Disome-seq reveals sequence-mediated coupling of translational pauses and protein structures

Taolan Zhao[1,2,†,*], Yanming Chen[1,2,3,†], Jia Wang[1,2,3], Siyu Chen[1,2,3], and Wenfeng Qian[1,2,3,*]

[1] State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[2] Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[3] University of Chinese Academy of Sciences, Beijing 100049, China

[†]Taolan Zhao and Yanming Chen contributed equally to this work.

*Correspondence: tlzhao@genetics.ac.cn or wfqian@genetics.ac.cn

ABSTRACT

The regulation of translation elongation plays a vital role in protein folding; an adequate translational pause provides time and cellular environments for the co-translational folding of nascent peptides. However, the genomic landscape, sequence determinants, and molecular consequences of translational pausing remain mostly unknown. In this study, we performed disome-seq that sequenced mRNA fragments protected by two consecutive ribosomes – a product of severe translational pauses during which the upstream ribosome collides into the paused one. We detected severe translational pauses on ~75% of yeast genes. These pauses were often explained by one of the three mechanisms: 1) slow ribosome releasing at stop codons, 2) slow peptide formation from proline, glycine, asparagine, and cysteine, and 3) slow leaving of polylysine from the exit tunnel of ribosomes. Notably, these amino acids also terminate the α-helical conformation. Such dual roles of amino acids establish an inborn coupling between the synthetic completion of a structural motif and a translational pause. Furthermore, paused ribosomes often recruit chaperones to assist protein folding. As a consequence, emergent protein structures during evolution should be ready to be correctly folded. Collectively, our study shows widespread translational pauses and sheds lights on a better understanding of the regulation of co-translational protein folding.

Keywords: Translation elongation; co-translational protein folding; disome-seq; ribosome collisions; translational pausing; ribosome releasing; ribosome-associated chaperones

INTRODUCTION

Translation elongation is a crucial process through which the genetic information in a transcript is sequentially decoded to a peptide chain by ribosomes. The mRNA sequence of coding regions can harbor more information than the amino-acid sequence; the local elongation rate is fine-tuned (Stein and Frydman, 2019). A change in translation elongation rate has been reported to lead to developmental abnormalities (Braat et al., 2004; Clark et al., 2000), neurologic diseases (Darnell et al., 2011; Ishimura et al., 2014), and cancers (Faller et al., 2015; Pelletier et al., 2015).

Translation elongation regulates biological processes likely through co-translational protein folding (Faure et al., 2017; Li et al., 2012; Sharma and O'Brien, 2018; Stein and Frydman, 2019). For example, it has been reported that bisulfate bond formation and protein conformation of crystallins were affected by the rate of translation elongation (Buhr et al., 2016). Consistently, accelerating translation elongation by exclusively using optimal synonymous codons could reduce the enzyme activity (Agashe et al., 2013). Furthermore, accumulating evidence supported that the CAG expansion in Huntington disease led to incorrect translational pauses and therefore, incorrect subcellular localization of the Htt protein (Nissley and O'Brien, 2016). During evolution optimal and non-optimal codons formed clusters (Cannarozzi et al., 2010; Chen et al., 2017; Clarke and Clark, 2008), creating fast and slow-translation regions in the same gene, which may participate the regulation of co-translational protein folding (Yu et al., 2015).

Despite the importance of the translation elongation, the knowledge of the causes and consequences for translational pausing remain highly limited, mainly because the detection of translational pauses was technically challenging at the genomic scale. The development of ribo-seq, an approach that sequences ribosome protected mRNA fragments in a high-throughput manner, significantly increased our knowledge on translation elongation in endogenous genes (Ingolia et al., 2009; Ingolia et al., 2011; Li et al., 2012). Accumulation of ribosome footprints at a site indicates slow translation elongation or translational pausing; based on this idea, sequence determinants of translation elongation have been discovered or disputed, such as synonymous codon usage, anti-Shine-Dalgarno sequence, positive charged amino acids, and mRNA secondary structures (Charneski and Hurst, 2013; Hussmann et al., 2015; Ingolia et al., 2009; Li et al., 2012; Mohammad et al., 2016; Qian et al., 2012; Weinberg et al., 2016; Yang et al., 2014).

However, ribo-seq misses the information of ribosome collisions – an extreme

form of translational pausing that the upstream ribosome chases up and collides into the downstream pausing ribosomes. Such severe translational pauses may play a more critical role in co-translational protein folding. Consecutive ribosomes (disomes) were previously detected by sucrose gradient centrifugation (Diament et al., 2018). In this study, we precisely captured disomes and systematically characterized the landscape of severe translational pauses in the yeast genome. Taking advantage of such information, we identified the sequence determinants of severe translational pauses. We proposed a new model on the coevolution of the protein structure and translational pausing.

RESULTS

Disomes were resistant to RNase digestion

We extracted mRNA from yeast cells together with ribosomes, digested the unprotected mRNA using RNase I, and performed sucrose gradient ultracentrifugation to separate particles with different densities (**Figure 1A**). In addition to the abundant monosome particles, we observed a significant amount of particles (indicated by the red arrow in **Figure 1A**) whose density was similar to the pre-digestion transcripts bound by two ribosomes, suggesting that the particle was composed of two ribosomes.

Although such "two-ribosome" particles could be explained by the incomplete digestion of RNase I, it is unlikely the case for the following reason. As we increased the dosage of RNase I from 100 U to 500 U, the amount of monosome reduced and that of free ribonucleoproteins (RNP) increased, suggesting the disruption of ribosomes by the extra dose of RNase I (**Figure 1B**). However, the two-ribosome particles persisted (**Figure 1B**), indicating that they were not an experimental artifact of incomplete digestion. Instead, the particle likely contained mRNA fragments protected by two consecutive ribosomes; the mRNA fragment between them was resistant to the RNase digestion likely due to the steric effect. We hereafter referred such RNase-resistant two-ribosome particles as to disomes – a consequence of ribosome collisions.

Disome-seq accurately detected severe translational pauses

To determine if the high-throughput sequencing of the disome-protected mRNA (disome-seq, **Figure 2A**) could detect severe translational pauses, we induced

translational pauses with 3-Amino-1,2,4-triazole (3-AT), an inhibitor of histidine biosynthesis (Klopotowski and Wiater, 1965); the histidine deprivation stalled ribosomes at the histidine codons, which could further cause a ribosome collision with the upstream (5′-)ribosome in the transcript.

In ribo-seq, the 28 and 29-nucleotide (nt) footprints were most abundant (**Figure 2B**), consistent with previous observations (Ingolia et al., 2009). In contrast, the most abundant footprints in disome-seq were 58 or 59 nts (**Figure 2B**). We mapped the monosome and disome footprints to the yeast genome. The 5′-end of the 28-nt monosome footprints displayed an apparent 3-nt periodicity (**Figure 2B**), which is consistent with the non-overlapping 3-nt genetic code – ribosome moves three nucleotides per step during translation. Notably, the 58-nt disome footprints displayed a similar 3-nt periodicity (**Figure 2B**), indicating that a 58-nt disome footprint likely contained a 28-nt monosome footprint at the 5′-end. As a negative control, mRNA-seq reads did not display the 3-nt periodicity (**Figure S1A**).

During histidine starvation, ribosomes should be stalled by histidine codons CAC and CAT when they are at the A-site, where codon-anticodon recognition takes place. The 5′-end of the monosome footprints exhibited the main peak 15-nt upstream of the histidine codons in the yeast genome, indicating that the A-site of the ribosome locates at the 16th–18th nts in the 28-nt monosome footprints (**Figure 2C**); this is consistent with previous observations (Ingolia et al., 2009). Notably, disome footprints exhibited a sharp peak 45-nt upstream of the histidine codons, which was 30-nt upstream of the 15-nt peak in ribo-seq (**Figure 2C**). The 30-nt space between the two peaks perfectly fits one in-frame ribosome. Therefore, we proposed that the 58-nt disome footprints were composed of two 28-nt monosome footprints (**Figure 2C**). The 2-nt space between them did not allow a further move of the 5′-ribosome, which requested three nucleotides, validating disome footprints as the product of ribosome collisions.

Disome-seq detected the translational pauses that were missed in ribo-seq

To identify the genomic locations of severe translational pauses in fast-growing cells, we applied disome-seq to yeast cells cultivated in the rich medium (**Figure 3A** and **Figure S2**). Disome footprints were observed in a total of 3929 genes, which represented over 70% of the 5349 translated genes (**Figure 3B**), indicating widespread translational pauses in the yeast genome.

Disome-seq often detected translational pauses undetectable in ribo-seq (**Figure 3C**). Take *CYC8*, a gene encoding a general transcriptional co-repressor which can form the prion [OCT+] (Patel et al., 2009), for example, a translational pause downstream of polyglutamine (polyQ) was detected by disome-seq instead of ribo-seq (**Figure 3D**). Similarly, TEF4p, the γ subunit of translational elongation factor eEF1B, contains a lysine-rich region that is often ubiquitinylated or succinylated (Fang et al., 2014; Weinert et al., 2013); the translational pause downstream of it was uniquely detected by disome-seq (**Figure 3D**). The traditional ribo-seq lacks sufficient power in detecting all translational pauses (**Figure 3C**) probably due to the omission of disome protected fragments that are at least partially resistant to RNase digestion (**Figure 3E**).

Severe translational pauses were induced by stop codons

To identify the causes of severe translational pauses, we searched for sequence features at the A-site, the P-site, and the exit tunnel of the 3′-ribosome of a disome, respectively. We defined an A-site pausing score for each of the 64 codons as the enrichment of a codon at the A-site of the 3′-ribosome of a disome footprint (i.e., the propensity of a codon to induce severe translational pauses, **Figure 4A**). Taking the codon GAA for example (**Figure 4A**), all disome fragments on one gene were classified into two categories based on the codon at the A-site, GAA or others. The odds ratio was estimated for each gene with the corresponding codon frequency in mRNA as a control; the common odds ratio across all genes was calculated with the Mantel-Haenszel procedure and was defined as the A-site pausing score (**Figure 4A**). A significantly >1 A-site pausing score implies a slow recognition of the codon.

Three stop codons exhibited highest A-site pausing scores (**Figure 4B**). Consistently, disome reads were accumulated in individual genes (**Figure 4C**) as well as globally (**Figure 4D**), suggesting that ribosome releasing is slow. The stop codon TAA showed the smallest A-site pausing score among the three (**Figure 4B)**, suggesting TAA releases ribosomes at the highest rate. Notably, TAA was also the most frequent stop codon in the yeast genome, used in 48% genes (**Figure 4B inset**), suggesting that genes are likely under natural selection to reduce ribosome sequestering through using TAA as the stop codon. We predicted that the selection should be stronger in more highly expressed genes as the impact of severe translational pauses scales the expression level. Indeed, the proportion of genes using TAA increased to 65% in the top 10% most highly expressed genes (**Figure 4B inset**).

Severe translational pauses were not induced by non-optimal codons

Codons encoding amino acids are recognized by tRNA species with different copy numbers (Hani and Feldmann, 1998). Codons recognized by low-copy tRNA species, or non-optimal codons, potentially slow down translation elongation (Curran and Yarus, 1989; Sorensen et al., 1989) and lead to ribosome collisions. To test this possibility, we estimated codon optimality from the relative synonymous codon usage (RSCU) in the yeast genome (Sharp and Li, 1987). RSCU was not correlated with the A-site pausing score ($\rho = -0.02$, $P = 0.9$, Spearman's correlation, **Figure 4E**), indicating that the usage of non-optimal codons is not a significant cause of severe translational pauses in the rich media.

Severe translational pauses frequently occurred at the amino acids that terminate α-helixes

We similarly estimated the P-site pausing score for each of the 20 amino acids since codon identity unlikely matters at the P-site. Some amino acids (e.g., proline, glycine, asparagine, cysteine, and lysine) showed higher P-site pausing scores (**Figure 5A**). Among these amino acids, proline and glycine are poor substrates for the formation of a peptide bond (Nelson et al., 2008; Rodnina, 2016), and therefore, may induce translational pauses. It remains unclear how other amino acids induce translational pauses. Nevertheless, these amino acids share the same property that weakens the stability of the α-helix, a prominent secondary structure of proteins (Nelson et al., 2008). Proline and glycine are conformational either too inflexible or flexible to form an α-helix. The bulk and shape of asparagine and cysteine also destabilize the α-helix. These four amino acids rank top 4 among 20 amino acids in the propensity to terminate an α-helical conformation ($P = 0.0004$, Mann-Whitney U test, **Figure 5B**) (Bryson et al., 1995; Myers et al., 1997). Besides, positively charged residues (e.g., polylysine) repel each other and also prevent the formation of an α-helix (Nelson et al., 2008).

We speculate that a prolonged translational pause occurring right after the translational completion of an α-helix provides sufficient time for the folding of the motif. Such coupling was observed at the genomic scale – disomes induced by prolines took place after the termination of an α-helix (**Figure 5C**), and in individual genes (*TDH3* as an example in **Figure 5D**). Furthermore, such pausing signals were

often evolutionarily conserved despite within the intrinsically disordered regions that often evolutionarily dynamic (Brown et al., 2010). For example, the disome-inducing proline in TDH3p was conserved across eukaryotic species (**Figure 5D**). Collectively, these observations raise the possibility that translational pauses at the end of an α-helix are favored by natural selection to avoid the interference of protein folding by the downstream peptides.

Severe translational pauses were induced by polylysine in the exit tunnel

We further estimated the exit-tunnel pausing score from the enrichment of each of the (20^3 =) 8000 amino-acid 3-mers in the 20-amino-acid region upstream of the P-site of the 3′-ribosome. Triple-lysine scored the highest, along with a few other positively charged 3-mers (**Figure 6A**). It has been proposed that the positively charged amino acids of the nascent peptide could hinder translation elongation by interacting with the negatively charged exit tunnel (Charneski and Hurst, 2013). However, lysine appeared to be more efficient in inducing translational pauses than the other two positively charged amino acids, arginine and histidine (**Figure 6A**).

To validate the difference in inducing translational pauses among the three positively charged amino acids, we constructed a dual-luciferase reporter system with firefly and renilla-luciferases expressed from two different promoters. The sequence to be tested was inserted after the start codon of the firefly-luciferase (**Figure 6B** and **Table S1**), and therefore, a reduction in the firefly/renilla ratio indicates a translational pause induced by the inserted sequence. We observed more significant reductions in the firefly/renilla ratio as the number of the consecutive lysine increases (**Figure 6B**, left panel); the reduction was not explained by mRNA levels (**Figure S3**), suggesting a cumulative inhibitory effect of lysines on translation. Furthermore, the firefly/renilla ratio reduced more when lysine was inserted, than arginine and histidine (**Figure 6B**, middle and right panels), indicating that, in addition to the positive charge, the bulk and shape of the lysine residue may boost translational pauses.

The consecutive lysine often appears in linker regions of domains because the repellence among the positively charged residues prevents the formation of a domain (Nelson et al., 2008). As an example, polylysine-induced translational pausing took places in the linker regions between two domains of *RPS31* (**Figure 6C**). Despite in an intrinsically disordered region, these positively charged amino acids (five lysines and one arginine) were conserved across eukaryotes kingdoms (**Figure 6C**). The translational pauses induced by the polylysine may benefit the cell by providing time

for the folding of protein domains.

Severe translational pauses recruited chaperones

Ribosomes are heterogeneous (Shi and Barna, 2015; Simsek et al., 2017). To understand the outcomes of the severe translational pauses on the composition of ribosomes, we identified disome-specific ribosome components by labeling ribosome proteins with stable isotopes (**Figure 7A**). We separated disomes from monosomes with sucrose density gradient centrifugation, mixed heavy-labeled disome with light-labeled monosome (or light-labeled disome with heavy-labeled monosome in a replicate), and performed tandem mass spectrometry (MS/MS, **Figure 7A**). In addition to ribosomal proteins, some highly expressed metabolic enzymes were also identified; we speculated that they were nascent peptides of incomplete protein translation and were attached to ribosomes. Consistently, the peptides of these enzymes captured by MS/MS tended to be in the first half of coding sequences (common odds ratio = 1.9, $P = 0.001$, Mantel-Haenszel test). For example, all detected peptides of FAS2p, a fatty acid synthetase, were in the first half (**Figure S4B**). By contrast, the peptides of ribosomal proteins captured by MS/MS were evenly distributed between the first and second half of coding sequences (common odds ratio = 1.0, $P = 0.7$, Mantel-Haenszel test).

Most proteins showed similar abundance in disomes and monosomes and therefore, were on the diagonal in the scatter plot showing protein intensities in disomes vs. in monosomes (**Figure 7B** and **Figure S4A**); many were the constitutive ribosome components (**Figure 7B** and **Figure S4A**). Proteins above the diagonal line were disome-enriched proteins (**Figure 7B** and **Figure S4A**; **Tables S2**, **S3**). Notably, all of the 11 chaperones identified in MS/MS had higher abundance in disomes than in monosomes ($P = 2.1\times10^{-4}$, Fisher's exact test, **Figure 7B**; $P = 6.0\times10^{-3}$, **Figure S4A**), indicating that chaperones are recruited during severe translational pauses. For example, whereas SSB1p was reported as a ribosome-associated chaperone (Doring et al., 2017) how it was distributed among ribosomes remained unclear. We showed that SSB1p were enriched in the disome fraction (**Figure 7B-C**), likely promoting the co-translational folding of the nascent peptide upstream to a translational pause (**Figure 7D**).

DISCUSSION

Translation elongation plays a central role in regulating protein function, subcellular localization, and aggregation (Nissley and O'Brien, 2016; Sharma and O'Brien, 2018; Stein and Frydman, 2019). The key to studying such regulation is to identify the positions of ribosomes in transcripts. Ribo-seq solves the problem to a large extent (Ingolia et al., 2009) but misses collided ribosomes, which provide crucial information about severe translational pausing. In this study, we performed disome-seq that specifically and sensitively detected the genomic positions of severe translational pauses. Such analyses enabled us to identify the leading sequence features regulating translation elongation.

It is worth noting that disome footprints studied here are fundamentally different from the transcripts binding to two ribosomes, in which two ribosomes are usually sparsely distributed on an mRNA, and the mRNA fragments between them are not resistant to RNase digestion. To illustrate this point experimentally, we collected the transcripts binding to two ribosomes from the sucrose gradient centrifugation, digested them using RNase I, and performed ribo-seq (**Figure S5A**). As expected, the footprints observed in the disome-seq did not show up in such experiments (**Figure S5B**).

In addition to the 58-nt disome fragments, the 61-nt fragments also exhibited an apparent 3-nt periodicity (**Figure 2B**), indicating the second conformation of disomes. As indicated by the disomes accumulated at the stop codons, 61-nt disome fragments were composed of two 28-nt monosome fragments spaced by five nucleotides (**Figure S6A**). Although in principle the 5′-ribosome in such disomes could make another move, two ribosomes were sufficiently close to each other in such quasi-collision that the mRNA fragment between them was at least partially resistant to RNase digestion. The P-site pausing scores estimated from 58 or 61-nt disome fragments were highly correlated ($\rho = 0.88$, $P = 1.2\times10^{-6}$, Spearman's correlation, **Figure S6B**), indicating that the 61-nt quasi-collisions also reflected severe translational pauses. Therefore, they were combined with the 58-nt collisions in this study.

Severer translational pauses may further lead to the collision of three consecutive ribosomes – trisomes. Indeed, after RNase digestion a small number of particles was observed at ~62 mm of the sucrose gradient centrifugation (**Figure 1A**); the density of these particles was higher than disomes and was similar to transcripts binding to three ribosomes, indicating they were trisomes. Although the abundance of these trisomes was insufficient for a trisome-seq, the presence of trisomes was detectable in disome-

seq – footprints with the 5′-end ~75-nt upstream to the stop codon (**Figure S6C**). These disome footprints were generated likely through RNase digestion between the mid- and the 3′-ribosome in a trisome (**Figure S6C**) since the mRNA fragments between consecutive ribosomes were only partially resistant to RNase digestion.

In addition to the sequence features identified at the A-site, P-site, and exit tunnel, others may also play a role. For example, mRNA secondary structure has been shown to cause translational pausing (Faure et al., 2017; Tang et al., 2016; Yang et al., 2014), which was also detected in this study; mRNA downstream of disomes exhibited stronger secondary structure (**Figure S7**). Relatedly, RNA helicase was recruited by disomes (**Figure 7** and **Tables S2**, **S3**), potentially removing mRNA secondary structures to relief translational pauses.

Non-optimal codons have been reported to be translated slower (Curran and Yarus, 1989; Sorensen et al., 1989) likely due to the limited supply of cognate tRNAs. However, non-optimal codons did not appear to induce severe translational pauses in the rich media (**Figure 4E**). One possible explanation is that the balanced use of tRNA and codons has minimized the difference in the elongation rate among optimal and non-optimal codons (Qian et al., 2012), to the level that the non-optimal codons are not sufficiently slow to trigger ribosome collisions in the rich media. By contrast, when treated with 3-AT that leads to histidine starvation, the non-optimal codon of histidine, CAT, induced severer translational pauses, as indicated by the taller blue peaks in **Figure 2C**. CAT and CAC are recognized by the same tRNA, suggesting that the difference in the propensity to induce translational pauses does not arise from the concentration of isoaccepting tRNAs. Instead, the difference between synonymous codons in the binding affinity to the anticodon GUG probably plays a role.

Yeast genes often use non-optimal codons at the 5′ of the coding sequences, which has been speculated to serve as a “ramp” – ribosomes collide early during translation instead of late to mitigate ribosomes sequestering (Tuller et al., 2010). We failed to detect the accumulation of disomes around the start codon in either ribo-seq or disome-seq in this study (**Figure 4D**). Previously observed accumulation of ribosomes around the start codon (Ingolia et al., 2009) has been shown as a byproduct of the pre-treatment of cycloheximide (Hussmann et al., 2015; Pelechano et al., 2015; Weinberg et al., 2016), which partially inhibited translation elongation but did not block translational initiation (Pestova and Hellen, 2003).

Ribosome release often led to severe translational pauses (**Figure 4B**) probably because all three stop codons are recognized by the same release factor, eRF1. The

glutamic acid at the 55th position (Glu55) of eRF1 plays a role in forming a hydrogen bond with the 2nd and/or 3rd adenine of the stop codon (Brown et al., 2015). The "wobble effect" in both positions (versus at most one wobble position for codon-anticodon recognition) may slow down the recognition between a stop codon and eRF1. Consistently, the stop codon TAA that forms two hydrogen bonds with Glu55 releases ribosomes fastest among the three (**Figure 4B**). Although the slow ribosome release was missed in the traditional ribo-seq studies (Gerashchenko et al., 2012; Ingolia et al., 2009), the phenomenon was observed by 5′P sequencing that traced the location of the most upstream translating ribosome in a degraded mRNA (Pelechano et al., 2015).

On the one hand, it is costly when ribosomes are sequestered upstream of the stop codon waiting for releasing since they are not recycled for a new round of protein synthesis. On the other hand, the slow ribosome release may benefit the cell by providing the newly synthesized peptide sufficient time to fold within the exit tunnel of ribosomes, rather than in the complex cytoplasmic environment crowded with all kinds of proteins (**Figure 7D**).

In addition to the stop codons, amino-acid sequence features that can lead to severe translational pauses were identified, proline and glycine at the P-site as well as polylysine in the exit tunnel. They are also well-known signals for the termination of α-helixes during protein folding (**Figure 5B**). Therefore, these amino acids play two roles: terminating an α-helical conformation and inducing a translational pause (**Figure 7D**). The latter provides sufficient time for the folding of the newly synthesized α-helixes.

Although some sequence determinants of translational elongation have been identified, it remains unclear how these determinants were placed during evolution to the genomic locations that a pause is required, and especially, as protein structures evolve, how proteins correctly fold when the translation pausing signals are yet to be placed. Our "dual-role" model does not require evolutionary time for placing a pause signal after a change in structure – the amino-acid substitution that results in a new protein structure on the same time generates a translational pause, making the new protein structure ready to be correctly folded. Together with the translational pauses induced by stop codons and chaperones recruited by collided ribosomes, the co-translational protein folding is warranted (**Figure 7D**).

METHODS

Polysome profiling and disome-seq

The laboratory strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 uraΔ0*) was cultivated at 30ºC in the rich media YPD (1% yeast extract, 2% peptone, and 2% dextrose) or in the synthetic complete media with histidine dropped out and with 100 mM 3-AT added (SC–His+3-AT). Cells were collected at OD_{660} ~0.6. Ribosomes were extracted with the polysome lysis buffer (PLB), which contained 200 mM Tris-HCl (pH 8.0), 200 mM KCl, 35 mM $MgCl_2$, 1% (v/v) Triton X-100, 5 mM DTT, and 50 μg/mL cycloheximide.

For the polysome profiling experiments, the extracted ribosomes were pelleted through a 30 mL sucrose cushion containing 400 mM Tris-HCl (pH 8.0), 200 mM KCl, 30 mM $MgCl_2$, 1.75 M sucrose, 5 mM DTT, and 50 μg/mL cycloheximide, by ultracentrifugation at 4°C overnight (33,500 rpm, Beckman, 70Ti rotor). The ribosome pellet was dissolved in 300 μL resuspension buffer (20 mM Tris-HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, and 50 μg/mL cycloheximide), and separated by ultracentrifugation at 4°C for 3 hours (35,300 rpm, Beckman, SW41 rotor) through a 5-50% sucrose gradient (40 mM Tris-HCl pH 8.4, 20 mM KCl, 10 mM $MgCl_2$, and 50 μg/mL cycloheximide) prepared by Gradient Master (Biocomp). The profiling signals were recorded by Piston Gradient Fractionator (Biocomp).

For the disome-seq experiments, the ribosome dissolved in the PLB buffer were treated with RNase I (Ambion, AM2294) at 25°C for 2 hours. RNA was extracted from the enzyme reaction system with hot phenol and was separated on a 17% (w/v) 7 M urea denaturing polyacrylamide gel in a 0.5×Tris-borate-EDTA (TBE) electrophoresis buffer. RNA fragments with the length of 25-30 nts or 50-80 nts were extracted by gel crushing and further incubated with an RNA gel extraction buffer (300 mM NaOAc pH 5.2, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) overnight. The purified RNA fragments were subjected to small RNA library construction for Illumina sequencing (Gnomegen, k02420). The 5′-RNA adaptor contained a 3-nt random sequence at the 3′-end to avoid biased ligation. Monosome and disome footprints were sequenced with single-end 50 and paired-end 100 modes on BGISEQ-500 (BGI Group), respectively.

Mapping reads to the yeast genome

The 3-nt random sequence at the 5′-end of each sequencing read was removed. The sequence identical to the 3′-sequencing adaptor was removed in each read using cutadapt V1.16 (http://gensoft.pasteur.fr/docs/cutadapt/1.6/index.html). The trimmed

reads shorter than 20 nts were excluded for further analysis. Reads without 3′-adaptor sequence were also removed since their lengths (>46 nts for ribo-seq or >96 nts for disome-seq) were much greater than the expectation (25-30 nts for ribo-seq and 50-80 nts for disome-seq).

The *S. cerevisiae* genome (SGD R64-1-1) (https://www.yeastgenome.org/) was used as the reference. The trimmed reads were mapped against rRNA with bowtie V1.2.2 (Langmead et al., 2009) (http://bowtie-bio.sourceforge.net/manual.shtml); the mapped reads were eliminated to avoid rRNA contamination. The non-rRNA reads were further aligned against the coding sequences with the –no-novel-juncs parameter using Tophat V2.1.1 (Kim et al., 2013) (https://ccb.jhu.edu/software/tophat/manual.shtml). Reads with multiple alignments or of mapping quality < 30 were discarded. Biological replicates were highly correlated (**Figure S1B**) and were combined in the majority of our analyses.

The definition of the A-site in a disome footprint

For 58-nt disome footprints, if the 5′-end of a read was mapped to frame 0, the 46th-48th nts were defined as the A-site of the 3′-ribosome; if mapped to frame 1 (or frame 2), the 45th-47th (or 47th-49th) nts were defined as the A-site since the footprint shifted by +1 and –1 nt during RNase I digestion. The A-site of the 3′-ribosome was also defined for the 59-nt disome footprints: if the 5′-end mapped to frame 0, meaning that one more nt was kept at the 3′-end during RNase I digestion, the 46th-48th nts were the A-site. If mapped to frame 2, meaning that one more nt was kept at the 5′-end during RNase I digestion, the 47th-49th nts were the A-site; if mapped to frame 1, most likely one more nt was kept at the 3′-end during RNase I digestion, and therefore, the 45th-47th nts were the A-site. For 61-nt and 62-nt disome footprints, the A-site was defined assuming that an extra codon was inserted into the space between two ribosomes.

The measure of codon optimality

Synonymous codon optimality was assessed by RSCU, which was retrieved from a previous study (Sharp and Li, 1987). It was further normalized by the number of synonymous codons of an amino acid (Qian et al., 2012), which mad the values of codons of different amino acids comparable.

The calculation of pausing scores

The A-site pausing score of each of the 64 codons was defined as the common odds ratio calculated by the Mantel-Haenszel test. A 2×2 table was generated for each gene. We divided disome footprints into two categories, the concerning codon at the

A-site of the 3′-ribosome and others. mRNA reads were used as the background to control for codon frequency in the gene; the 16th-18th nts of a 28-nt mRNA-seq read was considered as the A-site. The P-site pausing score was estimated similarly, with the amino acid at the P-site in consideration. The exit-tunnel pausing score was estimated for each of the amino-acid 3-mers in the 20 amino-acid upstream of the P-site; flanking 3-mers may hitchhike on the causal 3-mer, and therefore, we combined multiple disome reads mapped to the same genomic position into 1 to reduce such hitchhiking effect.

Statistical analyses were performed with *R* (https://www.r-project.org/), and plots were generated with an *R* package, ggplot2 (https://ggplot2.tidyverse.org/).

The definition of α-helixes

The protein secondary structure annotation of yeast protein was retrieved from UniProt (UniProt, 2019) (https://www.uniprot.org/). Adjacent α-helixes without a gap were concatenated. Short α-helixes composed of 4 amino acids or less were removed. The protein secondary structure of TDH3p was predicted with the online secondary structure prediction method, GOR IV (Garnier et al., 1996) at https://npsa-prabi.ibcp.fr/.

Prediction of the minimum free energy of an mRNA fragment

The minimum free energy of the 30-nt mRNA downstream of a disome footprint was calculated with the RNAfold V2.4.13 of the Vienna RNA package (Lorenz et al., 2011) (http://rna.tbi.univie.ac.at/).

Sequence alignment of genes from different species

Orthologous genes were identified by BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Amino-acid sequences from multiple species were aligned by Clustal X (Larkin et al., 2007). Species trees were constructed with TimeTree (Kumar et al., 2017) at http://www.timetree.org.

Dual-luciferase assays

The plasmid for the dual-luciferase assay, pTH727-CEN-RLuc/staCFLuc, was obtained from Addgene (http://www.addgene.org). The sequences of K_2 (2×AAG), K_4 (4×AAG), K_5 (5×AAG), K_6 (6×AAG), K_8 (8×AAG), K_{10} (10×AAG), R_5 (5×AGA), R_{10} (10×AGA), H_5 (5×CAC), H_{10} (10×CAC), A_5 (5×GCT), and A_{10} (10×GCT) were inserted into the plasmid, respectively, right downstream of the start codon of the firefly-luciferase. Individual plasmids were transformed into the BY4742 strain. The yeast clones containing plasmids were selected on an SC–uracil plate.

The level of luciferase was detected as described in a previous study (Merritt et al., 2010). Briefly, transgenic yeast strains were individually cultivated in the SC–uracil medium (150 μL) at 30°C overnight. 10 μL of each yeast culture was transferred into a fresh SC–uracil medium (140 μL) and was cultivated for another 4 hours in a 96-well plate. Passive lysis buffer (40 μL, Promega, E1910) was added per well for cell lysis. 25 μL of the suspension was mixed with the firefly luciferase substrate (25 μL, Promega, E1910) and was incubated for 20 minutes at 25°C. Firefly luciferase activity was measured by the Synergy multi-mode reader (BioTek). Stop-and-Glo reagent (25 μL, Promega, E1910) was added and mixed per reaction and was incubated for another 20 minutes before the measurement of renilla-luciferase activity.

The mRNA level of luciferase was detected by quantitative-PCR using the total RNA extracted from each transgenic yeast strain. The primer sequences were listed in **Table S4**.

Mass spectrometry

Stable isotope labeling with amino acids in cell culture (SILAC) was performed as described previously (Chen et al., 2019). Briefly, a strain modified from BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 arg4Δ0::kanMX4 car1Δ0::LEU2*) was cultured at 30°C in the SC media with arginine and lysine dropped out (SC–Arg–Lys) in the presence of the light amino-acid isotopes (30 mg/L Lys0 and 20 mg/L Arg0) or the heavy amino-acid isotopes (37.25 mg/L Lys8 and 20.94 mg/L Arg10).

Cells were harvested at the mid-log phase (OD_{660} ~0.6), and polysomes were extracted. After RNase I treatment, polysome profiling was performed, and proteins were precipitated from the monosome or disome fraction with a double volume of 95% ethanol, respectively. The protein precipitants were dissolved with the urea buffer (8 M urea, 1 mM sodium orthovanadate, 1 mM sodium fluoride, 2.5 mM sodium pyrophosphate, 1 mM B-glycerophosphate, 0.2% tablet of protease inhibitor, and 1 mM PMSF). 150 μg heavy-labeled monosome protein was mix with 150 μg light-labeled disome protein; 150 μg light-labeled monosome protein was mix with 150 μg heavy-labeled disome protein. The protein mixtures were subject to the MS/MS analysis on Orbitrap Elite (Thermo Fisher Scientific).

The raw data were processed with MaxQuant V1.5.8.3 (Cox and Mann, 2008) (https://www.biochem.mpg.de/5111795/maxquant) using default parameters. All raw data were searched against the yeast proteome with the addition of potential contaminants. The protease was set as trypsin/P and trypsin. Two missed cuts were

allowed.

If the peptides of metabolic enzymes were detected by MS/MS on ribosome particles because they were nascent peptides of incomplete protein translation, these peptides should be enriched in the first half (N-terminus) of a protein. To determine if it is true, we divided peptides detected from each gene into two categories, according to whether the first amino acid of the peptide belonged to the first half of the gene or not. With the theoretical digested peptide (arginine or lysine at the C terminus) as a background, a 2×2 table was generated for each gene. The common odds ratio among genes was calculated by the Mantel-Haenszel test.

ACKNOWLEDGMENTS

We thank Ting Li (IGDB, CAS) for technical support in ribosome profiling, Xiahe Huang and Yingchun Wang (IGDB, CAS) for technical support in proteomics, Addgene (http://www.addgene.org/) for providing plasmids, Rongxin Yang (IGDB, CAS) for technical help in molecular biological experiments, and Jinzhong Lin (Fudan University), Lucas Carey (Peking University), Ning Gao (Peking University), Xiaofeng Cao (IGDB, CAS) and Haotian Guo (Huazhong University of Science and Technology) for discussion.

This work was supported by grants from the National Natural Science Foundation of China (31900455 to T.Z. and 91331112 to W.Q.).

AUTHOR CONTRIBUTIONS

T.Z., Y.C., and W.Q. designed the experiments. T.Z. and J.W. performed the disome-seq. T.Z. and S.C. performed the mass spectrometry. T.Z. performed the mRNA-seq, luciferase assays, and molecular biological experiments. Y.C. and T.Z. performed data analyses. T.Z., Y.C., and W.Q. wrote the manuscript.

REFERENCES

Agashe, D., Martinez-Gomez, N.C., Drummond, D.A., and Marx, C.J. (2013). Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol Biol Evol *30*, 549-560.

Braat, A.K., Yan, N., Arn, E., Harrison, D., and Macdonald, P.M. (2004). Localization-dependent oskar protein accumulation; control after the initiation of translation. Dev Cell *7*, 125-131.

Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. Nature *524*, 493-496.

Brown, C.J., Johnson, A.K., and Daughdrill, G.W. (2010). Comparing models of evolution for ordered and disordered proteins. Mol Biol Evol *27*, 609-621.

Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T., and DeGrado, W.F. (1995). Protein design: a hierarchic approach. Science *270*, 935-941.

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M.V., and Komar, A.A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Mol Cell *61*, 341-351.

Cannarozzi, G., Schraudolph, N.N., Faty, M., von Rohr, P., Friberg, M.T., Roth, A.C., Gonnet, P., Gonnet, G., and Barral, Y. (2010). A role for codon order in translation dynamics. Cell *141*, 355-367.

Charneski, C.A., and Hurst, L.D. (2013). Positively charged residues are the major determinants of ribosomal velocity. PLoS Biol *11*, e1001508.

Chen, S., Li, K., Cao, W., Wang, J., Zhao, T., Huan, Q., Yang, Y.F., Wu, S., and Qian, W. (2017). Codon-Resolution Analysis Reveals a Direct and Context-Dependent Impact of Individual Synonymous Mutations on mRNA Level. Mol Biol Evol *34*, 2944-2958.

Chen, Y., Chen, S., Li, K., Zhang, Y., Huang, X., Li, T., Wu, S., Wang, Y., Carey, L.B., and Qian, W. (2019). Overdosage of Balanced Protein Complexes Reduces Proliferation Rate in Aneuploid Cells. Cell Syst.

Clark, I.E., Wyckoff, D., and Gavis, E.R. (2000). Synthesis of the posterior determinant Nanos is spatially restricted by a novel cotranslational regulatory mechanism. Curr Biol *10*, 1311-1314.

Clarke, T.F.t., and Clark, P.L. (2008). Rare codons cluster. PLoS One *3*, e3412.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Curran, J.F., and Yarus, M. (1989). Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J Mol Biol *209*, 65-77.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., *et al.* (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.

Diament, A., Feldman, A., Schochet, E., Kupiec, M., Arava, Y., and Tuller, T. (2018). The extent of ribosome queuing in budding yeast. PLoS Comput Biol *14*, e1005951.

Doring, K., Ahmed, N., Riemer, T., Suresh, H.G., Vainshtein, Y., Habich, M., Riemer, J., Mayer, M.P., O'Brien, E.P., Kramer, G., *et al.* (2017). Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. Cell *170*, 298-311 e220.

Faller, W.J., Jackson, T.J., Knight, J.R., Ridgway, R.A., Jamieson, T., Karim, S.A., Jones, C., Radulescu, S., Huels, D.J., Myant, K.B., *et al.* (2015). mTORC1-mediated translational elongation limits intestinal tumour initiation and growth. Nature *517*, 497-500.

Fang, N.N., Chan, G.T., Zhu, M., Comyn, S.A., Persaud, A., Deshaies, R.J., Rotin, D., Gsponer, J., and Mayor, T. (2014). Rsp5/Nedd4 is the main ubiquitin ligase that targets cytosolic misfolded proteins following heat stress. Nat Cell Biol *16*, 1227-1237.

Faure, G., Ogurtsov, A.Y., Shabalina, S.A., and Koonin, E.V. (2017). Adaptation of mRNA structure to control protein folding. RNA Biol *14*, 1649-1654.

Garnier, J., Gibrat, J.F., and Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol *266*, 540-553.

Gerashchenko, M.V., Lobanov, A.V., and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc Natl Acad Sci U S A *109*, 17394-17399.

Hani, J., and Feldmann, H. (1998). tRNA genes and retroelements in the yeast genome. Nucleic Acids Res *26*, 689-696.

Hussmann, J.A., Patchett, S., Johnson, A., Sawyer, S., and Press, W.H. (2015). Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. PLoS Genet *11*, e1005732.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789-802.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455-459.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Klopotowski, T., and Wiater, A. (1965). Synergism of aminotriazole and phosphate on the inhibition of yeast imidazole glycerol phosphate dehydratase. Arch Biochem Biophys *112*, 562-566.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol *34*, 1812-1819.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Li, G.W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538-541.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26.

Merritt, G.H., Naemi, W.R., Mugnier, P., Webb, H.M., Tuite, M.F., and von der Haar, T. (2010). Decoding accuracy in eRF1 mutants and its correlation with pleiotropic quantitative traits in yeast. Nucleic Acids Res *38*, 5479-5492.

Mohammad, F., Woolstenhulme, C.J., Green, R., and Buskirk, A.R. (2016). Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. Cell Rep *14*, 686-694.

Myers, J.K., Pace, C.N., and Scholtz, J.M. (1997). Helix propensities are identical in proteins and peptides. Biochemistry *36*, 10923-10929.

Nelson, D.L., Lehninger, A.L., and Cox, M.M. (2008). Lehninger Principles of Biochemistry (W. H. Freeman).

Nissley, D.A., and O'Brien, E.P. (2016). Altered Co-Translational Processing Plays a Role in Huntington's Pathogenesis-A Hypothesis. Front Mol Neurosci *9*, 54.

Patel, B.K., Gavin-Smyth, J., and Liebman, S.W. (2009). The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. Nat Cell Biol *11*, 344-349.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400-1412.

Pelletier, J., Graff, J., Ruggero, D., and Sonenberg, N. (2015). Targeting the eIF4F translation initiation complex: a critical nexus for cancer development. Cancer Res

75, 250-263.

Pestova, T.V., and Hellen, C.U. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. Genes Dev *17*, 181-186.

Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet *8*, e1002603.

Rodnina, M.V. (2016). The ribosome in action: Tuning of translational efficiency and protein folding. Protein Sci *25*, 1390-1406.

Sharma, A.K., and O'Brien, E.P. (2018). Non-equilibrium coupling of protein structure and function to translation-elongation kinetics. Curr Opin Struct Biol *49*, 94-103.

Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res *15*, 1281-1295.

Shi, Z., and Barna, M. (2015). Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. Annu Rev Cell Dev Biol *31*, 31-54.

Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., Chang, H.Y., and Barna, M. (2017). The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. Cell *169*, 1051-1065 e1018.

Sorensen, M.A., Kurland, C.G., and Pedersen, S. (1989). Codon usage determines translation rate in Escherichia coli. J Mol Biol *207*, 365-377.

Stein, K.C., and Frydman, J. (2019). The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. J Biol Chem *294*, 2076-2084.

Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2016). Protein Structure Is Related to RNA Structural Reactivity In Vivo. J Mol Biol *428*, 758-766.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344-354.

UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res *47*, D506-D515.

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., and Bartel, D.P. (2016). Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. Cell Rep *14*, 1787-1799.

Weinert, B.T., Scholz, C., Wagner, S.A., Iesmantavicius, V., Su, D., Daniel, J.A., and Choudhary, C. (2013). Lysine succinylation is a frequently occurring modification in

prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep *4*, 842-851.

Yang, J.R., Chen, X., and Zhang, J. (2014). Codon-by-codon modulation of translational speed and accuracy via mRNA folding. PLoS Biol *12*, e1001910.

Yu, C.H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M.S., and Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. Mol Cell *59*, 744-754.

FIGURES

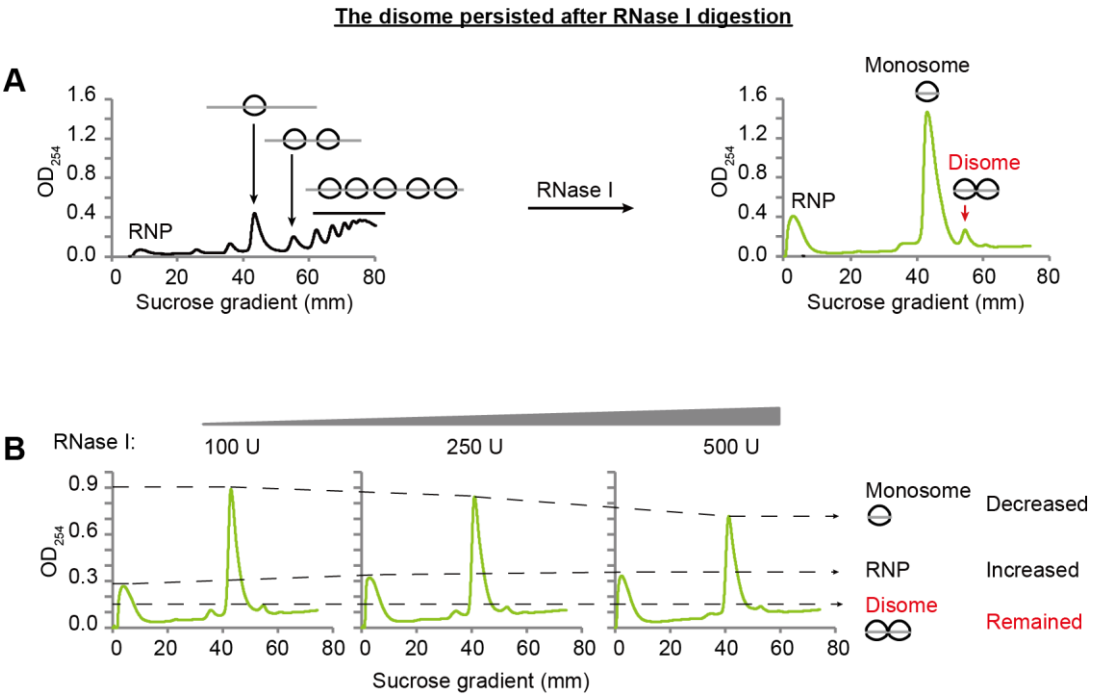

Figure 1. Disome protected fragments were RNase I resistant.

(A) The disome persisted after RNase I digestion. Sucrose gradient profiles of ribosome-mRNA without (black) and with (green) RNase I digestion were shown. The *x*-axis displays the detecting position of the 5-50% sucrose gradient. The *y*-axis indicates the RNA abundance inferred from UV absorption (OD_{254}). RNP, free ribonucleoprotein.

(B) The disomes remained as the dosage of RNase I increased. Samples containing an equal amount of ribosome-mRNA were treated with 100 U, 250 U, and 500 U RNase I, respectively.

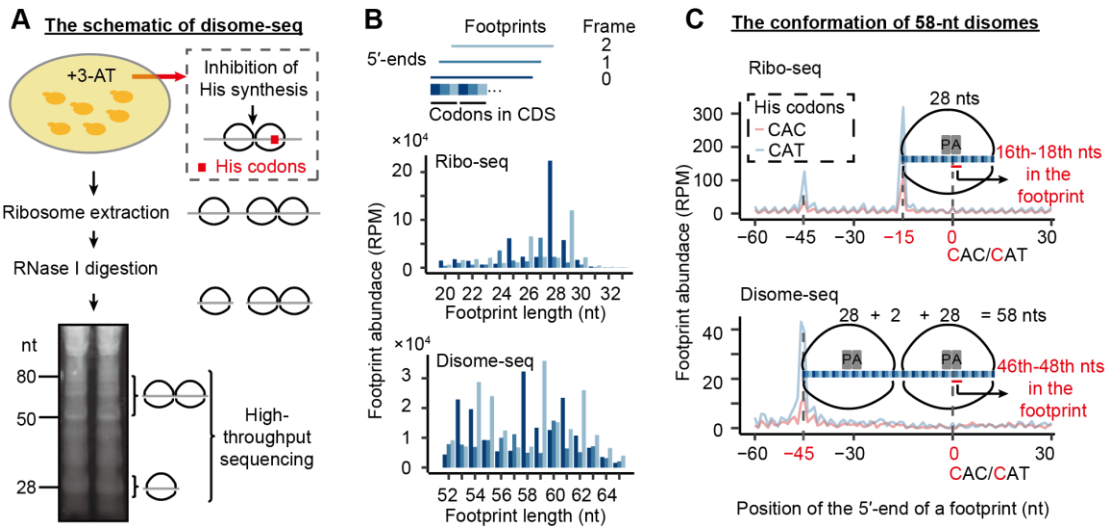

Figure 2. Disome-seq specifically detected severe translational pauses.

(A) The schematic of disome-seq for 3-AT treated yeast cells.

(B) The length distribution of monosome or disome footprints. The inset shows the definition of the reading frame of a footprint – it is inferred from the reading frame of the 5′-end of a footprint. The footprint abundance is in the unit of reads per million (RPM).

(C) Determination of the disome conformation. The distribution of the 5′-end of monosome footprints (the top panel) and disome footprints (the bottom panel) around the histidine codons (from the –60th nt to the +30th nt). The two histidine codons are distinguished by color. P and A represent the P-site and A-site of a ribosome, respectively.

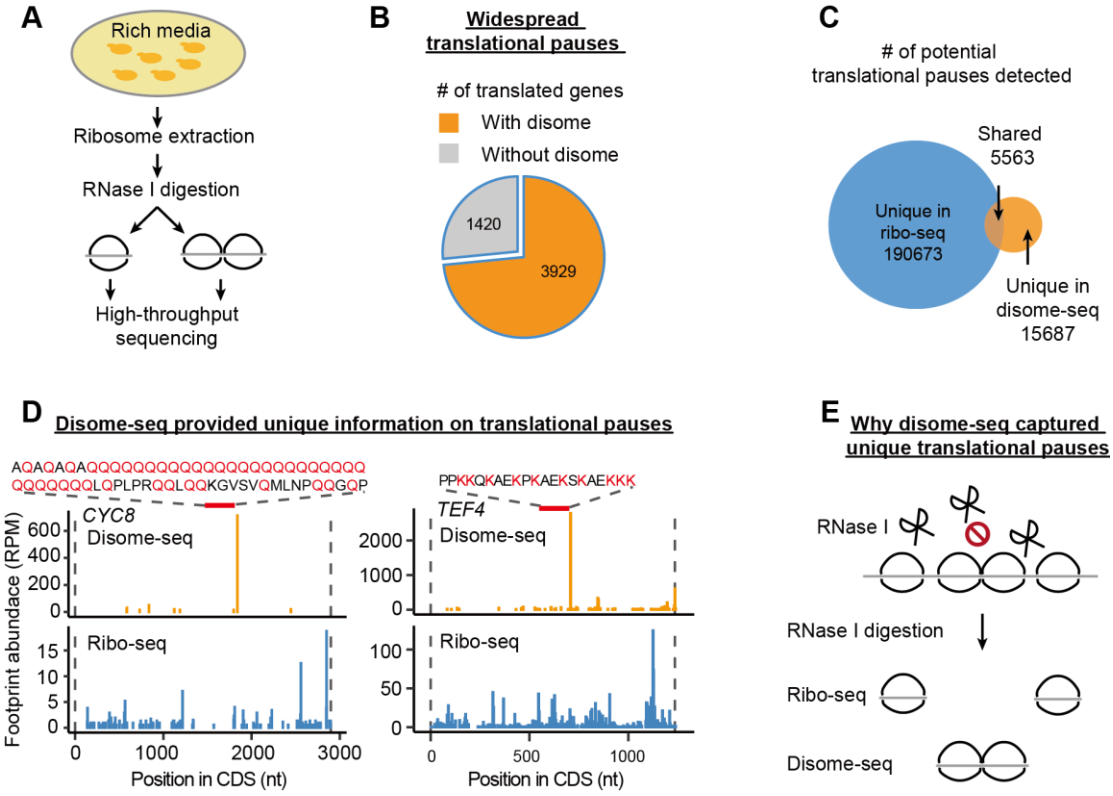

Figure 3. Disome-seq detected the translational pauses missed in ribo-seq.

(A) The schematic of disome-seq for yeast cells cultivated in the rich media.

(B) Widespread translational pauses were detected in the yeast genome. Translated genes with and without disome footprints are shown in orange and grey, respectively.

(C) The positions of monosome footprints (blue) and those of disome footprints (orange) rarely overlapped. The average read count per codon site was calculated for each gene; only those codon sites with read count greater than the average was considered in this panel. The A-site of a monosome footprint was compared with the A-site of the 3′-ribosome of a disome footprint.

(D) Two genes exemplify the unique information obtained by disome-seq. The distributions of disome (orange) and monosome (blue) footprints are shown. The *x*-axis shows the A-site of a monosome footprint or that of the 3′-ribosome of a disome footprint in the coding sequence (CDS). The start and stop codons are denoted with grey dashed lines.

(E) The schematic graph explains why traditional ribo-seq missed severe translational pauses.

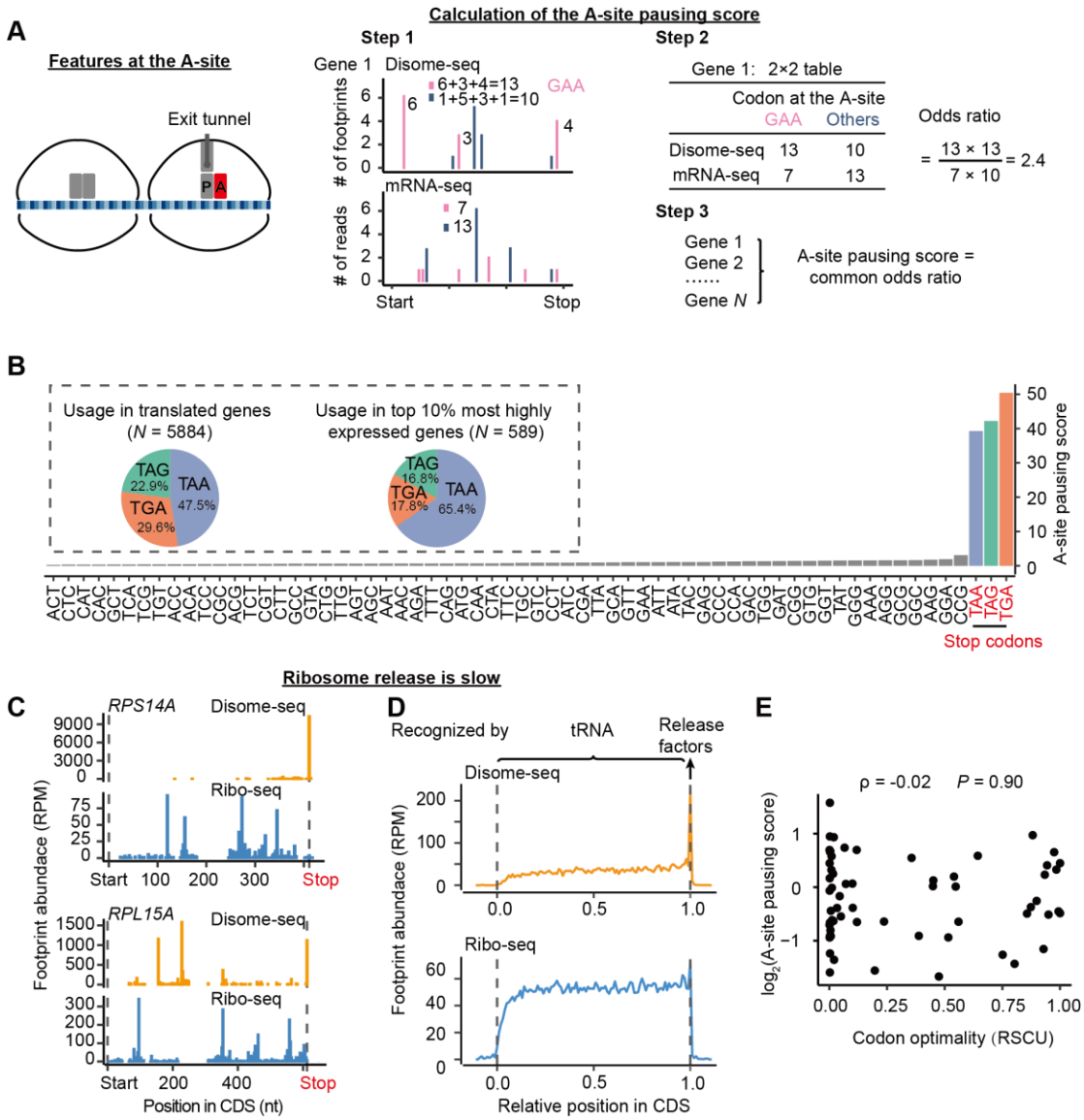

Figure 4. Translational pauses were induced by stop codons.

(A) Schematic of the calculation of the A-site pausing score. We counted the number of disome footprints with A-site having the concerning codon (GAA as an example here) and having other codons, respectively. The mRNA-seq reads were used to control for the codon frequency in the transcript. The A-site pausing score is defined as the common odds ratio among genes. A significantly greater-than-1 pausing score indicates a propensity to induce translational pausing.

(B) The stop codons show the highest A-site pausing scores. The A-site pausing score of each of the 64 codons is shown. The usages of three stop codons are displayed among all yeast genes or among the top 10% most highly expressed genes (inset).

(C) Two genes exemplify translational pauses at stop codons.

(D) Metagene analysis of disome (orange) or monosome (blue) footprints. The *x*-axis shows the normalized coding regions from 0 (start codon) to 1 (stop codon) with the bin width = 0.01. The footprint distribution of each gene was normalized by the total

footprint abundance of the corresponding gene; the distribution was then added up among genes.

(E) An indicator of codon optimality, RSCU, is poorly correlated with the A-site pausing score. The *P*-value was given by the Spearman's correlation.

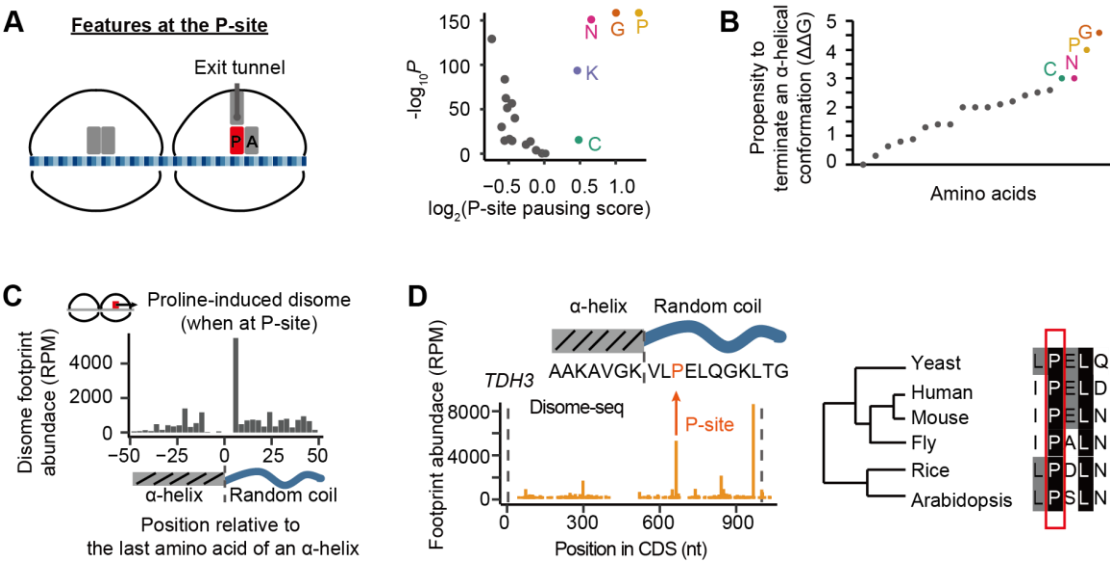

Figure 5. Translational pauses occurred at the amino acids that terminate α-helixes.

(A) The P-site pausing score of each amino acid is shown. Five amino acids with significant greater-than-1 P-site pausing scores are labeled.

(B) The amino acids are ordered by their propensity to terminate an α-helical conformation.

(C) Proline induced translational pauses often occur right after the termination of an α-helix at the genomic scale. All α-helixes in the proteome were aligned to their last amino acids, and the total abundance of disome footprints (RPM) was shown in the *y*-axis.

(D) A translational pause occurs right after an α-helix of TDH3p (the left panel). The proline is conserved among eukaryotic species (the right panel).

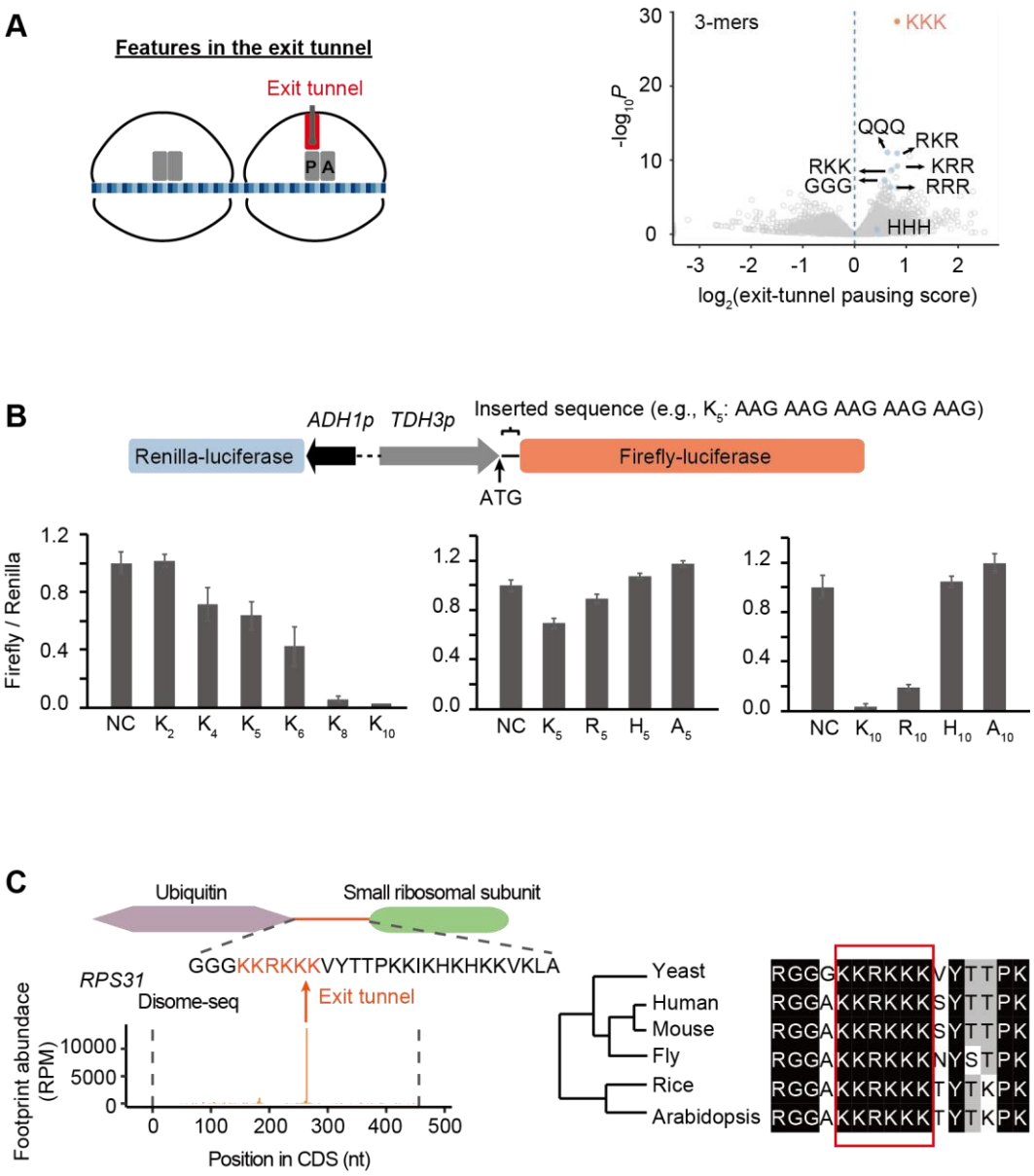

Figure 6. Translational pauses were induced by polylysine in the exit tunnel.

(A) The triple-lysine is significantly enriched in the exit tunnel of the 3′-ribosome of a disome. The *x*-axis displays the exit-tunnel pausing score of amino-acid 3-mers.

(B) The dual-luciferase assay. Sequences to be tested were inserted right after the start codon of firefly-luciferase. The relative translation elongation rate of the inserted sequence is shown by the ratio of enzyme activities (firefly/renilla). Ratios were normalized to NC (the negative control vector without insertion).

(C) A translational pause induced by polylysine took places in the linker regions between two protein domains of RPS31p (the left panel). The polylysine sequence is conserved among eukaryotic species (the right panel).

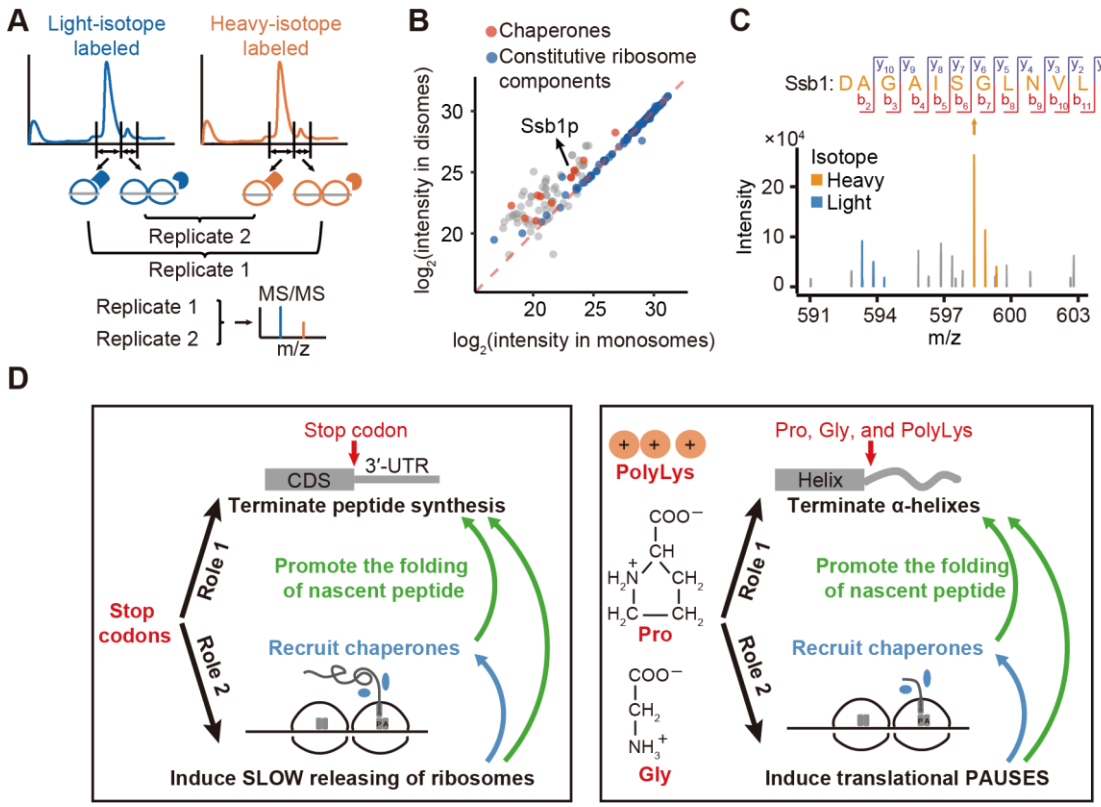

Figure 7. Translational pauses recruited chaperones.

(A) The schematic of the proteomic detection of the disome associated proteins.

(B) Chaperones (red dots) are enriched in disomes. Constitutive ribosomal proteins (blue dots) are evenly distributed between monosomes and disomes.

(C) The mass spectrometry result of a chaperone, SSB1p, shows its enrichment in disomes.

(D) The dual-role played by stop codons and certain amino acids in translational pausing and co-translational protein folding.