ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 3 (1997) 37–41

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

The evolution of a non universal codon as detected in *Candida rugosa* lipase

Lilia Alberghina *, Marina Lotti

Department of General Physiology and Biochemistry, University of Milano, I-20133 Milan, Italy

Abstract

In several *Candida* species belonging to the same monophyletic group, CUG — one of the six triplets coding leucine in the universal genetic code — is read as an additional codon for serine. CUG for serine is rarely employed but in the lipase genes of *Candida rugosa*, where it is the major serine codon. This yeast secretes multiple lipase isoenzymes (CRLs) tightly related in their amino acid sequence. CRL proteins contain 16 to 19 CUG-serines comprehensive of Ser 209 in the catalytic center and other serines having an obvious structural role. In this paper, results obtained from sequence analysis and mutagenesis are discussed and shown to be consistent with an evolutionary pathway in which the codon was reassigned via ambiguous decoding by a novel seryl-tRNA whose abundance conferred positive selection pressure for the extensive use of CUG as a serine codon.

Keywords: *Candida*; Lipase; Genetic code; Evolution

One of the most exciting recent findings in molecular biology has been the discovery that the genetic code, previously thought to be unchanging in all living beings (universal), is instead flexible to a certain degree in that a few codons may change their meaning during the course of evolution. The first evidence of such codon reassignment was detected in mitochondria, where some universal stop codons code for amino acids (e.g., Met and Trp) [1]. In the following years, similar changes were discovered in nuclear genomes of both prokaryotes and eukaryotes, as reviewed in [2] and summarized in Table 1. It is worth to be mentioned that the reassignment of CUG from leucine to serine observed in a number of *Candida* species, is so far the only known change in an amino acid codon, i.e. non involving a stop codon.

The occurrence of an unusual serine codon in *Candida* yeasts was first suggested by the work of Kawaguchi and colleagues on the cloning of a *C. rugosa* (*C. cylindracea*) lipase mRNA [3]. These authors noticed a discrepancy between the number of serines/leucines predicted on the basis of the gene sequence and the experimental determination of amino acid composition. The ability of cell-free extracts of *C. rugosa* to translate a synthetic poly-CUG to poly-serine suggested a 'non universal' decoding of the triplet CUG. The subsequent isolation of a $tRNA^{Ser}$ with anticodon CAG definitely assigned this codon to serine and ruled out the possibility that codon change evolved via a

* Corresponding author. Tel.: +39-2-70644801; fax: +39-2-70632811; e-mail: l.alberg@imiucca.csi.unimi.it.

PII S1381-1177(96)00051-3

Table 1
Deviations from the universal genetic code in nuclear genomes

Organism	Codon	Assignment	'Universal' assignment
Prokaryotes			
Mycoplasma	UGA	Trp	STOP
Spiroplasma	UGA	Trp	STOP
Eukaryotes			
Candida	CUG	Ser	Leu
Ciliate protozoans	UAR	Gln	STOP

R = A or G; adapted from [2].

mutation in the corresponding aminoacyl-tRNA synthetase [4]. Later on, the same codon change was reported also in other related *Candida* species [5,6]. Phylogeny based on 18S rRNA sequences showed that the genus *Candida* is polyphyletic with a monophyletic group enclosing all the species having CUG as a serine codon [7]. At the present, taking together all data available in literature and in sequence databases, ten different species are included in the group of CUG-serine organisms (Table 2).

Codon changes have been interpreted on the basis of the codon capture theory proposed by Osawa and Jukes in 1989 [8]. In this context, the force driving the evolution of the genetic code is the directional mutation pressure, that is the tendency of a genome to accumulate either C + G (CG pressure) or A + T (AT pressure) nucleotides in processes of neutral mutation. Mutation pressure would cause codon frequency to change and the occasional disappearance from the genome of disfavored codons, with a consequent loss of the corresponding tRNA. Unassigned codons might reappear by mutation and be captured by new tRNAs with mutated anticodons. Accordingly, CUG may have become unassigned under strong selective AT pressure, causing in turn a loss of the corresponding leucyl-$tRNA_{CAG}$. At a later stage, the emergence of a new tRNA, made it possible the reassignment of the codon to serine. The codon capture theory however, implies a stage during evolution where genes containing the unassigned codons at non dispensable sites become non functional (pseudogenes). Moreover, since two mutations are required to change UCN to CUG, also the intermediate UUG and CCG codons (leucine and proline) should be lost. Therefore codon changes are expected to involve rare codons because otherwise the process of reassignment would be extremely harmful, if not lethal, for the cell.

Table 2
Candida species decoding CUG as serine

Species	Ref.
Candida parapsilosis	[5]
Candida zeylanoides	[5]
Candida albicans	[5]
Candida cylindracea JMC1613	[5]
Candida rugosa ATCC 14830	[3]
Candida melibiosica	[5]
Candida maltosa	[6]
Candida guillermondii	$tRNA^{Ser}$CAG in database
Candida tropicalis	$tRNA^{Ser}$CAG in database
Candida viswananthii	[7]

CUG–Ser is very rare in most *Candida* proteins. Thus, for example in over 60 *C. albicans* coding sequences available in databases, CUG accounts for only 2% of the serine codons. Accordingly, a number of genes cloned from CUG–Ser *Candidas* have been successfully expressed in universal host organisms that introduce leucines at CUG sites. In such cases one can assume that CUGs occur at positions of minor structural and functional importance within the protein. This picture would be consistent with the mechanism of reassignment proposed by the codon capture theory, although doubts have been recently raised by Santos and Tuite on the CUG reassignment in *C. albicans* via codon disappearance and two-steps mutations [9].

1. CUG–serines in *C. rugosa* lipase genes

C. rugosa is unique in the group of CUG–Ser *Candida* yeasts for its strong preference in the use of CUG. We have isolated from strain ATCC 14830 five genes coding for lipase where CUG is the predominant serine codon and it is

used for residues with a central functional and structural role [10]. This exceptional bias for CUG–Ser raises issues about both the mechanism of CUG reassignment and the evolution of CRL genes as we discuss in the following.

In the genes coding for the lipase isoenzymes, CUG accounts for about 40% of all serine codons [10]. All six universal serine triplets occur, although UCA and UCU only rarely. Otherwise, only three codons codes leucine since, besides the lack of CUG, even UUA and CUA are never employed, at least in lipase sequences (Table 3). The high frequency use of CUG relates well with the high abundance of the seryl-tRNA$_{CAG}$. The *C. rugosa* genome contains five copies of tRNA$_{CAG}$ genes per haploid genome [11]. In contrast, a single gene encodes seryl-tRNA$_{CAG}$ in the other CUG–serine *Candidas*, where CUG is rare [9]. Thus, it is obvious that the use of the new codon was not disfavored. This emerges not only from the codon usage, but also from the comparative analysis of the sequences of CRLs and related proteins (Fig. 1) Within the CRLs family, 13 CUG–serines are conserved in all five lipase genes sequenced to date [10]. Some of these residues are conserved at equivalent positions in other highly related enzymes, such as the lipases produced by *Geotrichum candidum* (GCLs) and/or several esterases and acetylcholinesterases belonging — as CRLs do — to the structural family of esterases/lipases [1]. This strict requirement of serine at some CUG sites suggests that CUG–Ser evolved from a previous codon for serine and not for any other amino acid. The alternative hypothesis that the lipase activity evolved in *C. rugosa* after CUG reassignment, can be rejected because of the

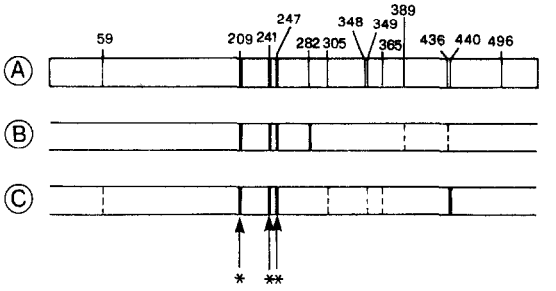

Fig. 1. Evolutionary conservation of serines within the esterases/lipases family. (A) Location of CUG–serine sites conserved in five *C. rugosa* lipase genes. Conservation of serines in *G. candidum* lipases (B) and in a number of esterases/acetylcholinesterases (C). In B and C bold lines represent serines conserved in all proteins of the family, whereas dashed lines mark a partial conservation. Arrows and stars indicate serines occurring in all related proteins.

Table 3
Codon usage in *C.rugosa* genes

Gly	GGG	0.12	Glu	GAG	0.92	Asp	GAU	0.07
	GGA	0.92		GAA	0.08		GAC	0.91
	GGU	0.15						
	GGC	0.71						
Val	GUG	0.8	Ala	GCG	0.35	Ser	AGU	0.06
	GUA	0.02		GCA	0.07		AGC	0.28
	GUU	0.06		GCU	0.13		UGC	0.14
	GUC	0.11		GCC	0.46		UCA	0.01
							UCU	0.02
							UCC	0.09
							CUG	0.39
Lys	AAG	0.96	Asn	AAU	0.04	Met	AUG	1.00
	AAA	0.04		AAC	0.96			
Ile	AUA	0.01	Thr	ACG	0.22	End	UGA	0
	AUU	0.22		ACA	0.02		UAG	0.8
	AUC	0.77		ACU	0.06		UAA	0.2
				ACC	0.69			
Trp	UGG	1.00	Cys	UGU	0.37	Tyr	UAU	0.08
				UGC	0.69		UAC	0.92
Leu	UUG	0.44	Phe	UUU	0.39	Arg	AGG	0.01
	UUA	0		UUC	0.61		AGA	0.04
	CUA	0					CGG	0.30
	CUU	0.14					CGA	0
	CUC	0.42					CGU	0.11
							CGC	0.54
Gln	CAG	0.96	His	CAU	0.09	Pro	CCG	0.49
	CAA	0.04		CAC	0.91		CCA	0.03
							CCU	0.09
							CCC	0.39

Calculated based on the sequences on the five *C.rugosa* genes sequenced to date, all coding for lipase.

[1] Sequences found on ESTHER at http://www.montpellier.inra.fr:70/cholinesterase.

high homology between CRL and the lipases secreted by the universal-code mould *G. candidum* (40% identity in the amino acid sequences).

Considering their occurrence in structurally and functionally related proteins it is reasonable to assume that at least some of the CUG–serines play an important structural/functional role in the CRL proteins. This hypothesis was tested by a mutational analysis on the CRL1 enzyme that contains 19 serines CUG encoded. The *LIP1* gene was overexpressed in the yeast *Saccharomyces cerevisiae*, this process of heterologous expression in universal code cells causing the insertion of leucine at every CUG site. The recombinant protein was inactive, since 19 serines enclosed the catalytic Ser 209 were misdecoded as leucine [12]. CUGs were stepwise replaced with universal codons for serine; after every step the mutated genes were expressed in *S. cerevisiae* and the protein tested for activity. Interestingly, restoring serines 209, 241 and 247 — the residues strictly conserved in the lipase/esterase family — was not sufficient to obtain an active protein, neither was it to reintroduce a group of other six serines selected on the basis of their position in the enzyme three-dimensional structure (unpublished results of our laboratory). Though recombinant proteins were correctly processed and glycosylated by the host cells, they accumulated intracellularly in an inactive form, possibly due to misfolding caused by the replacement of serine with the hydrophobic leucine. Once more, the existence of strict sequence requirements in regions of the protein not obviously involved in the maintainement of the three-dimensional structure, points to the importance of CUG serines in *C. rugosa* proteins.

The information presently available on *C. rugosa* sequences — unfortunately limited to the only lipase encoding genes — is hardly compatible with a change in CUG assignment through subsequent lost-reassignment of this codon, even taking into account the redundancy of lipase genes [10]. Recently, Schlutz and Yarus proposed an alternative mechanism of codon reassignment that better frames our experimental observations on *C. rugosa* genes [13]. This mechanism is based on the finding that mutations at sites outside the anticodon can affect tRNA specificity. Mutated adapters in fact have been shown to read codons usually not allowed by the base-pairing and wobble rules [14]. Since mutated tRNAs can mediate ambiguous decoding (i.e., the same adaptor recognizes both nor-

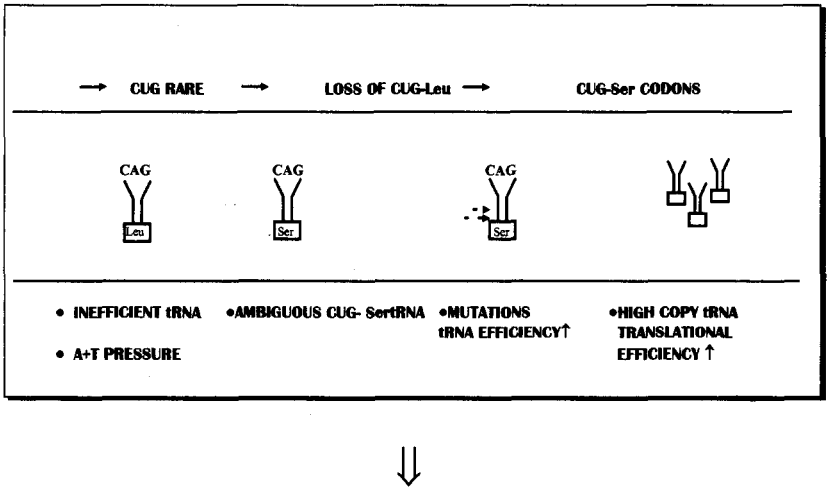

Fig. 2. A model for the reassignment of CUG to serine in *C. rugosa*.

mal and unusual codons), it is reasonable to assume that the same mechanism could catalyze codon reassignment. One could imagine an initial background with CUG codons sparely used because of the scarcity or low translational efficiency of the cognate leucyl-tRNA. The appearance by mutation of a novel tRNA (seryl-$tRNA_{CAG}$) able to translate CUG ambiguously, possibly let the leucyl-tRNA to succumb to competition and disappear. Mutations both in the tRNA and in the UCN serine codons should have gradually improved translational efficiency until reassignment was accomplished (Fig. 2). This scenario is consistent with all available experimental observations: (i) the strong bias in the use of leucine codons by *C. rugosa* shown in Table 2; (ii) the close sequence relatedness of seryl-$tRNA_{CAG}$ with other seryl-tRNAs [11,7] and (iii) the unusual structure of seryl-$tRNA_{CAG}$ and its ability of equivocal decoding in vitro [9]. At this stage one can reasonably assume the occurrence of a process of adaptation of the *Candida* protein sequences towards an optimal use of codons, conferring an improved fitness for the mutated biochemical background. Mutations in the UCN serine codons that further improved their translability by the new abundant tRNA, should have favored their conversion to CUG. Though two sequential mutations have to accumulate to change UCN to CUG, the process of reassignment should be tolerable for cells since it never implies loss of protein function. On the contrary, a positive selection for functional proteins may be considered, exerting a sort of biochemical pressure on the evolution of the genetic code.

In the above framework, the key role in codon reassignment is played by tRNA molecules, whose abundance and efficiency is the major selection tool for codon choice. Experimental support to this hypothesis is provided by the finding that heterologous expression of eukaryotic coding sequences using rare codons is improved if the level of the corresponding tRNA is increased [15]. In this perspective, codon reassignment may be considered a special case of codon bias, i.e. the preferential use of certain codons observed both in prokaryotes and eukaryotes. Codon bias in fact relates very well with the abundance in the cells of specific tRNA molecules [16]. The competition among isoacceptor adapters would be the basis for the enrichment in those codons for which more (or more efficient) tRNAs are available.

References

[1] B.G. Barrell, A.T. Bankier and J. Drouin, Nature 282 (1979) 189–194.
[2] S. Osawa, Evolution of the genetic code (Oxford University Press, Oxford, 1995).
[3] Y. Kawaguchi, H. Honda, J. Taniguchi-Morimura and S. Iwasaki, Nature 341 (1989) 164–166.
[4] T. Yokogawa, T. Suzuki, T. Ueda, M. Mori, T. Ohama, Y. Kuchino, S. Yoshinari, I. Motoki, K. Nishikawa, S. Osawa and K. Watanabe, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 7408–7411.
[5] T. Ohama, T. Suzuki, M. Mori, S. Osawa, T. Ueda, K. Watanabe and T. Nakase, Nucl. Acid Res. 21(17) (1993) 4039–4045.
[6] H. Sugiyama, M. Ohkuma, Y. Masuda, S.-M. Park, A. Ohta and M. Takagi, Yeast 11 (1995) 43–52.
[7] G. Pesole, M. Lotti, L. Alberghina and C. Saccone, Genetics 141 (1995) 903–907.
[8] S. Osawa and T.H. Jukes, J. Mol. Evol. 28 (1989) 271–278.
[9] M.A.S. Santos and M.F. Tuite, Nucl. Acids Res. 23(9) (1995) 1481–1486.
[10] M. Lotti, R. Grandori, F. Fusetti, S. Longhi, S. Brocca, A. Tramontano and L. Alberghina, Gene 124 (1993) 45–55.
[11] T. Suzuki, T. Ueda, T. Yokogawa, K. Nishikawa and K. Watanabe, Nucl. Acids Res. 22(2) (1994) 115–123.
[12] F. Fusetti, S. Brocca, D. Porro and M. Lotti, Biotechnol. Lett. 18(3) (1996) 281–286.
[13] D.W. Schultz and M. Yarus, J. Mol. Biol 235 (1994) 1377–1380.
[14] D.W. Schultz and M. Yarus, J. Mol. Biol. 235 (1994) 1381–1394; 1395–1405.
[15] P.M. Schenk, S. Baumann, R. Mattes and H.-H.- Steinbiss, Biotechniques 19 (1995) 196–200.
[16] P.M. Sharp and G. Matassi, Curr. Opin. Genet. Dev. 4 (1994) 851–860.