# Endogenous polybasic polypeptides cause slower movement of ribosomes but are not targets of the ribosome quality control (RQC) complex

Géssica C. Barros[1#], Rodrigo D. Requião[1#], Rodolfo L. Carneiro[1], Claudio A. Masuda[2], Mariana H. Moreira[1], Silvana Rossetto[3], Tatiana Domitrovic[4*] and Fernando L. Palhano[1*]

[1] Programa de Biologia Estrutural, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil.
[2] Programa de Biologia Molecular e Biotecnologia, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil.
[3] Departamento de Ciência da Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil.
[4] Departamento de Virologia, Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-902, Brazil.

* To whom correspondence should be addressed: Tel.: +55 21 3938-6761; Email: palhano@bioqmed.ufrj.br or domitrovic@micro.ufrj.br

# These authors contributed equally.

**ABSTRACT**

Highly positively charged protein segments are known to result in poor translation efficiency by the action of the ribosome quality control (RQC) complex. This effect may be explained by ribosome stalling caused by electrostatic interactions between the nascent peptide and the negatively charged ribosome exit tunnel, leading to translation termination followed by the protein degradation by the RQC complex. These polybasic peptides are mainly studied with reporter systems, where artificial sequences are introduced into heterologous genes. The unique example of an endogenous protein targeted by the RQC complex is Rqc1, a protein essential for RQC function. The polybasic motif in Rqc1 N-terminal is thought to activate the RQC, leading to Rqc1 down-regulation. We aimed to determine whether the RQC complex acts as a regulatory mechanism for other endogenous proteins containing polybasic segments. Here, we show by bioinformatics, ribosome profiling data, and western blot protein quantification that endogenous proteins containing polybasic motifs similar to or even more positively charged than Rqc1 are not targeted by the RQC complex, suggesting that endogenous polybasic motifs do not induce degradation by this complex. We further demonstrate that Rqc1 levels are not regulated by the RQC complex but by Ltn1 alone in a posttranslational fashion.

Keywords: ribosome quality control, ribosome profiling, poly(A) tracts, polybasic domains, stalled polypeptides, protein quality control

## INTRODUCTION

The first study to propose that the translation of positive amino acids could lead to translational stalling was published in 2008 by Lu and Deutsch (1). It was already known by that time that the ribosomal exit tunnel, a narrow passage through which the nascent peptide passes as soon as it is synthesized, is negatively charged (2) and that the translation of the poly-A tail from constructs lacking a stop codon, which leads to the incorporation of lysine residues (the codon AAA translates into lysine), causes translational repression (3-5). This led to the hypothesis that a high concentration of positively charged residues inside the exit tunnel could lead to slower translation rates due to the presence of electrostatic interactions. Using a cell-free approach, Lu and Deutsch demonstrated that the presence of lysine or arginine repeats is sufficient to decrease translation rates. They observed that the presence of as few as four nonconsecutive arginines in the construct was already enough to diminish the translational speed, while an equal concentration of negatively charged or uncharged residues did not affect the translation. This phenomenon was also observed with lysines and was independent of codon choice, suggesting that the retardation effect was explicitly caused by the charge itself, not the amino acid structure or tRNA availability (1).

These early experiments prompted the use of polybasic stretches as stalling sequences in many subsequent papers. Most of these publications focused on the characterization of the ribosome quality control (RQC) complex, a cellular pathway responsible for the cotranslational degradation of mRNA and nascent peptides (6, 7). Polybasic motifs were added to or removed from a reporter protein, and the amount of translated protein was detected by western blot (6, 8, 9). Using this approach, several authors have confirmed that the presence of poly-arginine (R), poly-lysine (K), or mixtures of both amino acids can reduce protein production (10-12) and that this effect was linked to the action of the quality control machinery (6, 13-16).

The RQC complex removes stalled nascent polypeptides from the ribosome and activates stress signals in *S. cerevisiae* (6, 16). A defective RQC complex causes proteotoxic stress in yeast cells (9) and neurodegeneration in mice (17). The yeast RQC complex components are Rqc1, Rqc2, and Ltn1, all of which have a relatively low copy number when compared with the number of ribosomes in a cell, thus indicating that the system could easily reach saturation (6, 13, 18-20). The RQC complex is recruited by Asc1 and Hel2, proteins that detect ribosome stalling and promote early termination, leading to ribosomal subunit dissociation (15). After ribosome dissociation, the RQC complex can bind to the 60S subunit and degrade the nascent polypeptide through Ltn1 activity, an E3 ubiquitin ligase involved in cotranslational degradation (8). Rqc1 and Ltn1 lead to the recruitment of Cdc48 and its co-factors, Npl4 and Ufd1, which extract the stalled ubiquitylated polypeptide, through ATP hydrolysis, leaving the 60S subunit free to engage in another round of translation. Rqc2 is involved in stress signaling through the formation of the C-terminal alanine-threonine (CAT) tail. After attaching to the 60S subunit, Rqc2 can attach alanine and threonine residues to the stalled peptide's C-terminal, creating the CAT tail (9, 21-24). This incorporation is random and independent of mRNA and is important to induce stress and expose lysine resides trapped inside the ribosome tunnel to the ubiquitin ligase activity of Ltn1 (23).

In addition to the action as a salvage pathway for ribosomes stuck in unproductive translation events, it was suggested that the RQC pathway could regulate the expression of transcripts containing stalling features. This idea was put forward by an observation that in yeast, Rqc1 levels could be controlled by RQC activity, providing negative feedback for the system. The work showed that the deletion of Asc1, Hel2, Ltn1, or Rqc2 led to an increase in Rqc1 protein expression (6). Rqc1 contains a well-conserved K/R polybasic domain in its N-terminal that, when substituted by alanine residues, led to increased protein expression. Based on these data, it was suggested that the RQC complex regulates Rqc1 protein levels through its polybasic domain during Rqc1 translation (6). This observation led to the question of whether other endogenous proteins with similar structural characteristics would suffer the same kind of regulation. More recently, it has been proposed that the trigger for RQC activity is ribosome collision, with the formation of di-ribosomes or disomes with empty A sites (14, 25-27). Different studies have shown that the existence of stalling features in endogenous sequences seems to be accompanied by unique structural characteristics, such as low initiation efficiency (28, 29) and/or the positioning of the stall sequence at the beginning of the transcript (30, 31). These adaptations are thought to result in fewer ribosomes per transcript, which would avoid ribosome collisions at the challenging segments, thus preventing RQC activation.

To test whether the RQC complex can regulate the expression of endogenous proteins through their polybasic domains, we used a bioinformatics screen to identify the proteins with higher concentrations of K/R (polybasic domains) of the yeast proteome. Interestingly, the extremely polybasic proteins caused slower ribosome movement during translation, but this was insufficient to promote the accumulation of ribosomes with an empty A site. These data suggest that endogenous polybasic proteins do not lead to RQC complex activation. To test this hypothesis, we inactivated the RQC complex by deleting *LTN1* and *ASC*1, expecting to see the increase in the final protein output that is usually observed with reporters containing stall sequences. However, we observed that the protein levels of the top polybasic proteins did not increase, suggesting that RQC does not restrict the protein production of endogenous genes with polybasic domains. Moreover, in our hands, Rqc1 upregulation was induced by only *LTN1* deletion but remained unperturbed when the complex was disabled upstream (*ASC1*, *HEL2*) or at CAT tail step (*RQC2*). This result is inconsistent with the current model of Rqc1 cotranslational RQC regulation. Our data indicate that endogenous proteins with polybasic domains, including Rqc1, are not targets of the RQC complex. Most likely, the RQC machinery is dedicated to recycling ribosomes involved in aberrant mRNA translation and plays a small role in controlling endogenous protein levels under normal growth conditions.

## MATERIALS AND METHODS

| **Data** | **Source** | **Identifier** |
|---|---|---|
| Translation efficiency | Weinberg et al., 2016 | SRR1049521 |
| *S. cerevisiae* drug-free Riboseq data | Pop et al., 2014 | SRR1688545 |
| *S. cerevisiae* + CHX + ANS Riboseq data | Wu et al., 2019 | SRR7241922 |
| *S. cerevisiae* + CHX + TIG Riboseq | Wu et al., 2019 | SRR7241919 |

| | | |
|---|---|---|
| data | | |
| Kozak score 50 pb | Cuperus et al., 2017 | |
| Kozak score 5 pb | Cuperus et al., 2017 | |
| Codon stabilization coefficient gene (CSCg) | Carneiro et al, 2019 | |
| Time for translation initiation | Siwiak and Zielenkiewicz, 2010 | |
| 40S ribosomal small subunit (SSU) footprint profiling | Archer et al., 2016 | |
| **Software and algorithms** | | |
| Net charge calculation | Requião et al, 2017 | |
| Poly (A) finder | This paper | https://github.com/mhoyerm/adenosine_sequence |
| Consecutive K and/or R finder | This paper | https://github.com/mhoyerm/arg_lys_sequence |
| Cluster of amino acids finder | This paper | https://github.com/mhoyerm/search_discrete_stretches |
| Stretches of K and/or R finder | This paper | https://github.com/mhoyerm/arg_lys_stretches |
| Protein shuffler | This paper | https://github.com/mhoyerm/protein_shuffler |

Data sources

Coding sequences and the annotation of *S. cerevisiae* were obtained from the *Saccharomyces* Genome Database (SGD; https://www.yeastgenome.org) and Ensembl Genomes (http://ensemblgenomes.org/). We excluded dubious ORFs as defined in the *Saccharomyces* Genome Database from our analysis.

Statistical analyses and raw data

The raw data used to create Figures 1A, 2A, 3A, and 3B and the statistical analysis t-test calculations are presented in Supplementary Tables S1-3. All statistical analyses were performed with GraphPad Prism 7.

Net charge calculation

We developed a program that screens the primary sequence of a given protein and calculates the net charge in consecutive frames of a 30-amino-acid window, the approximate number of amino acids that fill the ribosome exit tunnel (32, 33). For the net charge determination, K and R were considered +1; D and E were considered -1; and every other residue was considered 0. The N- and C-termini charges were disregarded. In a previous publication, we have shown that these simplified parameters are equivalent to a calculation using partial charges of individual amino acids at pH 7.4, according to their p*K*a values and the Henderson-Hasselbach equation (33). The algorithm initially establishes a stretch containing a predetermined number of amino acids (#1 to #30, for example). The stretch charge is calculated, and the charge value and the position of the first amino acid are temporarily saved to the memory. Then, our algorithm advances to the next stretch, from amino acid #2 to amino acid #31, and performs the same analysis until the end of the protein.

Codon stabilization coefficient (CSC) calculation

Based on Pearson's correlation between the frequency of occurrence of each codon in each mRNA and the mRNA half-lives, Coller and colleagues created the codon occurrence to mRNA stability coefficient (CSC) (34). We designed and implemented an algorithm that calculated the mean value of the CSC for each gene (CSCg) (35).

Cluster of amino acid finder

An algorithm was created and implemented to find any amino acid clusters on the proteome. A list with the amino acid cluster (i.e., KKKKKK) is used as input, and the algorithm returns as output the protein name and the position of the cluster found for each protein.

The fold change calculation of Figure 1A was obtained by Eq. 1.

$$FC = \frac{X_i}{E(i,n)} \tag{1}$$

where *FC* is the fold of change, $X_i$ is the observed occurrence, and *E(i,n)* is the expected occurrence corresponding to each amino acid *i* in a stretch of *n* repeated amino acids.

The expected values of amino acid sequences were obtained through two approaches. The first approach was based on the calculation using Eq. 2,

$$E(i,n) = X_i \cdot P(i,n) \tag{2}$$

where *P(i,n)* is the probability of finding the repeated sequence, and its calculation is given by Eq. 3,

$$P(i,n) = \frac{X_i! \cdot (N-n)!}{(X_i - n)! \cdot N! \cdot n!} \tag{3}$$

where N is the total amount of amino acids in the proteome.

For the second approach (Fig S1), using the Protein shuffler algorithm, the positions of the amino acids were randomly changed in the yeast proteome, keeping each protein size and amino acid occurrence. The mixed proteome was analyzed by the algorithm Cluster of amino acids finder, which counted the occurrences for a chosen stretch, and the value obtained with this algorithm is equivalent to *E(i,n)* as obtained in the first approach.

Poly (A) finder

An algorithm was developed to locate the position of stretches containing only adenine (A) in an ORF. The algorithm finds adenine stretches of a chosen size for each ORF and generates as output a file containing the gene name, the stretch size, and its position in the ORF. For our analysis, we considered stretches of size equal to or larger than 10 adenines.

Stretches of K and/or R finder

We developed an algorithm that searches for stretches containing a certain number of arginines and/or lysines in a stretch of 10 amino acids in a protein. The chosen number of K and/or R in the stretches was equal to or greater than 8. The input file of the algorithm was a proteome fasta file, and the output file returned the protein name and the position of the stretch found for each protein.

Consecutive K and/or R finder

An algorithm was developed to find the location of stretches containing consecutive arginines and/or lysines in a protein. The stretches analyzed had a size equal to or larger than 6. The input file of the algorithm was a proteome fasta file, and the output file returned the protein name, the stretch size, and the position of the stretch in the protein.

Yeast strains and growth conditions

Since no commercial antibodies are available for the proteins used herein, we used a TAP-TAG collection of yeast (36). In this collection, each ORF was tagged with a C-terminal TAG, and the same antibody could be used to detect the full collection. For each potential target of the RQC complex described before, a knockout of the *LTN1* or *ASC1* strain was created by homologous recombination. For the Rqc1 TAP-TAG strain, a knockout of *RQC2* and *HEL2* was also created. After the creation of the knockout strains, the level of the TAG proteins was compared with the wild-type strain by western blot analyses. The TAP-TAG collection strains were cultivated in YPD medium (1% yeast extract, 2% peptone, 2% glucose). The *S. cerevisiae* BY4741 strain transformed with GFP-R12-RFP vector was cultivated in minimum synthetic medium, uracil dropout (0.67% Difco Yeast Nitrogen Base without amino acids, 2% glucose, and 0.003% methionine, leucine, and histidine).

Yeast strain constructions

For knockout creation, *S. cerevisiae* BY4741 yeast (TAP-TAG collection) was transformed using the PEG/Lithium Acetate method (37). The product of the PCR used in the transformation was obtained by the amplification of the KanMX6 gene contained in the pFA6a-KanMx6 vector. All genes were knocked out by the insertion of the geneticin resistance gene into their original ORFs. The knockouts were confirmed using two sets of PCR. One PCR set combined the primer for the target gene flank region with a resistance gene primer. Another PCR set combined a primer for the target gene flank region and an ORF gene target primer. All the primer pairs used in the construction and confirmation of the mutants are depicted Support Information.

Protein extraction and SDS-PAGE

The protein extraction of the strains was conducted in the exponential phase of growth, treating equal quantities of cells using the NaOH/TCA method (38). For the SDS-PAGE, equal sample volumes were applied on the acrylamide gels and submitted to an electrophoretic run. For the YHR131C TAP-TAG strain, protein extraction was conducted using a glass bead protein extraction

protocol (39). Total protein quantification was conducted using the BCA Protein Assay kit (Thermo Fisher), and equal protein quantities were applied on SDS-PAGE. The acrylamide gel concentration changed between 10 and 12.5% according to the size of the analyzed protein.

Western blotting

The proteins separated by SDS-PAGE were transferred to PVDF membranes using the TRANS-BLOT semidry system from BIO-RAD in transfer buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, 20% methanol, pH 8.3). The membranes were blocked using Li-Cor blocking buffer with overnight incubation at room temperature. Incubation with primary antibodies was performed at 4ºC for at least 1 hour. The membranes were incubated in Li-Cor Odyssey secondary antibodies at room temperature for at least 1 hour and then revealed in an Odyssey scanner. The primary antibodies used were mouse anti-PGK1 (Invitrogen; dilution 1:10,000), rabbit anti-TAP-Tag (Thermo Scientific; dilution 1:10,000), mouse anti-GFP (Sigma; dilution 1:2,000), and rabbit anti-Asc1 (dilution 1:5,000) (40). The secondary antibodies that were used were anti-mouse 680 LT Li-Cor (dilution 1:10,000) and anti-rabbit 800 CW Li-Cor (dilution 1:1,000).

Cycloheximide (CHX) treatment

For the experiments using CHX, concentrations ranging from $10^{-6}$ to $10^{-1}$ mg/ml were used, with concentration steps of 10 times the dilution factor. Strains grew until the exponential phase and were treated with CHX for 10 hours under shaking at 28ºC.

Ribosome profiling data

*S. cerevisiae* ribosome profiling data were treated as described previously (33, 41). The data were analyzed as described by Ingolia and collaborators (42), except that the program used here was Geneious R11 (Biomatter Ltd., New Zealand) instead of the CASAVA 1.8 pipeline. The data were downloaded from GEO (43, 44), and the adaptor sequence was trimmed. The trimmed FASTA sequences were aligned to *S. cerevisiae* ribosomal and noncoding RNA sequences to remove rRNA reads. The unaligned reads were aligned to the *S. cerevisiae* S288C genome deposited in the *Saccharomyces* Genome Database. First, we removed any reads that mapped to multiple locations. Then, the reads were aligned to the *S. cerevisiae* coding sequence database, allowing two mismatches per read. For 27-29 nt footprint analyses, the ribosome profiling data were obtained without any drug, and just fragments with 27-29 nt were used (43). For 20-22 nt footprint analyses, the ribosome profiling data were obtained with cells lysed with buffer containing 0.1 mg/ml of cycloheximide (CHX) and tigecycline (TIG) or 0.1 mg/ml of cycloheximide (CHX) and anisomycin (ANS) (44). We normalized the coverage within the same transcript by dividing each nucleotide read by the sum of the number of reads for each gene in a given window. A window of 150 nucleotides before and after the beginning of a stretch of interest was used to calculate the number of reads of each gene.

Clustering analysis

The clustering analyse of Figure 3B was performed by the Euclidean distance using Orange 3 software (45).

Translation initiation score

For the Figure 3B, the genome was divided into eight equal groups based on their absolute values, from highest to lowest. The groups with the highest and the lowest values were further divided, making a total of ten groups. Each group was normalized individually. Their normalized values were then adjusted to fit the whole genome, meaning that the percentages from the group with the lowest values went from 0-100% to 0-6.25%, and the percentages from the group with the highest values, from 0-100% to 93.75-100%, for example. This methodology was used to avoid the influence of a few genes with extremely high and low values that can make the visualization of the difference between the values of most proteins difficult.

## RESULTS

### Repetitive stretches of arginine and lysine are relatively frequent and slow down ribosomes without compromise translation.

To determine whether repetitive stretches of arginine or lysine are avoided in the yeast proteome, we determined the amino acid repeat occurrence in the *S. cerevisiae* protein sequences. We searched for consecutive clusters of 2 up to 6 identical amino acids (for example, RR, RRR, RRRR, etc.). Then, we determined the ratio between the observed and the expected frequency of repeats. This was calculated considering the representation of each amino acid in the total proteome (bar graph Fig 1A) and assuming that the occurrences are independent (null model) (Fig 1A, the panel on the left). We observed that all amino acid repeats were more frequent than expected by random choice. Some amino acids, such as tryptophan (W) and methionine (M), had no more than 3 or 4 repeats, which can be partially explained by the fact that they are relatively rare in the proteome (Fig 1A). Conversely, valine (V) and isoleucine (I) are relatively common amino acids, but no more than 6 repeats were found (Fig 1A). Poly-lysines (K) and poly-arginines (K) are found in frequencies similar to those of other noncharged or negatively charged amino acids, suggesting that there is no particular bias against repetitions of these two amino acids (Fig 1A, arrows) (46). We obtained a similar result when the observed number of repeats was divided by the number of repeats obtained from a scrambled version of the proteome data set (Fig S1A). The absolute number of repeats for each amino acid is shown in Supplemental Table S1. Next, we calculated whether the repeats were preferentially located in the N- or C-terminal extremities of the proteins. Some repetitive stretches of residues such as phenylalanine (F) and tyrosine (Y) had a strong bias towards the N-terminal, while clusters formed by valine (V) repeats were mostly located in the C-terminal of the proteins (Fig 1A). Arginine and lysine showed discrete location preferences towards the N- and C-terminals, respectively.

To test if the repeats of the amino acids have any effect on translation, we took advantage of previously published ribosome footprint profiling (RP) data. This methodology is based on the deep sequencing of ribosome-protected mRNA fragments (42). During translation, each ribosome encloses a portion of mRNA, protecting it against RNAase digestion. This enclosure allows protected fragments to be sequenced, generating a map, at nucleotide resolution, of translational activity for transcribed genes. This map can report the occurrence of ribosome pausing or stalling, revealed by the accumulation of deep sequencing reads at specific positions of the transcript. Recently, it was shown that ribosomes with open A sites produce shorter mRNA fragments of 20-22 nucleotides than ribosomes with occupied A sites, which produce longer mRNA fragments (27-29 nucleotides) (44). Structural and biochemical evidence (47) demonstrated that stalled ribosomes are in the open A site conformation, suggesting that the 20-22 fragments could more accurately reflect translation efficiency.

In Figure 1B and 1C, we show the ribosome footprint density on the codons coding for 4 or more amino acid repeats (position 0). The footprint counts were normalized by the total number of reads for each transcript. We also compared the amino acid repeats with a group of 1800 genes containing inhibitory codon pairs (IP) that are known to impair translation with substantial downregulation in protein output (e.g., CGA-CGA and CGA-CCG) (48). As expected, the IP sequences showed increased ribosome density both at the 27-29 and the 20-22 populations, indicating severe translation arrest. We observed increased density on poly-lysine (K), arginine (R), phenylalanine (F), and methionine (M). However, for all the repeats except tryptophan, the stalling effect was less pronounced than that observed with the IP group and was observed only at the 27-29 nt population. Virtually no increase in poly K and poly R was observed in the 20-22 nt footprint. We performed the same analysis as Figure 1C with 20-22 fragments obtained with a different combination of translation inhibitors (cycloheximide and anisomycin), and the same result was observed (Fig S1C). We conclude that repetitions of R and K are not particularly rare in the proteome, and even though they may cause a slight translational pause, as indicated by the increased footprint density of 27-29 nt (Fig 1B), there was no evidence that these sequences impose a severe translation arrest leading to open A site ribosomes (Fig 1C).

Positively charged protein domains are thought to interact with the negatively charged ribosome exit tunnel (2), arresting translation and leading to ribosome stalling. Therefore, our next step was to find endogenous yeast proteins containing sequences with different arrangements of positively charged amino acids. Next, we describe the four distinct, but possibly overlapping, selected architectures: a) protein segments with net charge ≥ +12 in a stretch of 30 residues. The search was made using a program that calculates the electrostatic net charge in consecutive windows of 30 residues, which is the polypeptide length enclosed by the ribosome exit tunnel. The program takes into account the charge of the amino acid lateral chain at pH 7.4 and screens the protein sequence (32, 33) b) More than 8 K and/or R in 10 amino acids and c) 6 consecutive K and/or R. These two architectures were selected because they are often present in RQC activation reporters and lead to the accumulation of reads in ribosome profiling data. d) Adenine repeats longer than 10 nucleotide residues (poly (A)). With this criterion, we screened for 2 or 3 K repeats coded from consecutive AAA. This mRNA

composition can lead to ribosome sliding, worsening the translation efficiency of these sequences (49, 50).

We found that 354 proteins (approximately 6% of the yeast proteome) contained at least one characteristic described above (Fig 2A). The features most commonly found were a high net charge and 10 poly(A), with 170 and 121 sequences, respectively. Moreover, the Venn diagram (Fig 2A) shows that just 91 proteins have more than one feature at the same time, and only 1 protein (YNL143C) has all the characteristics (see also Fig 2C, D and E). We selected YNL143C and other proteins with extreme polybasic features (indicated in Fig 2C, D, and E) for further experimental analysis in the next section.

We repeated the localization and footprint density analysis of Figure 1 using the polybasic data set (Fig 2B, F, G). We found that the inhibitory pairs of nucleotides, 6 consecutive K and/or R, and 10 poly(A) sequences tend to be located at the N-terminal, while 8 K and/or R in 10 amino acids are mostly located at the C-terminal. No bias was found for the 30 residues ≥ +12 group (Fig 2B). The footprint counts with 27-29 nucleotides indicated a translational arrest for all polybasic sequences except for the group containing 30 residues ≥ +12 (Fig 2F). When we examined the 20-22 nt footprints, no accumulation was observed for any group other than the IP control, suggesting that, similar to the amino acid repeats, these polybasic domains do not cause a translation pause that is strong enough to result in empty A site ribosomes (Fig 2G). Again, we repeated this analysis with 20-22 fragments obtained with a different combination of translation inhibitors (cycloheximide and anisomycin), and the same result was observed (Fig S2).

If the increase in 27-29 footprint counts observed with some of the polybasic sequences (Fig 2F) represents an important obstacle to translation, we expected to find other signatures of poor translatability. Therefore, we asked whether the polybasic data set and the IP data set would be significantly different from the rest of the yeast genes in the following experimental measurements that reflect translation activity, such as translation efficiency (TE), as determined by ribosome profiling and calculated from the ratio of ribosome-protected footprint read counts to the total mRNA read counts for each gene (51), and the codon stabilization coefficient (CSC), a metric derived from mRNA half-life that positively correlates with TE and protein abundance (35). Another characteristic associated with genes containing stall sequences is a slower or less efficient translation initiation. This is thought to be an adaptation to avoid ribosome collision during elongation (29). Therefore, we analyzed the time for translation initiation (52), the Kozak efficiency score of the 5 nt- and 50 nt-long sequence (53), and the 5'UTR footprint density of the 40S ribosomal small subunits (54). The cumulative distributions for each of the 6 independent measurements described above are shown in Figure S3. In Figure 3A, we show the p values obtained with the t-test statistical analyses comparing the individual data sets with the rest of the yeast genes. As expected, the IP group had not only worse TE and CSC values but also slower initiation times associated with lower Kozak's efficiency scores, indicating that slower translation rates were associated with a less efficient initiation. When considered as a unique group (Fig 3A, polybasic group), polybasic sequences had lower TE and CSC values than the other yeast genes, but no significant effect was observed when considering initiation

parameters. The 5'UTR footprint density of the 40S ribosomal small subunits yielded no significant difference against IP and polybasic group, probably because this new technique still provides low coverage of the genome.

To evaluate the characteristics of each gene from the polybasic group, we created a unified rank for all the measurements above. Using the values obtained with the entire yeast data set, each score system was organized in 10 bins; next, the maximum and minimum bin values were used to normalize the scale. We created a heat map where 1 refers to the maximum translation activity (TE, CSC) or optimal initiation (translation initiation time, Kozak's efficiency, and the 5'UTR footprint density of the 40S ribosomal small subunits). A cluster analysis of the 354 polybasic genes revealed that the group is rather heterogeneous and displays all spectra of characteristics, with some sequences having excellent translatability indicators with poor initiation scores and vice versa (Fig 3B). Notably, the sequences with extreme polybasic characteristics were also diverse (Fig 3B, lower panel).

**Proteins with extremely polybasic domains in yeast are not targets of the ribosome quality control (RQC) complex.**

RQC pathway activation leads to 80S dissociation, mRNA degradation, and nascent protein ubiquitination followed by digestion through the proteasome. This process can decrease or completely suppress gene expression, as observed with reporters containing stall sequences or lacking stop codons. In either case, protein levels can be restored by the deletion of key players of the RQC pathway, such as *LTN1* and *ASC*1. Next, we examined whether these RQC components have any effect on the final protein output of endogenous genes containing polybasic sequences, thus implying a regulatory role for the RQC complex on the expression of these genes. Among the 354 genes with one or more polybasic features identified in our bioinformatics screening, we selected the four top genes for subsequent analyses: YBR054W/Yro2 (net charge = + 17), YGL078C/Dbp3 (net charge = + 16), YHR131C (14 consecutive R) and YNL143C (32 consecutive adenines coding for 10 K repeats) (Figs 2 and 3). We also included genes with the two longest stretches with 6 or more K and/or R in 10 amino acids of the yeast proteome: YLR197W/Nop56 and YOR310C/Nop58, which have a polybasic domain of 42 and 46 residues, respectively. Tap-tagged versions of these 6 genes were selected from a yeast library containing a chromosomally integrated TAP-tag at the 3' region of each gene, thus preserving the natural promoter region and the 5'UTR of the transcripts (36). Each TAP-tag strain was then used to generate the Δ*ltn1* or Δ*asc1* variants, confirmed by PCR and western blot (Fig S4). In Figure 4, we show the protein expression of each gene as probed by anti-tap western blot in the 3 different backgrounds and the relative quantification of at least 3 independent experiments. For each gene, we also show the complete ribosome profile and the protein TE value in comparison to the yeast proteome. Blue marks indicate the basic domain position in the transcript or the gene TE value in the distribution. We observed that Δ*ltn1* caused no regulation of protein levels for any of the transcripts except YNL143C. The results with Δ*asc1* revealed no regulation for YGL078C/Dbp3 and YNL197W/Nop56; downregulation for YBR054W/Yor2, YHR131C, and YNL143C; and an increase for

YOR310C/Nop58 (Fig 4). This pattern could not be explained by any feature, such as ribosome density on the polybasic site, TE, or initiation time. All of these regulation models differed from what was observed with reporters containing stall sequences that suffered upregulation in Δ*asc1* strains (see also Fig 5B). Therefore, we conclude that even though Asc1 seems to control the expression of some of these constructs, the complete RQC pathway is not actively degrading any of the endogenous polybasic sequences tested.

To our knowledge, the only endogenous protein considered to have its levels regulated by the activity of the ribosome control pathway is Rqc1 (6). Moreover, because the stall sequence of Rqc1 is thought to be a conserved polybasic domain (8 K and/or R in 10 aa, net charge = +13), it is critical to check how this protein behaves in our experimental setting (6). In Figure 5A, we analyzed Rqc1-tap expression in different backgrounds (Δ*ltn1*, Δ*asc1*, Δ*rqc2*, and Δ*hel2*). We reproduced the data of increased levels of Rqc1 in a Δ*ltn1* strain, but in our hands, the deletion of other RQC components did not lead to Rqc1 increase, suggesting that even this protein should not be considered a canonical RQC target. To probe our experimental conditions, we tested the expression of the well-described GFP-R12-RFP reporter (20). The stall sequence (R12) between GFP and RFP has 12 consecutive arginines with 6 residues coded by 3 CGA repeats, the least used codon in yeast. Therefore, this reporter contains not only a polybasic motif but also a strong IP sequence (31). As predicted, the deletion of *LTN1* stabilized the polypeptide just before the R12 stalling point (Fig 5B, GFP), while the deletion of *ASC1* and *HEL2* led to the stabilization of the full-length construct (Fig 5B, GFP-RFP).

Another way to check whether a transcript is a target for the RQC pathway is through an increase in protein targets caused by the saturation of the RQC machinery. This can be accomplished by treatment with low concentrations of an elongation inhibitory drug, such as cycloheximide (CHX) (6, 25). Accordingly, Figure 5C shows that the GFP signal from the GFP-R12-RFP reporter increased at low CHX concentrations and that this effect was Ltn1 dependent (Fig 5C). However, when we tested the Rqc1-TAP strain, we could not observe the same result, further indicating that RQC does not control the final Rqc1 protein output during translation (Fig 5D). Instead, the increase in Rqc1 expression observed in the Δ*ltn1* strain should be interpreted as evidence for a posttranslational regulatory interaction. Moreover, despite the apparent challenge that polybasic protein motifs impose on translation, we conclude that the RQC machinery is not constantly targeting these sequences in yeast.

## DISCUSSION

The hypothesis that the translation of positive amino acids leads to ribosome stalling and translation repression has been disputed since shortly after its proposal. The Grayhack group was the first to propose that codon bias was the major factor behind ribosome stalling observed with positive sequences (30, 31, 55, 56). They examined the influence of all sequences of ten repeated codons on luciferase activity, and they demonstrated that the codon choice had a stronger effect than the final

charge. Of particular interest is the codon CGA, which leads to a 99% reduction in luciferase activity in comparison with the codon AGA, both coding for arginine (30). In more recent works, they proposed that the CGA codon does not need to be highly concentrated to cause translational repression: the presence of as few as four consecutive CGA codons (56, 57) or specific combinations of CGA with a second codon, the so-called inhibitory pairs (48, 58), are sufficient to inhibit translation. In the same token, the Green and Djuranovic groups proposed that poly-A sequences are the major cause of translation repression observed with poly-lysine stretches (11, 47, 49, 59). They showed that the presence of six AAA codons in tandem leads to an approximately 30% to 70% reduction in protein expression in comparison to AAG in different organisms and that this reduction is proportional to the quantity of AAA (11, 49). Letzring et al., 2010, had already observed that the codon AAA is worse than AAG (30), but what both the Green and the Djuranovic groups demonstrated is that the translation disruption is caused by the presence of poly-adenine in the mRNA, independent of codon choice. Sequences of AAA codons followed or preceded by codons rich in adenine (preceded by CAA and followed by AAC, for example) can cause greater translation repression than the same sequences of AAA codons alone (11). The presence of an extended poly-A sequence can cause ribosomal sliding, a slow ribosome frameshifting movement that leads to the incorporation of extra lysine residues in the nascent peptide. More importantly, it leads to the translation of premature termination codons, which are targeted by nonsense-mediated decay and lead to translation termination (49). This may be caused by the helical conformation that poly-A sequences can adopt, which allows them to enter the ribosomal A site and inhibit translation (47).

On the other hand, there is still evidence that charge alone can play a role in translation dynamics. Positively charged sequences, independently of codon choice, are translated at slower rates than negatively charged or noncharged sequences (33, 41), which is still considered a hallmark of ribosome stalling. Additionally, proteins with a high density of negatively charged amino acids in a 30-residue stretch are far more common than proteins with a high concentration of R and/or K in 30 residues, suggesting a negative selection pressure against long polybasic stretches (33, 46). Moreover, even in the papers that highlight the effect of codon choice on the inhibitory effect observed with polybasic sequences, we can see evidence that charge alone is still a relevant factor for translation efficiency. Both Arthur et al., 2015, and Koutmou et al., 2015, show that the translation of twelve AGA codons leads to a reduction in protein amount of approximately 50% (11, 49). Taken together, the literature suggests that the translation of rare codons (CGA) and poly-A sequences is detrimental for translation but does not rule out the effect of positively charged segments as RQC complex activation factors.

Our search for endogenous targets of the RQC complex was motivated by the possibility of translation repression caused by positively charged segments, as described above. Rqc1 was the only known example of an endogenous protein with posttranscriptional regulation promoted by the RQC complex (6). Our results show that polybasic domains similar to Rqc1 are relatively common (Figs 1 and 2), and even though they are translated at slower rates (41), it seems that the stalling is not sufficient to cause a severe translation pause, as characterized by the stabilization of 20 to 22 nucleotide

footprints (44, 47) (Fig 2G). Consistent with our results that polybasic domains are not avoided and do not cause premature termination, not even the most positively charged domain could be identified as an endogenous target of the RQC complex (Fig 4). Based on the lack of evidence of endogenous targets, we believe that the RQC complex works mainly on aberrant translation products and plays an insignificant role in protein expression regulation.

Similar to what was observed with other proteins with polybasic domains, we showed that the RQC complex does not inhibit Rqc1 translation. The only background associated with Rqc1 overexpression was Δ*ltn1*. In our hands, Δ*rqc2*, Δ*hel2*, and Δ*asc1* did not significantly affect relative Rqc1 levels, which speaks against an RQC complex-dependent regulation (Fig 5A). The result with Δ*ltn1* agrees with the result from Brandman et al. 2012, which, at the time, was interpreted as evidence that the Rqc1-Flag nascent protein was being targeted by the RQC complex. However, the deletion of *LTN1* alone is usually associated with the stabilization of translation intermediates (up to the stalling point), not full-length constructs, as we reproduced using a stalling reporter system (Fig 5B). Brandman et al. also used an Rqc1-Flag construct where K and/or R was substituted by alanine and observed an increase in protein expression in RQC mutants that was not accompanied by an increase in mRNA levels (6). Taking our data into account, we speculate that the K-rich basic domain can function as the ubiquitylation site involved in Rqc1 regulation. Since it was previously reported that Ltn1 could ubiquitylate substrates independently of the RQC complex (60, 61), we speculate that Ltn1 might regulate Rqc1 in a process that is uncoupled from Rqc1 translation.

## DATA AVAILABILITY

The algorithms used in this work are available in the GitHub repository (https://github.com/ mhoyerm).


## ACKNOWLEDGMENT

We thank Dr. Wendy Gilbert for providing us with the Asc1 antibody and Dr. Onn Brandman for providing us with a GFP-R12RFP stalling reporter. We also thank Santiago Alonso e Ramon Cid for technical support.

## FUNDING

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

## CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

**Supplemental data for this article can be accessed on the publisher's website.**

## REFERENCES


1. Lu,J. and Deutsch,C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.*, **384**, 73–86.

2. Lu,J., Kobertz,W.R. and Deutsch,C. (2007) Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.*, **371**, 1378–1391.

3. Frischmeyer,P.A., van Hoof,A., O'Donnell,K., Guerrerio,A.L., Parker,R. and Dietz,H.C. (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*, **295**, 2258–2261.

4. van Hoof,A., Frischmeyer,P.A., Dietz,H.C. and Parker,R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science*, **295**, 2262–2264.

5. Inada,T. and Aiba,H. (2005) Translation of aberrant mRNAs lacking a termination codon or with a shortened 3'-UTR is repressed after initiation in yeast. *EMBO J.*, **24**, 1584–1595.

6. Brandman,O., Stewart-Ornstein,J., Wong,D., Larson,A., Williams,C.C., Li,G.-W., Zhou,S., King,D., Shen,P.S., Weibezahn,J., *et al.* (2012) A Ribosome-Bound Quality Control Complex Triggers Degradation of Nascent Peptides and Signals Translation Stress. *Cell*, **151**, 1042–1054.

7. Defenouillère,Q., Yao,Y., Mouaikel,J., Namane,A., Galopier,A., Decourty,L., Doyen,A., Malabat,C., Saveanu,C., Jacquier,A., *et al.* (2013) Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5046–5051.

8. Bengtson,M.H. and Joazeiro,C.A.P. (2010) Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature*, **467**, 470–473.

9. Choe,Y.-J., Park,S.-H., Hassemer,T., Körner,R., Vincenz-Donnelly,L., Hayer-Hartl,M. and Hartl,F.U. (2016) Failure of RQC machinery causes protein aggregation and proteotoxic stress. *Nature*, **531**, 191–195.

10. Ito-Harashima,S., Kuroha,K., Tatematsu,T. and Inada,T. (2007) Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. *Genes Dev.*, **21**, 519–524.

11. Arthur,L.L., Pavlovic-Djuranovic,S., Koutmou,K.S., Green,R., Szczesny,P. and Djuranovic,S. (2015) Translational control by lysine-encoding A-rich sequences. *Sci. Adv.*, **1**, e1500154.

12. Dimitrova,L.N., Kuroha,K., Tatematsu,T. and Inada,T. (2009) Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. *J. Biol. Chem.*, **284**, 10343–10352.

13. Defenouillère,Q., Zhang,E., Namane,A., Mouaikel,J., Jacquier,A. and Fromont-Racine,M. (2016) Rqc1 and Ltn1 Prevent C-terminal Alanine-Threonine Tail (CAT-tail)-induced Protein Aggregation by Efficient Recruitment of Cdc48 on Stalled 60S Subunits. *J. Biol. Chem.*, **291**, 12245–12253.

14. Juszkiewicz,S., Chandrasekaran,V., Lin,Z., Kraatz,S., Ramakrishnan,V. and Hegde,R.S. (2018) ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Mol. Cell*, **72**, 469–481.

15. Joazeiro,C.A.P. (2019) Mechanisms and functions of ribosome-associated protein quality control. *Nat. Rev. Mol. Cell Biol.*, **20**, 368–383.

16. Joazeiro,C.A.P. (2017) Ribosomal Stalling During Translation: Providing Substrates for Ribosome-Associated Protein Quality Control. *Annu. Rev. Cell Dev. Biol.*, **33**, 343–368.

17. Chu,J., Hong,N.A., Masuda,C.A., Jenkins,B.V., Nelms,K.A., Goodnow,C.C., Glynne,R.J., Wu,H., Masliah,E., Joazeiro,C.A.P., *et al.* (2009) A mouse forward genetics screen identifies LISTERIN as an E3 ubiquitin ligase involved in neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2097–2103.

18. Gamerdinger,M. (2016) Protein quality control at the ribosome: focus on RAC, NAC and RQC. *Essays Biochem.*, **60**, 203–212.

19. Juszkiewicz,S. and Hegde,R.S. (2017) Initiation of Quality Control during Poly(A) Translation Requires Site-Specific Ribosome Ubiquitination. *Mol. Cell*, **65**, 743–750.

20. Sitron,C.S., Park,J.H. and Brandman,O. (2017) Asc1, Hel2, and Slh1 couple translation arrest to nascent chain degradation. *RNA*, **23**, 798–810.

21. Shen,P.S., Park,J., Qin,Y., Li,X., Parsawar,K., Larson,M.H., Cox,J., Cheng,Y., Lambowitz,A.M., Weissman,J.S., *et al.* (2015) Protein synthesis. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. *Science*, **347**, 75–78.

22. Yonashiro,R., Tahara,E.B., Bengtson,M.H., Khokhrina,M., Lorenz,H., Chen,K.-C., Kigoshi-Tansho,Y., Savas,J.N., Yates,J.R., Kay,S.A., *et al.* (2016) The Rqc2/Tae2 subunit of the ribosome-associated quality control (RQC) complex marks ribosome-stalled nascent polypeptide chains for aggregation. *Elife*, **5**, e11794.

23. Kostova,K.K., Hickey,K.L., Osuna,B.A., Hussmann,J.A., Frost,A., Weinberg,D.E. and Weissman,J.S. (2017) CAT-tailing as a fail-safe mechanism for efficient degradation of stalled nascent polypeptides. *Science*, **357**, 414–417.

24. Sitron,C.S. and Brandman,O. (2019) CAT tails drive degradation of stalled polypeptides on and off the ribosome. *Nat. Struct. Mol. Biol.*, **26**, 450–459.

25. Simms,C.L., Yan,L.L. and Zaher,H.S. (2017) Ribosome Collision Is Critical for Quality Control during No-Go Decay. *Mol. Cell*, **68**, 361–373.

26. Simms,C.L., Yan,L.L., Qiu,J.K. and Zaher,H.S. (2019) Ribosome Collisions Result in +1 Frameshifting in the Absence of No-Go Decay. *Cell Rep.*, **28**, 1679–1689.

27. Ikeuchi,K., Tesina,P., Matsuo,Y., Sugiyama,T., Cheng,J., Saeki,Y., Tanaka,K., Becker,T., Beckmann,R. and Inada,T. (2019) Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. *EMBO J.*, **38**, e100276.

28. Heyer,E.E. and Moore,M.J. (2016) Redefining the Translational Status of 80S Monosomes. *Cell*, **164**, 757–769.

29. Park,H. and Subramaniam,A.R. (2019) Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLoS Biol.*, **17**, e3000396.

30. Letzring,D.P., Dean,K.M. and Grayhack,E.J. (2010) Control of translation efficiency in yeast by codon-anticodon interactions. *RNA*, **16**, 2516–2528.

31. Letzring,D.P., Wolf,A.S., Brule,C.E. and Grayhack,E.J. (2013) Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. *RNA*, **19**, 1208–1217.

32. Fedyukina,D.V. and Cavagnero,S. (2011) Protein folding at the exit tunnel. *Annu. Rev. Biophys.*, **40**, 337–359.

33. Requião,R.D., Fernandes,L., de Souza,H.J.A., Rossetto,S., Domitrovic,T. and Palhano,F.L. (2017) Protein charge distribution in proteomes and its impact on translation. *PLoS Comput. Biol.*, **13**, e1005549.

34. Presnyak,V., Alhusaini,N., Chen,Y.-H., Martin,S., Morris,N., Kline,N., Olson,S., Weinberg,D., Baker,K.E., Graveley,B.R., *et al.* (2015) Codon Optimality Is a Major Determinant of mRNA Stability. *Cell*, **160**, 1111–1124.

35. Carneiro,R.L., Requião,R.D., Rossetto,S., Domitrovic,T. and Palhano,F.L. (2019) Codon stabilization coefficient as a metric to gain insights into mRNA stability and codon bias and their relationships with translation. *Nucleic Acids Res.*, **47**, 2216–2228.

36. Puig,O., Caspary,F., Rigaut,G., Rutz,B., Bouveret,E., Bragado-Nilsson,E., Wilm,M. and Séraphin,B. (2001) The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. *Methods*, **24**, 218–229.

37. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.

38. Volland,C., Urban-Grimal,D., Géraud,G. and Haguenauer-Tsapis,R. (1994) Endocytosis and degradation of the yeast uracil permease under adverse conditions. *J. Biol. Chem.*, **269**, 9833–9841.

39. Boucherie,H. and Monribot-Espagne,C. (2006) Two-Dimensional Gel Electrophoresis of Total Yeast Proteins. In: Xiao W. (eds) *Yeast Protocol*, Methods in Molecular Biology, Humana Press, Totowa, NJ, pp. 47–64.

40. Coyle,S.M., Gilbert,W.V. and Doudna,J.A. (2009) Direct link between RACK1 function and localization at the ribosome in vivo. *Mol. Cell. Biol.*, **29**, 1626–1634.

41. Requião,R.D., de Souza,H.J.A., Rossetto,S., Domitrovic,T. and Palhano,F.L. (2016) Increased ribosome density associated to positively charged residues is evident in ribosome profiling experiments performed in the absence of translation inhibitors. *RNA Biol.*, **13**, 561–568.

42. Ingolia,N.T., Brar,G.A., Rouskin,S., McGeachy,A.M. and Weissman,J.S. (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.*, **7**, 1534–1550.

43. Pop,C., Rouskin,S., Ingolia,N.T., Han,L., Phizicky,E.M., Weissman,J.S. and Koller,D. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.*, **10**, 770–770.

44. Wu,C.C.-C., Zinshteyn,B., Wehner,K.A. and Green,R. (2019) High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Mol. Cell*, **73**, 959–970.

45. Demšar,J., Curk,T., Erjavec,A., Gorup Č., Hočevar, T. Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., *et al*. (2013) Orange: data mining toolbox in Python. *J. Mach. Learn. Res.,***14**, 2349-2353.

46. Karlin,S., Brocchieri,L., Bergman,A., Mrazek,J. and Gentles,A.J. (2002) Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 333–338.

47. Tesina,P., Lessen,L.N., Buschauer,R., Cheng,J., Wu,C.C.-C., Berninghausen,O., Buskirk,A.R., Becker,T., Beckmann,R. and Green,R. (2019) Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts. *biorXiv*, 10.1101/755652, 04-September-2019, pre-print: not peer-reviewed.

48. Gamble,C.E., Brule,C.E., Dean,K.M., Fields,S. and Grayhack,E.J. (2016) Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell*, **166**, 679–690.

49. Koutmou,K.S., Schuller,A.P., Brunelle,J.L., Radhakrishnan,A., Djuranovic,S. and Green,R. (2015) Ribosomes slide on lysine-encoding homopolymeric A stretches. *Elife*, **4**, e05534.

50. Habich,M., Djuranovic,S. and Szczesny,P. (2016) PATACSDB—the database of polyA translational attenuators in coding sequences. *Peer J. Comput. Sci.*, **2**, e45.

51. 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.

52. Siwiak,M. and Zielenkiewicz,P. (2010) A Comprehensive, Quantitative, and Genome-Wide Model of Translation. *PLoS Comput Biol*, **6**, e1000865–15.

53. Cuperus,J.T., Groves,B., Kuchina,A., Rosenberg,A.B., Jojic,N., Fields,S. and Seelig,G. (2017) Deep learning of the regulatory grammar of yeast 5′ untranslated regions from 500,000 random sequences. *Genome Res.*, **27**, 2015–2024.

54. Archer,S.K., Shirokikh,N.E., Beilharz,T.H. and Preiss,T. (2016) Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*, **535**, 570–574.

55. Dean,K.M. and Grayhack,E.J. (2012) RNA-ID, a highly sensitive and robust method to identify cis-regulatory sequences using superfolder GFP and a fluorescence-based assay. *RNA*, **18**, 2335–2344.

56. Wolf,A.S. and Grayhack,E.J. (2015) Asc1, homolog of human RACK1, prevents frameshifting in yeast by ribosomes stalled at CGA codon repeats. *RNA*, **21**, 935–945.

57. Wang,J., Zhou,J., Yang,Q. and Grayhack,E.J. (2018) Multi-protein bridging factor 1(Mbf1), Rps3 and Asc1 prevent stalled ribosomes from frameshifting. *Elife*, **7**, 934.

58. Ghoneim,D.H., Zhang,X., Brule,C.E., Mathews,D.H. and Grayhack,E.J. (2018) Conservation of location of several specific inhibitory codon pairs in the Saccharomyces sensu stricto yeastsreveals translational selection. *Nucleic Acids Res.*, **59**, 149–14.

59. Arthur,L.L. and Djuranovic,S. (2018) PolyA tracks, polybasic peptides, poly-translational hurdles. *WIREs RNA*, **9**, e1486–12.

60. Maurer,M.J., Spear,E.D., Yu,A.T., Lee,E.J., Shahzad,S. and Michaelis,S. (2016) Degradation Signals for Ubiquitin-Proteasome Dependent Cytosolic Protein Quality Control (CytoQC) in Yeast. *G3 (Bethesda)*, **6**, 1853–1866.

61. Zuzow,N., Ghosh,A., Leonard,M., Liao,J., Yang,B. and Bennett,E.J. (2018) Mapping the mammalian ribosome quality control complex interactome using proximity labeling approaches. *Mol. Biol. Cell.*, **29**, 1258–1269.

**FIGURES**

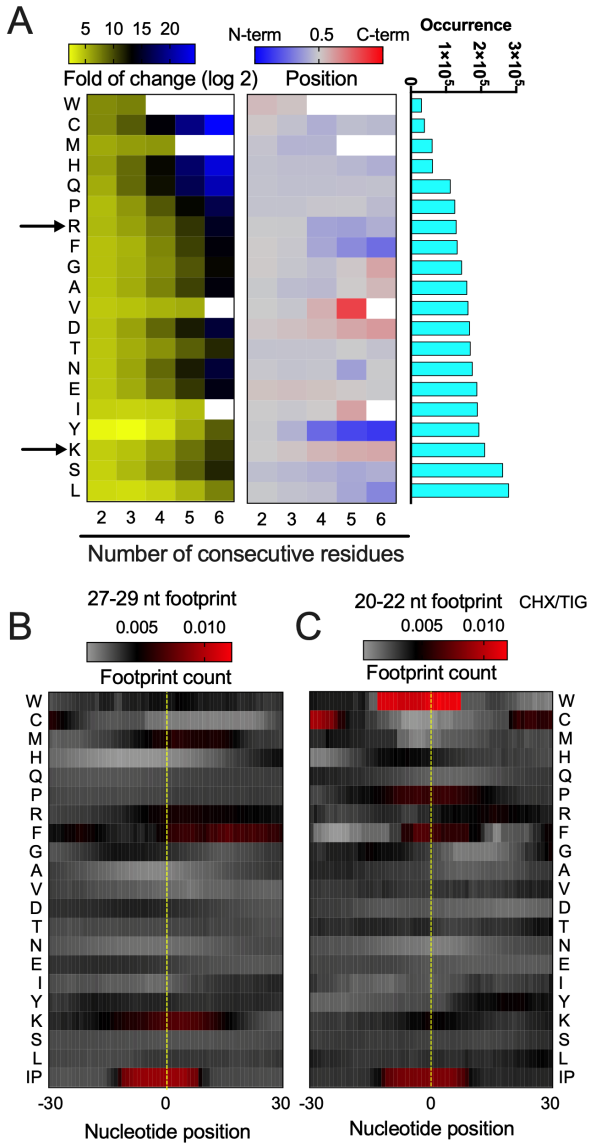


**Figure 1. The frequency, distribution, and effect on the translation of consecutive clusters of amino acids in the yeast genome.** (A) Amino acid clusters from 2 to 6 consecutive residues of yeast protein sequences show a higher frequency of occurrence than expected by a radon distribution. The raw data are presented in Supplemental Table S1. The middle panel represents the mean of the normalized localization of each cluster of amino acids. For each gene, the desired amino acid cluster was localized, and the position was determined in relation to the length of the gene. The panel on the right (blue bars) shows the absolute number of occurrences of each amino acid in the yeast proteome. (B) 27-29 and (C) 20-22 nucleotide footprint analyses of genes containing a cluster of ≥4 (except for C, M, W, and Y, where the cutoff used was ≥3) consecutive amino acids. The dotted yellow line represents the start of the consecutive amino acid cluster. As a control, genes with one of the 17 combinations of inhibitory codon pairs (IP) were used. The 20-22 nt footprint was obtained by the use of cycloheximide and tigecycline (CHX/TIG).

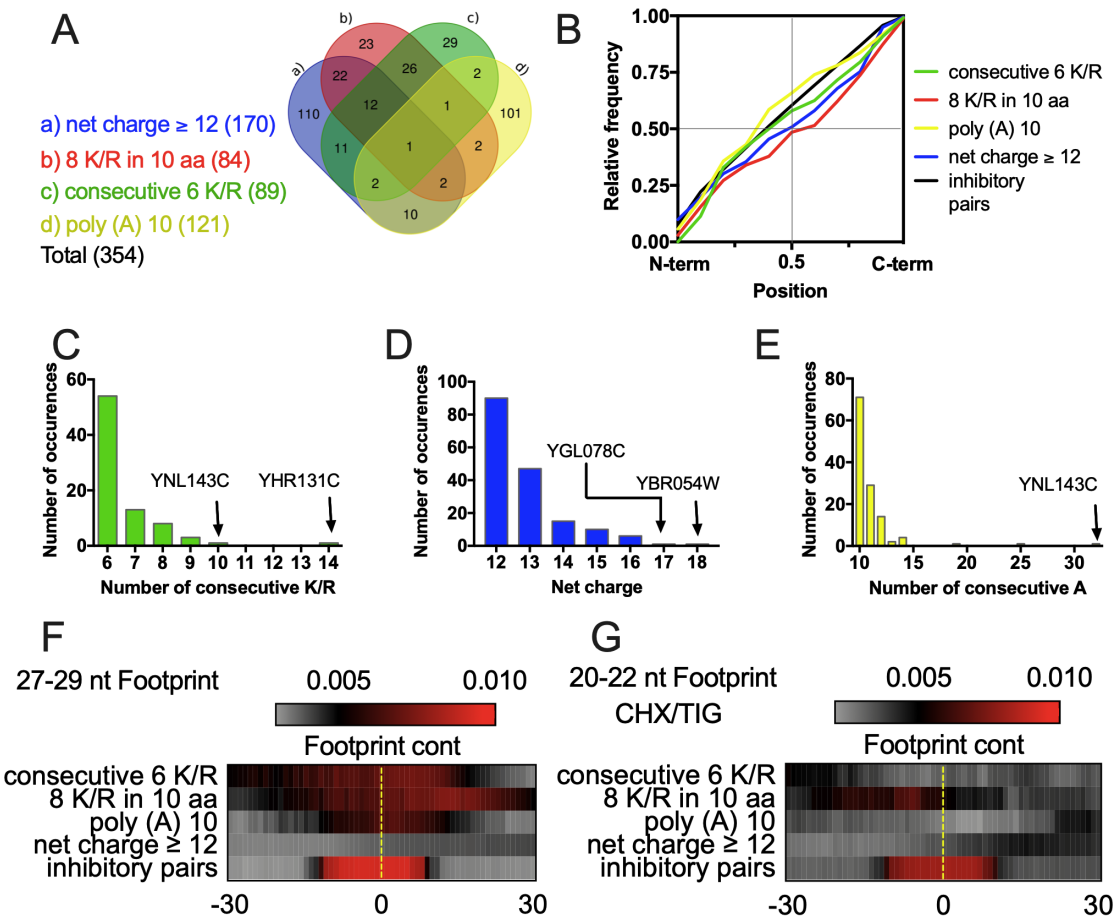


**Figure 2. Bioinformatics screening to identify the polybasic domains in the yeast proteome.** (A) Venn diagram showing the genes identified as potential targets of the RQC complex. The list of genes identified is presented in Supplemental Table S2. (B) Cumulative frequency distribution of the localization of each group of polybasic domains. For each gene, the desired attribute was localized, and the position was determined in relation to the length of the gene. The frequency of distribution of genes is depicted in the following categories: consecutive 6 K and/or R (C, consecutive 6 K/R), net charge equal to or higher than + 12 (D, net charge 12) and 10 or more consecutive adenines (A) (E, poly (A) 10). Highlighted are the genes with the top attribute in each analysis. (F) 27-29 and (G) 20-22 nucleotide footprint analyses of genes containing polybasic domains. The dotted yellow line represents the start of the feature analyzed. As a control, genes with one of the 17 inhibitory codon pairs (IP) were used. The 20-22 nt footprint was obtained by the use of cycloheximide and tigecycline (CHX/TIG).

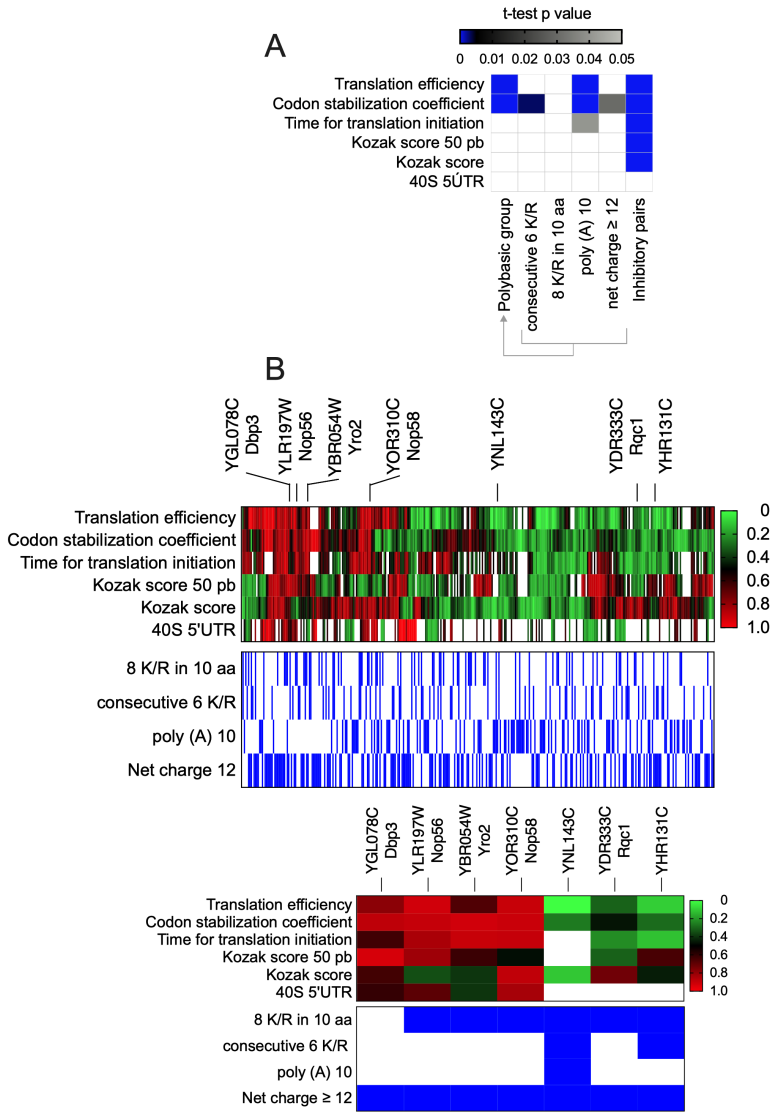


**Figure 3. Translation initiation rate of genes with polybasic domains.** (A) The different groups of genes with some polybasic domains were compared regarding parameters that, to some extent, reflect translation initiation rate with the genome. The Kolmogorov-Smirnov test p values are plotted for each comparison. A white rectangle means that a nonsignificant difference was found (p values >0.05). As a control, genes with one of the 17 inhibitory codon pairs (IP) were used. The parameters used to infer initiation gene rate were translation efficiency (51), time for translation initiation (52), Kozak score of the 50 first nucleotides (53), Kozak score of the first 5 nucleotides (53), 40S ribosomal small subunit (SSU) footprint profiling (SSU 5'UTR: start codon) (54) and the codon stabilization coefficient (35). (B) The values of the parameter described in panel A were normalized, where 0 and 1 represent the lowest and highest initiation rates of the full genome, respectively. Then, the 354 genes with some extremely polybasic features were clustered by the Euclidean distance. The lower panel is the genes with the top polybasic domains, according to the criterion set up in Figure 2A. YDR333C/Rqc1 was included for comparison. The raw data, sample size, and p values are presented in Supplemental Table S3.

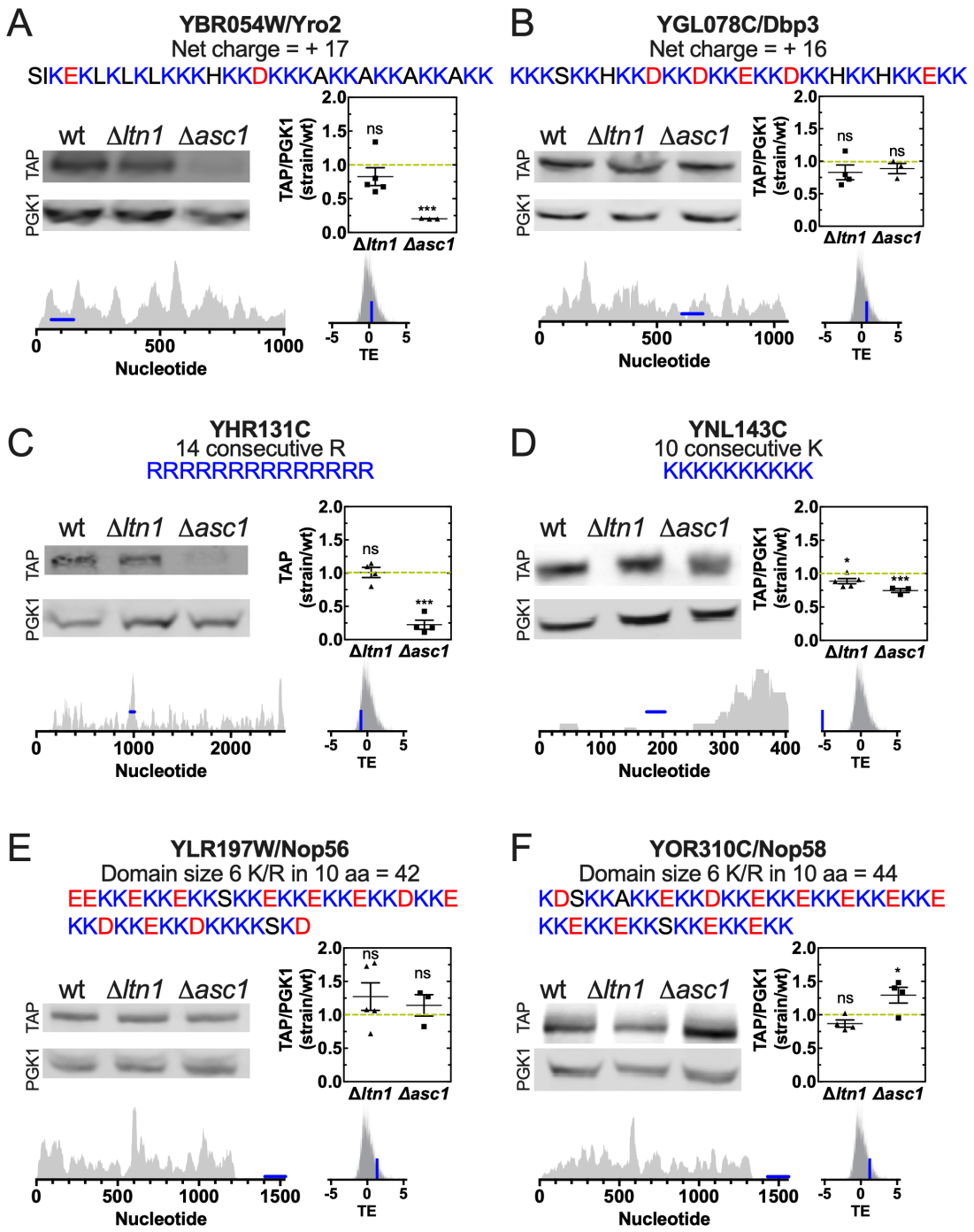


**Figure 4. Proteins with extremely polybasic domains in yeast are not targets of the ribosome quality control (RQC) complex.** Proteins with the top polybasic domains of yeast were measured by western blot in wild type (wt), Δ*lnt1*, and Δ*asc1* strains. For each protein, the primary sequence of the polybasic domain, ribosome profiling, and translation efficiency (TE) is presented. The blue bar in the ribosome profiling data represents the polybasic stalling sequence, while for TE, the value is compared to all genome. The band intensity was quantified and normalized in relation to the housekeeping gene phosphoglycerate kinase (PGK1). One-way ANOVA with Bonferroni's multiple comparison test was used for statistical analyses.

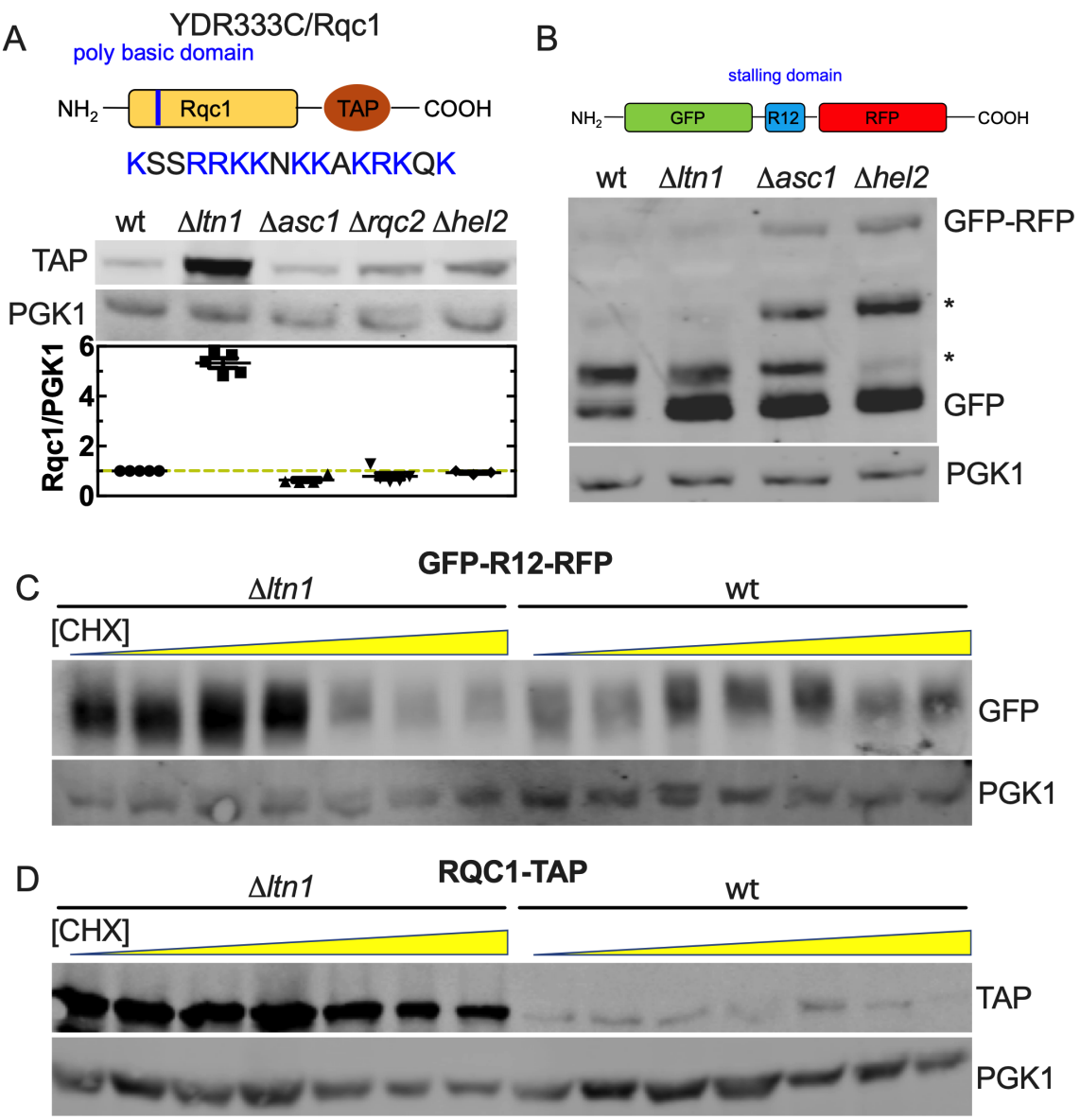


**Figure 5. Rcq1 is not co-translationally degraded by the RQC complex.** (A) A TAP-tagged version of Rqc1 was used to measure the protein levels in the wt, Δ*lnt1*, Δ*asc1*, Δ*rqc2*, and Δ*hel2* strains. The band intensity was quantified and normalized in relation to the housekeeping gene phosphoglycerate kinase (PGK1). (B) Whole-cell immunoblots of the indicated yeast strains expressing the stalling reporter (GFP-R12-RFP). GFP-RFP indicates the full-length translation product, and GFP indicates the arrest product. Stars denote products of the stalling reporter that likely to result from proteolytic cleavage. (C) GFP levels in strains expressing a co-translationally degraded polybasic reporter (GFP-R12-RFP) subjected to 10 hours cycloheximide (CHX) treatment at the increased concentrations. Note that under lower concentrations of CHX, the amount of GFP increases. The same experiment was performed with a Δ*lnt1* strain. (D) The same experiment as panel C was performed with an Rqc1-TAP strain. However, in this case, no increase in protein levels was detected upon CHX incubation.