The importance of arginine codons AGA and AGG for the expression in *E. coli* of triosephosphate isomerase from seven different species

Beatriz Aguirre-López, Nallely Cabrera, Marietta Tuena de Gómez-Puyou,
Ruy Perez-Montfort*, Armando Gómez-Puyou

Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Av. Universidad 3000, Coyoacán, 04510 México DF, Mexico

ABSTRACT

Rare arginine codons AGA and AGG affect the heterologous expression of proteins in *Eschericha coli*. The tRNAs necessary for protein synthesis are scarce in *E. coli* strain BL21(DE3) pLysS and plentiful in strain BL21(DE3) CodonPlus −RIL. We evaluated in both bacterial strains the effect of these rare codons on the expression of triosephosphate isomerases from 7 different species, whose sequences had different dispositions of rare arginine codons. The ratio of expressed protein (CP/Bl21) correlated with the number of rare codons. Our study shows that the number, position and particularities of the combination of rare Arg codons in the natural non-optimized sequences of the triosephosphate isomerases influence the synthesis of heterologous proteins in *E. coli* and could have implications in the selection of better sequences for engineering enzymes for novel or manipulated metabolic pathways or for the expression levels of non enzymatic proteins..

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Escherichia coli is a popular organism for producing high quantities of proteins from other species. However, it is accepted that in *E. coli* heterologous protein synthesis [1] and function [2] may be strongly affected when the rare codons (for *E. coli*) in the mRNA are translated. In many instances the relatively low levels of protein expression has been ascribed to pauses in translation that can be related to the presence of codons for which there is a low abundance of cognate tRNAs [3]; these are generally referred to as rare codons. Mechanistically, it is thought that during protein elongation ribosomes wait longer for tRNA that are in low concentrations, than for those in which there is an abundance of tRNA. In addition it has been shown that such pauses increase the probability of mistranslation [4,5] or modifications of folding kinetics [6]. But also, "codon harmonization" has been successfully used to influence translational rate and reduce unfolded or misfolded proteins by pausing transcription rate at selected important codons [3].

A particularly problematic amino acid is Arg; this is because in *E. coli* the tRNAs for two of the six Arg codons, AGA AGG, are in low abundance [7]. The impact of these codons on protein expression has been extensively documented [7], however most of the studies have been carried out with single proteins in which the AGA and AGG codons had been introduced at predetermined positions of a coding sequence. In previous work we reported on the importance of the AGA and AGG codons for heterologously expressed triosephosphate isomerase from *Homo sapiens* (HsTIM), which produced two different proteins because of the substitutions of an Arg for a Lys in one of the monomers of HsTIM [4]. Here we made additional experiments, exploring the rate of expression of seven homologous TIMs from different species with the novelty of naturally having a different number of the rare codons for *E. coli*. Many expression studies using *E. coli* to produce proteins use codon optimization to increase the yield to achieve the best results possible, but there is relatively little information regarding the expression of those proteins using the codon biases that occur naturally in each species. Since our interest was to investigate the variations of rare codons for Arg in the wild type sequences of TIMs occurring in nature, we chose not to use any codon optimization to influence the expression of these proteins in *E. coli*. The genes that encode these proteins differ in the number of Arg residues that form the polypeptide of the enzymes (from 8 to 13); however, in the coding sequences, the enzymes have markedly different

numbers of rare and frequent codons. The latter ranges from zero rare codons for the 9 Arg of TIM from *Trypanosoma brucei* (TbTIM) to that of *Saccharomyces cerevisiae* (ScTIM) in which its 8 Arg are encoded by rare codons (Table 1A). In the seven TIMs that we studied, the Arg with or without rare codons are often in different positions of the amino acid sequence (Table 1B). Thus, the variety of homologous TIMs allowed us to explore if the level of expression of similar enzymes depends on the number of rare codons, or whether the effect of a rare codon depends on a particular position of the coding sequence.

TIM is a homodimer in which the monomers of the enzymes studied are formed by 249–251 amino acids, except TIM from *Giardia lamblia* (GlTIM), which is formed by 258 amino acids. The identities of the amino acid sequences for the seven TIMs vary from 41 to 73% (Table 2). For the purpose of this work, in addition to their similarity, the TIMs chosen have a further advantage. The CCC, CUA and AUA codons that respectively code for Pro, Leu and Ile are also rare for *E. coli* (**7**); however, and although the TIMs studied have between 5 and 8 Pro and a high number of Leu and Ile, in very few cases are these residues coded by CCC, CUA or AUA codons. Thus, in principle, we are practically only studying the impact of the rare Arg codons AGA and AGG on the expression of homologous proteins in a system that has a low abundance of the tRNAs that recognize those codons.

Other factors that we investigated were the existence of rare codons among the first 25 positions in the sequence and also the presence of rare adjacent codons. Both of these factors have been shown to produce pauses during the translation of mRNA [7–10].

Finally, since the sequence of HsTIM has 4 rare codons, of which 2 occur in positions 5 and 18, and positions 99 and 100 have rare adjacent codons for Arg; we used this TIM to change the position of the first codon (and subsequent codons) by attaching a histidine tag (His-tag) to the amino-terminal end. As expected, this shift had the effect of increasing the level of protein expression.

2. Materials and methods

2.1. Cloning and expression of the triosephosphate isomerases

The TIMs from different species used in this work were *Homo sapiens* (HsTIM) with His-tag [4,11], *Rhipicephalus (Boophilus) microplus* (BoTIM) [12], *Trypanosoma brucei* (TbTIM) [13,14], *T. cruzi* (TcTIM) [15], *Giardia lamblia* (GlTIM) [16] and *Saccharomyces cerevisiae* (ScTIM) [17]. The genes for the previous TIMs were transformed in the expression plasmid pET3a and expressed in *E. coli* strain BL21(DE3) pLysS (Novagen). The gene for TIM from *Plasmodium falciparum* (PfTIM) [18] was in the expression plasmid pTrc 99a that uses the same promoter as pET3a. For some experiments we used the gene for HsTIM without His-tag, which was in the plasmid pARSH-3.

As stated above, for some experiments a HsTIM which had the initial sequence: MHHHHHHSSGRENLYFQGH consisting of a hexahistidine tag and a tobacco etch virus protease recognition

sequence, totaling 18 additional amino acids after the initial Met of the wild type HsTIM sequence, was used for some experiments[4]. This enzyme is referred to as His-tag HsTIM.

We measured the expression of the seven TIMs in the BL21 (DE3) pLysS strain and the BL21-CodonPlus (DE3)-RIL strain of *E. coli*. The former cells have low amounts of the tRNAs that recognize the AGA and AGG codons. The CodonPlus (DE3)-RIL strain (CP (DE3)-RIL), on the other hand, contains extra copies of the genes that encode tRNAs that recognize the AGA and AGG codons; it also has tRNAs that recognize the Ile and Leu rare codons [19]. The difference in TIM expression in the two strains (generally higher in the CP (DE3)-RIL strain) was considered to represent the extent of the detrimental effect of the AGA and AGG codons on the synthesis of the TIM assessed. The latter effect would be exclusively due to the rare Arg codons of which none is within the first 25 positions of the sequence and none is adjacent, and not to Ile and Leu rare codons, since the latter are absent in all the TIMs studied.

For some experiments involving the expression of HsTIM, we used *E. coli* strain Rosetta ™ (DE3) pLysS, which is derived from strain BL21(DE3) pLysS to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli* such as, AGA and AGG (Arg), AUA (Ile), CUA (Leu), CCC (Pro), CGA (Gly), CGG and CGA (Arg) and has a ColE1 chloramphenicol resistant plasmid [20].

All three strains were grown in the presence of 0.1 mg/L ampicillin and 0.034 mg/L chloramphenicol.

Experimentally, the *E. coli* BL21(DE3) pLysS strain, the CP (DE3)-RIL strain and the Rosetta ™ (DE3)pLysS strain, transformed with the desired enzyme, were allowed to grow at 37 °C in 200 mL of LB media to an optical density of 0.6; at this time three 50 mL aliquots of each of the three cultures were transferred to Erlenmeyer flasks and induced with 0.4 mM isopropyl β-ᴅ-thiogalactopyranoside (IPTG). For the determination of endogenous *E. coli* TIM we followed the same protocol described above using the expression plasmid pET3a but without a gene for TIM. After an incubation time of 30 min at 37 °C under constant shaking, the cells were collected by centrifugation. The pellet was suspended in 15 mL of 100 mM triethanolamine, 10 mM EDTA and 1 mM dithiothreitol (pH 7.4) and sonicated for 2 min with 30 s intervals at 4 °C. To assess the extent and reproducibility of the amount of cells broken by sonication, 50 μL of suspensions with bacteria were quantified in separate experiments, done in triplicate, using an Attune Acoustic Focusing Cytometer (Applied Biosystems) set to the high sensitivity parameters. A violet laser at 504 nm and an excitation blue laser at 488 nm were used for measuring the forward scatter and the side scatter with the BL1 detector. The results were analyzed with the Attune Cytometric 1.2 (BA) software, and normalized against a preparation of the corresponding intact bacteria. In all cases, and for all *E. coli* strains expressing TIM from different species, the results varied between 96.06–100 percent broken cells with a maximal standard deviation 6.55%. Aliquots of the latter suspension of disrupted cells were used for determination of TIM activity with 1 mM glyceraldehyde 3-phosphate (G3P) as substrate (see below). The latter was used to calculate the amount of expressed

Table 1A
Rare codons for Arg and their distribution sequences of different TIMs.

TIM	Total codons for Arg	Rare codons for Arg (AGA, AGG)	Rare adjacent codons	Rare codons within the first 25 amino acids of the sequence
TbTIM	9	0	0	0
TcTIM	13	2	0	0
BoTIM	12	3	0	1
GlTIM	12	3	0	0
HsTIM	8	4	1	1
PfTIM	8	7	1	1
ScTIM	8	8	1	1

Table 1B
Positions of arginines and their respective codons in different TIMs.

TIM (total Arg/rare Arg*)	Position and sequence of the triplet of the codon for Arg																														
Tb TIM (9/0)								54			98	99				134	138					191				207		220	226		
								cgt			cgc	cgc				cgt	cgc					cgc				cgc		cgc	cga		
Tc TIM (13/2)								**55**		**71**	98	99				135	139		162			192	193		196	208		221	227		
								agg		**agg**	cgg	cgt				cgg	cgc		cgc			cgc	cgt		cgc	cgc		cgc	cgc		
Bo TIM (14/3)	4	**5**		22		52					98	99				134	**138**	145		189						**205**		222			
	cgc	**agg**		cgc		cgg					cgc	cga				cga	**agg**	cgg		cgt						**agg**		cgg			
Gl TIM (12/3)	4	5							**61**		**98**	99	**100**		115	135	138							194			210				256
	cgt	cgc							**agg**		**aga**	cgc	**aga**		cgt	cgc	cgc							cga			cgt				cgt
Hs TIM (8/4)		**5**	18				53					**99**	**100**			**134**					190					205					
		agg	cgg				cgg					**aga**	**agg**			**agg**					cga					cgt					
Pf TIM (8/7)	**3**					**52**					**98**	**99**		110		**134**				**189**						**205**					
	aga					**agg**					**aga**	**aga**		cgt		**aga**				**aga**						**aga**					
Sc TIM (8/8)	**3**				**26**						**98**	**99**						**145**		**189**						**205**				**247**	
	aga				**aga**						**aga**	**aga**						**aga**		**aga**						**aga**				**aga**	

*Rare Arg codons for E. coli are in bold characters.

Table 2
Percent amino acid identity between the sequences of different TIMs*.

TIM	Tb TIM	Tc TIM	Bo TIM	Gl TIM	Hs TIM	Pf TIM	Sc TIM
Tb TIM	100	73.6	49.3	45.6	53.4	43.0	49.8
Tc TIM		100	47.7	46.4	52.4	44.3	46.9
Bo TIM			100	45.7	67.3	41.7	54.6
Gl TIM				100	45.5	41.3	48.8
Hs TIM					100	43.3	53.8
Pf TIM						100	41.7
Sc TIM							100

*Identities were calculated using the Clustal W software at the following url http://www.ebi.ac.uk/Tool/msa/clustalw2. The GenBank accession numbers for the 7 sequences used were Tb TIM: EAN79515.1; Tc TIM: XP_818253; Bo TIM: ABK76308.1 (12 residues were added (VSGIRCYKPERS) after the amino acid in position 248 yielding a sequence of 260 amino acids used in this alignment); Gl TIM: 4BI5_L (2 residues were added (RT) after the last amino acid in position 255 yielding a sequence of 257 amino acids used in this alignment); Hs TIM: CAA49379.1, Pf TIM: AAA18799.1 (4 residues were added (NLLI) after the amino acid in position 238 yielding a sequence of 242 amino acids used in this alignment; Sc TIM: EDV08247.1.

enzyme. The suspensions of disrupted cells were then centrifuged at $5524 \times g$ for 20 min and the TIM activity was also measured in the supernatant of each sample. Since the activity with 1 mM G3P of all the enzymes tested is known, the amount of expressed enzyme in milligrams was calculated. Each experimental condition was studied in triplicate. The data shown are the average of at least three independent experiments performed in triplicate.

2.2. Protein concentration

The protein concentration of each TIM was calculated from the quotient of the total activity of the supernatant divided by the specific activity of each enzyme. For the determination of protein concentration for polyacrylamide gel electrophoresis analysis, we used the Bradford method with bovine serum albumin as a standard.

2.3. Enzyme activity

Enzyme activity was measured in the direction G3P to dihydroxyacetone phosphate. The assay was performed in a volume of 1 mL, which contained 100 mM triethanolamine, 10 mM EDTA, 1 mM G3P, 0.2 mM NADH, and 0.9 units of α-glycerophosphate dehydrogenase (pH 7.4). Activity was calculated from the decrease of NADH absorbance at 340 nm in a Hewlett-Packard spectrophotometer with a multi-cell attachment at 25 °C. The reaction was started by the addition of enzyme, always 5 μL of a 1:200 dilution of the supernatants described above. The specific activities of each enzyme used for the calculation of the protein concentrations were: TbTIM 2858 μmol min^{-1}mg^{-1}, TcTIM 2950 μmol min^{-1}mg^{-1}, BoTIM
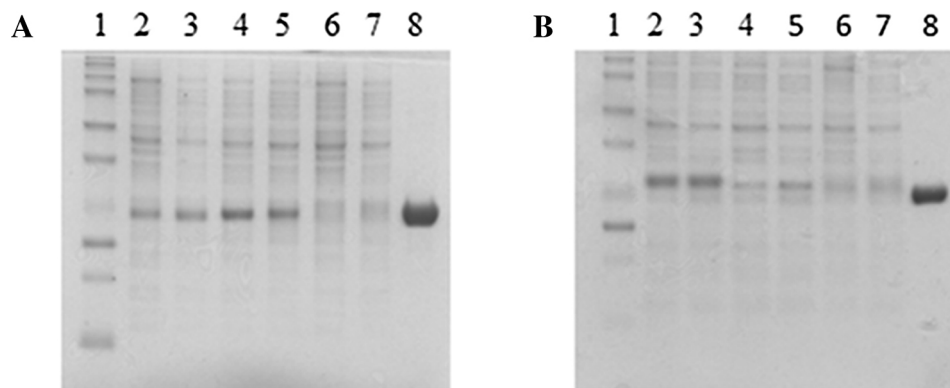
Fig. 1. Native SDS-PAGE analysis of 10 μg of protein of supernatant containing TIM from different species expressed in E. coli strains BL21(DE3) pLysS and CP (DE3)-RIL. In the gel shown in A) lanes 2 and 3 are Tb TIM; 4 and 5 are Tc TIM and 6 and 7 are Bo TIM. In the gel shown in B) lanes 2 and 3 are Gl TIM, 4 and 5 are Pf TIM and 6 and 7 are Sc TIM. TIMs expressed in BL21(DE3) pLysS are in lanes 2, 4 and 6 and TIMs expressed in CP (DE3)-RIL are in lanes 3, 5 and 8. Lanes 1 and 8 correspond in both panels A) and B) to molecular weight standards (Precision Plus Protein Kaleidoscope from Biorad) and to purified TcTIM used as a control, respectively.

3852 μmol min^{-1}mg^{-1}, GlTIM 3800 μmol min^{-1}mg^{-1}, HsTIM 5200 μmol min^{-1}mg^{-1}, PfTIM 6081 μmol min^{-1}mg^{-1} and ScTIM 7265 μmol min^{-1}mg^{-1}, respectively.

2.4. Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis was performed following the method of Schägger and von Jagow [21].

3. Results

Table 2 shows that the identity of the sequences of the seven TIMs chosen for this study ranges from 41 to 73%. As shown in Table 1A, the number of Arg in these seven enzymes ranges from 8 to 13 per monomer. The differences in the number of Arg in the enzymes is not surprising, however, Tables 1A and B, show that there is a wide variety in the codons that code for Arg; for example, TbTIM has 9 Arg, but none of them is coded by AGA or AGG codons, in contrast, AGA codes for the 8 Arg of ScTIM. In the other TIMs, the number of rare Arg codons is between these two extremes.

Thus by using the aforementioned TIMs, we determined if the number of rare Arg codons in the seven TIMs affects their rate of expression. It is noted that the amount of enzyme formed after 30 min of induction has been considered as an index of the rate of TIM synthesis. This consideration was based on SDS-PAGE analysis of 10 μg protein of the supernatant from all expressed TIMs in the different bacterial strains (Fig. 1). Using densitometry and adjusting to 100% the respective control bands of TcTIM in both lanes 8 of panels A and B of Figure 1, the relative density of the bands corresponding to the molecular mass of TIM in the gels were: for panel A lane 2 (TbTIM BL21(DE3)pLysS) 28.3%, lane 3 (TbTIM CP (DE3)-RIL) 26.2%, lane 4 (TcTIM BL21(DE3)pLysS) 43.0%, lane 5 (TcTIM CP(DE3)-RIL) 42.3%, lane 6 (BoTIM BL21(DE3)pLysS) 23.3%, lane 7 (BoTIM CP(DE3)-RIL) 23.8% and for panel B lane 2 (GlTIM BL21(DE3)pLysS) 57.1%, lane 3 (GlTIM CP(DE3)-RIL) 63.7%, lane 4 (PfTIM BL21(DE3)pLysS) 38.0%, lane 5 (PfTIM CP(DE3)-RIL) 39.6%, lane 6 (ScTIM BL21(DE3)pLysS) 37.0% and lane 7 (ScTIM CP(DE3)-RIL) 40.0%, respectively.

Fig. 2 shows the amount of TIM formed by BL21 (DE3) pLysS and CP (DE3)-RIL *E. coli* cells (white and black bars, respectively) after 30 min of induction. In BL21(DE3) pLysS *E. coli*, with the exception of GlTIM, which has a peculiar distribution of rare codons, that is, it

has no rare codons in the first 25 amino acids of its sequence [8,9] and has no adjacent codons (see Table 1A), the amount of enzyme formed ranges from 0.17 to 0.95 mg. However, it is noteworthy that the smallest ratio of enzyme expressed in CP (DE3)-RIL over BL21 (DE3)pLysS cells was obtained with TbTIM in which all of its 8 Arg are coded by frequent codons, and that the highest ratio was that of ScTIM in which its 8 Arg are coded by the AGA codon. Another exception is HsTIM, which has a total of 4 rare codons, with one within the first 25 amino acids of its sequence and one pair of adjacent rare codons, which have been studied before [4]. Thus, it would appear that, in general, the level of expression has a tendency to decrease as the number of rare Arg codons in the coding sequence increases.

The inset in Fig. 2 shows that the expression of the endogenous TIM from *E. coli* does not vary in both the BL21(DE3) pLysS (white bar) and CP (DE3)-RIL (black bar) strains and that it is, at least, one order of magnitude lower than the heterologous TIM with the lowest level of expression: ScTIM.

Fig. 2 shows that when the enzymes were expressed in the *E. coli* CP (DE3)-RIL strain (black bars), the amount of enzyme protein formed was higher than in the *E. coli* BL21(DE3) pLysS strain, the exceptions were TbTIM and, again, GlTIM. Since TbTIM lacks rare codons for its 8 Arg, it would be expected that excess tRNA for AGA and AGG codons would not affect its rate of formation. In regard to the other TIMs, it is also significant that the plot of the ratio of enzyme formed in the CP (DE3)-RIL strain to that synthesized in BL21(DE3) pLysS showed a clear increase as the number of rare codons in the coding sequence becomes higher (Fig. 3).

Trying to optimize the expression of HsTIM in *E. coli* we searched for a strain that would contain all the tRNAs for the rare codons of low abundance in its sequence and found strain Rosetta™ (DE3)pLysS, which could complement the rare codons found in the sequence of HsTIM. We also decided to test the effect of shifting the codon reading framework by adding a sequence, that had a total length of 18 amino acids (His-tag), in and modifying the expression of the enzyme by the absence or presence of the induction agent IPTG.

Fig. 4A shows the relative positions of the rare codons in HsTIM without and with the His-tag. The levels of expression of the wild
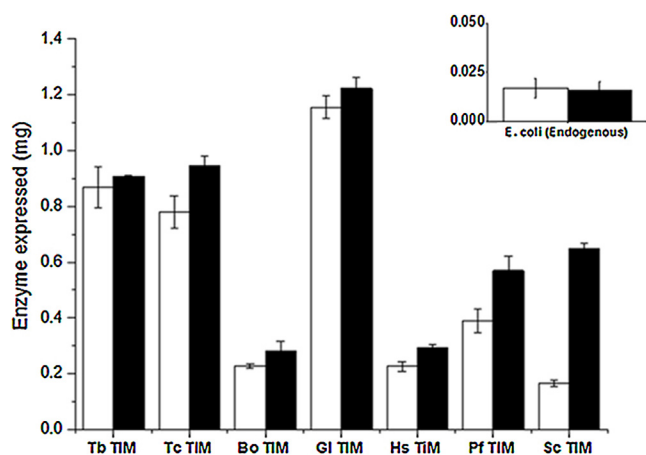

Fig. 2. Amount of native enzyme expressed and quantified in the supernatants for TIMs from different species. White and black bars show the amount of TIM expressed in *E. coli* strains BL21(DE3) pLysS and CP (DE3)-RIL, respectively. The inset shows the amount of endogenous *E. coli* TIM in both strains.
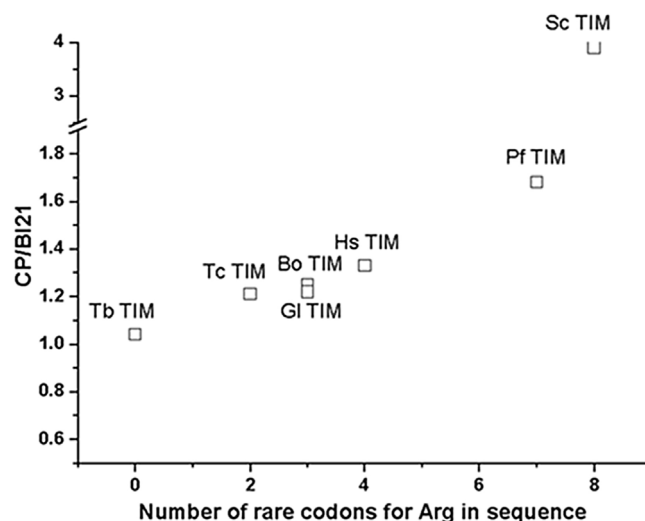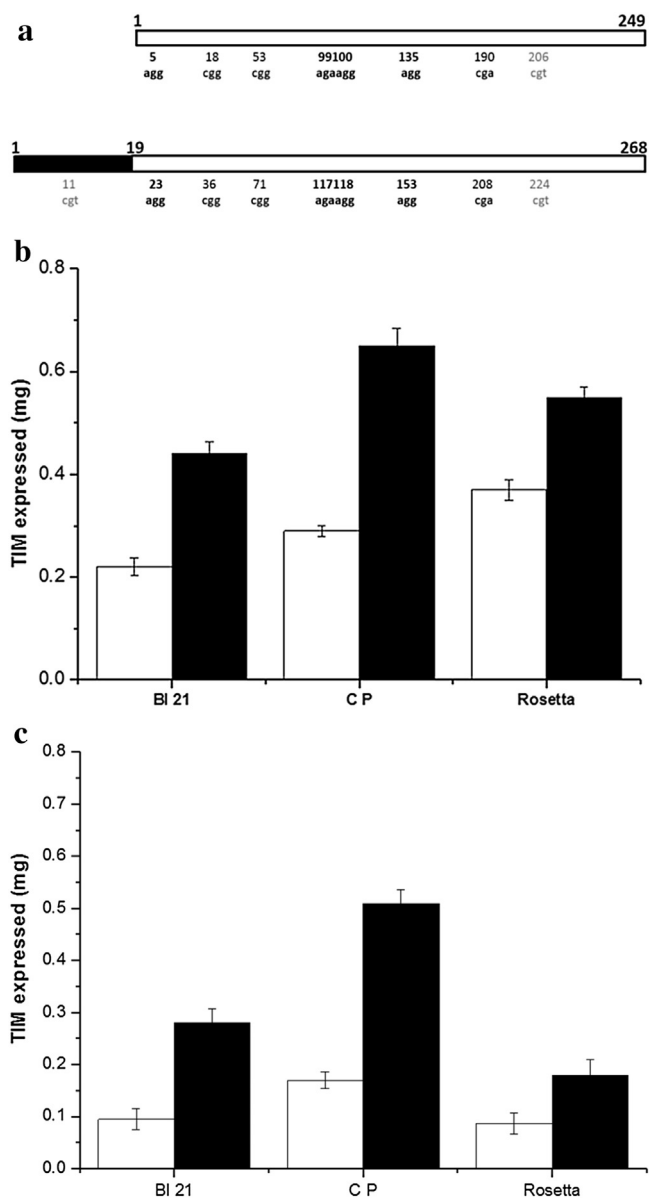

Fig. 3. Plot of the ratio of the amounts of enzyme expressed in *E. coli* strain CP (DE3)-RIL and *E. coli* strain BL21(DE3) pLysS versus the number of rare codons for Arg in the sequence of TIMs for different species. The relative amount of enzyme expressed in *E. coli* strain CP (DE3)-RIL grows as the number of rare codons for Arg in the sequence increases.

a

b

c

Fig. 4. A) Number of the position and relative abundance of the rare codons for Arg (bold black, very rare: AGA and AGG; regular black, rare: CGG and CGA; and light gray, normal abundance: CGT) in the sequence of HsTIM and His-tag HsTIM. B) Comparison of the effect of shifting the position of the rare codons for Arg by adding the sequence of 18 amino acids (His-tag) to Hs TIM. White bars represent the wild type HsTIM expressed after induction with IPTG and black bars represent the His-tag HsTIM expressed after induction with IPTG in three different strains of *E. coli*. C) Expression of wild type HsTIM (white bars) and His-tag HsTIM (black bars) in the absence of induction with IPTG in three different strains of *E. coli*.

type HsTIM in three strains of *E. coli* in the presence of IPTG increased as expected (Fig. 4B): averaging 0.22 mg, 0.28 mg and 0.35 mg for the Bl21(DE3) pLysS, the CP (DE3)-RIL and the Rosetta [TM] (DE3)pLysS strains, respectively. The insertion of the His-tag (black bars) increased the levels of expression, approximately doubling the amount of expressed protein for the Bl21(DE3) pLysS and CP (DE3)-RIL strains, but had less effect with the Rosetta [TM] (DE3)pLysS strain. This result, although surprising and unexpected, could be due to some of the factors responsible for the lower efficiency of expression that has been observed for some proteins with the Rosetta [TM] (DE3)pLysS strain [20].

To find out if induction with IPTG had an effect on the levels of background expression, the amount of expressed protein in the absence of the inducing agent was also determined. Fig. 4C shows that, again, the levels of expression were higher for the HsTIM with the His-tag (black bars), in this case, more than doubling the amount of protein in the case of the BL21 (DE3) pLysS and CP(DE3)-RIL strains, but the Rosetta [TM] (DE3)pLysS strain had a lower level of expression than the Bl21(DE3) pLysS strain (see above, [20]).

4. Discussion

In consonance with many studies on the detrimental effect of Arg codons on protein synthesis in *E. coli* in proteins modified by molecular techniques, our studies with homologous enzymes with marked similarity show that in native sequences there is a clear positive relation between the amount of protein synthesized in the presence of the tRNAs that recognize the AGA and AGG codons and the number of rare codons in the coding sequence. Although the trend is clear, no strict correlation for the effect of rare codons can be determined. This is clearly shown by the data with ScTIM that has 8 rare codons; in this enzyme the ratio of enzyme synthesized in CP (DE3)-RIL to that in BL21(DE3) pLysS was several-fold higher than that observed with HsTIM and PfTIM that have 4 and 7 rare codons, respectively. These observations suggest that the effect of rare codons depends on factors other than their mere frequency in a coding sequence. The latter is also well illustrated by GlTIM and BoTIM. Each of the two enzymes has three rare codons and yet the amount of GlTIM protein formed in the BL21 (DE3) pLysS *E. coli* is nearly four times higher than that BoTIM. Moreover, the synthesis of BoTIM, but not that of GlTIM, is increased when expressed in CP (DE3)-RIL strain.

The latter findings suggest that the position of the rare codons is instrumental to their effect on protein expression. An examination of the position of the rare codons shows that BoTIM, but not GlTIM has an AGG codon at position 5 of the amino acid sequence. Moreover, the four TIMs (BoTIM, HsTIM, PfTIM and ScTIM) whose synthesis is clearly higher in the CP (DE3)-RIL than in the BL21 (DE3) pLysS strains have rare codons in either positions 3, 4, or 5 of the amino acid sequence. This may relevant to TIM expression, since it has been documented that AGA and AGG codons at the beginning of a coding sequence have an adverse effect, and may even interrupt protein synthesis [9].

We do not know if there is any evolutionary significance or effect on protein activity, which is influenced by the frequency and position of Arg codons in the gene sequence, but there is an additional feature in the position of rare codons that apparently affect the synthesis of TIM proteins. These are the Arg at position 98 and 99 (in some TIM, such as TcTIM these correspond to Arg 99 and 100). These two Arg residues are conserved in all TIMs so far studied. As shown in Fig. 2, the TIMs whose expression is significantly higher in the CP (DE3)-RIL than in the BL21(DE3) pLysS strains (HsTIM, PfTIM and ScTIM) have a pair of rare codons in those positions. GlTIM does not have this pair of rare codons, albeit it has them in positions 98 and 100. Since there are reports that indicate that adjacent rare codons are more effective than when they are separated [7,10], it is probable that the two non-adjacent rare codons of GlTIM do not affect the rate of protein synthesis.

In respect to the latter codons, it has been shown that the expression of HsTIM in *E. coli* BL21 (DE3) pLysS yields two proteins: one corresponds to the expected HsTIM, whilst the other corresponds to a protein in which one of the monomers has Lys in position 99 and the other the expected Arg [4]. This mistranslation of Arg for Lys has been extensively documented in other proteins [22–25]. Therefore, it would seem that there are two positions for rare codons that may be critical for the

expression of TIM, those at the beginning and the adjacent codons at positions 99 and 100.

The results of the expression levels of His-tag HsTIM with the Rosetta ™ (DE3)pLysS strain (Fig. 4C) were very surprising, considering it has all the relevant tRNAs to compensate for the rare Arg codons. The amount of HsTIM expressed after induction with IPTG, was slightly higher than in strain CP (DE3)-RIL and for the case of His-tag HsTIM it was somewhat lower. But, in both cases, it was comparable to the expression level of strain CP (DE3)-RIL and, as expected, higher than the levels of strain BL21 (DE3) pLysS. The unexpected result was the low expression levels in the absence of induction with IPTG, in which the levels of expression that are even lower than those of strain BL21 (DE3) pLysS were obtained. As previously observed for a few other human proteins in which the expression levels in strain Rosetta ™ (DE3)pLysS are lower than those in strain BL21 (DE3) pLysS (Tegel et al., 2010), this could possibly be due to the extra metabolic burden of the extra pRARE plasmid in the strain Rosetta ™ (DE3)pLysS. Even though this occurs, it is noted that the relative amount of His-tag HsTIM expressed in this strain was always higher than the amount of HsTIM, pointing to the importance of the rare codon in position 5 in the sequence of wild type HsTIM.

In sum the data of this work show that there are different influences in expression level depending on the number, position and combination of rare Arg codons on the synthesis of heterologous proteins in *E. coli*. An important additional factor is the strain in which one chooses to express the protein.

5. Conclusions

- For the seven TIMs from different species studied in this work, the ratio of protein synthesized in the CP (DE3)-RIL strain of *E. coli* to that synthesized in BL21(DE3) pLysS increases as the number of rare codons in the coding sequence becomes higher (Fig. 3).
- The amount of wild type HsTIM expressed in strains BL21(DE3) pLysS, CP (DE3)-RIL and Rosetta ™ (DE3)pLysS of *E. coli*, upon induction with IPTG, increases by shifting the position of the rare codons for Arg when adding the sequence of 18 amino acids (His-tag) to HsTIM (Fig. 4B).
- The insertion of a sequence consisting of a hexahistidine tag and a tobacco etch virus protease recognition sequence, totaling 18 additional amino after the initial Met of the wild type HsTIM sequence, called a His-tag, increases the levels of expression, practically doubling the amount of expressed protein in the cases of the Bl21(DE3) pLysS and CP (DE3)-RIL strains (Fig. 4C).
- In some cases the Rosetta ™ (DE3)pLysS strain had a lower level of expression than the Bl21(DE3) pLysS strain, as has been observed for some other proteins (Figs. 4 B and 4C).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

M.T.G.-P., A.G.-P., R.P.-M., designed research and experiments; B. A., and N.C. performed research; M.T.G.-P., A.G.-P., R.P.-M., analyzed the data and A.G.-P. and R.P.-M. wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

References

[1] R. Hershberg, D.A. Petrov, Selection on codon bias, Annu. Rev. Genet. 42 (2008) 287.
[2] A.A. Komar, T. Lesnik, C. Reiss, Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation, FEBS Lett. 462 (1999) 387.
[3] E. Angov, C.J. Hillier, R.L. Kincaid, J.A. Lyon, Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host, PLoS One 3 (2008) e2189.
[4] B. Aguirre, M. Costas, N. Cabrera, G. Mendoza-Hernández, D.L. Helseth Jr., P. Fernández, M. Tuena de Gómez-Puyou, R. Perez-Montfort, A. Torres-Larios, A. Gómez Puyou, A ribosomal misincorporation of Lys for Arg in human triosephosphate isomerase expressed in Escherichia coli gives rise to two protein populations, PLoS One 6 (2011) e21035.
[5] E.B. Kramer, P.J. Farabaugh, The frequency of translational misreading errors in E: coli is largely determined by tRNA competition, RNA 13 (2007) 87.
[6] P. Cortazzo, C. Cerveñansky, M. Marín, C. Reiss, R. Ehrlich, A. Deana, Silent mutations affect in vivo folding in Escherichia coli, Biochem. Biophys. Res. Commun. 293 (2002) 537.
[7] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli, Curr. Opin. Biotechnol. 6 (1995) 494.
[8] G.F. Chen, M. Inouye, Suppression of the negative effect of minor arginine codons on gene expression; preferential use of minor codons within the first 25 codons of the *Escherichia coli* genes, Nucleic Acids Res. 18 (1990) 1465.
[9] J.J. Olivares-Trejo, J.G. Bueno-Martínez, G. Guarneros, J. Hernández-Sanchez, The pair of arginine codons AGA AGG close to the initiation codon of the lambda *int* gene inhibits cell growth and protein synthesis by accumulating peptidyl-tRNA[Arg4], Mol. Microbiol. 49 (2003) 1043.
[10] C. Kurland, J. Gallant, Errors of heterologous protein expression, Curr. Opin. Biotechnol. 7 (1996) 489.
[11] C. Rodríguez-Almazán, R. Arreola, D. Rodríguez-Larrea, B. Aguirre-López, M. Tuena de Gómez-Puyou, R. Pérez-Montfort, M. Costas, A. Gómez-Puyou, A. Torres-Larios, Structural basis of triosephosphate isomerase deficiency: mutation E104D is related to alterations of a conserved water network at the dimer interface, J. Biol. Chem. 283 (2008) 23254.
[12] J. Moraes, R. Arreola, N. Cabrera, L. Saramago, D. Freitas, A. Masuda, I. da Silva Vaz Jr., M. Tuena de Gómez-Puyou, R. Perez-Montfort, A. Gómez-Puyou, C. Logullo, Structural and biochemical characterization of a recombinant triosephosphate somerase from Riphicephalus (Boophilus) microplus, Insect Biochem. Mol. Biol. 41 (2011) 400.
[13] T.V. Borchert, K. Pratt, J.P. Zeelen, M. Callens, M.E.M. Noble, F.R. Opperdoes, P.A. Michels, R.K. Wierenga, Overexpression of trypanosomal triosephosphate isomerase in Escherichia coli and characterization of a dimer interface mutant, Eur. J. Biochem. 211 (1993) 703.
[14] G. Garza-Ramos, N. Cabrera, E. Saavedra-Lira, M. Tuena de Gómez-Puyou, P. Ostoa-Saloma, R. Perez-Montfort, A. Gómez-Puyou, Sulfhydryl reagent susceptibility in proteins with high sequence similarity: triosephosphate isomerase from Trypanosoma brucei, Trypanosoma cruzi and Leishmania mexicana, Eur. J. Biochem. 253 (1998) 684.
[15] P. Ostoa-Saloma, G. Garza-Ramos, J. Ramírez, I. Becker, M. Berzunza, A. Landa, A. Gómez-Puyou, M. Tuena de Gómez-Puyou, R. Perez- Montfort, Cloning, expression, purification and characterization of triosephosphate isomerase from Trypanosoma cruzi, Eur. J. Biochem. 244 (1997) 700.
[16] G. López-Velázquez, D. Molina-Ortíz, N. Cabrera, G. Hernández-Alcántara, J. Peon-Peralta, L. Yépez-Mulia, R. Perez-Montfort, H. Reyes-Vivas, An unusual triosephosphate isomerase from the early divergent eukaryote Giardia lamblia, Proteins 55 (2004) 824.
[17] E.B. Nickbarg, J.R. Knowles, Triosephosphate isomerase; energetics of the reaction catalyzed by the yeast enzyme expressed in Escherichia coli, Biochemistry 27 (1988) 5939.
[18] J. Ranie, V.P. Kumar, H. Balaram, Cloning of the triosephosphate isomerase gene of Plasmodium falciparum and expression in Escherichia coli, Mol. Biochem. Parasitol. 61 (1993) 159.
[19] T. Kleber-Janke, W.M. Becker, Use of modified (DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage, Protein Expr. Purif. 19 (2000) 419.
[20] H. Tegel, S. Tourle, J. Ottosson, A. Persson, Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta (DE3), Protein Expr. Purif. 69 (2010) 159.
[21] H. Schägger, G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Anal. Biochem. 166 (1987) 368.
[22] R. Seetharam, R.A. Heeren, E.Y. Wong, S.R. Braford, B.K. Klein, S. Aykent, C.E. Kotts, K.J. Mathis, B.F. Bishop, M.J. Jennings, C.E. Smith, N.R. Siegel, Mistranslation in IGF-1 during over-expression of the protein in Escherichia

coli using a synthetic gene containing low frequency codons, Biochem. Biophys. Res. Commun. 155 (1988) 518.

[23] T.L. Calderone, R.D. Stevens, T.G. Oas, High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli, J. Mol. Biol. 262 (1996) 407.

[24] A.J. Day, R.T. Aplin, A.C. Willis, Overexpression, purification, and refolding of link module from human TSG-6 in Escherichia coli: effect of temperature, media, and mutagenesis on lysine misincorporation at arginine AGA codons, Protein Expr. Purif. 8 (1996) 1.

[25] M.D. Forman, R.F. Stack, P.S. Masters, C.R. Hauer, S.M. Baxter, High level, context dependent misincorporation of lysine for arginine in Saccharomyces cerevisiae a1 homeodomain expressed in Escherichia coli, Protein Sci. 7 (1998) 500.