Journal of Biotechnology 343 (2022) 15–24

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Strong negative correlation between codon usage bias and protein structural disorder impedes protein expression after codon optimization

Kunshan Liu, Yaqi Ouyang, Ru Lin, Chenyu Ge, Mian Zhou *

State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Keywords:
Codon usage bias
Codon optimization
Protein expression
Pichia pastoris

ABSTRACT

As a common phenomenon existing in almost all genomes, codon usage bias has been studied for a long time. Codon optimization is a frequently used strategy to accelerate protein synthesis rate. Besides regulating protein translation speed, codon usage bias has also be reported to affect co-translation folding and transcription. *P. pastoris* is a well-developed expression system, whose efficiency is tightly correlated with commercial value. However, few studies focus on the role of codon usage bias in affecting protein expression in *P. pastoris*. Besides, many genes in *P. pastoris* genome show significant negative correlation between codon usage bias and protein structural disorder tendency. It's not known whether this feature is important for their expression. In order to answer these questions, we picked 4*P. pastoris* gene candidates with strong negative correlation between codon usage bias and protein structural disorder. We then performed full-length codon optimization which completely eliminated the correlation. Protein and RNA assays were then used to compare protein and mRNA levels before and after codon optimization. As a result, codon optimization failed to elevate their protein expression levels, and even resulted in a decrease. As represented by the trypsin sensitivity assays, codon optimization also altered the protein structure of 0616 and 0788. Besides protein, codon optimization also affected mRNA levels. Shown by *in vitro* and *in vivo* RNA degradation assays, the mRNA stability of *0616*, *0788* and *0135* were also altered by codon optimization. For each gene, the detailed effect may be related with its specific sequence and protein structure. Our results suggest that codon usage bias is an important factor to regulate gene expression level, as well as mRNA and protein stabilities in *P. pastoris*. "Extreme" codon optimization in genes with strong negative correlation between codon usage bias and protein structural disorder tendency may not be favored. Compromised strategies should be tried if expression is not successful. Besides, codon optimization may affect protein structural conformation more severely in structural disordered proteins.

1. Introduction

Except tryptophan and methionine, all other amino acids are coded by 2–6 synonymous codons in the nuclear genome. Different species have different codon preferences (Lehmann and Libchaber, 2008), which is referred as codon usage bias. The effect of codon usage bias on protein translation speed is a phenomenon that has been recognized for a long time. Optimal codons are always thought to accelerate translation, while non-optimal codons are known to reduce the translation speed (Weinberg et al., 2016). Optimal codons usually correspond to highly abundant tRNAs, which would recognize the ribosome A site faster. Therefore, the codon optimization strategy has been frequently used to increase protein expression efficiency, by replacing non-optimal codons with optimal ones without changing the amino acid sequence.

In many cases codon optimization successfully improved protein expression levels, however, sometimes unexpected results including expression failure were also identified (Gustafsson et al., 2004; Burgess-Brown et al., 2008). This may be explained by recent discoveries that besides translation speed, codon usage bias also regulates protein folding and function. For example, synonymous mutations on *E. coli* EgFABP1 coding gene changed the protein solubility ranging from 52% to 95% (Cortazzo et al., 2002) and codon optimization of Suf1 affected the efficiency of protein folding, which could be rescued by low temperature treatment (Zhang et al., 2009). In addition to *E. coli*, codon optimization of cyanobacteria circadian clock genes *KaiB* and *KaiC* also affected their functions in circadian clock regulation (Xu et al., 2013). Besides prokaryotes, the effect of codon bias on protein structure and function in eukaryotes has also been reported. Early studies in some

* Corresponding author.
E-mail address: mianzhou@ecust.edu.cn (M. Zhou).

https://doi.org/10.1016/j.jbiotec.2021.11.001
Received 7 June 2021; Received in revised form 14 October 2021; Accepted 4 November 2021
Available online 8 November 2021

gene candidates associated single-nucleotide silent mutation with protein functions, such as human period genes and sleep preference, MDR1 gene and substrate specificity (Kimchi-Sarfaty et al., 2007; Carpen et al., 2006). In 2013, codon usage bias was shown to regulate co-translational folding and protein function by taking the *N. crassa* circadian clock core component FRQ as an example (Zhou et al., 2013). After that, the influence of codon bias on protein structure and function was subsequently confirmed on *D. melanogaster* circadian clock core component dPER and *N. crassa* Cross Pathway Control Protein 1 (CPC-1) (Fu et al., 2016; Lyu and Liu, 2020). In addition, a significant negative correlation between codon usage bias and protein structural disorder was identified in the genomes of five model organisms: *E. coli*, *N. crassa*, *S. cerevisiae*, *C. elegans* and *D. melanogaster*. The use of non-optimal codons in some regions may be a mechanism to reduce translation speed and provide enough folding time (Zhou et al., 2015).

Compared with translation, the influence of codon usage bias on transcription has been overlooked for years. Although synonymous codon substitution does not alter amino acid sequence, the nucleotide sequence is then changed and is likely to affect the mRNA level in terms of transcription activity and mRNA stability (Hia and Takeuchi, 2021). It has been pointed out there was a potential correlation between gene sequence and transcription speed. Specific nucleotide pentamers (5 consecutive nucleotide bases) tended to have a specific transcription rate which was highly conserved in different genes (Cohen et al., 2018). Besides, a correlation between codon usage bias and nuclear RNA levels has also been reported in the genome of *N. crassa*. A batch of transcription factors or histone transmethylases might be involved in supporting this correlation (Zhao et al., 2021). *In vivo* studies in *N. crassa* showed that the transcriptional activity of luciferase increased with codon optimization, which was unrelated with translation. The inner mechanism was further revealed as reduced gene silencing marker H3K9me3 levels in codon optimized sequences (Larrondo et al., 2012; Zhou et al., 2016). In addition to histone modification, previously we also identified the potential correlation among codon choice, cytosine methylation sites and mRNA levels (Xing et al., 2020). In addition to transcription elongation, codon bias may also play a regulatory role at the step of termination. Since the "AAUAAA" transcription termination signal is poorly conserved in fungi, possible combinations of A/U rich codons may mimic this signal and terminate transcription prematurely (Zhou et al., 2018).

The choice of synonymous codons is also likely to affect mRNA half-life (Gustafsson et al., 2004; Horstick et al., 2015). Replacing the optimal codons at the 5′ end of *S. cerevisiae* genes with non-optimal codons resulted in only 1/10 of original protein level and 1/3 of original mRNA level (Hoekema et al., 1987). Presnyak et al. (2015) found in the *S. cerevisiae* genome that high tAI (tRNA Adaption Index) codons always appeared with a higher frequency in mRNAs with a long half-life. However, there are still some debates on the effect of codon usage bias on mRNA stability. Studies in *N. crassa* showed no significant difference in mRNA stability before and after codon optimization in the several genes tested (Horstick et al., 2015; Yang et al., 2019). Therefore, more researches are needed to reveal the principles clearly.

As one of the most frequently used eukaryotic expression system, *P. pastoris* has been used to express a lot of valuable recombinant protein successfully (Ahmad et al., 2014). The expression efficiency of *P. pastoris* system is tightly correlated with industrial output value. Therefore, research topics including promoter engineering, secretion enhancement and induction module have become hotspots to conduct modification and optimization of the *P. pastoris* expression system. However, few existing researches focus on the mechanism of how codon usage bias regulates protein expression in *P. pastoris*, as well as the development of codon optimization strategy. Compared with other organisms, the success rate of codon optimization in *P. pastoris* seems not high. Although good in some cases, codon optimization was reported to elevate α-amylase and PAS_chr2–2_0376 expression levels by only 40%. Some target proteins even failed to be detected after codon optimization (Huang et al., 2017; Xu et al., 2021). Being opposite to *N. crassa*, the *P. pastoris* genome prefers to A/U rich codons over G/C rich ones. A/U rich codons may have some pre-translational disadvantage including premature transcription termination and attenuated mRNA stability. However, more studies should be performed to elucidate the role of codon optimization on protein expression in *P. pastoris*.

Similar as in many other organisms, many *P. pastoris* genes also showed significant negative correlation between codon usage bias and protein structural disorder (Xu et al., 2021). In order to examine whether this negative correlation is necessary and important for the proper expression of these genes, here we picked four gene candidates with different CAI value and disorder score. These four genes were *PAS_FragD_0013*, *PAS_chr1–1_0135*, *PAS_chr1–4_0616* and *PAS_chr4_0788*. We then performed full-length codon optimization and compared protein levels as well as mRNA levels before and after optimization. Trypsin sensitivity assay was then performed to reveal possible protein structural change caused by codon optimization. mRNA stabilities of sequences with and without codon optimization were also compared. We hope that this study would add more information on how codon optimization affects *P. pastoris* protein expression profile and provide some guidance on codon optimization strategy development.

2. Materials and methods

2.1. Strains and culture conditions

E. coli DH5α strain was used for plasmid construction and propagation. DH5α cells were cultured in LB medium, containing 0.5% yeast extract, 1% tryptone and 1.0% NaCl at 37 °C. 100 μg/mL of ampicillin or kanamycin was added to the medium when required.

P. pastoris strain GS115 was used as the wild-type and the host to make transgenic strains. All cells were shaking cultured at 220 rpm, 30 °C. For seed preparation, yeast cells were inoculated into YPD medium (2% tryptone, 1% yeast extract, 2% glucose) until OD600 reached 6–8. If with *GAP* promoter, cells were then inoculated into YPD again with initial OD600 = 1.0.

2.2. Plasmid construction and strain generation

All gene sequences without codon optimization were PCR amplified from GS115 genome and codon optimized sequences were synthesized by Generay. Each sequence was cloned into the pGAPZα vector first, right after the *GAP* promoter. After that, the whole promoter-gene-terminator cassette was cloned into the pPIC 8 K vector to replace the P_{AOX1} expression cassette. pPIC 8 K was previously built based on pPIC 3.5 K by deleting the Amp gene (vector map shown in Fig. S1). One-step ligation kit from Vazyme was used for fragment ligation. All primer sets used are listed in Table S1, and synthesized by Genewiz.

Each final construct was linearized inside *his4* by *Sal*I and electroporated into the GS115 wild-type strain. Positive transformants were selected on histidine deficient plates and strains with single copied transgene were verified by PCR.

2.3. Data resources and codon bias analysis

The genomic information of *P. pastoris* GS115 was retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/genome). ω and CAI values were calculated as described (Sharp and Li, 1987) and codon usage frequency in *P. pastoris* was downloaded from the Codon Usage Database (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4922). Protein structural disorder score was calculated by IUPred (https://iupred2a.elte.hu/). When plotting CAI and disorder curves, a sliding window size of 10 codons were used. Correlation between codon usage bias and disorder tendency were calculated as described previously (Zhou et al., 2015) with a 10-codon sliding window as well. For codon optimization, synonymous codons with highest ω

Table 1
CAI value, disorder tendency score and correlation of 4 gene candidates.

GeneName	CAI	Disorder	Length (Protein)	Correlation	Mw (Protein)
PAS_FragD_0013	0.766	0.247	111	- 0.724	15.7kD
PAS_chr1–1_0135	0.817	0.294	266	- 0.552	28.7kD
PAS_chr1–4_0616	0.704	0.679	227	- 0.539	24.9kD
PAS_chr4_0788	0.725	0.181	255	- 0.630	31.7kD

value were used to replace original codons. All codon optimized and original sequences in this study are shown in Table S2.

2.4. SDS-PAGE and Western blot

Cells were cultured to steady state. For different strains, cell density (represented by OD_{600}) were normalized first and same amount of culture were used. Cells were separated from medium by centrifugation. The cell-free medium was then collected and concentrated for 20 times to generate extracellular protein sample. After breaking cells, RIPA buffer was used to extract intracellular proteins. The suspension was centrifuged at 3000 rpm to get rid of unbroken cells, and 12,000 rpm to separate soluble and insoluble fractions. For soluble samples, protein concentration was measured by Bradford assay and 50 μg total protein was loaded on SDS-PAGE. For extracellular and insoluble samples, same amount of samples were loaded. For western blot, primary antibody was his-tag antibody (Beyotime AH367), and secondary antibody was HRP conjugated (Beyotime A0216). Band intensities were quantified by Image J.

2.5. Trypsin sensitivity assay

Intracellular total proteins were diluted to a concentration of 7 mg/mL. For each strain, 150 μL protein extract was treated with trypsin (final concentration 1–20 μg/mL) at 25 °C. A 20 μL sample was taken from the reaction at indicated time points, boiled with protein loading buffer immediately and resolved on a SDS-PAGE gel. Western blot was performed afterwards to examine target protein levels.

2.6. RNA extraction and quantitative real-time PCR

Cells were cultured to steady state and total RNA was extracted by trizol reagent. RNA concentration was measured by NanoDrop Microvolume Spectrophotometer. For each sample, 2 μg total RNA was used for cDNA synthesis (FastKing RT Kit, TIANGEN) and subsequent quantitative PCR (SYBR Green, TIANGEN). qPCR program was set as following: step 1, 95 °C for 5 min; step 2, 58 °C for 30 s; step 3, 72 °C for 2 min, repeat step 1–3 for 31 cycles; step 4, 72 °C for 2 min. Primers for amplification were designed by the help of Beacon designer 7.9 and sequences are listed in Table S1. Primer amplification efficiency was measured and considered when comparing original and codon optimized sequences.

2.7. RNA degradation assay

2.7.1. In vitro

For each time point, 1 μg total RNA from the *ori* strain and 1 μg total RNA from the *opt* strain were mixed on ice. RNase A (BEYOTIME, ST578) and DEPC water were then added with a final concentration of 0.2 μg/μL and final volume of 10 μL. Immediately the tubes were transferred into 37 °C water bath to allow RNA degradation for different time length. Reverse transcription was then performed with the existence of RNaseA inhibitor. After cDNA synthesis, subsequent quantitative PCR was carried out similarly in 2.6.

2.7.2. In vivo

Cultured cells (OD_{600} = 1.0) were treated 1, 10-phenanthroline (0.15 mg/mL) to inhibit transcription and allow RNA degradation. For each time point, cells were harvested by quick centrifugation and RNA was extracted by trizol reagent. cDNA synthesis and subsequent quantitative PCR was carried out similarly in 2.6.

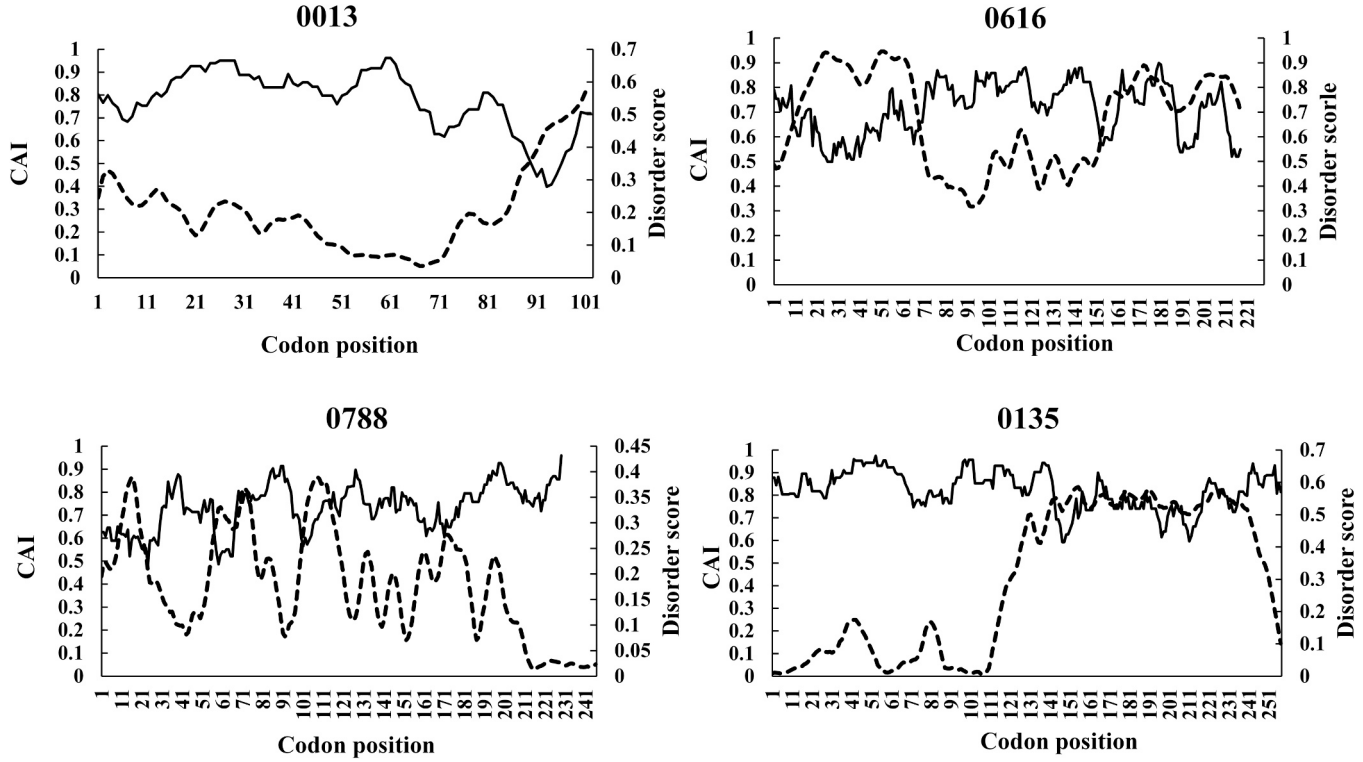

Fig. 1. CAI (solid line) and disorder score (dashed line) curves of the 4 gene candidates in this study.

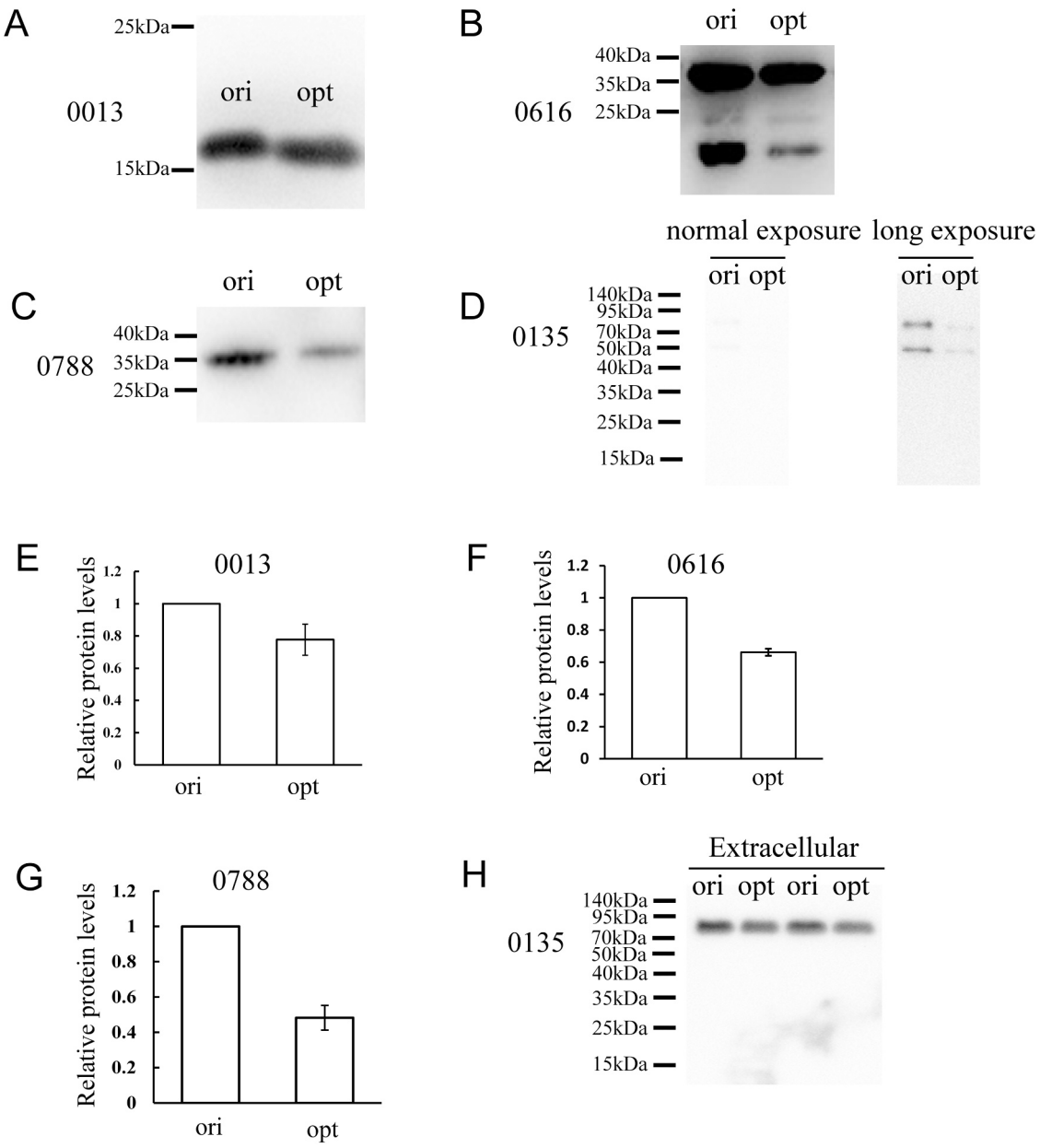

Fig. 2. Protein levels of the 4 gene candidates. (A–D) Western blot analysis showing intracellular protein levels before and after codon optimization. (E–G) Quantification of western blot results of A, B and C, respectively. (H) Western blot analysis showing extracellular 0135 levels before and after codon optimization.

3. Results

3.1. Analyze the codon usage bias and protein structural disorder of the 4 gene candidates

In order to study the necessity of the negative correlation between codon usage bias and protein structural disorder, we picked 4 gene candidates with strong negative correlation. The four genes are *PAS_FragD_0013, PAS_chr1–1_0135, PAS_chr1–4_0616* and *PAS_chr4_0788*, abbreviated as *0013*, *0135*, *0616* and *0788*. Both 0013 and 0788 are predicted to be ribosomal proteins. The other two candidates are annotated as hypothetical proteins without a defined function.

CAI value, disorder tendency score and their correlation are shown in Table 1. The genome average correlation is 0.0383 (Xu et al., 2021), and all these four genes have a correlation below − 0.5 (ranked within top 1% of all the 2084 genes with negative correlation, Table S3). Codon Adaptation Index (CAI) and disorder curves along the ORF are plotted in Fig. 1. Among the 4 gene candidates, *0616* seems to have highest disorder tendency while *0013, 0135,* and *0788* seem to have lower disorder tendency.

3.2. Codon optimization affects both protein and mRNA expression levels of the gene candidates studied

Codon optimization was then performed on these genes, which completely eliminated the negative correlation. All original (named *0616ori*, *0013ori*, *0135ori* and *0788ori*) and codon optimized (named *0616opt*, *0013opt*, *0135opt* and *0788opt*) sequences were cloned into the pPIC 8 K vector and electroporated into yeast wild-type strain GS115. For each sequence, single copied transgenic yeast strain was selected. The expression of these sequences were driven by *GAP* promoter, and a 6 ×His tag was added at C-terminus for antibody recognition. The 6×His tag also distinguished the manually introduced sequence from endogenous sequence.

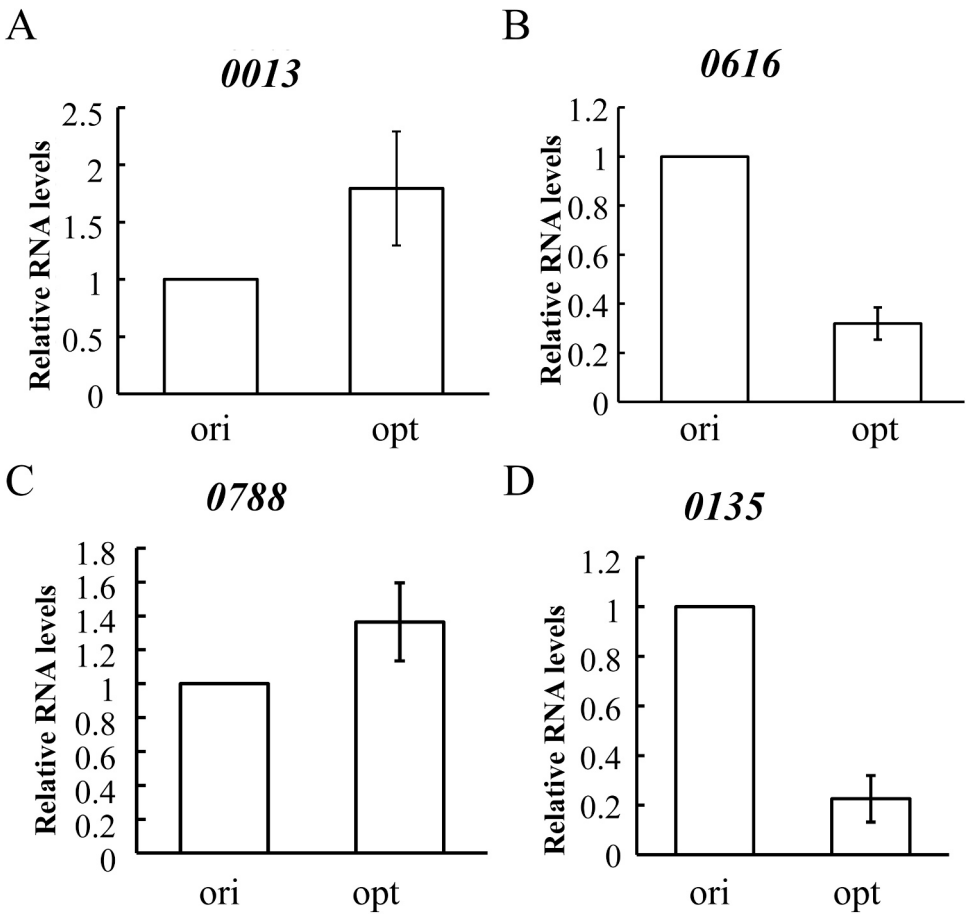

Fig. 3. mRNA levels of the 4 gene candidates in this study, examined by qRT-PCR.

Then intracellular target protein levels were examined in strains bearing original or codon optimized sequences. Western blot suggests that 0013, 0616 and 0788 proteins were successfully expressed (Fig. 2), and the sizes were similar as predicted. The smaller bands in 0616 strains may be partial degradation products (Fig. 2B). Interestingly, codon optimization of the three genes failed to elevate protein expression levels, and even resulted in a decrease. The level of 0013, 0616 and 0788 proteins were decreased by around 20% (Fig. 2A&E), 40% (Fig. 2B&F) and 60% (Fig. 2C&G), respectively. As for 0135, we failed to detect any protein bands in both *ori* and *opt* strains at the beginning with a normal exposure. After a very long exposure, some bands showed up with the size much larger than expected (Fig. 2D). Since 0135 was predicted to have a secretion signal, we then checked its extracellular level. Shown by Fig. 2H, it was indeed detected in culture medium. The size change may be caused by some extensive modifications or unknown mechanisms. Anyway, it seems that codon optimization failed to elevate the 0135 protein level as well (Fig. 2H).

mRNA levels in these strains were then checked by RT-qPCR. As shown by Fig. 3, mRNA levels of *0013* and *0788* were elevated by around 70% (Fig. 3A) and 40% (Fig. 3C) after codon optimization, respectively. However, *0616* mRNA level was decreased to 35% (Fig. 3B) and *0135* mRNA level was decreased to only 20% (Fig. 3D). For *0013* and *0788*, mRNA levels were not consistent with their protein levels. Protein stabilities were examined later to further look into this question.

3.3. Codon optimization affects protein stability and solubility of 0616 and 0788

We then performed trypsin digestion assay to compare the structural stability of proteins translated from original and codon-optimized sequences. Trypsin cleaves proteins after basic amino acids (Lys, Arg). Proteins with different three-dimensional structures may have different extent of basic amino acids exposure on the surface, thus exhibiting different trypsin sensitivity. After adding trypsin, protein samples were taken at time points 0, 5, 15, and 30 min and examined by SDS-PAGE and Western blot. As shown in Fig. 4, codon optimization significantly altered the stability of 0616 (Fig. 4B), slightly decreased the stability of 0788 (Fig. 4C) and did not affect 0013 (Fig. 4A). This indicates that codon optimization possibly affected the conformation of 0616 and 0788. The compromised stability of 0788opt may be one contributing factor towards its lower protein level. As for *0013*, the inconsistent mRNA and protein levels still cannot be explained. However, trypsin sensitivity only measures *in vitro* stability, which may not represent real *in vivo* stability accurately.

Besides stability, solubility of a protein is also tightly correlated with its correct conformation. In addition to soluble fraction, we analyzed target protein levels from the insoluble fraction of intracellular protein sample. As a result, codon optimization decreased the solubility of 0616 significantly (Fig. 5B) and 0788 slightly (Fig. 5C). Besides, solubility of 0013 was improved by codon optimization (Fig. 5A), and no 0135 proteins were detected in the insoluble fraction (Fig. 5D).

3.4. Codon optimization affects 0788 and 0616 mRNA stability

Similar as trypsin digestion, we then performed RNase A treatment to compare mRNA stability of genes before and after codon optimization *in vitro*. For each gene, same amount of total RNA from *ori* and *opt* strains were mixed and treated with RNase A for 0, 5, 10, 15, 20, and 30 min. Samples were then reverse transcribed to cDNA and examined by qPCR to compare *ori* and *opt* mRNA levels. RNase A inhibitor was added during

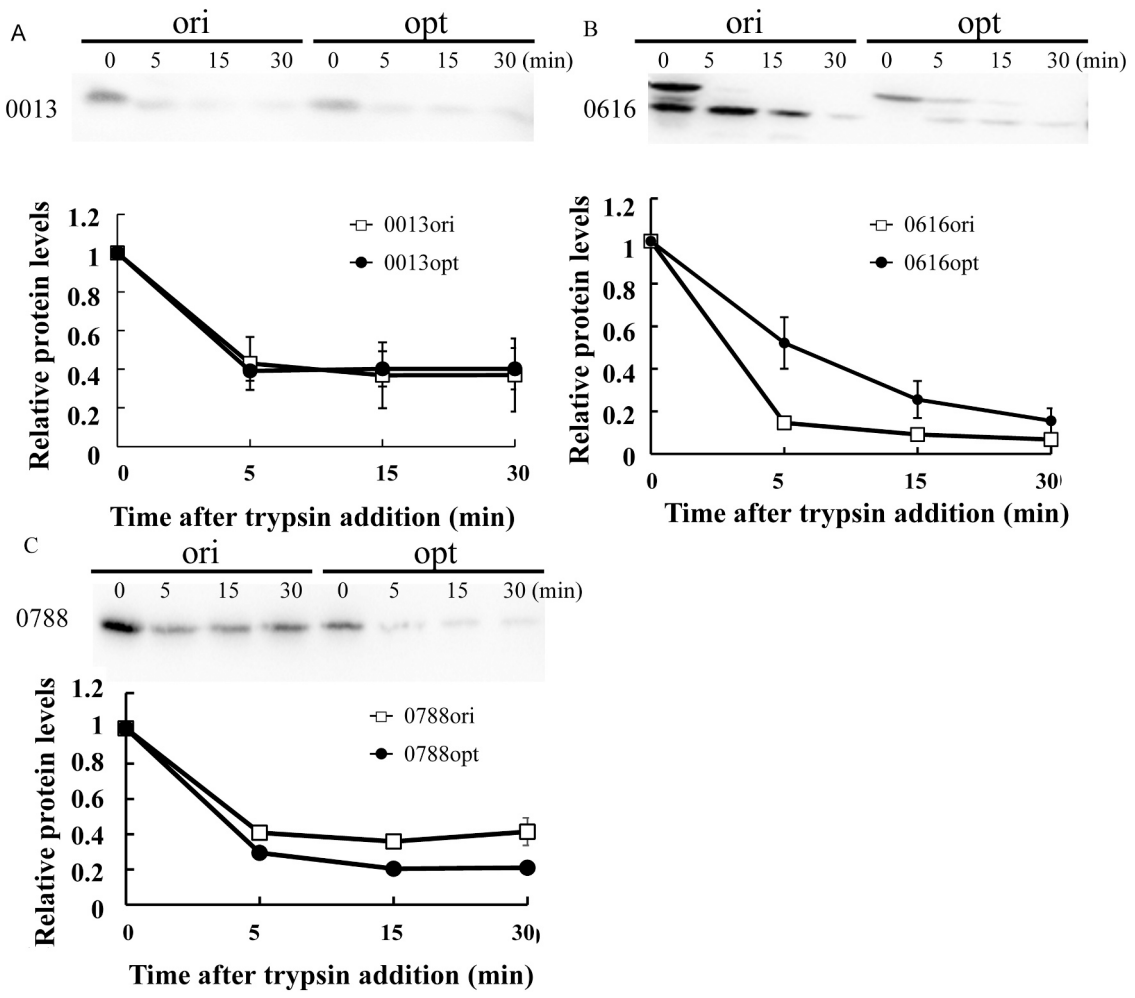

Fig. 4. Trypsin sensitivity of 0013 (A), 0616 (B) and 0788 (C). For each panel, top is western blot analysis of target protein under trypsin digestion; bottom is the quantification of western blot results.

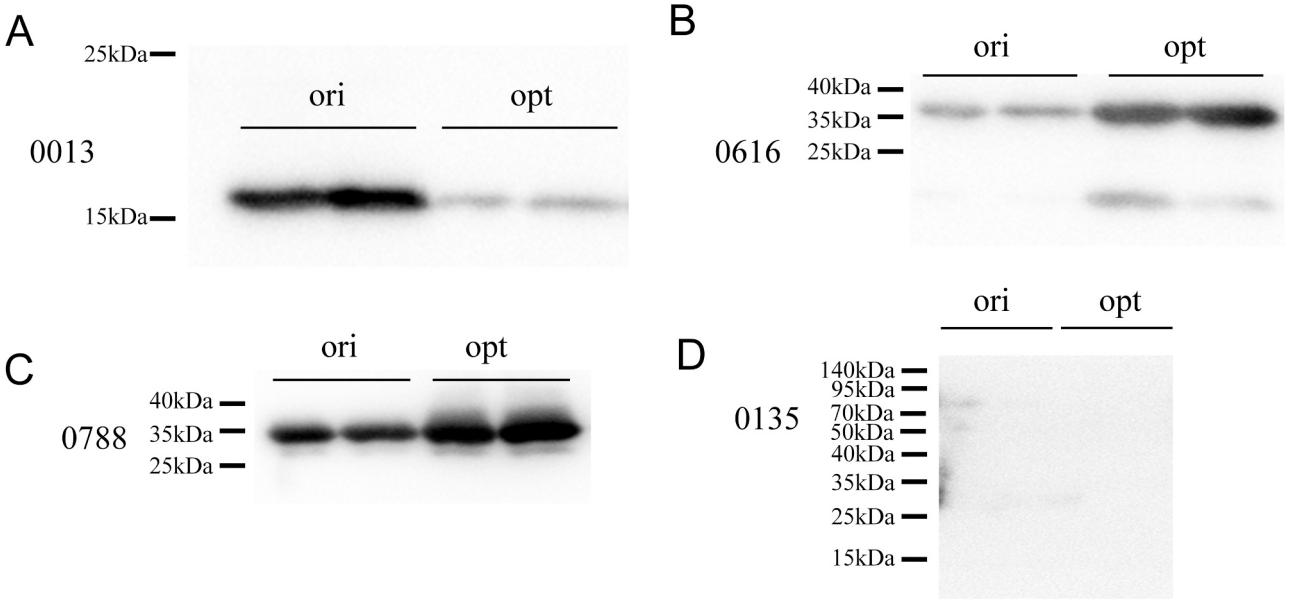

Fig. 5. Protein levels of the 4 gene candidates in insoluble fractions. (A–D) Western blot analysis showing indicated protein levels before and after codon optimization.

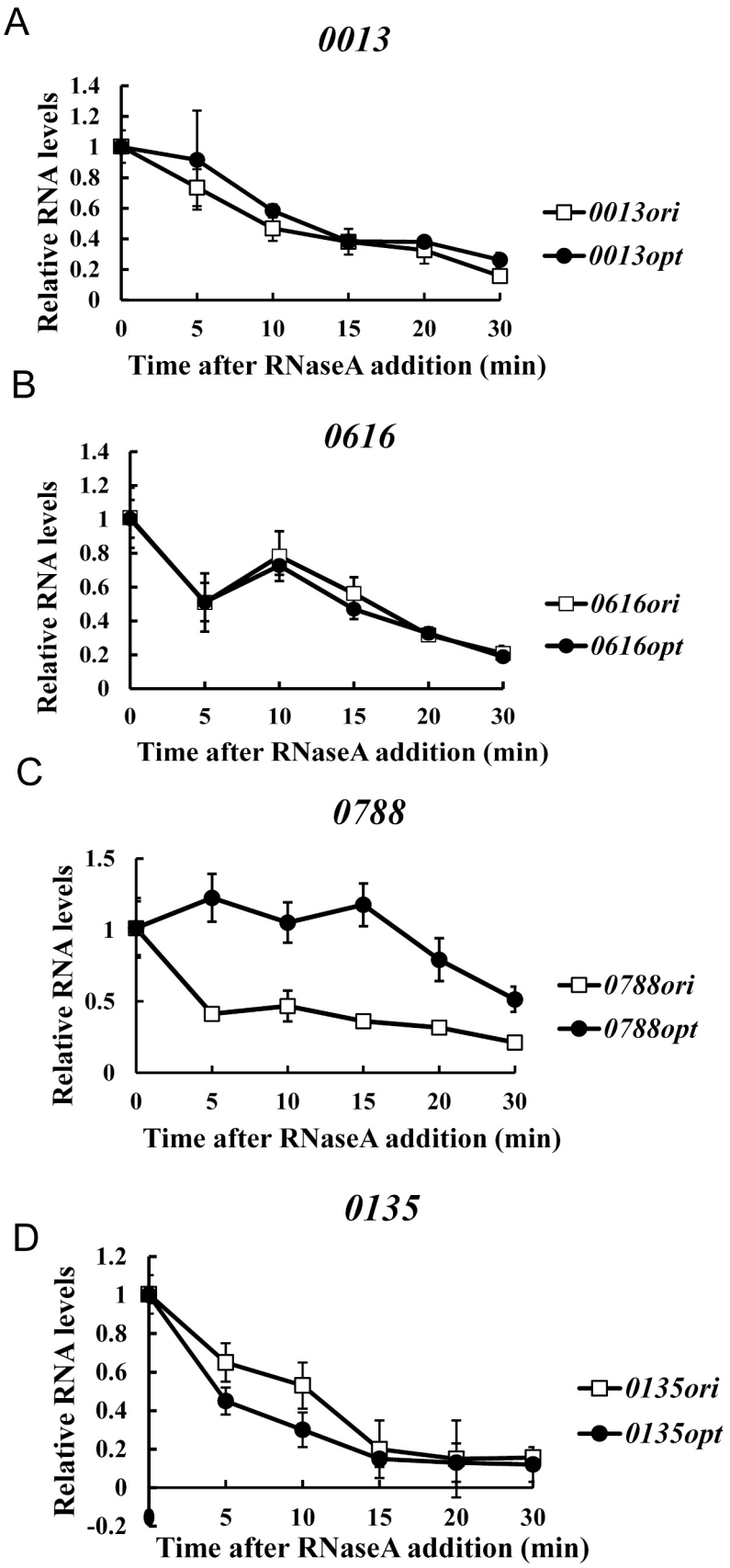

Fig. 6. RNase A sensitivity of 0013 (A), 0616 (B), 0788 (C) and 0135 (D) at different time points.

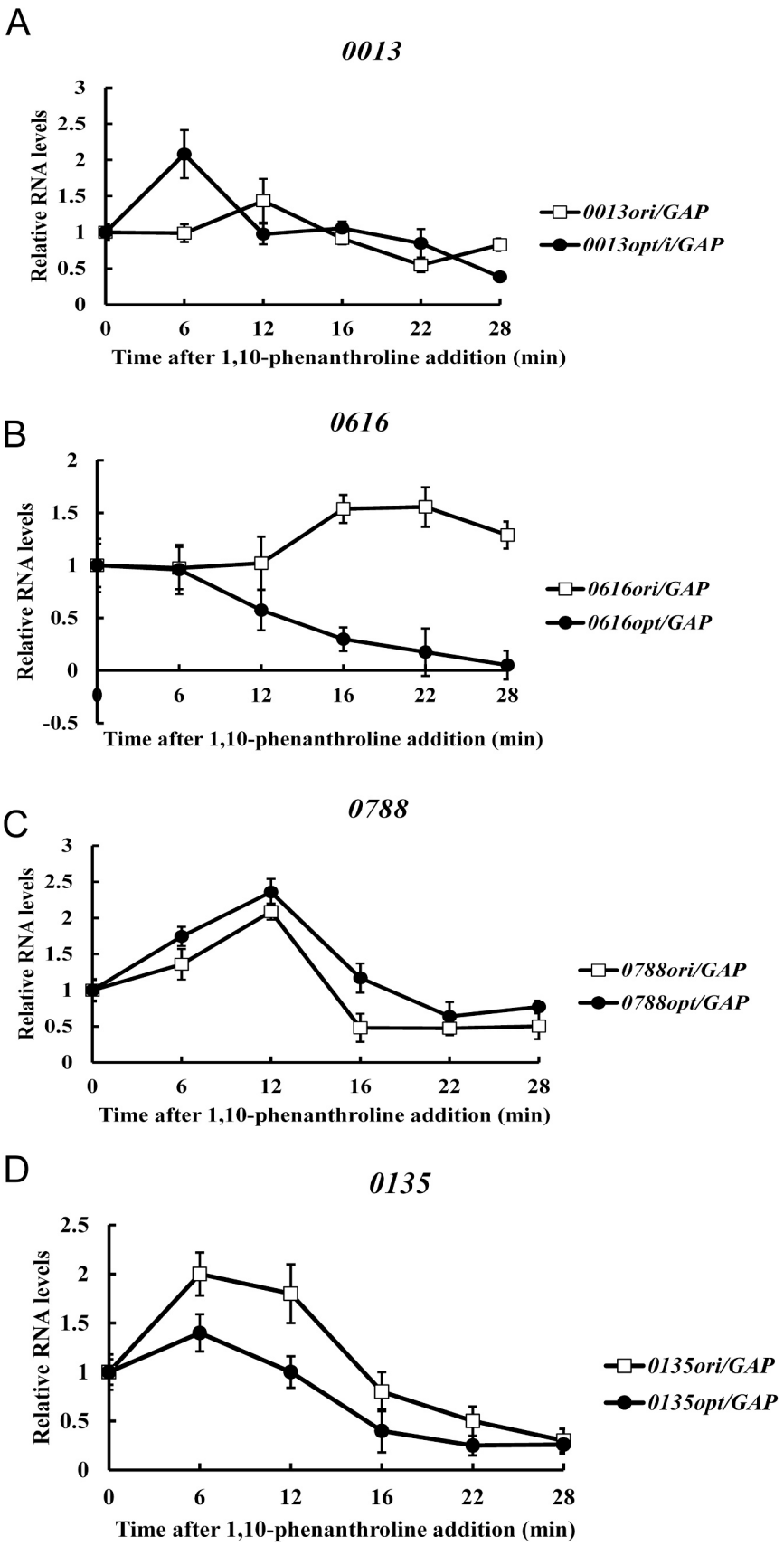

Fig. 7. *In vivo* degradation curves of 0013 (A), 0616 (B), 0788 (C) and 0135 (D) at different time points. For each gene, the relative degradation rate with the *GAP* gene was calculated, and the ratio at time point 0 was normalized to 1.

reverse transcription to avoid further degradation. For each mRNA, the level at 0 min was set as 1, and the relative ratio at other time points were calculated. As shown by Fig. 6A&B, codon optimization has little effect on the stability of *0013* and *0616* mRNA. However, mRNA stability of 0788 was significantly strengthened (Fig. 6C) while that of 0135 (Fig. 6D) was compromised in the codon optimized strains.

Since RNase A sensitivity may not represent real mRNA stability inside the cell, we then measured mRNA degradation rate *in vivo* by 1, 10-phenanthroline treatment. 1, 10-phenanthroline inhibits transcription, and relative RNA degradation rate with *GAP* mRNA was plotted at 0, 6, 12, 16, 22, 28 min time points (Fig. 7). Consistent with RNase A degradation, the *in vivo* degradation rate of *0013* was comparable between *ori* and *opt* sequences (Fig. 7A). *0788opt* was slightly more stable than *0788ori* (Fig. 7C). Although the different was smaller than *in vitro* study (Fig. 6B), this might partially explain its elevated mRNA level (Fig. 3C). Surprisingly, the *in vivo* stability of *0616* was quite different from *in vitro*, with a much faster degradation rate after codon optimization (Fig. 7B). Being consistent with *in vitro* stability, the *in vivo* stability of *0135opt* was also lower. Again, the faster degradation rates of *0616opt* and *0135opt* may be a contributing factor for their compromised mRNA levels (Fig. 3B&D).

4. Conclusion and discussion

In this study, we picked 4 gene candidates in *P. pastoris* with extreme negative correlation between codon usage bias and structural disorder tendency. In order to study the importance of the negative correlation, we performed codon optimization on these 4 genes to completely eliminate it. As a result, codon optimization failed to elevate, and even decreased their protein levels. Codon optimization also altered mRNA level, as well as protein and mRNA stabilities in some genes. The detailed effect might be related with specific gene sequence and protein structure, which differed in each gene.

As reported by many studies in *P. pastoris*, codon optimization has also been applied in a couple of endogenous and heterologous proteins including PAS_chr2–2_0376 (Xu et al., 2021), lysozyme (He et al., 2020), α-amylase (Huang et al., 2017), polygalacturonase (Karaoğlan and Erden-Karaoğlan, 2020) and GFP (data not shown). The effects were not bad, with an elevation by different extents instead of reduction. Compared with the 4 gene candidates here, these genes are with much weaker negative correlation between codon usage bias and secondary structural disorder. Therefore, the poor optimization phenotype in gene studied here may be correlated with the strong negative correlation. However, more examples still need to be given to strengthen this correlation and reveal more inner mechanisms under it. Besides, combining all trypsin sensitivity and solubility phenotypes so far, codon optimization seems to have a larger effect on structure alternation in genes whose protein product have larger structural disorder tendency (*0376*, disorder score 0.656; *0616*, disorder score 0.679). Therefore, when dealing with this group of proteins, special codon optimization strategy may be needed.

Here we only focus on “extreme” codon optimization, in which all codons are replaced by the most optimal synonymous codons. “Extreme” codon optimization may work well in most genes, however, may not be suitable for genes with strong negative correlation. “Weaker” codon optimization strategies by using secondary optimal codons or keeping some negative correlation patches may be tried to solve the problems.

CRediT authorship contribution statement

Mian Zhou: Project planning, Experimental design, Data analysis and Writing the manuscript. **Kunshan Liu:** Project planning, Experimental design, Data analysis and Writing the manuscript, Performing most of the experiments. **Yaqi Ouyang:** Performing some experiments and offering suggestions. **Ru Lin:** Performing some experiments and offering suggestions. **Chenyu Ge:** Performing some experiments and offering suggestions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was sponsored by the National Key Research and Development Program of China [2018YFA0900300] and Shanghai Science and Technology Committee Rising-Star Program [19QA1402600].

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jbiotec.2021.11.001.

References

Ahmad, M., Hirz, M., Pichler, H., Schwab, H., 2014. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Appl. Microbiol. Biotechnol. 98 (12), 5301–5317. https://doi.org/10.1007/s00253-014-5732-5.

Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., Gileadi, O., 2008. Codon optimization can improve expression of human genes in Escherichia coli: a multi-gene study. Protein Expr. Purif. 59 (1), 94–102. https://doi.org/10.1016/j.pep.2008.01.008.

Carpen, J.D., von Schantz, M., Smits, M., Skene, D.J., Archer, S.N., 2006. A silent polymorphism in the PER1 gene associates with extreme diurnal preference in humans. J. Hum. Genet. 51 (12), 1122–1125. https://doi.org/10.1007/s10038-006-0060-y.

Cohen, E., Zafrir, Z., Tuller, T., 2018. A code for transcription elongation speed. RNA Biol. 15 (1), 81–94. https://doi.org/10.1080/15476286.2017.1384118.

Cortazzo, P., Cerveñansky, C., Marín, M., Reiss, C., Ehrlich, R., Deana, A., 2002. Silent mutations affect in vivo protein folding in Escherichia coli. Biochem. Biophys. Res. Commun. 293 (1), 537–541. https://doi.org/10.1016/S0006-291X(02)00226-7.

Fu, J., Murphy, K.A., Zhou, M., Li, Y.H., Lam, V.H., Tabuloc, C.A., Chiu, J.C., Liu, Y., 2016. Codon usage affects the structure and function of the Drosophila circadian clock protein PERIOD. Genes Dev. 30 (15), 1761–1775. https://doi.org/10.1101/gad.281030.116.

Gustafsson, C., Govindarajan, S., Minshull, J., 2004. Codon bias and heterologous protein expression. Trends Biotechnol. 22 (7), 346–353. https://doi.org/10.1016/j.tibtech.2004.04.006.

He, H., Wu, S., Mei, M., Ning, J., Li, C., Ma, L., Zhang, G., Yi, L., 2020. A combinational strategy for effective heterologous production of functional human lysozyme in Pichia pastoris. Front. Bioeng. Biotechnol. 8, 118. https://doi.org/10.3389/fbioe.2020.00118.

Hia, F., Takeuchi, O., 2021. The effects of codon bias and optimality on mRNA and protein regulation. Cell. Mol. Life Sci. 78 (5), 1909–1928. https://doi.org/10.1007/s00018-020-03685-7.Epub 2020 Oct 30.

Hoekema, A., Kastelein, R.A., Vasser, M., de Boer, H.A., 1987. Codon replacement in the PGK1 gene of Saccharomyces cerevisiae: experimental approach to study the role of biased codon usage in gene expression. Mol. Cell. Biol. 7 (8), 2914–2924. https://doi.org/10.1128/mcb.7.8.2914-2924.1987.

Horstick, E.J., Jordan, D.C., Bergeron, S.A., Tabor, K.M., Serpe, M., Feldman, B., Burgess, H.A., 2015. Increased functional protein expression using nucleotide sequence features enriched in highly expressed genes in zebrafish. Nucleic Acids Res. 43 (7), e48 https://doi.org/10.1093/nar/gkv035.

Huang, M., Gao, Y., Zhou, X., Zhang, Y., Cai, M., 2017. Regulating unfolded protein response activator HAC1p for production of thermostable raw-starch hydrolyzing α-amylase in Pichia pastoris. Bioprocess Biosyst. Eng. 40 (3), 341–350. https://doi.org/10.1007/s00449-016-1701-y.

Karaoğlan, M., Erden-Karaoğlan, F., 2020. Effect of codon optimization and promoter choice on recombinant endo-polygalacturonase production in Pichia pastoris. Enzym. Microb. Technol. 139, 109589 https://doi.org/10.1016/j.enzmictec.2020.109589.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., Gottesman, M.M., 2007. A “silent” polymorphism in the MDR1 gene changes substrate specificity. Science 315 (5811), 525–528. https://doi.org/10.1126/science.1135308.

Larrondo, L.F., Loros, J.J., Dunlap, J.C., 2012. High-resolution spatiotemporal analysis of gene expression in real time: in vivo analysis of circadian rhythms in Neurospora crassa using a FREQUENCY-luciferase translational reporter. Fungal Genet. Biol. 49 (9), 681–683. https://doi.org/10.1016/j.fgb.2012.06.001.

Lehmann, J., Libchaber, A., 2008. Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. RNA 14 (7), 1264–1269. https://doi.org/10.1261/rna.1029808.

Lyu, X., Liu, Y., 2020. Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. mBio 11 (5), e02605–e02620. https://doi.org/10.1128/mBio.02605-20.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., Coller, J., 2015. Codon optimality is a major determinant of mRNA stability. Cell 160 (6), 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029.

Sharp, P.M., 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. https://doi.org/10.1093/nar/15.3.1281.

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., Bartel, D.P., 2016. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Rep. 14 (7), 1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043.

Xing, Y., Gong, R., Xu, Y., Liu, K., Zhou, M., 2020. Codon usage bias affects α-amylase mRNA level by altering RNA stability and cytosine methylation patterns in Escherichia coli. Can. J. Microbiol. 66 (9), 521–528. https://doi.org/10.1139/cjm-2019-0624.

Xu, Y., Liu, K., Han, Y., Xing, Y., Zhang, Y., Yang, Q., Zhou, M., 2021. Codon usage bias regulates gene expression and protein conformation in yeast expression system P. pastoris. Microb. Cell Fact. 20 (1), 91. https://doi.org/10.1186/s12934-021-01580-9.

Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., Johnson, C.H., 2013. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature 495 (7439), 116–120. https://doi.org/10.1038/nature11942.

Yang, Q., Yu, C.H., Zhao, F., Dang, Y., Wu, C., Xie, P., Sachs, M.S., Liu, Y., 2019. eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. Nucleic Acids Res. 47 (17), 9243–9258.

Zhang, G., Hubalewska, M., Ignatova, Z., 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. 16 (3), 274–280. https://doi.org/10.1038/nsmb.1554.

Zhao, F., Zhou, Z., Dang, Y., Na, H., Adam, C., Lipzen, A., Ng, V., Grigoriev, I.V., Liu, Y., 2021. Genome-wide role of codon usage on transcription and identification of potential regulators. Proc. Natl. Acad. Sci. USA 118 (6), e2022590118. https://doi.org/10.1073/pnas.2022590118.

Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.H., Fu, J., Chen, S., Liu, Y., 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc. Natl. Acad. Sci. USA 113 (41), E6117–E6125. https://doi.org/10.1073/pnas.1606724113.

Zhou, Z., Dang, Y., Zhou, M., Yuan, H., Liu, Y., 2018. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. Elife 7, e33569. https://doi.org/10.7554/eLife.33569.

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

Zhou, M., Wang, T., Fu, J., Xiao, G., Liu, Y., 2015. Nonoptimal codon usage influences protein structure in intrinsically disordered regions. Mol. Microbiol. 97 (5), 974–987. https://doi.org/10.1111/mmi.13079.