Biotechnology Reports 13 (2017) 42–48

Contents lists available at ScienceDirect

Biotechnology Reports

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

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

ARTICLE INFO

Article history:
Received 10 November 2016
Received in revised form 15 December 2016
Accepted 5 January 2017
Available online 8 January 2017

Keywords:
Rare codons
Protein expression
Escherichia coli
Triosephosphate isomerase

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..

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

Abbreviations: mRNA, messenger ribonucleic acid; tRNA, transfer ribonucleic acid; TIM, triosephosphate isomerase; IPTG, isopropyl β-D-thiogalactopyranoside.

* Corresponding author.
E-mail address: ruy@ifc.unam.mx (R. Perez-Montfort).

http://dx.doi.org/10.1016/j.btre.2017.01.002

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-D-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