a pre-cooled 50 ml Falcon tube and centrifuged for 10 min at 1800 xg (4°C). The supernatant was discarded and the cell pellet was resuspended on ice in 5 ml ice-cold 1x wash buffer (included in the Mix&Go Kit) with a cut 1 ml pipette tip (to reduce shearing forces). The cell suspension was again centrifuged for 10 min, 1800 xg at 4°C and the cell pellet was resuspended again on ice in 5 ml ice cold 1x competent buffer (included in the Mix&Go Kit) with a cut 1 ml pipette tip. These cells were aliquoted into 20 µl, 50 µl and 100 µl fractions into pre-cooled 1.5 ml reaction tubes on ice with a cut 100 µl pipette tip. All cells were directly stored at –80°C (no snap freezing in liquid nitrogen).

Heat shock transformation (E. coli) - One aliquot of competent *E. coli* cells was thawed for 10 min on ice. Competent XL1-blue *E. coli* cells were used for general molecular cloning approaches, whereas Rosetta cells were used to overexpress human proteins (Table 7). The amount and ratio of Plasmid and competent cells is critical for the efficacy. Re-transformations with purified plasmids were done with 20 µl of competent *E. coli* cells and 50-100 ng of purified plasmid. In the case of a transformation with heat inactivated ligation mixtures different amounts were chosen (50 µl competent cells + 1/3/5 µl Ligation mix). All tubes were slightly tapped by hand to mix the cell suspension and incubated for 20 min on ice, followed by a 35 - 45 second heat shock at 42°C on a thermoblock (Eppendorf). The sample was immediately transferred back on ice and incubated for 2 minutes. Afterwards 350 µl LB medium were added and incubated for 1-2 hours at 37°C and 550 rpm on a thermoblock (Thermomixer comfort, Eppendorf). The cell suspension was centrifuged for 3 min at 3000 xg (room temperature) and 250 µl of the supernatant was discarded. The cell pellet was resuspended in the remaining LB medium and transferred to LB plates with the correct antibiotics (pre-warmed to 37°C, Table 15). Glass beads were used to streak cells all over the plate and were incubated at 37°C overnight.

Colony PCR (E. coli) - A single colony was picked from the selection plate with a pipette tip and directly resuspended in the respective PCR mixture. The tip was then transferred into a test tube with 4 ml LB with the required antibiotics (Table 15) and incubated at 37°C in a rotation wheel. The PCR reaction was done with OptiTaq polymerase (Roboklon) and the provided 10x buffer B or C (Roboklon) or with an in-house purified Taq-Polymerase and the 1x Thermo Pol-buffer (20 mM Tris pH 8.8, 10 mM Ammonium sulfate, 10 mM KCl, 0.1% (v/v) Triton and 2 mM magnesium sulfate). All reaction components were mixed in PCR tubes (Sarstedt) on ice and quickly transfer to a thermo cycler (T Personal Thermocycler, Biometra) pre-heated to the denaturation temperature (94°C).

Plasmid sequencing - All generated plasmids were sequenced with own primers (Table 16) or primers provided by Eurofins/GATC. For sequencing plasmids, a 20 µl sample with a final concentration of approximately 50 ng/µl was prepared and send to GATC-Biotech/Eurofins.

Glycerol stocks (E. coli) - 700 µl of a dense overnight culture were mixed with 300 µl of 50% sterile glycerol within a screw cap tube (Biozym) by inverting the tube several times. All cells were frozen in liquid nitrogen before they were stored at -80°C. For reactivation of the glycerol stock a small amount of the frozen stock was scratched off with a micro loop (EcoLab) and transferred to LB plates supplemented with the respective antibiotics (pre-warmed to 37°C, Table 15). These plates were incubated at 37°C for at least 24 hours.

Site directed mutagenesis - Two kinds of site directed mutagenesis were used, in the one hand to substitute single nucleotides by designing a primer pair that incorporate the nucleotide exchange in the center, including at least 10 complementary nucleotides on both sites of the mismatch. On the other hand, primers were designed that flank a planned deletion region. For all site directed mutagenesis a gradient PCR was employed with Phusion polymerase with GC buffer. All PCR samples were combined, mixed and a fraction of it directly used for a combined PNK (Fermentas), Dpn 1 (New England BioLabs), T4 ligase (New England BioLabs) reaction in 1x PNK buffer A (50 mM Tris-HCL pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) supplemented with 1 mM ATP. This reaction mix was incubated for at least 1 hour at 22°C and then used for a normal heat shock transformation.

Detailed plasmid generation - All primer sequences used for cloning are available in Table 16 and all plasmids are listed in Table 10. As described in Bertolini *et al.* 2021: For DiSP experiments, *LMNA* residues 31-542, corresponding to lamin C lacking the unstructured head domain, was PCR-amplified from a self-made U2OS cDNA library (SuperScript III first-strand synthesis kit, Thermo Fisher Scientific). The employed PCR primers (MB143 + MB144) added a NdeI restriction site followed by a splitFlAsH tag (SF: MAGSCCGG) at the 5' end and a TwinStrep tag (TS: GGSGSAWSHPQFEKGGGSGGGS GGSAWSHPQFEKGA) with a BamHI overhang at the 3' end of the construct (final sequence named SFLMNCTS available in Table S4 of Bertolini *et al.* 2021). T4 DNA ligase was used to ligate the gel-purified PCR fragment into a BamHI/NdeI restricted pET3a vector. The resulting plasmid was sequenced with standard Eurofins primers (T7 forward and pET reverse primers) and custom primers (MB37, MB75, MB76 or MB78). It was found out that this plasmid had a frameshift right in front of the C-terminal twin strep tag. This plasmid was corrected by Jiri Koubek to generate a pET3a-FLAG-SF-LMNC-TS plasmid, which was then used for all SeRP experiments in *E. coli*.
This plasmid was further used as template for amplification of coil 1B (MB212 + *MB213*). The pET3a-SF-coil1B*-mcherry-TS plasmid was ordered via BioCat, with a SpeI and XhoI restriction site flanking the mutated coil 1B* sequence (SF_Coil1B_Mut_mCherry_TS, Table S4 of Bertolini *et al.* 2021). This plasmid was used to substitute the mutated coil 1B* sequence by the PCR amplified wild type coil 1B sequence via restriction and ligation (SF_Coil1B_WT_mCherry_TS, Table S4 of Bertolini *et al.* 2021). In addition, this plasmid was further used to substitute the mutated coil 1B* sequence by the PCR amplified wild type coil 2AB sequence (MB214 + MB215) via restriction and ligation. The resulting plasmid was sequenced with standard Eurofins primers (T7 forward and pET reverse primers) and custom primers that bind to the mCherry sequence (MB216 + MB217).
The SF-LMNC-TS plasmid was also used for the coil 1A deletion generated by site directed mutagenesis with primers flanking coil 1A for the deletion (MB184 + MB189). Positive colonies were identified by a band shift after colony PCR with MB36 and MB108.
DCTN1 was PCR amplified (MB209 + MB210) from a pENTR221-*DCTN1* ($p150^{glued}$) plasmid ordered from the DKFZ vector and clone repository. Gibson assembly *(Gibson et al. 2009)*

was used to transfer the PCR amplified *DCTN1* sequence from the ordered plasmid into a pET3d-vector, flanked by an N-terminal splitFlAsH tag and a C-terminal TwinStrep tag (MB205 + MB206). The resulting plasmid (SF_DCTN1_TS, Table S4 of Bertolini *et al.* 2021) was sequenced with standard Eurofins primers (M13 forward and re-verse primers) and custom primers (MB197 + MB198).
Plasmids used for the dimerization assay were generated by PCR amplification of coil1A (MB159 + MB160), coil1B (MB161 + MB162), coil1A1B (MB159+ MB162), coil2AB (MB163 + MB164), coil1A-1B-2AB (MB159 + MB164) and head-domain-coil1A-1B-2AB (MB158 + MB164), each flanked by homologous regions to the target vector, with the SF-LMNC-TS plasmid as template. Gibson assembly was used to clone each fragment into a SalI/BamHI digest pJH391-TS plasmid containing a C-terminal TwinStrep tag. The resulting plasmids were sequenced with custom primers (#1229 + #1230).

SDS-PAGE protein separation - Precast 12% ExpressPlus PAGE gels with 12 or 15 wells (GeneScript) were used for most protein separations. In some cases, bigger gels with 18 or 24 wells were used from Bio-Rad (PROTEAN® TGX™ precast protein gels), each with the provider's recommended 1x SDS running buffer. Bradford protein measurements were used to determine the total protein amount and to then load of most samples 10 μg or 20 μg per pocket. All samples were heated up to 99°C for 10 min at 750 rpm on a Thermomixer comfort (Eppendorf) mixed 1:4 with 4x SDS sample buffer (240 mM Tris pH 6.8, 40% (v/v) glycerol, 8% (w/v) SDS, 0.04% (v/v) bromophenol blue, 5% (v/v) 2-mercaptoethanol). The PageRuler Prestained Protein ladder (Fermentas) was used as protein marker. As running buffer 1x MOPS buffer (provided by GeneScript, 0.05 M Tris, 0.05 M MOPS, 1% (w/v) SDS, 1 mM EDTA) or 1x Tris/glycine/SDS (provided by Bio-Rad, 25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3) was used. The protein separation was done in general for 1.5 hours at 140 V (Power PAC1000, Bio-Rad) in a MiniProtean TetraSystem chamber (Bio-Rad). After the separation all gels were subsequently used for transfer to a membrane or Coomassie staining.

Semi-dry Transfer - To assemble the transfer stack four layers of Whatman paper (soaked in transfer buffer) were transferred on the cathode–side (+ pol). The next layer contains the 0.45 μm PVDF membrane (Amersham Hybond), which was activated in 100% methanol for 5 min before use. The SDS gel was transferred on top of the PVDF membrane and finally 4 layers of Whatman paper (soaked in transfer buffer) were added on top. The transfer was done with the Trans-Blot Turbo transfer system (Bio-Rad) or the Pierce Power station transfer system (Thermo Fisher Scientific) in 1x Trans-Blot Turbo transfer buffer (Bio-Rad) at 1.3A, 25 V for 10 min.

Wet Transfer - The transfer stack was assembled in the same way as described in the section before. The transfer was done within a Transfer chamber in which the assembled stack was completely submerged in 1x Wet Transfer Buffer (0.2 M glycine, 0.025 M Tris, 10% (v/v) ethanol). The transfer was done for 2 hours at 100 V at 4°C.

Coomassie staining - To stain proteins in an SDS PAGE or to control the transfer efficacy all SDS gels were stained for one hour in a 1:1 dilution of Instant blue solution (Expedeon) with water on a shaking table. Afterwards all gels were destained in water over night on the shaking table.

Immunoblot – All membranes were blocked after the transfer for at least 1 hour at room temperature in 1x TBST supplemented with 5% (w/v) skim milk powder (Fluka) or 5% (w/v) BSA on a shaking table. Afterwards all membranes were transferred into 50 ml Falcon tubes (Sarstedt) with 10 ml 5% (w/v) BSA dissolved in 1x TBST and a dilution of the respective antibody (see Table 6). The incubation with the primary antibody was usually done overnight

at 4°C on a rotation or shaking table or if required at room temperature for at least 1 hour. All antibody dilutions in 1x TBST and 5% (w/v) BSA were stored at -20°C and reused up to 10 times. On the next day all membranes were washed three times with 1x TBST (~10 min incubation on shaking table). Afterwards 10 ml 5% (w/v) BSA dissolved in 1x TBST were added with 1:20.000 dilution of alkaline phosphatase anti-mouse IgG (made in horse) or anti-rabbit (made in goat) secondary antibody (Vector Laboratories, Table 6). The secondary antibody treatment was done for at least 1 hour at room temperature on a shaking table. Afterwards all blots were washed three times with 1x TBST (approximately 10 min incubation on shaking table) and the membranes were transferred into a transparent foil. To each membrane 1.5 ml of 1x ECF solution (GE Healthcare) diluted with water was added and incubated for 10 min at room temperature. The detection was done with the Luminescent Image Analyzer LAS 4000. The Image Reader Las-4000 software was used with the following settings:

Method:	Fluorescence, GFP/515 nm/2.8
Tray Position:	1-3
Filter:	Y515D
Iris:	F2.8
Exposure type:	Precision
Sensitivity:	Standard or High resolution

The chosen exposure time varied between 1/30 up to 20 seconds per membrane with these settings to obtain good exposed pictures without saturation (the software automatically indicated saturated samples with a red bar).

Antibody stripping **-** To reuse the same membrane with another primary antibody a purchased antibody stripping solution named "Restore Western Blot Stripping-Puffer" (Thermo Fisher Scientific) was used according to the manufacture's manual.
The following protocol was used if it was necessary to harshly strip away any traces of previously bound antibodies. The membrane was washed two times with 1x TBST (10 min on shaking table) after the visualization with ECF solution. For harsh antibody stripping a self-made stripping buffer (125 mM Tris, 85 mM 2-mercaptoethanol, 2% (w/v) SDS) was added into a plastic box so that the membrane was slightly covered by the solution. The box was sealed with Parafilm® and incubated for 30 min at 50°C in a water bath. The stripping buffer was discarded into the organic solvent waste and the membrane was washed two times with 1x TBST (5 min on shaking table). Afterwards the membrane was blocked again with 5% (w/v) skim milk powder (Fluka) dissolved in 1x TBST for 2 hours on the shaking table.

7.5 Other methods

Dimerization Assay **-** As described in Bertolini *et al.* 2021: This protocol is based on (Schaefer *et al.* 2016) and aims to combine (i) OD_{600} measurement, (ii) cell permeabilization, (iii) ONPG breakdown, and (iv) kinetic OD_{420} quantification into a single step. The required FI8202 *E. coli* strain [ΔntrBCfadAB101::Tn10 laqIq lacL8/λ202] (Fiedler *et al.* 1995, Table 7) has a lac repressor (lac1q) deletion, therefore it is galactosidase positive. Strains transformed with a plasmid expressing an active dimerization domain fused to the N-terminal part of the lambda repressor (residues 1 to 102 of λ repressor) will have reduced galactosidase activity. The pKH101 plasmid (expressing only N-terminal part of the lambda repressor) (Hu *et al.* 1990) was used as negative control, and pFG157 (expressing the full-length lambda repressor) as positive control (Hu *et al.* 1990). Freshly transformed FI8200 colonies were picked from LB plates for overnight cultures in LB media. 80 µl of each overnight culture were transferred into a 96-well Greiner flat bottom microplate (transparent), 120 µl freshly prepared

master-mix (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 36 mM β-mercaptoethanol, 6.70% (v/v) PopCulture Reagent, 1.1 mg/ml ONPG, Lysozyme) were quickly added and the measurement started using SPECTROstar Nano Microplate Reader (program: OD_{600} and OD_{420} readings taken every 60 sec for 1 hour, at room temperature, shook at 500 rpm (double orbital shaking) for 30 seconds before each cycle). The linear slope of OD_{420} over time (OD_{420}/min) was multiplied by 5000, and adjusted for the OD_{600} reading at the first time point (defined as Miller units). OD_{600} was assumed to be constant, since lysis of cells had only minor effect on the OD_{600} values over the measured time period. Repression efficiencies were calculated as in (Hu *et al*. 1990).
To be able to compare the repression between different fusion constructs it is necessary to obtain equal steady state protein levels. Therefore SDS-PAGE samples were isolated from all used overnight cultures and analyzed by immunostaining. In the case of too low protein steady state levels, the experiment was repeated with minor amounts of IPTG in the growth media (5 -100 μM) of an overnight culture, to obtain similar expression levels.

Affinity purification of TwinStrep tagged lamin C - One confluent T75 flask of HEK293-T cells (wild type or heterozygous $LMNC^{WT/TS}$, Table 7) was harvested by trypsinization and washed in 1x PBS. Cells were pelleted at 1000 xg for 3 min and then resuspended in 0.5 ml hypotonic buffer (10 mM HEPES pH 7, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM EDTA, 0.05% (v/v) NP-40). Cell suspensions were incubated for 15 min on ice to allow cells to lyse (samples were vortex occasionally). Nuclei were pelleted at 3300 ×g for 10 min at 4°C and washed in 0.5 ml hypotonic buffer and pelleted again. Nuclei were lysed in 200 µl lamin extraction buffer (25 mM Tris pH 8.6, 1% (v/v) NP-40, 0.5% (w/v) DOC, 0.1% (v/v) SDS, 500 mM NaCl, 1 µl Benzonase (Millipore), EDTA Free protease inhibitor tablet (Roche)). Nuclear lysates were incubated for 15 min on ice with occasional vortexing and cleared by centrifugation for 10 min at 20000 ×g. Each cleared lysate was subjected to affinity purification with 40 µl MagStrep "type3" XT beads (iba, Table 13)) according to provider's instructions. Elution was performed by incubating beads with 20 µl lamin extraction buffer supplemented with 1x Buffer BXT (iba) for at least 10 min at room temperature. Input, flow-through and elution samples were analyzed by Western blotting using anti-Lamin A/C antibody (Table 6).

Disome footprint ligation - HEK293-T cells were lysed as described for DiSP. In one case a normal RNase treatment with 150 U RNase I per 40 μg RNA was conducted for 15 or 30 min at 4°C. In another sample set 75 U RNase I per 40 μg RNA was employed for 15 or 30 min at 4°C, to obtain only a partial RNA digestion. All RNA digested samples were loaded on a 5-45% (w/v) sucrose gradient and monosome as well as disome fractions were isolated, as described for the standard DiSP approach (see above). Based on the measured A_{260} of all isolated fractions, the amount corresponding to approximately 10 pmol of ribosomes were used for a sucrose cushion (1.5 hours centrifugation at 100 000 rpm 4°C in a TLA100.2 rotor). Each ribosome pellet was diluted in 100 μl ligation buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.1 mM GTP, 1 mM $MnCl_2$, 15% (v/v) PEG8000). Each sample was splitted in two 50 μl fractions and only one half was mixed with 1 μl RtcB ligase. All samples were incubated for 3.5 h at 37°C. Afterward all samples were employed for a normal hot acid phenol extraction, followed by 60 nt footprint isolation.

7.6 Data analysis and visualization

Processing of raw sequencing data - All employed tools for data analysis are listed in Table 8. U20S samples were sequenced with a HiSeq 2000 sequencer and contained therefore different cDNA library linkers. The raw data processing commands can be found described in great detail in Döring *et al.* 2017.

All HEK293-T samples were sequenced with a NextSeq550 sequencer (single end sequencing with 51 sequencing cycles for "Read1"). Libraries that were not processed according to our in-house cDNA library preparation (based on McGlincy *et al.* 2017, Galmozzi *et al.* 2019) are labeled in Table 19. The raw data processing commands for all in-house library prepared samples can be found described in great detail in Bertolini *et al.* 2021. In brief, the raw sequencing data were demultiplexed with bcl2fastq according to the i7 index that was sequenced with an independent primer in addition to the read 1 primer.
For all libraries prepared with our in-house library preparation the output fastq files were first trimmed with cutadapt to remove the L1 in-house 3' adaptor sequence ("ATCGTAGATCGGAAGAG-CACACGTCTGAACTCCAGTCAC"), the employed sequence removed also the 5 nt barcode that belongs to the linker L1 specific barcode ("ATCGT").

Structure of "Read1" sequenced with a NextSeq550

NN-(FFFFFFFFFF)n-NNNNN-BBBBB-AA(AAA)n

N = random nucleotides implemented to prevent ligation biases, 2 nt at 5'- and 5 nt at 3'-end
F = ribosome protected mRNA footprint (for one ribosome around 30 nt long)
B = 5 nt barcode form linker L1 ("ATCGT")
A = 3' linker adapter sequence

Libraries that were prepared with the TAKARA SMARTer smRNA-Seq Kit for Illumina or with the NEXTflex small RNA-Seq Kit v3 had their own read1 structure and were processed with cutadapt according to the user manuals.
The output from cutadapt for samples of our in-house library preparation were further processed to generate a unique molecular identifier (UMI) from each read using a custom Julia script (Script 1, see Supplementary Materials; Bertolini *et al*. 2021). Each UMI consists of the five random 3' and two random 5' nucleotides implemented in the library preparation to prevent ligation biases. The output of this custom Julia script is another fastq file, which contains the UMI in the read name and trimmed these nucleotides away from all read1 sequences. This fastq file was then aligned to human or *E. coli* rRNA sequences with bowtie2 and reads that did not align to rRNA were employed for human genome alignments (GRCh38p10) or *E. coli* BL21 (DE3) genome alignments (GCA_000022665.2, modified to include additional chromosomes consisting of plasmid-encoded gene sequences, see Table S5 for sequences and respective gene names in Bertolini *et al*. 2021 or here in Supplemental Material) using STAR. For each gene, the transcript with the longest coding sequence was selected and reads were P-site assigned via a custom Julia script, this script also filtered out all reads that had the same UMI at one alignment position, thereby excluding read accumulations caused by PCR amplification biases (Script 2, see Supplementary Materials; Bertolini *et al*. 2021). This output of this script was a HDF5 file, which contains one data frame per detected gene and certain attributes. For each gene the gene name as well as protein names, the transcript ID, the coding sequence length and the chromosome and strand position are saved. The gene-specific data frame consists of a 2-row matrix, the first row contains each nucleotide position within the CDS (starting from position 1), and the second row the number of detected unique P-site aligned reads at each nucleotide position. This file was the input for all further data analysis and visualizations with RiboSeqTools (https://github.com/ilia-kats/RiboSeqTools) and for custom R or python scripts (see Supplement Material).

Single mRNA profiles - All single mRNA profiles and enrichment plots were visualized with the "RiboSeqTools" R package (https://github.com/ilia-kats/RiboSeqTools). The input file contained always P-site assigned reads exclusively aligned to the coding sequence of the longest transcript of each gene, no UMI duplicates were allowed and therefore only unique

reads are plotted. For single mRNA plots the position-wise 95% Poisson confidence interval was plotted as vertical bars for each codon of a single mRNA. The plotted bars are corrected for library size and smoothed with a sliding window of 15 codons (more details are given in Bertolini *et al*. 2021). For DiSP data sets the disome over monosome ratio was plotted, whereas for SeRP data sets the affinity purified translatome to total translatome ratio. The intensity of each vertical bar indicates the number of aligned reads (faint bars represent 1-10 reads per sliding window position, whereas dark bars represent more than 30 reads per sliding window position).

Metagene profiles - For all human DiSP 30 nt metagene profiles only genes with a summed coverage (monosome + disome raw counts in two replicates) higher than 0.5 reads per codon were included. The here employed read normalization with RiboSeqTools used additional total translatome data to divide the monosome and disome read densities of each gene by the read density of the total translatome. Thereby all metagenes were calculated based on position-wise read density normalized to their expression level. The plotted metagene indicates a solid line, which resembles the average read density along all detected mRNAs. The vertical bars in the background indicate the 95% confidence interval of 100 bootstrap samples (more details are given in Bertolini *et al*. 2021). Profiles were calculated each time separately for individual replicates. Metagenes were not smoothed if not explicitly indicated in the figure legend.

Sigmoidal fitting - The sigmoidal fitting tool was developed by Ilia Kats and is described in great detail in Bertolini *et al*. 2021 (Script 3, see Supplementary Materials; Bertolini *et al*. 2021). In brief, three different models were pre-defined to be fitted to a single gene disome over monosome enrichment profile: (i) a flat line without any enrichment, (ii) a single sigmoidal enrichment or (iii) a double sigmoidal enrichment. For each single gene enrichment profile, different parameters were estimated by maximum likelihood, and the best model was selected using the Bayesian Information Criterion (described in great detail in Bertolini *et al*. 2021). All genes with no enrichment were classified as "non-co-co candidates", whereas single and double sigmoidal genes were classified as possible "co-co candidates". The onset of assembly was determined based on the inflection point of the sigmoid curve, for double sigmoidal curves the first inflection point was chosen as onset of assembly.

Protein-localization and -complex annotation - The annotation of the oligomeric state and localization of all detected human proteins was done in close collaboration with Frank Tippman (Bertolini *et al.* 2021). In brief, four different databases were used to retrieve the oligomeric state of a protein, including UniprotKB (The UniProt Consortium 2019), PDB (Berman *et al.* 2000), Corum (Giurgiu *et al.* 2019), Swissmodel (Waterhouse *et al.* 2018). The oligomeric state was retrieved from human, mouse and rat proteins to obtain information for as many proteins as possible (in case of multiple annotations the human annotations were always ranked before mouse and rat annotations). Several proteins had different oligomeric state annotations according to the four databases. Therefore, a hierarchical selection criterion was implemented to avoid multiple annotations for single proteins (ranking order: UniprotKB > PDB > Corum > Swissmodel). In addition, multiple oligomeric state annotations were also observed within a single database. Therefore, the following ranking order was applied: “homomer” > “heteromer” > “monomer”. Proteins annotated as "heteromer of homomers" were excluded to avoid noise from subunits whose assembly partner is uncertain.
Four different databases were used to obtain annotations of the subcellular localization of human proteins, including the Human Proteome Atlas (Thul *et al.* 2017), UniProtKB (The UniProt Consortium 2019), LOCATE (Sprenger *et al.* 2008), and the benchmark dataset of iLoc-Euk (Chou *et al.* 2011). The subcellular localization was retrieved in all cases possible for human proteins, and for mouse and rat homologs if no annotation was available for the human protein. All proteins were further classified as ‘cyto-nuclear’ if at least one of the

following keywords was present in the final localization file: 'cytosol', 'nucleoplasm', 'nucleus', 'cytoplasm', 'nucleoli', 'nucleolus', 'perinuclear region of cytoplasm'. In addition, it was necessary for all classified ‘cyto-nuclear’ proteins that no TMD annotation was present in the UniprotKB retrieved data.
The annotations for *E. coli* proteins were directly retrieved from the EcoCyc webpage (Keseler *et al.* 2017).

High- and low-confidence classification - All single and double sigmoidal fitted gene profiles were further ranked into a high confidence or low confidence class. The intimal classification of a co-co candidates and the onset of assembly was determined based on the HEK293-T high salt DiSP experiment. To be classified as high confidence candidate two criteria must be fulfilled. (i) It was necessary that the sigmoidal disome enrichment is significantly decreased after PK and puromycin treatment (both control experiments were compared to their corresponding "untreated control", see Bertolini *et al.* 2021). To determine the significance of the decrease for the PK treated samples it was necessary to employ a "dose-response model" for all PK conditions (explained in detail in Bertolini *et al.* 2021). In brief, all read counts were summed up for each single co-co candidate before and after the determined onset of co-co assembly (termed response), in all "PK treated" and the "untreated" data sets (termed dose), respectively. The corresponding dose-response plot indicated a non-linear loss of the disome/monosome ratio after onset of co-co-assembly with increasing PK concentrations. To employ a linear fitting model it was required to include an optimized value X, instead of actual PK concentration, which forces a linear relationship between loss of sigmoidal enrichment and PK concentration (developed by Ilia Kats, see Bertolini *et al.* 2021). This allowed to fit each single gene enrichment plot to a beta-binomial generalized linear model (GLM). A loss of sigmoidal enrichment was defined as those single gene profiles, which showed a negative coefficent of the interaction term at an FDR ≤ 0.01. Similar to the PK samples also the effect of puromycin on the disome/monosome ratio after onset of co-co-assembly was determined. (ii) All proteins, which were not classified as "cyto-nuclear" (see section above) were automatically classified as low confidence candidates.
All single gene profiles that remained their sigmoidal enrichment after one of both treatments or are classified as membrane or translocated protein were ranked as low confidence candidates. All single gene profiles that showed no change of the sigmoidal footprint shift by both treatments were classified as "non-co-co candidates", even though an sigmoidal enrichment was detected.

Crystal structure analysis - This has been done by Frank Tippman and the method is published in Bertolini *et al.* 2021. In brief, all x-ray structures of human proteins were downloaded from the Protein Data Bank (PDB). Only structures with the highest sequence coverage (at least more than 10 amino acids) and highest resolution were chosen. The interface analysis for exclusive homomeric and exclusive heteromeric proteins was performed as described in Natan *et al.* 2018. The plotted interface enrichment represents the normalization of each data point by the arithmetic mean of all data points. The same analysis was repeated for proteins of the high confidence co-co assembly class to calculate the onset-aligned interface enrichment. Therefore, the onsets of co-co assembly for each candidate was set to position zero and only interfaces located in a window of 500 amino acids around the onset were considered (250 amino acids up- and 250 amino acids downstream). Analysis of heteromeric subunits was limited to subunits interfaces, where at least one subunit was enclosed in the high confidence list.

Domain enrichment calculation - This has been done in collaboration with Matilde Bertolini and the method is published in Bertolini *et al.* 2021. In brief, all annotations of protein domains and the respective positions in protein sequences was retrieved from UniprotKB. A domain

was classified as fully ribosome exposed before onset of assembly, if the position of the last domain residue was 30 residues away from the determined onset of assembly (to account for the length of the ribosomal exit tunnel). To obtain a fair analysis of domain enrichment in the high confidence class it was necessary to account for different factors. The comparison of exposed domains at assembly onset to all annotated domains in the human proteome would lead to an artificial enrichment of N-terminal protein domains (since co-co assembly occurs mostly on the N-terminal half). Therefore, all non-co-co classified proteins received an randomized assembly onsets (see below). This defined an appropriate "background" for the enrichment calculation with a similar distribution of rather N-terminal domains for the human "non-co-co" proteome. The high confidence class consists of only cytosolic or nuclear proteins, therefore also the "background" was restricted to the same localization.
The final domain enrichment calculation consists of the following steps. First the fraction of candidates exposing a domain of interest was determined in the high confidence list. Then a background sample of the same size as the high confidence class (829 candidates) was drawn from all “cyto-nuclear” proteins from the "non-co-co" proteome with randomly assigned onsets and the fraction of proteins exposing the same domain of interest was determined. This step was repeated 10^5 times for each analyzed domain. Finally, the fraction of the high confidence list exposing a domain of interest was divided by the median of the fraction of the background list exposing the same domain (see Bertolini et al. 2021). The used R code can be found at our DataKramer Server, see Supplemental Material.

Homomer, heteromer and monomer enrichment - The “frequency enrichment” was calculated as the frequency of proteins annotated as monomers, homomers or heteromers in the low- or high confidence class divided by their frequency in their respective background proteome. The high confidence class includes exclusively “cyto-nuclear” proteins and were therefore compared to all “cyto-nuclear” proteins as background. The frequency in the low confidence class was instead calculated with all human proteins as background. The subset of proteins detected by DiSP and included in high- and low confidence classes are biased towards highly expressed genes. The "goseq" R package was used to perform a bias-corrected significance calculation of all frequency enrichments. This was necessary because our high- and low confidence lists showed a bias towards highly expressed genes (the used R code can be found at our DataKramer Server, see Supplemental Material).

Random onset determination - It was important to assign an onset to candidates without detectable co-co assembly to compare specific features of the co-co candidates with an assigned onset to non-co-co candidates. To achieve a fair comparison, it was important to not randomly assign onsets that are not within the boundaries of the observed onsets of the respective protein class (e.g. onset are not observed for the first translated amino acid of a protein nor the last amino acid). Such a semi-randomized onset was determined based on the linear regression fitted to the correlation of log(CDS length) and log(onset) for the respective co-co assembly class (explained in great detail in Bertolini *et al*. 2021). The semi-random onset generator is based on the "truncnorm" R package to determine an onset for a specific CDS length (the used R code can be found at our DataKramer Server, see Supplemental Material).

Monosome depletion - To calculate the monosome depletion for each gene, it was necessary to first determine the number of monosome reads before and after the onset of assembly for all high- and low-confidence candidates. A mild ribosome density loss towards the C-terminus was observed in total translatome data. Therefore, all monosome read counts were normalized to total translatome read counts to quantify the depletion of monosomes after onset of co-co assembly. The following equation was used to calculate the final monosome depletion after onset of assembly per gene: (monosome reads before onset/total translatome

reads before onset) / (mono reads after onset/total translatome reads after onset). This value was subtracted from 1 to obtain the reduction of monosomes after onset of assembly, to obtain the percentage of monosome depletion this value was multiplied by 100 (the used R code can be found at our DataKramer Server, see Supplemental Material).

Alpha-helical and coiled coil predictions - Coiled coil and alpha-helical prediction were based on the amino acid sequence of lamin C coil 1B, coil 2AB (coil segment positions and amino acid sequences were retrieved from UniProt) and the mutated coil 1B*. The coil prediction was performed with "https://embnet.vital-it.ch/software/COILS_form.html" (Lupas *et al.* 1991) and the alpha-helix prediction with "https://predictprotein.org" (Bernhofer *et al.* 2021).

Co-co vs non-co-co TMD analysis - All TMD containing proteins of the low confidence list with a single exposed TMD before the onset of assembly were classified as "co-co TMDs" (258 candidates). Candidates that exposed multiple TMDs before onset were excluded, because the close proximity of TMDs does not allow to determine which TMD leads to the onset of co-co assembly. All TMD containing proteins that were classified as non-co-co assembling proteins were used to randomly assign an onset of assembly (see "Random onset determination"). The "non-co-co TMD" class included all candidates that exposed a single TMD before the assigned onset of assembly. First the hydrophobicity of both classes of TMD domains were analyzed, by employing the Kyte-Doolittle scale. A Welch two sample t-test was employed to calculate the significance between the determined mean hydrophobicity.
To generate amino acid frequency plots the "logomaker" function of the "Logolas" R package was employed. The difference in the apparent free energy prediction (ΔG_{app}) for insertion of a putative TMD sequence into the ER membrane by the Sec61 translocon (Hessa *et al.* 2007) was calculated on the following webpage "https://dgpred.cbr.su.se/index.php?p=TMpred".

Sucrose gradient simulation algorithm - The used coefficients for the calculation that indicate the relationship between the viscosity of each sucrose solution and temperature are only valid for 0 - 1 mol/L sucrose solutions (0 - 34.2% (v/v) sucrose solutions) (Dingman 1972). The here tested range of sucrose gradients takes 5% to 45% sucrose solutions into account (0.146 - 1.314 g/mol), which exceeds the 1 mol/L sucrose limit. The abbreviation from the 1 mol/L sucrose solution resulted in a curved line in the predicted running behavior (**Fig. 40**), however the predicted sedimentation behavior could be proven valid for a 5-45% gradient (**Fig. 41**). The commented python script for the prediction calculation can be found on our Data Kramer server (see Supplemental Material), which also indicates the fixed parameters required for the calculations (e.g. acceleration and deacceleration times of the used ultracentrifuge, dimensions of the used SW40 tubes in combination with the used short lids, SW40Ti rotor specifications, assumed ribosome density and measured sucrose densities). In general, the script can be easily adaptable to all rotor and tube dimensions, initial and final solvent concentrations, temperatures (from 0 - 20°C) and particle densities. These calculations are only valid for sucrose gradients containing non-significant amounts of salt, high salt gradients require a precise determination of the solute densities, since it is known that a 0.1% error in the assumed density results in an approximate 0.2% error in the calculated sedimentation behavior (Dingman 1972). For using the output of the predictions, it is important to keep in mind that the sucrose gradient fractionator can only enter the gradient tube till 8.1 cm and not the full end of the tube, which is 8.4 cm.

8. References

Almén, M. S., Nordström, K. J. V., Fredriksson, R., & Schiöth, H. B. (2009). Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*, *7*, 50. https://doi.org/10.1186/1741-7007-7-50

Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., … Gevaert, K. (2009). Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(20), 8157–8162. https://doi.org/10.1073/pnas.0901931106

Arpat, A. B., Liechti, A., de Matos, M., Dreos, R., Janich, P., & Gatfield, D. (2019). Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing. *BioRxiv*. https://doi.org/10.1101/710061

Aviram, N., Ast, T., Costa, E. A., Arakel, E. C., Chuartzman, S. G., Jan, C. H., … Schuldiner, M. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature*, *540*(7631), 134–138. https://doi.org/10.1038/nature20169

Bañó-Polo, M., Baeza-Delgado, C., Tamborero, S., Hazel, A., Grau, B., Nilsson, I. M., … Mingarro, I. (2018). Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07554-7

Becker, A. H., Oh, E., Weissman, J. S., Kramer, G., & Bukau, B. (2013). Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nature Protocols*, *8*(11), 2212–2239. https://doi.org/10.1038/nprot.2013.133

Behnke, J., Feige, M. J., & Hendershot, L. M. (2015). BiP and Its Nucleotide Exchange Factors Grp170 and Sil1: Mechanisms of Action and Biological Functions. *Journal of Molecular Biology*, *427*(7), 1589–1608. https://doi.org/10.1016/j.jmb.2015.02.011

Bengtson, M. H., & Joazeiro, C. A. P. (2010). Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature*, *467*(7314), 470–473. https://doi.org/10.1038/nature09371

Bergman, L. W., & Kuehl, W. M. (1979). Co-translational modification of nascent immunoglobulin heavy and light chains. *Journal of Supramolecular Structure*, *11*(1), 9–24. https://doi.org/10.1002/jss.400110103

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bernhofer, M., Dallago, C., Karl, T., Satagopam, V., Heinzinger, M., Littmann, M., … Rost, B. (2021). PredictProtein – Predicting Protein Structure and Function for 29 Years. *BioRxiv*, 2021.02.23.432527. https://doi.org/10.1101/2021.02.23.432527

Bertolini, M., Fenzl, K., Kats, I., Wruck, F., Tippmann, F., Schmitt, J., … Kramer, G. (2021). Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science*, *371*(6524), 57–64. https://doi.org/10.1126/science.abc7151

Bezanson, J., Karpinski, S., Shah, V. B., & Edelman, A. (2012). Julia: A Fast Dynamic Language for Technical Computing, 1–27. Retrieved from http://arxiv.org/abs/1209.5145

Bhushan, S., Gartmann, M., Halic, M., Armache, J. P., Jarasch, A., Mielke, T., … Beckmann, R. (2010). α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nature Structural and Molecular Biology*, *17*(3), 313–317.

https://doi.org/10.1038/nsmb.1756

Blobel, G., & Sabatini, D. (1971). Dissociation of Mammalian Polyribosomes into Subunits by Puromycin I7 . I ,' I ' ', *68*(2), 390–394.

Boulay, F., Doms, R. W., Webster, R. G., & Helenius, A. (1988). Posttranslational oligomerization and cooperative acid activation of mixed influenza hemagglutinin trimers. *Journal of Cell Biology*, *106*(3), 629–639. https://doi.org/10.1083/jcb.106.3.629

Brandman, O., & Hegde, R. S. (2016). Ribosome-associated protein quality control. *Nature Structural & Molecular Biology*, *23*(1), 7–15. https://doi.org/10.1038/nsmb.3147

Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C. C., Li, G. W., … Weissman, J. S. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell*, *151*(5), 1042–1054. https://doi.org/10.1016/j.cell.2012.10.044

Brandt, F., Carlson, L., Hartl, F. U., Baumeister, W., & Gru, K. (2010). Article The Three-Dimensional Organization of Polyribosomes in Intact Human Cells. https://doi.org/10.1016/j.molcel.2010.08.003

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., … Komar, A. A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Molecular Cell*, *61*(3), 341–351. https://doi.org/10.1016/j.molcel.2016.01.008

Bulleid, N. J., & Freedman, R. B. (1988). Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. *Nature*, *335*(6191), 649–651. https://doi.org/10.1038/335649a0

Cabantous, S., Terwilliger, T. C., & Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnology*, *23*(1), 102–107. https://doi.org/10.1038/nbt1044

Calvo, S. E., & Mootha, V. K. (2010). The Mitochondrial Proteome and Human Disease. *Annual Review of Genomics and Human Genetics*, *11*(1), 25–44. https://doi.org/10.1146/annurev-genom-082509-141720

Chang, L., Shav-Tal, Y., Trcek, T., Singer, R. H., & Goldman, R. D. (2006). Assembling an intermediate filament network by dynamic cotranslation. *Journal of Cell Biology*, *172*(5), 747–758. https://doi.org/10.1083/jcb.200511033

Chapman-Smith, A., Forbes, B. E., Wallace, J. C., & Cronan, J. E. (1997). Covalent modification of an exposed surface turn alters the global conformation of the biotin carrier domain of Escherichia coli acetyl-CoA carboxylase. *Journal of Biological Chemistry*, *272*(41), 26017–26022. https://doi.org/10.1074/jbc.272.41.26017

Chen, G., Mulla, W. A., Kucharavy, A., Tsai, H. J., Rubinstein, B., Conkright, J., … Li, R. (2015). Targeting the adaptability of heterogeneous aneuploids. *Cell*, *160*(4), 771–784. https://doi.org/10.1016/j.cell.2015.01.026

Chen, J., Petrov, A., Johansson, M., Tsai, A., O'Leary, S. E., & Puglisi, J. D. (2014). Dynamic pathways of −1 translational frameshifting. *Nature*, *512*(7514), 328–332. https://doi.org/10.1038/nature13428

Chou, K.-C., Wu, Z.-C., & Xiao, X. (2011). iLoc-Euk: A Multi-Label Classifier for Predicting the Subcellular Localization of Singleplex and Multiplex Eukaryotic Proteins. *PLoS ONE*, *6*(3), e18258. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3068162/

Ciryam, P., Morimoto, R. I., Vendruscolo, M., Dobson, C. M., & O'Brien, E. P. (2013). In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proceedings of the National Academy of Sciences of the United*

States of America, *110*(2), 396–397. https://doi.org/10.1073/pnas.1213624110

Collart, M. A., & Weiss, B. (2020). Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Research*, *48*(3), 1043–1055. https://doi.org/10.1093/nar/gkz763

Defenouillère, Q., Yao, Y., Mouaikel, J., Namane, A., Galopier, A., Decourty, L., … Fromont-Racine, M. (2013). Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(13), 5046–5051. https://doi.org/10.1073/pnas.1221724110

del Alamo, M., Hogan, D. J., Pechmann, S., Albanese, V., Brown, P. O., & Frydman, J. (2011). Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biology*, *9*(7). https://doi.org/10.1371/journal.pbio.1001100

de Leeuw, R., Gruenbaum, Y., & Medalia, O. (2018). Nuclear Lamins: Thin Filaments with Major Functions. *Trends in Cell Biology*, *28*(1), 34–45. https://doi.org/10.1016/j.tcb.2017.08.004

Deng, J. M., & Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Research*, *4*(4), 264–269. https://doi.org/10.1007/BF01969120

Deng, S., McTiernan, N., Wei, X., Arnesen, T., & Marmorstein, R. (2020). Molecular basis for N-terminal acetylation by human NatE and its modulation by HYPK. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14584-7

Denic, V., Dötsch, V., & Sinning, I. (2013). Endoplasmic reticulum targeting and insertion of tail-anchored membrane proteins by the GET pathway. *Cold Spring Harbor Perspectives in Biology*, *5*(8), 1–12. https://doi.org/10.1101/cshperspect.a013334

Dingman, W. (1972). Program for the Coefficients Rapid Calculation Salt of Sedimentation in Linear or Sucrose A number of reports in the past few years have dealt with various tech-niques for measuring sedimentation coefficients in preparative ultracen- trifuges using stabi. *Analytical Biochemistry*, *49*, 124–133.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dodgson, S. E., Kim, S., Costanzo, M., Baryshnikova, A., Morse, D. L., Kaiser, C. A., … Amon, A. (2016). Chromosome-specific and global effects of aneuploidy in Saccharomyces cerevisiae. *Genetics*, *202*(4), 1395–1409. https://doi.org/10.1534/genetics.115.185660

Döring, K., Ahmed, N., Riemer, T., Suresh, H. G., Vainshtein, Y., Habich, M., … Bukau, B. (2017). Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. *Cell*, *170*(2), 298-311.e20. https://doi.org/10.1016/j.cell.2017.06.038

Drummond, D. A., & Wilke, C. O. (2008). Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell*, *134*(2), 341–352. https://doi.org/10.1016/j.cell.2008.05.042

Duc, K. D., Batra, S. S., Bhattacharya, N., Cate, J. H. D., & Song, Y. S. (2019). Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Research*, *47*(8), 4198–4210. https://doi.org/10.1093/nar/gkz106

Duncan, C. D. S., & Mata, J. (2011). Widespread cotranslational formation of protein complexes. *PLoS Genetics*, *7*(12). https://doi.org/10.1371/journal.pgen.1002398

Duncan, C. D. S., & Mata, J. (2014). Cotranslational protein-RNA associations predict protein-

protein interactions. *BMC Genomics*, *15*(1), 1–7. https://doi.org/10.1186/1471-2164-15-298

Duncan, R., & Hershey, J. W. (1983). Identification and quantitation of levels of protein synthesis initiation factors in crude HeLa cell lysates by two-dimensional polyacrylamide gel electrophoresis. *Journal of Biological Chemistry*, *258*(11), 7228–7235. https://doi.org/10.1016/S0021-9258(18)32356-1

Ellgaard, L., McCaul, N., Chatsisvili, A., & Braakman, I. (2016). Co- and Post-Translational Protein Folding in the ER. *Traffic*, *17*(6), 615–638. https://doi.org/10.1111/tra.12392

Erickson, H. P. (1993). of Talented Proteins in Search of Functions. *Current Opinion in Cell Biology*, *5*, 869–876.

Farías-Rico, J. A., Selin, F. R., Myronidi, I., Frühauf, M., & Von Heijne, G. (2018). Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(40), E9280–E9287. https://doi.org/10.1073/pnas.1812756115

Fiebiger, E., Tortorella, D., Jouvin, M. H., Kinet, J. P., & Ploegh, H. L. (2005). Cotranslational endoplasmic reticulum assembly of FcεRI controls the formation of functional IgE-binding receptors. *Journal of Experimental Medicine*, *201*(2), 267–277. https://doi.org/10.1084/jem.20041384

Fiedler, U., & Weiss, V. (1995). A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. *The EMBO Journal*, *14*(15), 3696–3705. https://doi.org/10.1002/j.1460-2075.1995.tb00039.x

Friedman, F. K., & Beychok, S. (1979). Probes of subunit assembly and reconstitution pathways in multisubunit proteins. *Annual Review of Biochemistry*, *48*, 217–250. https://doi.org/10.1146/annurev.bi.48.070179.001245

Friedrich, U. A., Zedan, M., Hessling, B., Fenzl, K., Gillet, L., Barry, J., … Bukau, B. (2021). Nα-terminal acetylation of proteins by NatA and NatB serves distinct physiological roles in Saccharomyces cerevisiae. *Cell Reports*, *34*(5). https://doi.org/10.1016/j.celrep.2021.108711

Fritch, B., Kosolapov, A., Hudson, P., Nissley, D. A., Woodcock, H. L., Deutsch, C., & O'Brien, E. P. (2018). Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *Journal of the American Chemical Society*, *140*(15), 5077–5087. https://doi.org/10.1021/jacs.7b11044

Frottin, F., Bienvenut, W. V., Bignon, J., Jacquet, E., Jacome, A. S. V., Van Dorsselaer, A., … Giglione, C. (2016). MetAP1 and MetAP2 drive cell selectivity for a potent anticancer agent in synergy, by controlling glutathione redox state. *Oncotarget*, *7*(39), 63306–63323. https://doi.org/10.18632/oncotarget.11216

Fujiki, M., & Verner, K. (1993). Coupling of cytosolic protein synthesis and mitochondrial protein import in yeast. Evidence for cotranslational import in vivo. *Journal of Biological Chemistry*, *268*(3), 1914–1920. https://doi.org/10.1016/S0021-9258(18)53941-7

Fujiwara, K., Katagi, Y., Ito, K., & Chiba, S. (2020). Proteome-wide Capture of Co-translational Protein Dynamics in Bacillus subtilis Using TnDR, a Transposable Protein-Dynamics Reporter. *Cell Reports*, *33*(2), 108250. https://doi.org/10.1016/j.celrep.2020.108250

Fulton, A. B., & L'Ecuyer, T. (1993). COMMENTARY: Cotranslational assembly of some cytoskeletal proteins: implications and prospects. *Journal of Cell Science*, *105*(4), 867–871. Retrieved from http://www.biologists.com/JCS/105/04/jcs8304.html

Fünfschilling, U., & Rospert, S. (1999). Nascent polypeptide-associated complex stimulates protein import into yeast mitochondria. *Molecular Biology of the Cell*, *10*(10), 3289–

3299. https://doi.org/10.1091/mbc.10.10.3289

Galmozzi, C. V., Merker, D., Friedrich, U. A., Döring, K., & Kramer, G. (2019). *Selective ribosome profiling to study interactions of translating ribosomes in yeast*. *Nature Protocols* (Vol. 14). Springer US. https://doi.org/10.1038/s41596-019-0185-z

Gamerdinger, M., Kobayashi, K., Wallisch, A., Kreft, S. G., Sailer, C., Schlömer, R., … Deuerling, E. (2019). Early Scanning of Nascent Polypeptides inside the Ribosomal Tunnel by NAC. *Molecular Cell*, *75*(5), 996-1006.e8. https://doi.org/10.1016/j.molcel.2019.06.030

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gilmore, R., Coffey, M. C., Leone, G., McLure, K., & Lee, P. W. (1996). Co-translational trimerization of the reovirus cell attachment protein. *The EMBO Journal*, *15*(11), 2651–2658. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=450200&tool=pmcentrez&rendertype=abstract

Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., … Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Research*, *47*(D1), D559–D563. https://doi.org/10.1093/nar/gky973

Gloge, F., Becker, A. H., Kramer, G., & Bukau, B. (2014). Co-translational mechanisms of protein maturation. *Current Opinion in Structural Biology*, *24*, 24–33. https://doi.org/10.1016/j.sbi.2013.11.004

Gold, V. A., Chroscicki, P., Bragoszewski, P., & Chacinska, A. (2017). Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Reports*, *18*(10), e201744261. https://doi.org/10.15252/embr.201744261

Goldman, D. H., Kaiser, C. M., Milin, A., Righini, M., Tinoco, I., & Bustamante, C. (2015). Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science*, *348*(6233), 457–460. https://doi.org/10.1126/science.1261909

Griffin, B. A., Adams, S. R., & Tsien, R. Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. *Science*, *281*(5374), 269–272. https://doi.org/10.1126/science.281.5374.269

Grudnik, P., Bange, G., & Sinning, I. (2009). Protein targeting by the signal recognition particle. *Biological Chemistry*, *390*(8), 775–782. https://doi.org/10.1515/BC.2009.102

Halbach, A., Zhang, H., Wengi, A., Jablonska, Z., Gruber, I. M. L., Halbeisen, R. E., … Dichtl, B. (2009). Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *The EMBO Journal*, *28*(19), 2959–2970. https://doi.org/10.1038/emboj.2009.240

Han, P., Shichino, Y., Schneider-Poetsch, T., Mito, M., Hashimoto, S., Udagawa, T., … Iwasaki, S. (2020). Genome-wide Survey of Ribosome Collision. *Cell Reports*, *31*(5), 107610. https://doi.org/10.1016/j.celrep.2020.107610

Harrington, H. R., Zimmer, M. H., Chamness, L. M., Nash, V., Penn, W. D., Miller, T. F., … Schlebach, J. P. (2020). Cotranslational folding stimulates programmed ribosomal frameshifting in the alphavirus structural polyprotein. *Journal of Biological Chemistry*, *295*(20), 6798–6808. https://doi.org/10.1074/jbc.RA120.012706

Harris, N. J., Reading, E., Ataka, K., Grzegorzewski, L., Charalambous, K., Liu, X., … Booth, P. J. (2017). Structure formation during translocon-unassisted co-translational membrane protein folding. *Scientific Reports*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-08522-9

Hartl, F. U., & Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in

vivo. *Nature Structural and Molecular Biology*, *16*(6), 574–581. https://doi.org/10.1038/nsmb.1591

Heissmeyer, V., Krappmann, D., Hatada, E. N., & Scheidereit, C. (2001). Shared Pathways of IκB Kinase-Induced SCFβTrCP-Mediated Ubiquitination and Degradation for the NF-κB Precursor p105 and IκBα. *Molecular and Cellular Biology*, *21*(4), 1024–1035. https://doi.org/10.1128/mcb.21.4.1024-1035.2001

Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., … Von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, *450*(7172), 1026–1030. https://doi.org/10.1038/nature06387

Hickey, K. L., Dickson, K., Cogan, J. Z., Replogle, J. M., Schoof, M., D'Orazio, K. N., … Kostova, K. K. (2020). GIGYF2 and 4EHP Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. *Molecular Cell*, *79*(6), 950-962.e6. https://doi.org/10.1016/j.molcel.2020.07.007

Ho, C. Y., & Lammerding, J. (2012). Lamins at a glance. *Journal of Cell Science*, *125*(9), 2087–2093. https://doi.org/10.1242/jcs.087288

Hsieh, H. H., Lee, J. H., Chandrasekar, S., & Shan, S. ou. (2020). A ribosome-associated chaperone enables substrate triage in a cotranslational protein targeting complex. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19548-5

Hu, J. C., Shea, E. K. O., Kim, P. S., & Sauer, R. T. (1990). Sequence Requirements for Coiled-Coils : Analysis with λ Repressor- GCN4 Leucine Zipper Fusions Author (s): James C . Hu , Erin K . O ' Shea , Peter S . Kim and Robert T . Sauer Published by : American Association for the Advancement of Science Stable U. *Science*, *250*(4986), 1400–1403.

Ikeuchi, K., Tesina, P., Matsuo, Y., Sugiyama, T., Cheng, J., Saeki, Y., … Inada, T. (2019). Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. *The EMBO Journal*, *38*(5), 1–21. https://doi.org/10.15252/embj.2018100276

Inglis, A. J., Page, K. R., Guna, A., & Voorhees, R. M. (2020). Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex. *Cell Reports*, *30*(11), 3691-3698.e5. https://doi.org/10.1016/j.celrep.2020.02.084

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, *324*(5924), 218–223. https://doi.org/10.1126/science.1168978

Ingolia, Nicholas T, Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, *7*(8), 1534–1550. https://doi.org/10.1038/nprot.2012.086

Isaacs, W. B., Cook, R. K., Van Atta, J. C., Redmond, C. M., & Fulton, A. B. (1989). Assembly of vimentin in cultured cells varies with cell type. *Journal of Biological Chemistry*, *264*(30), 17953–17960.

Isaacs, W. B., & Fulton, a B. (1987). Cotranslational assembly of myosin heavy chain in developing cultured skeletal muscle. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(17), 6174–6178. https://doi.org/10.1073/pnas.84.17.6174

Isaacs, W. B., Kim, I. S., Struve, A., & Fulton, A. B. (1989). Biosynthesis of titin in cultured skeletal muscle cells. *Journal of Cell Biology*, *109*(5), 2189–2195. https://doi.org/10.1083/jcb.109.5.2189

Ismail, N., Hedman, R., Schiller, N., & Heijne, G. Von. (2013). Europe PMC Funders Group

Europe PMC Funders Author Manuscripts A bi-phasic pulling force acts on transmembrane helices during translocon-mediated membrane integration, *19*(10), 1018–1022. https://doi.org/10.1038/nsmb.2376.A

Ito, K., & Chiba, S. (2013, June). Arrest peptides: Cis-acting modulators of translation. *Annual Review of Biochemistry*. Annu Rev Biochem. https://doi.org/10.1146/annurev-biochem-080211-105026

Jacobs, W. M., & Shakhnovich, E. I. (2017). Evidence of evolutionary selection for cotranslational folding. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), 11434–11439. https://doi.org/10.1073/pnas.1705772114

Jacobson, G. N., & Clark, P. L. (2016). Quality over quantity: optimizing co-translational protein folding with non-'optimal' synonymous codons. *Current Opinion in Structural Biology*, *38*(1), 102–110. https://doi.org/10.1016/j.sbi.2016.06.002

Jaenicke, R., & Lilie, H. (2000). Folding and association of oligomeric and multimeric proteins. *Advances in Protein Chemistry*, *53*, 329–362. https://doi.org/10.1016/S0065-3233(00)53007-1

Jaiswal, H., Conz, C., Otto, H., Wolfle, T., Fitzke, E., Mayer, M. P., & Rospert, S. (2011). The Chaperone Network Connected to Human Ribosome-Associated Complex. *Molecular and Cellular Biology*, *31*(6), 1160–1173. https://doi.org/10.1128/mcb.00986-10

Jan, C. H., Williams, C. C., & Weissman, J. S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*, *346*(6210), 748–751. https://doi.org/10.1126/science.1257521

Joazeiro, C. A. P. (2019). Mechanisms and functions of ribosome-associated protein quality control. *Nature Reviews Molecular Cell Biology*, *20*(6), 368–383. https://doi.org/10.1038/s41580-019-0118-2

Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V., & Hegde, R. S. (2018). ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Molecular Cell*, *72*(3), 469-481.e7. https://doi.org/10.1016/j.molcel.2018.08.037

Juszkiewicz, S., Speldewinde, S. H., Wan, L., Svejstrup, J. Q., & Hegde, R. S. (2020). The ASC-1 Complex Disassembles Collided Ribosomes. *Molecular Cell*, *79*(4), 603-614.e8. https://doi.org/10.1016/j.molcel.2020.06.006

Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., … Tora, L. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. *Nature Communications*, *10*(1), 25–28. https://doi.org/10.1038/s41467-019-09749-y

Kampinga, H. H., Hageman, J., Vos, M. J., Kubota, H., Tanguay, R. M., Bruford, E. A., … Hightower, L. E. (2009). Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress and Chaperones*, *14*(1), 105–111. https://doi.org/10.1007/s12192-008-0068-7

Karin, M., & Lin, A. (2002). NF-κB at the crossroads of life and death. *Nature Immunology*, *3*(3), 221–227. https://doi.org/10.1038/ni0302-221

Kassem, S., Villanyi, Z., & Collart, M. A. (2017). Not5-dependent co-translational assembly of Ada2 and Spt20 is essential for functional integrity of SAGA. *Nucleic Acids Research*, *45*(3), 1186–1199. https://doi.org/10.1093/nar/gkw1059

Kendall, R. L., & Bradshaw, R. A. (1992). Isolation and characterization of the methionine aminopeptidase from porcine liver responsible for the co-translational processing of proteins. *Journal of Biological Chemistry*, *267*(29), 20667–20673. https://doi.org/10.1016/S0021-9258(19)36737-7

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R.,

… Karp, P. D. (2017). The EcoCyc database: Reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research*, *45*(D1), D543–D550. https://doi.org/10.1093/nar/gkw1003

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, H.-K., Liu, F., Fei, J., Bustamante, C., Gonzalez, R. L., & Tinoco, I. (2014). A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proceedings of the National Academy of Sciences*, *111*(15), 5538–5543. https://doi.org/10.1073/pnas.1403457111

Kim, S. J., Yoon, J. S., Shishido, H., Yang, Z., Rooney, L. A. A., Barral, J. M., & Skach, W. R. (2015). Translational tuning optimizes nascent protein folding in cells. *Science*, *348*(6233), 444–448. https://doi.org/10.1126/science.aaa3974

Kleiger, G., & Eisenberg, D. (2002). GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding rossmann folds through Cα-H···O hydrogen bonds and van der Waals interactions. *Journal of Molecular Biology*, *323*(1), 69–76. https://doi.org/10.1016/S0022-2836(02)00885-9

Kleiger, G., Grothe, R., Mallick, P., & Eisenberg, D. (2002). GXXXG and AXXXA: Common α-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry*, *41*(19), 5990–5997. https://doi.org/10.1021/bi0200763

Kolb, T., Maaß, K., Hergt, M., Aebi, U., & Herrmann, H. (2011). Lamin A and lamin C form homodimers and coexist in higher complex forms both in the nucleoplasmic fraction and in the lamina of cultured human cells. *Nucleus*, *2*(5). https://doi.org/10.4161/nucl.2.5.17765

Kosolapov, A., & Deutsch, C. (2009). Tertiary interactions within the ribosomal exit tunnel. *Nature Structural & Molecular Biology*, *16*(4), 405–411. https://doi.org/10.1038/nsmb.1571

Kramer, G., Boehringer, D., Ban, N., & Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Structural and Molecular Biology*, *16*(6), 589–597. https://doi.org/10.1038/nsmb.1614

Kramer, G., Shiber, A., & Bukau, B. (2019). Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins. *Annual Review of Biochemistry*, *88*(1), 337–364. https://doi.org/10.1146/annurev-biochem-013118-111717

Kraushar, M. L., Krupp, F., Turko, P., Ambrozkiewicz, M. C., Sprink, T., Imami, K., … Spahn, C. M. T. (2020). The architecture of protein synthesis in the developing neocortex at near-atomic resolution reveals Ebp1-mediated neuronal proteostasis at the 60S tunnel exit. *BioRxiv*, 2020.02.08.939488. https://doi.org/10.1101/2020.02.08.939488

L'Ecuyer, T. J., Noller, J. A., & Fulton, A. B. (1998). Assembly of Tropomyosin Isoforms into the Cytoskeleton of Avian Muscle Cells. *Pediatric Research*, *43*(6), 813–822. https://doi.org/10.1203/00006450-199806000-00016

Lakshminarayan, R., Phillips, B. P., Binnian, I. L., Gomez-Navarro, N., Escudero-Urquijo, N., Warren, A. J., & Miller, E. A. (2020). Pre-emptive Quality Control of a Misfolded Membrane Protein by Ribosome-Driven Effects. *Current Biology*, *30*(5), 854-864.e5. https://doi.org/10.1016/j.cub.2019.12.060

Lamriben, L., Graham, J. B., Adams, B. M., Hebert, D. N., & Biology, C. (2017). Control System : the Calnexin Binding Cycle, *17*(4), 308–326. https://doi.org/10.1111/tra.12358.N-glycan

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*

Methods, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Lee, Y., Zhou, T., Tartaglia, G. G., Vendruscolo, M., & Wilke, C. O. (2010). Translationally optimal codons associate with aggregation-prone sites in proteins. *Proteomics*, *10*(23), 4163–4171. https://doi.org/10.1002/pmic.201000229

Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B. J., Dempsey, C. E., & Engelman, D. M. (1992). Glycophorin A dimerization is driven by specific interactions between transmembrane α-helices. *Journal of Biological Chemistry*, *267*(11), 7683–7689. https://doi.org/10.1016/S0021-9258(18)42569-0

Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, *46*(D1), D493–D496. https://doi.org/10.1093/nar/gkx922

Li, E., Wimley, W. C., & Hristova, K. (2012). Transmembrane helix dimerization: Beyond the search for sequence motifs. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1818*(2), 183–193. https://doi.org/10.1016/j.bbamem.2011.08.031

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X., Chen, Y., Qi, H., Liu, L., & Shuai, J. (2016). Synonymous mutations in oncogenesis and apoptosis versus survival unveiled by network modeling. *Oncotarget*, *7*(23), 34599–34616. https://doi.org/10.18632/oncotarget.8963

Lin, L., DeMartino, G. N., & Greene, W. C. (2000). Cotranslational dimerization of the Rel homology domain of NF-kappaB1 generates p50-p105 heterodimers and is required for effective p50 production. *The EMBO Journal*, *19*(17), 4712–4722. https://doi.org/10.1093/emboj/19.17.4712

Liu, F., Jones, D. K., de Lange, W. J., & Robertson, G. A. (2016). Cotranslational association of mRNA encoding subunits of heteromeric ion channels. *Proceedings of the National Academy of Sciences*, *113*(17), 4859–4864. https://doi.org/10.1073/pnas.1521577113

Liu, K., Rehfus, J. E., Mattson, E., & Kaiser, C. M. (2017). The ribosome destabilizes native and non-native structures in a nascent multidomain protein. *Protein Science*, *26*(7), 1439–1451. https://doi.org/10.1002/pro.3189

Liutkute, M., Samatova, E., & Rodnina, M. V. (2020). Cotranslational folding of proteins on the ribosome. *Biomolecules*, *10*(1). https://doi.org/10.3390/biom10010097

Lu, J., & Deutsch, C. (2008). Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. *Journal of Molecular Biology*, *384*(1), 73–86. https://doi.org/10.1016/j.jmb.2008.08.089

Lu, J., Robinson, J. M., Edwards, D., & Deutsch, C. (2001). T1−T1 Interactions Occur in ER Membranes while Nascent Kv Peptides Are Still Attached to Ribosomes †. *Biochemistry*, *40*(37), 10934–10946. https://doi.org/10.1021/bi010763e

Luedtke, N. W., Dexter, R. J., Fried, D. B., & Schepartz, A. (2007). Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH, *3*(12), 779–784. https://doi.org/10.1038/nchembio.2007.49

Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, *252*(5009), 1162–1164. https://doi.org/10.1126/science.252.5009.1162

Lynch, M. (2012). The evolution of multimeric protein assemblages. *Molecular Biology and Evolution*, *29*(5), 1353–1366. https://doi.org/10.1093/molbev/msr300

Makarov, A. A., Zou, J., Houston, D. R., Spanos, C., Solovyova, A. S., Cardenal-Peralta, C., … Schirmer, E. C. (2019). Lamin A molecular compression and sliding as mechanisms behind nucleoskeleton elasticity. *Nature Communications*, *10*(1), 1–17. https://doi.org/10.1038/s41467-019-11063-6

Marsh, J. A., Hernández, H., Hall, Z., Ahnert, S. E., Perica, T., Robinson, C. V., & Teichmann, S. A. (2013). Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. *Cell*, *153*(2), 461–470. https://doi.org/10.1016/j.cell.2013.02.044

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martoglio, B., & Dobberstein, B. (1998). Signal sequences: More than just greasy peptides. *Trends in Cell Biology*, *8*(10), 410–415. https://doi.org/10.1016/S0962-8924(98)01360-9

McGlincy, N. J., & Ingolia, N. T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods*, *126*(May), 112–129. https://doi.org/10.1016/j.ymeth.2017.05.028

McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., … Selbach, M. (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell*, *167*(3), 803-815.e21. https://doi.org/10.1016/j.cell.2016.09.015

Mena, E. L., Kjolby, R. A. S., Saxton, R. A., Werner, A., Lew, B. G., Boyle, J. M., … Rape, M. (2018). Dimerization quality control ensures neuronal development and survival. *Science*, *362*(6411), eaap8236. https://doi.org/10.1126/science.aap8236

Meydan, S., & Guydosh, N. R. (2020a). A cellular handbook for collided ribosomes: surveillance pathways and collision types. *Current Genetics*, (0123456789). https://doi.org/10.1007/s00294-020-01111-w

Meydan, S., & Guydosh, N. R. (2020b). Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control. *Molecular Cell*, *79*(4), 588-602.e6. https://doi.org/10.1016/j.molcel.2020.06.010

Miettinen, T. P., & Bjorklund, M. (2015). Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Research*, *43*(2), 1019–1034. https://doi.org/10.1093/nar/gku1310

Miller, J. D., Tajima, S., Lauffer, L., & Walter, P. (1995). The β subunit of the signal recognition particle receptor is a transmembrane GTpase that anchors the α subunit, a peripheral membrane GTpase, to the endoplasmic reticulum membrane. *Journal of Cell Biology*, *128*(3), 273–282. https://doi.org/10.1083/jcb.128.3.273

Mingarro, I., Nilsson, I. M., Whitley, P., & von Heijne, G. (2000). Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biology*, *1*. https://doi.org/10.1186/1471-2121-1-3

Moir, R. D., Donaldson, A. D., & Stewart, M. (1991). Expression in Escherichia coli of human lamins A and C: Influence of head and tail domains on assembly properties and paracrystal formation. *Journal of Cell Science*, *99*(2), 363–372.

Mrazek, J., Toso, D., Ryazantsev, S., Zhang, X., Zhou, Z. H., Fernandez, B. C., … Rome, L. H. (2014). Polyribosomes are molecular 3D nanoprinters that orchestrate the assembly of vault particles. *ACS Nano*, *8*(11), 11552–11559. https://doi.org/10.1021/nn504778h

Natan, E., Endoh, T., Haim-Vilmovsky, L., Flock, T., Chalancon, G., Hopper, J. T. S., … Teichmann, S. A. (2018). Cotranslational protein assembly imposes evolutionary constraints on homomeric proteins. *Nature Structural and Molecular Biology*, *25*(3), 279–288. https://doi.org/10.1038/s41594-018-0029-5

Natan, E., Wells, J. N., Teichmann, S. A., & Marsh, J. A. (2017). ScienceDirect Regulation ,

evolution and consequences of cotranslational protein complex assembly. *Current Opinion in Structural Biology*, *42*, 90–97. https://doi.org/10.1016/j.sbi.2016.11.023

Natarajan, N., Foresti, O., Wendrich, K., Stein, A., & Carvalho, P. (2020). Quality Control of Protein Complex Assembly by a Transmembrane Recognition Factor. *Molecular Cell*, *77*(1), 108-119.e9. https://doi.org/10.1016/j.molcel.2019.10.003

Neidhardt, F. C., Bloch, P. L., & Smith, D. F. (1974). Culture medium for enterobacteria. *Journal of Bacteriology*, *119*(3), 736–747. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4604283

Nguyen, K. T., Mun, S. H., Lee, C. S., & Hwang, C. S. (2018). Control of protein degradation by N-terminal acetylation and the N-end rule pathway. *Experimental and Molecular Medicine*, *50*(7). https://doi.org/10.1038/s12276-018-0097-y

Nguyen, T. H., Law, D. T. S., & Williams, D. B. (1991). Binding protein BiP is required for translocation of secretory proteins into the endoplasmic reticulum in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(4), 1565–1569. https://doi.org/10.1073/pnas.88.4.1565

Nicholls, C. D., McLure, K. G., Shields, M. A., & Lee, P. W. K. (2002). Biogenesis of p53 involves cotranslational dimerization of monomers and posttranslational dimerization of dimers. Implications on the dominant negative effect. *Journal of Biological Chemistry*, *277*(15), 12937–12945. https://doi.org/10.1074/jbc.M108815200

Nilsson, O. B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., … von Heijne, G. (2015). Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Reports*, *12*(10), 1533–1540. https://doi.org/10.1016/j.celrep.2015.07.065

Noordeen, N. A., Carafoli, F., Hohenester, E., Horton, M. A., & Leitinger, B. (2006). A transmembrane leucine zipper is required for activation of the dimeric receptor tyrosine kinase DDR1. *Journal of Biological Chemistry*, *281*(32), 22744–22751. https://doi.org/10.1074/jbc.M603233200

O'Brien, E. P., Vendruscolo, M., & Dobson, C. M. (2012). Prediction of variable translation rate effects on cotranslational protein folding. *Nature Communications*, *3*(February 2014), 868. https://doi.org/10.1038/ncomms1850

Otto, H., Conz, C., Maier, P., Wolfle, T., Suzuki, C. K., Jeno, P., … Rospert, S. (2005). The chaperones MPP11 and Hsp70L1 form the mammalian ribosome-associated complex. *Proceedings of the National Academy of Sciences*, *102*(29), 10064–10069. https://doi.org/10.1073/pnas.0504400102

Panasenko, O. O., Somasekharan, S. P., Villanyi, Z., Zagatti, M., Bezrukov, F., Rashpa, R., … Collart, M. A. (2019). Co-translational assembly of proteasome subunits in NOT1-containing assemblysomes. *Nature Structural & Molecular Biology*, *26*(2), 110–120. https://doi.org/10.1038/s41594-018-0179-5

Pechmann, S., & Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Structural & Molecular Biology*, *20*(2), 237–243. https://doi.org/10.1038/nsmb.2466

Phillip, Y., & Schreiber, G. (2013). Formation of protein complexes in crowded environments-From in vitro to in vivo. https://doi.org/10.1016/j.febslet.2013.01.007

Phillips, B. P., & Miller, E. A. (2020). Ribosome-associated quality control of membrane proteins at the endoplasmic reticulum. *Journal of Cell Science*, *133*(22). https://doi.org/10.1242/jcs.251983

Pisareva, V. P., Skabkin, M. A., Hellen, C. U. T., Pestova, T. V., & Pisarev, A. V. (2011). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled

elongation complexes. *EMBO Journal*, *30*(9), 1804–1817. https://doi.org/10.1038/emboj.2011.93

Pradhan, P., Li, W., & Kaur, P. (2009). Translational Coupling Controls Expression and Function of the DrrAB Drug Efflux Pump. *Journal of Molecular Biology*, *385*(3), 831–842. https://doi.org/10.1016/j.jmb.2008.11.027

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., … Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, *174*(3), 744-757.e24. https://doi.org/10.1016/j.cell.2018.05.024

Raue, U., Oellerer, S., & Rospert, S. (2007). Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence. *Journal of Biological Chemistry*, *282*(11), 7809–7816. https://doi.org/10.1074/jbc.M611436200

Redick, S. D., & Schwarzbauer, J. E. (1995). Rapid intracellular assembly of tenascin hexabrachions suggests a novel co- translational process, 9. Retrieved from http://files/103/Redick and Schwarzbauer - Rapid intracellular assembly of tenascin hexabrach.pdf

Ree, R., Varland, S., & Arnesen, T. (2018). Spotlight on protein N-terminal acetylation. *Experimental and Molecular Medicine*, *50*(7). https://doi.org/10.1038/s12276-018-0116-z

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genome Viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754.Integrative

Rocha, E. P. C. (2004). Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*, *14*(11), 2279–2286. https://doi.org/10.1101/gr.2896904

Rodríguez-Galán, O., García-Gómez, J. J., Rosado, I. V., Wei, W., Méndez-Godoy, A., Pillet, B., … De La Cruz, J. (2021). A functional connection between translation elongation and protein folding at the ribosome exit tunnel in Saccharomyces cerevisiae. *Nucleic Acids Research*, *49*(1), 206–220. https://doi.org/10.1093/nar/gkaa1200

Rose, A., Schraegle, S. J., Stahlberg, E. A., & Meier, I. (2005a). Coiled-coil protein composition of 22 proteomes - Differences and common themes in subcellular infrastructure and traffic control. *BMC Evolutionary Biology*, *5*, 1–21. https://doi.org/10.1186/1471-2148-5-66

Rose, R., Weyand, M., Lammers, M., Ishizaki, T., Ahmadian, M. R., & Wittinghofer, A. (2005b). Structural and mechanistic insights into the interaction between Rho and mammalian Dia. *Nature*, *435*(7041), 513–518. https://doi.org/10.1038/nature03604

Ruan, J., Xu, C., Bian, C., Lam, R., Wang, J. P., Kania, J., … Zang, J. (2012). Crystal structures of the coil 2B fragment and the globular tail domain of human lamin B1. *FEBS Letters*, *586*(4), 314–318. https://doi.org/10.1016/j.febslet.2012.01.007

Ruiz-Canada, C., Kelleher, D. J., & Gilmore, R. (2009). Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms. *Cell*, *136*(2), 272–283. https://doi.org/10.1016/j.cell.2008.11.047

Schaefer, J., Jovanovic, G., Kotta-Loizou, I., & Buck, M. (2016). Single-step method for β-galactosidase assays in Escherichia coli using a 96-well microplate reader. *Analytical Biochemistry*, *503*, 56–57. https://doi.org/10.1016/j.ab.2016.03.017

Schreiber, K. H., & Kennedy, B. K. (2013). When Lamins Go Bad: Nuclear Structure and Disease. *Cell*, *152*(6), 1365–1375. https://doi.org/10.1016/j.cell.2013.02.015

Schwarz, A., & Beck, M. (2019). The Benefits of Cotranslational Assembly: A Structural Perspective. *Trends in Cell Biology*, *29*(10), 791–803. https://doi.org/10.1016/j.tcb.2019.07.006

Shao, S., Brown, A., Santhanam, B., & Hegde, R. S. (2015). Structure and assembly pathway of the ribosome quality control complex. *Molecular Cell*, *57*(3), 433–444. https://doi.org/10.1016/j.molcel.2014.12.015

Sharma, A. K., Sormanni, P., Ahmed, N., Ciryam, P., Friedrich, U. A., Kramer, G., & O'Brien, E. P. (2018). A Chemical Kinetic Basis for Measuring Translation Initiation and Elongation Rates from Ribosome Profiling data. *BioRxiv*, 1–27. https://doi.org/10.1101/490730

Shen, P. S., Park, J., Qin, Y., Li, X., Parsawar, K., Larson, M. H., … Frost, A. (2015). Elongation of Nascent Chains. *Science (New York, N.Y.)*, *1*(1), 1–2.

Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., … Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*, *561*(7722), 268–272. https://doi.org/10.1038/s41586-018-0462-y

Shieh, Y.-W., Minguez, P., Bork, P., Auburger, J. J., Guilbride, D. L., Kramer, G., & Bukau, B. (2015). Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science (New York, N.Y.)*, *350*(6261), 678–680. https://doi.org/10.1126/science.aac8171

Shoemaker, C. J., Eyler, D. E., & Green, R. (2010). Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science*, *330*(6002), 369–372. https://doi.org/10.1126/science.1192430

Sims, C. E., & Allbritton, N. L. (2007). Analysis of single mammalian cells on-chip. *Lab on a Chip*, *7*(4), 423. https://doi.org/10.1039/b615235j

Sinha, N. K., Ordureau, A., Best, K. M., Saba, J. A., Zinshteyn, B., Sundaramoorthy, E., … Green, R. (2020). EDF1 coordinates cellular responses to ribosome collisions. *ELife*, *9*, 1–84. https://doi.org/10.7554/ELIFE.58828

Sprenger, J., Lynn Fink, J., Karunaratne, S., Hanson, K., Hamilton, N. A., & Teasdale, R. D. (2008). LOCATE: A mammalian protein subcellular localization database. *Nucleic Acids Research*, *36*(SUPPL. 1), 230–233. https://doi.org/10.1093/nar/gkm950

Spring, J., Beck, K., & Chiquet-Ehrismann, R. (1989). Two contrary functions of tenascin: Dissection of the active sites by recombinant tenascin fragments. *Cell*, *59*(2), 325–334. https://doi.org/10.1016/0092-8674(89)90294-8

Stephani, M., Picchianti, L., Gajic, A., Beveridge, R., Skarwan, E., Hernandez, V. S. de M., … Dagdas, Y. (2020). A cross-kingdom conserved er-phagy receptor maintains endoplasmic reticulum homeostasis during stress. *ELife*, *9*, 1–105. https://doi.org/10.7554/ELIFE.58396

Stogios, P. J., Downs, G. S., Jauhal, J. J. S., Nandra, S. K., & Privé, G. G. (2005). Sequence and structural analysis of BTB domain proteins. *Genome Biology*, *6*(10), R82. https://doi.org/10.1186/gb-2005-6-10-r82

Sundaramoorthy, E., Leonard, M., Mak, R., Liao, J., Fulzele, A., & Bennett, E. J. (2017). ZNF598 and RACK1 Regulate Mammalian Ribosome-Associated Quality Control Function by Mediating Regulatory 40S Ribosomal Ubiquitylation. *Molecular Cell*, *65*(4), 751-760.e4. https://doi.org/10.1016/j.molcel.2016.12.026

Tanaka, N., Meineke, B., & Shuman, S. (2011). RtcB, a novel RNA ligase, can catalyze tRNA splicing and HAC1 mRNA splicing in vivo. *Journal of Biological Chemistry*, *286*(35), 30253–30257. https://doi.org/10.1074/jbc.C111.274597

Taylor, H. C., Lightner, V. A., Beyer, W. F., McCaslin, D., Briscoe, G., & Erickson, H. P. (1989).

Biochemical and structural studies of tenascin/hexabrachion proteins. *Journal of Cellular Biochemistry*, *41*(2), 71–90. https://doi.org/10.1002/jcb.240410204

The UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., … Lundberg, E. (2017). A subcellular map of the human proteome. *Science*, *356*(6340). https://doi.org/10.1126/science.aal3321

Trentini, D. B., Pecoraro, M., Tiwary, S., Cox, J., Mann, M., Hipp, M. S., & Hartl, F. U. (2020). Role for ribosome-associated quality control in sampling proteins for MHC class I-mediated antigen presentation. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(8), 4099–4108. https://doi.org/10.1073/pnas.1914401117

Trovato, F., & O'Brien, E. P. (2017). Fast Protein Translation Can Promote Co- and Posttranslational Folding of Misfolding-Prone Proteins. *Biophysical Journal*, *112*(9), 1807–1819. https://doi.org/10.1016/j.bpj.2017.04.006

Tsuboi, T., Kuroha, K., Kudo, K., Makino, S., Inoue, E., Kashima, I., & Inada, T. (2012). Dom34: Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3' End of Aberrant mRNA. *Molecular Cell*, *46*(4), 518–529. https://doi.org/10.1016/j.molcel.2012.03.013

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., … Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220). https://doi.org/10.1126/science.1260419

Uhlen, M., Tegel, H., Sivertsson, Å., Kuo, C., Jahir, M., Lewis, N. E., … Schwenk, J. M. (2018). The human secretome – the proteins secreted from human cells, 1–14.

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., … Lander, E. S. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. *Journal of Visualized Experiments*, (39), 1–7. https://doi.org/10.3791/1869

Varenne, S., Buc, J., Lloubes, R., & Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of Molecular Biology*, *180*(3), 549–576. https://doi.org/10.1016/0022-2836(84)90027-5

Veitia, R. A. (2007). Exploring the molecular etiology of dominant-negative mutations. *Plant Cell*, *19*(12), 3843–3851. https://doi.org/10.1105/tpc.107.055053

Volkmar, N., & Christianson, J. C. (2020). Squaring the EMC - How promoting membrane protein biogenesis impacts cellular functions and organismal homeostasis. *Journal of Cell Science*, *133*(8), 1–14. https://doi.org/10.1242/jcs.243519

Wagner, S., Herrmannová, A., Hronová, V., Gunišová, S., Sen, N. D., Hannan, R. D., … Valášek, L. S. (2020). Selective Translation Complex Profiling Reveals Staged Initiation and Co-translational Assembly of Initiation Factor Complexes. *Molecular Cell*, *79*(4), 546-560.e7. https://doi.org/10.1016/j.molcel.2020.06.004

Wang, F., Durfee, L. A., & Huibregtse, J. M. (2013). A cotranslational ubiquitination pathway for quality control of misfolded proteins. *Molecular Cell*, *50*(3), 368–378. https://doi.org/10.1016/j.molcel.2013.03.009

Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D., & von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics*, *15*(18), 3163–3168. https://doi.org/10.1002/pmic.201400441

Wang, R. (2005). Mass spectrometry of the M. smegmatis proteome: Protein expression

levels correlate with function, operons, and codon bias. *Genome Research*, *15*(8), 1118–1126. https://doi.org/10.1101/gr.3994105

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Waudby, C. A., Dobson, C. M., & Christodoulou, J. (2019). Nature and Regulation of Protein Folding on the Ribosome. https://doi.org/10.1016/j.tibs.2019.06.008

Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K., & Martoglio, B. (2002). Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science*, *296*(5576), 2215–2218. https://doi.org/10.1126/science.1070925

Wells, J. N., Bergendahl, L. T., & Marsh, J. A. (2015). Co-translational assembly of protein complexes. *Biochemical Society Transactions*, *43*, 1221–1226. https://doi.org/10.1042/BST20150159

Wells, J. N., Bergendahl, L. T., & Marsh, J. A. (2016). Operon Gene Order Is Optimized for Ordered Protein Complex Assembly. *Cell Reports*, *14*(4), 679–685. https://doi.org/10.1016/j.celrep.2015.12.085

Wild, K., Aleksić, M., Lapouge, K., Juaire, K. D., Flemming, D., Pfeffer, S., & Sinning, I. (2020). MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-14603-7

Williams, C. C., Jan, C. H., & Weissman, J. S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science*, *346*(6210), 748–751. https://doi.org/10.1126/science.1257522

Wilson, D. N., & Cate, J. H. D. (2012). The Structure and Function of the Eukaryotic Ribosome, 1–17.

Wolin, S. L., & Walter, P. (1989). Signal recognition particle mediates a transient elongation arrest of preprolactin in reticulocyte lysate. *Journal of Cell Biology*, *109*(6 I), 2617–2622. https://doi.org/10.1083/jcb.109.6.2617

Woolhead, C. A., McCormick, P. J., & Johnson, A. E. (2004). Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell*, *116*(5), 725–736. https://doi.org/10.1016/S0092-8674(04)00169-2

Wu, C. C. C., Peterson, A., Zinshteyn, B., Regot, S., & Green, R. (2020). Ribosome Collisions Trigger General Stress Responses to Regulate Cell Fate. *Cell*, *182*(2), 404-416.e14. https://doi.org/10.1016/j.cell.2020.06.006

Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O., & Pei, D. (2010). Protein N-terminal processing: Substrate specificity of escherichia coli and human methionine aminopeptidases. *Biochemistry*, *49*(26), 5588–5599. https://doi.org/10.1021/bi1005464

Xu, Y., Anderson, D. E., & Ye, Y. (2016). The HECT domain ubiquitin ligase HUWE1 targets unassembled soluble proteins for degradation. *Cell Discovery*, *2*, 1–16. https://doi.org/10.1038/celldisc.2016.40

Yanagitani, K., Juszkiewicz, S., & Hegde, R. S. (2017). UBE2O is a quality control factor for orphans of multiprotein complexes. *Science*, *357*(6350), 472–475. https://doi.org/10.1126/science.aan0178

Ye, Q., & Worman, H. J. (1995). Protein-Protein Interactions between Human Nuclear Lamins Expressed in Yeast. *Experimental Cell Research*, *219*(1), 292–298. https://doi.org/10.1006/excr.1995.1230

Young, J. C., & Andrews, D. W. (1996). Assembles Co-Translationally on the Endoplasmic Translation Pause in Vitro, *15*(1), 172–181.

Yu, C.-H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell*, *59*(5), 744–754. https://doi.org/10.1016/j.molcel.2015.07.018

Zhang, G., Hubalewska, M., & Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural and Molecular Biology*, *16*(3), 274–280. https://doi.org/10.1038/nsmb.1554

Zhang, G., & Ignatova, Z. (2011). Folding at the birth of the nascent chain: Coordinating translation with co-translational folding. *Current Opinion in Structural Biology*, *21*(1), 25–31. https://doi.org/10.1016/j.sbi.2010.10.008

Zhang, L., Paakkarinen, V., Van Wijk, K. J., & Aro, E. M. (1999). Co-translational assembly of the D1 protein into photosystem II. *Journal of Biological Chemistry*, *274*(23), 16062–16067. https://doi.org/10.1074/jbc.274.23.16062

Zhang, Y., De Laurentiis, E., Bohnsack, K. E., Wahlig, M., Ranjan, N., Gruseck, S., … Rospert, S. (2021). Ribosome-bound Get4/5 facilitates the capture of tail-anchored proteins by Sgt2 in yeast. *Nature Communications*, *12*(1), 1–17. https://doi.org/10.1038/s41467-021-20981-3

Zhao, T., Chen, Y. M., Li, Y., Wang, J., Chen, S., Gao, N., & Qian, W. (2020). Disome-seq reveals widespread ribosome collisions that recruit co-translational chaperones. *BioRxiv*, 1–32. https://doi.org/10.1101/746875

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., … Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, *495*(7439), 111–115. https://doi.org/10.1038/nature11833

Zipser, D. (1963). Studies on the ribosome-bound β-galactosidase of Escherichia coli. *Journal of Molecular Biology*, *7*(6), 739–751. https://doi.org/10.1016/S0022-2836(63)80120-5

9. Publications

M. Bertolini*, **K. Fenzl***, I. Kats, F. Wruck, F. Tippmann, J. Schmitt, J. Auburger, S. Tans, B. Bukau, G. Kramer, Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science, 371*(6524), 57–64, (2021), https://doi.org/10.1126/science.abc7151

*** equal contribution**

U.A. Friedrich*, M. Zedan*, B. Hessling, **K. Fenzl**, L. Gillet, J. Barry, M. Knop, G. Kramer, B. Bukau, N^{α}-terminal acetylation of proteins by NatA and NatB serves distinct physiological roles in *Saccharomyces cerevisiae*. *Cell Reports,* (2021), https://doi.org/10.1016/j.celrep.2021.108711

J. Bohlen, **K. Fenzl**, G. Kramer, B. Bukau, A. Teleman, Selective 40S Footprinting Reveals Cap-Tethered Ribosome Scanning in Human Cells. *Molecular Cell*, (2020), https://doi.org/10.1016/j.molcel.2020.06.005

J. Bohlen, L. Harbrecht, S. Blanco, K. Clemm von Hohenberg, **K. Fenzl**, G. Kramer, B. Bukau, A. Teleman, DENR promotes translation reinitiation via ribosome recycling to drive expression of oncogenes including ATF4. *Nat Commun* 11, 4676 (2020). https://doi.org/10.1038/s41467-020-18452-2

In preparation:

F. Wruck, J. Schmitt*, **K. Fenzl***, M. Bertolini*, A. Katranidis, B. Bukau, G. Kramer, S. Tans, The conformational basis of Lamin co-translational assembly.

*** equal contribution**

10. Supplemental Material

All custom-made Python, Julia and R scripts employed for this thesis can be found on our DataKramer Server (".../DataKramer/Matilde_Kai_backup/"). This folder contains in addition all raw and all processed data sets and the respective R markdown files with the employed commands to process the raw sequencing data of this thesis. Also, plasmid maps and sequences of all constructed plasmids are stored in this folder. In addition, I uploaded all experimental protocols and an electronic version of my lab book, sorted according to each experiment (see ReadMe.txt file in the folder for further explanation). It also contains all presentations of my time as a PhD student in the Bukau lab. In addition, I uploaded an excel sheet with the gene names of all high- and low-confidence candidates. The same excel sheet and all Julia Scripts (written by Ilia Kats) can be also found as supplementary material published in Bertolini *et al.* 2021

10.1 Optimization of DiSP

10.1.1 Sucrose gradient optimization

The enrichment factor of disome reads over monosome reads highly depends on the separation efficiency of both populations. To obtain the best separation along a sucrose gradient, we had to optimize three factors: (i) pick the correct initial sucrose concentrations to form a linear gradient (the usual range of sucrose is 5-60% (v/v)), (ii) choose the correct centrifugation time for maximal separation, and (iii) obtain a proper linear sucrose gradient. The density of the sucrose solutions is always smaller than the density of the particles of interest, therefore, all loaded ribosomes will reach the bottom of the tube after a certain centrifugation time. Sucrose is the most used sedimentation media for biological macromolecules because it has good solubility in water, electrical neutrality, and does not disturb the absorbance measurements of rRNA at 260 nm, which is the used absorbance wavelength to detect ribosomes.

To omit endless, try and error runs for this multidimensional problem, we tried to simulate the sedimentation behavior of monosomes and disomes in all possible sucrose gradient variations (based on the pre-installed gradient mixing programs of our sucrose gradient forming station (BioComp). The simulations were based on a published algorithm to calculate the sedimentation coefficient of a particle in linear sucrose gradients (Dingman 1972, more details in Material & Methods).

Two aspects became clear from the sedimentation simulation iterated over all potential gradients (**Fig. 40**). First, higher concentrated heavy sucrose solutions lead to longer centrifugation times to get a deep migration of disomes to obtain the best monosome-disome separation. Second, lower light sucrose concentrations lead to further separated monosome and disome peaks. Therefore, a 5-20% linear sucrose gradient allows, in theory, the best monosome-disome separation (calculated distance of 2.45 cm between monosome and disome peak maxima) and obtains this also in the shortest possible centrifugation time of three hours.

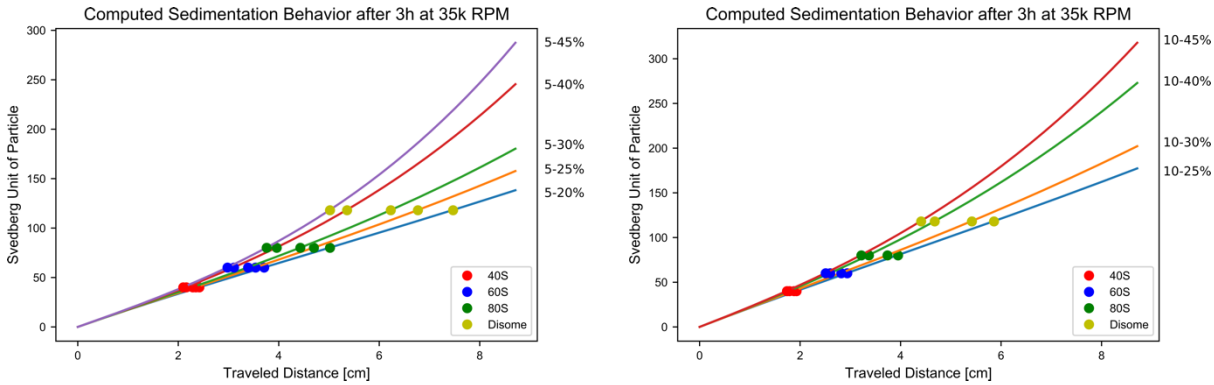

Fig. 40: Calculated sedimentation behavior of the ribosomal small subunit (40S), large subunit (60S), monosomes (80S) and disomes (118S) in different sucrose gradients. Ribosomal running behavior in a selection of potential sucrose gradients starting with a 5% (left) or 10% (right) light sucrose solution and varying heavy sucrose solutions from 20-45%. Both plots show the separation between monosomes and disomes in different sucrose gradients after three hours of centrifugation at 35k rpm with a SW40 rotor at 4°C (see Material & Methods for all fixed parameters required for the calculations and further explanations).

However, the use of a 5-20% sucrose gradient results in the pelleting of all higher ribosomal structures. For the optimization of DiSP it was not only important to get the best separation of monosomes and disomes, but also to observe higher polysome structures to judge the RNA digestion efficiency, which is another crucial factor to obtain good DiSP results (see result section **4.2.2**). A linear sucrose gradient of 5-45% fulfills both criteria (**Fig. 40**, calculated distance of 1.66 cm between monosome and disome peak maxima).

Only entirely linear sucrose gradients allow to obtain good particle separations in agreement with the calculated sedimentation profiles. Therefore, it is important to use the right mixing program. We observed that pre-installed mixing programs of our Gradient Station did not produce satisfying linear gradients (**Fig. 41**, left). Thus, we optimized in collaboration with David Coombs, the developer of the Gradient Station, the gradient mixing program for 5-45% sucrose gradients (**Fig. 41**, right).

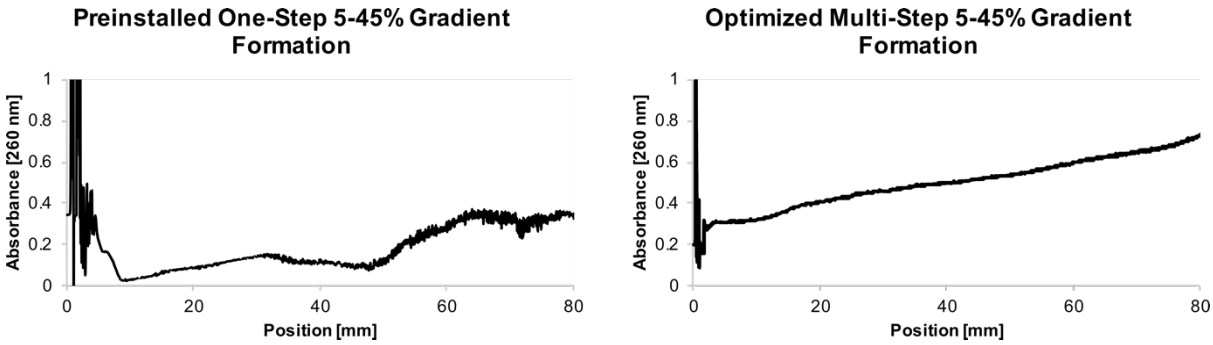

Fig. 41: Comparison of the pre-installed one-step gradient formation program for 5-45% sucrose gradients in comparison to an optimized multi-step mixing program. Both plots show the absorbance at 260 nm for two gradients that were supplemented with Trypan Blue in the heavy sucrose solution before gradient formation (see Material & Methods).

A missing factor in the sedimentation simulations was that it only calculated the distance between the peak maxima but not the overall peak shape. We observed that the peak shape became much broader in the 5-20% linear sucrose gradient in comparison to the 5-45% linear sucrose gradient (data not shown). A higher starting sucrose concentration allowed narrower peaks. Therefore, a 10-25% as well as 5-45% gradients with 3.5h of centrifugation were used for all further DiSP experiments. The 5-45% was used in the beginning to optimize the RNase digestion conditions (see below) and the 10-25% for later DiSP repetitions. The predicted sedimentation behavior was proven valid for both gradients (**Fig. 42**).

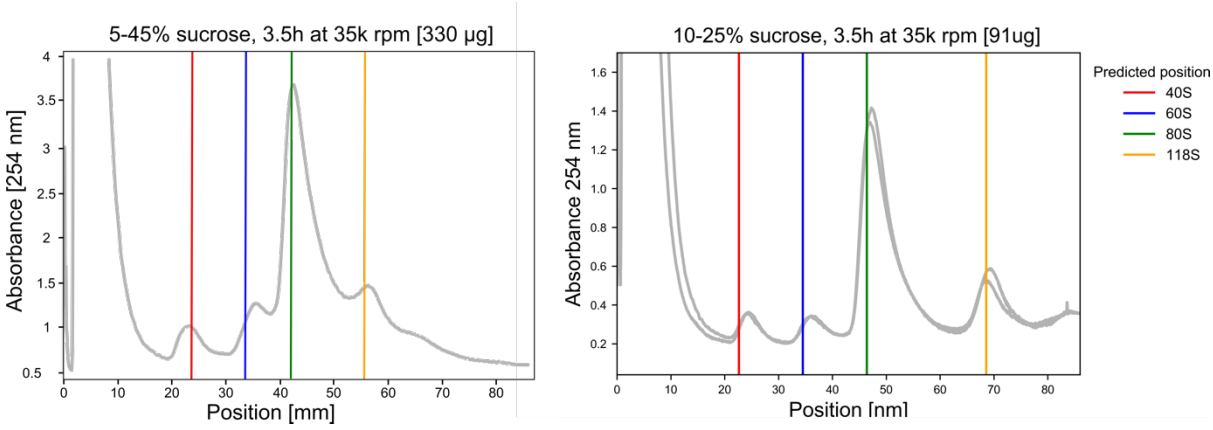

Fig. 42: Comparison of simulated and real sedimentation behavior of ribosomes. A 5-45% sucrose gradient (left) or a 10-25% sucrose gradient (right) after 3.5 hours of centrifugation at 35k rpm in a SW40 rotor. The colored lines indicate the calculated position of migration for the small and large ribosomal subunit as well as for monosomes (80S) and disomes (118S).

10.1.2 RNase digestion optimization

The last optimization that had to be established was the RNase amount to digest human polysomes into monosomes and disomes. It is known that "over-digestion" of human ribosomes can occur with RNase I, which can be observed in a sucrose gradient by a shrinkage of the monosome peak height, mostly caused by cleavage of ribosomal rRNA extension segments and not by ribosome disassembly (**Fig. 43A**). We observed that the phenomenon of monosome shrinkage was dependent on the RNase treatment time, temperature, as well as the total amount of employed RNase I (not all data shown). We also observed differences between HEK293-T and U2OS cells in the susceptibility to RNase I, which, to a certain extent, is caused by the two different harvest and lysis methods (**Fig. 43A**, **B,** see Material & Methods). These optimizations revealed that 60U RNase I for 40 μg total RNA was the optimal condition during U2OS cell harvest and lysis (see Material & Methods). U2OS lysates showed clear indications of ribosomal over-digestion with higher RNase amounts (**Fig. 43A**, left), whereas for the cell harvest and lysis of HEK293-T cells 30 min at 4°C with 150U RNase I for 40 μg total RNA led to an almost complete polysome digestion without strongly affecting the overall ribosome integrity (**Fig. 43B**). We also tested MNase in the same manner for U2OS cells and observed that the effect of ribosomal over-digestion was not detectable for human derived ribosomes, and that much higher amounts of MNase were required to obtain satisfactory polysome digestions (**Fig. 43A**, right). However, our first sequencing data of the U2OS monosome and disome fractions from both types of RNase digestion revealed that MNase digestions of eukaryotic ribosomes led to an increase in UTR- and a decrease of CDS-ribosomal footprints (**Fig. 43C**). This finding was in agreement with published data that describe the same phenomenon for MNase digested human samples (Miettinen and Björklund 2015). Therefore, we decided to use the optimized RNase I concentration for further experiments and did not focus on the artifacts produced by MNase.

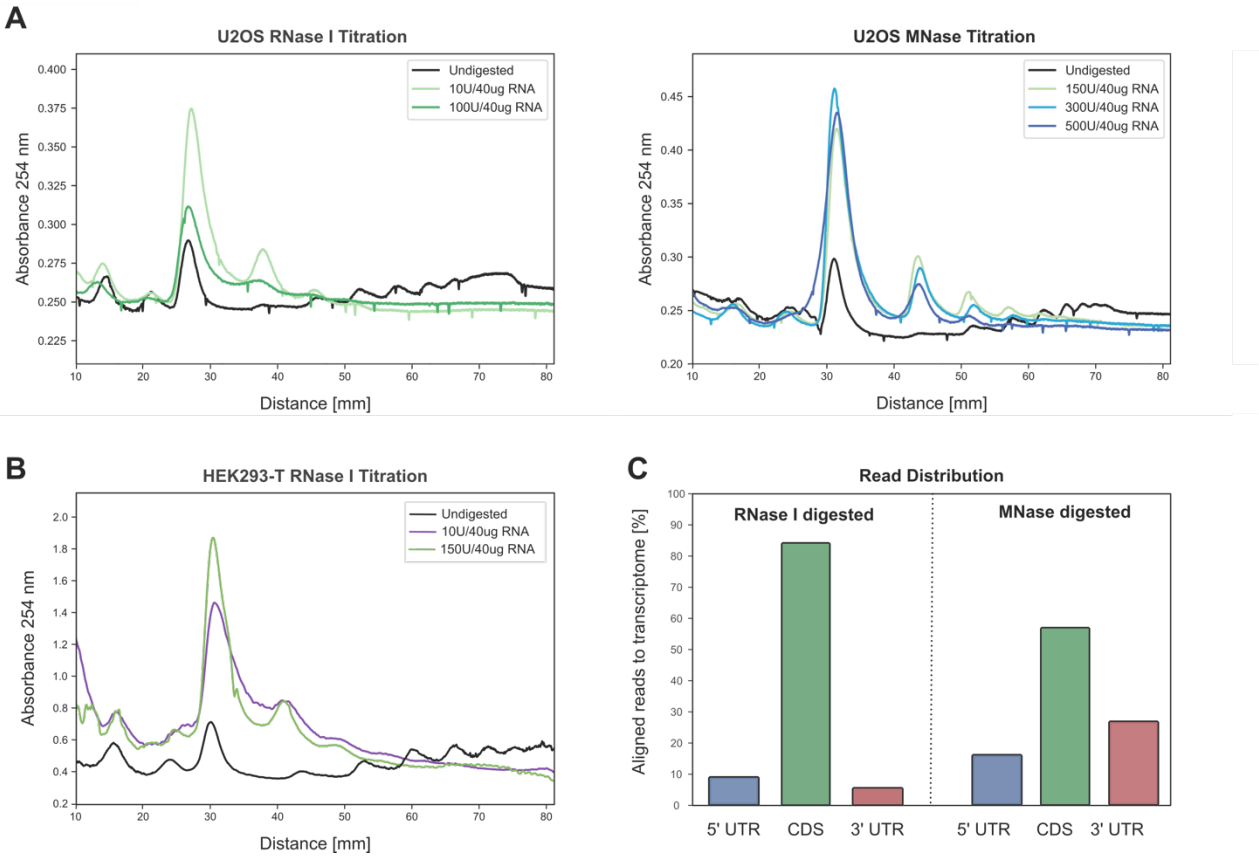

Fig. 43: Titration of RNase I or MNase for human lysate digestions. (A) 5-45% sucrose gradient profiles of undigested U2OS lysates (black solid lines) in comparison to RNase I digested (left) or MNase digested lysates (right). **(B)** 5-45% sucrose gradient profiles of undigested HEK293-T lysates (black solid lines) in comparison to RNase I digested lysates revealed a different susceptibility of ribosomes to RNase I in comparison to U2OS. **(C)** Aligned read distribution of RNase I and MNase digested monosome samples. RNase I digestion produces mostly reads that aligned to the coding sequence of all translated mRNAs (left), whereas MNase shows that over 40% of all reads aligned to untranslated mRNA regions (right).