Characterization of serine and leucine tRNAs in an asporogenic yeast *Candida cylindracea* and evolutionary implications of genes for tRNA^{Ser}CAG responsible for translation of a non-universal genetic code

Tsutomu Suzuki, Takuya Ueda*, Takashi Yokogawa¹, Kazuya Nishikawa¹ and Kimitsuna Watanabe

Department of Industrial Chemistry, Faculty of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo 113 and ¹Department of Biological Sciences, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 227, Japan

Received December 13, 1993; Accepted December 16, 1993 DDBJ accession nos D12939, D14894–D14900 (incl.)

ABSTRACT

Five serine and three leucine isoacceptor tRNAs were purified from the asporogenic yeast *Candida cylindracea*, in which codon CUG is translated as serine instead of leucine [1], and their primary structures were determined. From the wobble hypothesis [2], it was assumed that one of the tRNA^{Leu} species (Leu1), with the anticodon CmAA, corresponded to the UUG leucine codon, and that the remaining two leucine tRNAs (Leu2 and Leu3), with the same IAG anticodon sequence, would decode the CUU, CUC and CUA codons as leucine, but not the CUG codon; this was clarified by an *in vitro* translation experiment with *C.cylindracea* using synthetic mRNAs containing the CUA or CUG codons. One of the serine tRNAs (Ser1) has already been demonstrated to have the anticodon CAG and to be responsible for translation of the codon CUG in *C.cylindracea* [3]. Three of the other species of tRNA^{Ser} (Ser2,3 and 4), with the anticodon sequences cm⁵UGA, IGA and CGA, can translate all four codons in the UCN codon box, while the remaining species (Ser5), with the anticodon GCU, corresponds to AGU and AGC serine codons. The gene sequences for these five serine and three leucine tRNAs were also determined, with the finding that only tRNA^{Ser}CAG (Ser1) has an intron. At least five different types of tRNA^{Ser}CAG genes exist in the genome of *C.cylindracea*. The nucleotide sequences of the flanking regions of these tRNA^{Ser}CAG genes indicated that the tRNA^{Ser}CAG gene has duplicated at least three times on the genome. The existence of multiple genes for tRNA^{Ser}CAG on the genome may account for the observation that codon CUG is used very frequently in *C.cylindracea*. All of these tRNA^{Ser}CAG genes contain the CCA sequence in their 3' termini,

suggesting the possibility that during their multiplication process in the evolution of the *C.cylindracea* genome, the tRNA^{Ser}CAG molecule was integrated into DNA via reverse transcription.

INTRODUCTION

The universality of the genetic code in living organisms was once widely accepted, but in 1979 Barrell *et al.* first demonstrated the existence of non-universal genetic codes in some organisms through sequence analysis of human and bovine mitochondrial DNAs [4]. Subsequently, in the 1980s genetic code variations were reported not only in non-plant mitochondria but also in nuclear systems (reviewed in [5]). These findings have led to the concept that the genetic code is, in fact, changeable during the evolutionary process of living organisms.

In comparison with organelles, far fewer instances of deviation from the universal genetic code have been revealed in nuclear systems. Most genetic code changes in nuclear systems have been discovered in the codon boxes related to termination codons [5]. However, it has been demonstrated that in the lipase I gene of an asporogenic yeast, *C.cylindracea*, codon CUG is used as serine instead of leucine [1]. We have already revealed the decoding mechanism for the non-universal codon CUG in *C.cylindracea* by determining the primary structure of the serine tRNA and by examining its translation capacity toward the CUG codon in an *in vitro* translation system of *C.cylindracea* [3].

One problem in the *C.cylindracea* genetic code is that since no CUA codon is found in the lipase I gene, it has been unclear whether CUA is used as a leucine or serine codon [1]. To clarify this problem, all the major serine and leucine isoacceptor tRNAs were isolated, their sequences were determined, and the translational capacity of some of these tRNAs was examined in an *in vitro* translation system of *C.cylindracea* using a synthetic

*To whom correspondence should be addressed

mRNA containing the CUA codon. From this experiment, the amino acid assignments for the leucine and serine codons in *C. cylindracea* were largely clarified.

The sequences and organization of their genes were also determined to elucidate how tRNA genes responsible for this non-universal genetic code have evolved.

MATERIALS AND METHODS

Materials

Benzoylated DEAE-cellulose [6] was purchased from Boehringer, Japan. [^{14}C] serine and [^{14}C] leucine were from Amersham, Japan. Oligonucleotide primers used for the polymerase chain reaction (PCR) were synthesized with an automatic DNA synthesizer (Model 391, Applied Biosystems, Japan). N-hydroxysuccinimide-2-naphthoxyacetate was purchased from Aldrich. PCR was basically performed as described in the literature [7] using a Zymoreactor (Model AB 1800, ATTO, Japan). Enzymes for cloning and sequencing were purchased from Takara Shuzo, and Toyobo, Japan. Partially purified yeast seryl- and leucyl-tRNA synthetases were kindly provided by Dr Y Kumazawa of Nagoya University.

Methods

Preparation of seryl- and leucyl-tRNA synthetases from *C. cylindracea*

Forty grams of *C. cylindracea* cells harvested at the log-phase were ground with acid-washed quartz sand. The extract was centrifuged at $100,000\times g$ for 3 hours and the supernatant was applied onto a TSK gel DEAE-Toyopearl 650S column and eluted with a linear gradient (10mM–200mM) of potassium acetate in a buffer containing 20mM Hepes–KOH (pH 7.4), 2mM $\text{Mg}(\text{OAc})_2$, 10mM dithiothreitol, 10mM KOAc and 10% glycerol. The fractions containing seryl- and leucyl-tRNA synthetase activities were separated by this column chromatography, and each fraction was pooled and stored at -20°C . Both enzyme activities were stable for at least one year.

Purification of serine tRNAs by naphthoxyacetylation

Low molecular weight RNA was extracted from cells cultivated in YEPD medium (1% yeast extract, 2% bacto-peptone and 2% glucose) with phenol [8]. The fraction was applied onto a DEAE-cellulose column (Whatman DE 52) and tRNAs were eluted with a buffer consisting of 20mM Tris–HCl (pH 7.5), 10 mM MgCl_2 , and 1M NaCl. 1,650 A₂₆₀ units of tRNA were aminoacylated with serine in 3.5 ml of the reaction mixture consisting of 100mM Hepes–KOH (pH 7.4), 10mM $\text{Mg}(\text{OAc})_2$, 2mM ATP, 10 mM KCl, 10mM dithiothreitol and 2mM L-serine at 30°C for 30 minutes, using seryl-tRNA synthetases partially purified from *C. cylindracea* cells. seryl-tRNA was precipitated with ethanol and dissolved in a 4ml solution of 50mM sodium acetate (pH 4.5), to which 2ml of 1,4-dioxane, 0.8ml n-hydroxysuccinimide-2-naphthoxyacetate (100mg/ml in 1,4-dioxane) and 1.6 ml 1M Tris–HCl (pH 8.0) were added. The naphthoxyacetylation of seryl-tRNA was performed on ice for 7 minutes, and stopped by the addition of 120 ml of acetate and 1.2 ml of 2M NaOAc (pH 4.5). The naphthoxyacetyl-seryl-tRNA was recovered with ethanol precipitation [9].

The precipitate was dissolved in 50mM sodium acetate (pH 4.5) buffer containing 2M NaCl and applied onto a BD-cellulose column equilibrated with 50 mM NaOAc containing 2M NaCl. After the column was washed with a buffer containing 50mM

NaOAc (pH 4.5), 10mM MgCl_2 , and 6% ethanol, naphthoxyacetyl-seryl-tRNA was eluted with same buffer containing 15% ethanol. After naphthoxyacetyl-seryl-tRNA was recovered with ethanol precipitation, it was incubated at 37°C for 2hrs in 1ml of 0.45 M Tris–HCl (pH 8.5) for deacylation of the modified tRNA. The serine acceptance of the tRNA fraction thus recovered was 1,130 pmol/A₂₆₀ unit. The serine tRNA fraction was further purified using RPC-5 column chromatography [10] and polyacrylamide gel electrophoresis as described in the legend of figure 1.

Purification of serine and leucine tRNAs from class II tRNAs

In the lower eukaryotes including yeasts, only leucine and serine tRNAs are of the class II type with a long variable arm [11]. Taking advantage of this, serine and leucine tRNAs were separated from class I tRNAs by 10% polyacrylamide gel electrophoresis containing 7M urea. The serine and leucine tRNA-rich fractions thus obtained were further subjected to RPC-5 column chromatography and gel electrophoresis, as described above.

Preparation of S30 fraction from *C. cylindracea*

Ten grams of *C. cylindracea* cells cultured aerobically at the early log-phase in YEPD medium were ground with acid-washed quartz sand and suspended in 5ml of an extraction buffer (20mM Hepes–KOH (pH 7.6), 2mM $\text{Mg}(\text{OAc})_2$, 100mM KOAc and 10mM dithiothreitol). The extract was centrifuged at $21,000\times g$ for 30 min and the supernatant was re-centrifuged at $31,000\times g$ for 30 min. The low molecular weight components were removed by Sephadex G-25 column chromatography (medium particle size; column size, 1.5×20 cm). The protein fraction was concentrated by a Centriprep10 concentrator (10,000 MW cut-off, Amicon). The S30 fraction thus prepared was divided into 60- μl aliquots (5–10mg of protein per 1ml), rapidly frozen with liquid nitrogen, and stored at -70°C until use.

Preparation of mRNAs for *in vitro* translation system

Preparation of mRNAs containing the in-frame CUA or CUG codons was essentially the same as reported previously [12]. The template DNA fragments were constructed as follows. Two single-stranded DNA fragments consisting of 50 and 54 nucleotides, in which the eleven bases at their 3'-ends were complementary to each other, were chemically synthesized, and after annealing, the remaining single-stranded regions were filled up with the complementary strands using Klenow enzyme. The double-stranded DNA fragment thus obtained contained an appropriate sequence of five bases at the 5'-ends, the *ApaI* site, the 15 repeats of the TTT codon, the nine repeats of the CUA or CUG codon, the *HindIII* site, and an appropriate sequence of 9 bases at the 3'-end in its sense strand. After cutting at the *ApaI* and *HindIII* sites the DNA fragment was inserted into the respective complementary sites in the multi-cloning sites of a plasmid vector, pGEM-7Zf (Promega), downstream of the T7 promoter. The mRNA containing the (UUU)₁₅-(CUA)₉ or (UUU)₁₅-(CUG)₉ sequence was transcribed with T7 RNA polymerase by the run-off transcription on this template plasmid which had been linearized by digestion with *HindIII* and purified by QIAGEN-tips (see Fig. 4a).

In vitro translation assay with synthetic mRNAs

The *in vitro* translation assay was carried out as already reported [12] with a slight modification. The reaction mixture (70 μl)

contained 20mM Hepes-KOH (pH 7.4), 150mM KOAc, 10mM dithiothreitol, 0.5mM ATP, 0.1mM GTP, 25mM creatine phosphate, 2mM glucose-6-phosphate, 4.6 μ g creatine phosphokinase, 0.5mM CaCl₂ and the S30 fraction (equivalent to 100 μ g protein). To remove the endogeneous mRNAs, the reaction mixture was pre-incubated at 20°C for 30min with 8 units of micrococcal nuclease (Pharmacia). Following the stopping of the reaction by the addition of 2mM EGTA, 20 μ M each of non-labelled phenylalanine, [¹⁴C]serine (5.92GBq/mmol) or [¹⁴C]leucine (11.8 Gbq/mmol), 8.5mM Mg(OAc)₂, 20 μ g yeast tRNA^{Phe}(Boeringer) and various amounts (0–100 μ g) of the synthetic mRNA were added to the reaction mixture which was incubated at 30°C for 1hr.

In the reaction started with aminoacyl-tRNA, 10,000–20,000 cpm of [¹⁴C]labelled seryl- or leucyl-tRNA was added to the reaction mixture instead of [¹⁴C]amino acid as described previously [3]. Even if this [¹⁴C]amino acid were released from [¹⁴C]-labeled aminoacyl-tRNA, its concentration would be so low that the amount of aminoacyl-tRNA possibly composed of other endogeneous tRNAs contained in the S30 fraction, if any, would be too small to be detected in the translation assay.

The reaction was stopped by adding 5% TCA and the reaction mixture was boiled for 10min. The TCA-insoluble materials were adsorbed onto the nitrocellulose membrane and washed 3times with 1% TCA. After drying the membrane, the radioactivity retained on the membrane was measured with a liquid scintillation counter.

Cloning and amplification of tRNA genes

Genomic DNA was prepared from *C. cylindracea* cells according to the method reported by Rozijn and Tonino [13] and Rozijn *et al.* [14]. tRNA genes were amplified by PCR using the DNA primers synthesized on the basis of the tRNA sequences determined. The genes of tRNA^{Ser}CAG was cloned from a λ phage (XEMBL3) library of genomic DNA partially digested with *Sau3AI* [15]. Phages carrying the tRNA^{Ser}CAG gene were screened using an oligonucleotide probe complementary to the anticodon region including the intron. The inserts of the candidate phages were subcloned into pUC19 and sequenced [15].

Table 1. Codon-anticodon pairing with tRNA^{Ser} (Ser 1–5) and tRNA^{Leu} (Leu 1–3) in *C. cylindracea*

amino acid	codon	anticodon	tRNA
Leu	UUA	CmAA	Leu1
	UUG		
	CUU	IAG	Leu2
	CUC		
	CUA		
Ser	CUG	CAG	Ser1
	UCU	cm ⁵ UGA	Ser2
	UCC		
	UCA		
	UCG		
	AGU	GCU	Ser5
	AGC		

Sequencing of tRNAs and tRNA genes

Purified tRNAs were sequenced by Donis-Keller's method [16]. For the determination of the modified nucleotides, Kuchino's method was used [17]. After tRNA genes were cloned or amplified by PCR, their DNA sequences were determined by the dideoxy-termination method [18].

RESULTS

Isolation of major serine and leucine tRNAs in *C. cylindracea*

The naphtoxyacetylation method [6] was first applied for isolation of serine isoacceptor tRNAs. Seryl-tRNAs formed by aminoacylation of serine tRNAs from *C. cylindracea* cells with seryl-tRNA synthetases partially purified from *C. cylindracea* were chemically modified with N-hydroxysuccinimide-2-naphtoxyacetate. The naphtoxyacetyl-seryl-tRNAs were separated from unmodified tRNAs with BD-cellulose, as described in Materials and Methods. The serine tRNA-rich fractions were fractionated by RPC-5 column chromatography (Fig. 1a) and further purified by gel-electrophoresis. Finally, five serine tRNAs were recovered to homogeneity, which were used for sequencing and characterization.

To isolate leucine isoacceptor tRNAs, classII tRNAs having a long variable arm were separated first by gel-electrophoresis and then by RPC-5 column chromatography (Fig. 1b). After gel-electrophoresis, as well as three leucine tRNAs, five serine tRNAs were also recovered. The five serine tRNA isoacceptors were identical to those purified by the procedure using naphtoxyacetylation described above.

Serine and leucine tRNA sequences and their characteristics in codon recognition

The purified serine and leucine tRNAs were sequenced by the methods of Donis-Keller [16] and Kuchino *et al.* [17], as shown in Fig.2. The nucleotides at the anticodon first position (position34) of these tRNAs were identified by using two-dimensional thin-layer chromatography (Fig. 3). Ser2 and Ser3 were found to have modified nucleosides, 5-carbamoylmethyluridine (cm⁵U) and inosine (I), respectively, whereas, Ser1, Ser4 and Ser5 had the usual nucleosides, C, C, and G, respectively. All three leucine tRNA isoacceptors had modified nucleosides at position 34; Leu1 had 2'-O-methylcytidine (Cm), and Leu2 and Leu3 had I.

The anticodon sequences indicated that Ser2 (cm⁵UGA), Ser3 (IGA) and Ser4 (CGA) belong to the tRNA group corresponding to UCN codons; Ser2 corresponding to UCA and UCG, Ser3 to UCU, UCC and UCA according to the wobble rule [2], and Ser4 to UCG. Ser5, possessing G at position 34, is a single major isoacceptor tRNA for codons AGC and AGU. Ser1, having the anticodon sequence CAG, is a unique tRNA which can decode the unusual CUG codon as serine, as described previously [3]. These five tRNAs are major serine tRNA species, which could be purified through separation from either class II tRNA fractions or naphtoxyacetyl-seryl-tRNA fractions. Since this set of serine tRNAs is sufficient to translate all serine codons, including the unusual codon CUG in *C. cylindracea* [3], it can be concluded that the primary sequences of all major serine isoacceptor tRNAs in *C. cylindracea* have thus been determined.

Three species of leucine tRNAs were also detected and their primary sequences determined. The leucine tRNA having the anticodon CmAA (Leu1) can decode the UUG codon, a major leucine codon in the lipase I gene of *C. cylindracea*, whereas no

Figure 1. RPC-5 column chromatography of 5 serine isoacceptor tRNAs (a) and class II tRNAs (b). (a) Serine tRNAs were specifically isolated by BD-cellulose column chromatography by naphthoxyacetylation (see 'Materials and Methods') and 10A₂₆₀ units of them were applied onto a RPC-5 column after deacylation. tRNAs were eluted by a linear gradient of NaCl (0.4–1.0M) in 10mM Tris-HCl (pH 7.5), 10mM MgCl₂, 15mM β -mercaptoethanol. Serine isoacceptor tRNAs eluted in 5 main peaks were named Ser1, 2, 3, 4 and 5 according to the elution order. tRNA^{Ser} contained in each peak fraction was further purified to homogeneity by 15% polyacrylamide gel electrophoresis with and without 7M urea [17]. (b) Class II tRNAs having a long variable arm were separated as an intermediately moving band between the bands for class I tRNAs and 5S RNA by 10% polyacrylamide gel electrophoresis (data not shown), and further applied onto a RPC-5 column. tRNAs were eluted by a linear gradient of NaCl (0.4–1.0M) in 10mM Tris-HCl (pH 7.5), 10mM MgCl₂, 15mM β -mercaptoethanol. Leucine isoacceptor tRNAs were named Leu1, 2, and 3 according to the elution order. tRNA^{Ser} and tRNA^{Leu} were further purified to homogeneity by polyacrylamide gel electrophoresis.

leucine tRNA corresponding to the UUA codon was detected in this study. The remaining two leucine tRNAs Leu2 and Leu3, having the same anticodon IAG, should recognize the codons CUU, CUC and CUA according to the wobble rule [2]. However, since there is a report suggesting that inosine in tRNA^{Ser}IGA apparently does not interact with adenosine in *Schizosaccharomyces pombe* [19], the amino acid assignment of the CUA codon needed to be verified experimentally. A tRNA^{Leu} with the anticodon UAG, which should translate the CUA codon, and which is usually present in *S.cerevisiae* [11], was not detected in *C.cylindracea*.

In vitro translation assay for the amino acid assignment of the codon CUA

To clarify whether the CUA codon is translated as leucine or serine in *C.cylindracea*, an *in vitro* translation system was constructed using the S30 fraction prepared from *C.cylindracea* cells and a synthetic mRNA containing nine repeats of the CUA codon, or the CUG codon as a control, in-frame downstream of the 15 repeats of the UUU codon (Fig. 4a). As shown in Fig. 4b, the CUA codon was actually translated as leucine, whereas the CUG codon was translated as serine by the endogenous tRNAs. To confirm that tRNA^{Leu} having the anticodon IAG actually translates the CUA codon, purified Leu2 tRNA (anticodon IAG) charged with [14 C]leucine was examined for the translation of CUA-containing mRNA in the translation system mentioned above. As shown in Fig. 4c, it was clearly demonstrated that Leu2 tRNA actually translated the CUA codon, suggesting that inosine at the first letter of the anticodon IAG can recognize adenosine at the third letter of the codon CUA.

[14 C]seryl-tRNA^{Ser}CAG (Ser1) which was used as a reference, translated the CUG codons in the CUG-containing mRNA, as already reported [3].

Aminoacylation of tRNA^{Ser}CAG

Having clarified how all the leucine and serine codons are translated by the corresponding individual tRNA species in *C.cylindracea*, including the unusual translation of the codon CUG as serine by Ser1 tRNA, we turned our attention to elucidating how such a non-universal codon has emerged in the evolutionary process of *C.cylindracea*.

One possible process is that mutations could have accumulated in seryl- and leucyl-tRNA synthetases, thus causing a change in the aminoacylation specificities of these two enzymes: it is possible to imagine that while a certain amino acid change in the enzymes during evolution would have caused the loss of the charging activity of leucyl-tRNA synthetases toward tRNA^{Leu}CAG, at the same time, seryl-tRNA synthetases would have acquired charging activity towards tRNA^{Ser}CAG as well as other serine tRNA isoacceptors. In this pathway it is not necessary to take into consideration any mutation in the relevant tRNA genes.

To investigate this possibility, the aminoacylation capacity of tRNA^{Ser}CAG by seryl- and leucyl-tRNA synthetases from *S.cerevisiae* and *C.cylindracea* were examined. If the leucyl-tRNA synthetases from *S.cerevisiae* charges tRNA^{Ser}CAG with leucine, this pathway could have occurred. However, as shown in Table 2, none of the leucyl-tRNA synthetases from *S.cerevisiae* and *C.cylindracea* could aminoacylate tRNA^{Ser}CAG with leucine, whereas seryl-tRNA synthetases from both organisms

Figure 2. Nucleotide sequences of 5 serine and 3 leucine tRNAs in *C. cylindracea*. The purified serine and leucine tRNAs were labeled at the 3' or 5' termini and sequenced by Donis-Keller's method [16]. Locations of modified nucleotides were determined by Kuchino's method [17].

could aminoacylate tRNA^{Ser}CAG well. Thus, the hypothesis that mutational changes in aminoacyl-tRNA synthetases resulted in the amino acid assignment changes of codon CUG can be ruled out.

Genes for serine and leucine tRNA isoacceptors

The genes for the serine and leucine tRNAs thus characterized were amplified by PCR from genomic DNA of *C. cylindracea*, using primers corresponding to the 5' and 3' terminal regions of these individual tRNAs, and their sequences were determined. All the tRNA gene sequences except for that of Ser1 tRNA were identical to those of the corresponding tRNAs shown in Fig. 2. Only the Ser1 tRNA gene was found to possess an intron of 15 nucleotide residues (Fig. 5), as already reported [3]; none of the other serine and leucine tRNA genes had an intron.

The tRNA^{Ser}CAG sequence showed fewer similarities to the other serine tRNA isoacceptors, although the other serine tRNAs themselves have higher sequence similarities to each other. These features suggest that the tRNA^{Ser}CAG gene may not have been derived from these major tRNA isoacceptors, or that a number of mutations could have accumulated in the tRNA^{Ser}CAG gene, thus making it difficult to elucidate whether tRNA^{Ser}CAG originates from serine or leucine tRNA.

The existence of several tRNA^{Ser}CAG genes on the *C. cylindracea* genome and their characteristic features

The genes encoding tRNA^{Ser}CAG on the genome of *C. cylindracea* were analyzed by the Southern hybridization method using oligonucleotide probes complementary to the anticodon and the intron regions [15]. After digestion of the whole

Figure 3. Identification of the nucleotides at the first position (position 34) of the anticodon of 5 serine and 3 leucine tRNAs by thin-layer chromatography. The 5'-[³²P]labeled nucleotides at position 34 isolated by Kuchino's method [17] were analyzed by two-dimensional thin-layer chromatography (Avicel SF cellulose, 10×10cm, Funakoshi Pharmaceutical) by two solvent systems as follows [17]: System 1, first dimension: isobutylic acid:NH₄OH:H₂O = 66:1:33; second dimension: isopropanol:HCl:H₂O = 70:15:15; System 2, first dimension: isobutylic acid:NH₄OH:H₂O = 66:1:33; second dimension: 0.1M sodium phosphate(pH6.8) 100ml + ammonium sulfate 60g + n-propanol 2ml. The first letter of the anticodon of Ser2 was identified as 5-carbamoylmethyluridine by referring to the data of Keith *et al.* [24]. The dotted circles show the locations of unmodified nucleotides, pU, pC, pA and pG.

genomic DNA by restriction enzymes, *Pst* I and *Sac* I, at least five DNA fragments were hybridized to the probe complementary to the anticodon and the intron regions, indicating that more than five genes exist on the genome of *C. cylindracea* (Fig. 5).

DNA fragments containing tRNA^{Ser}CAG genes were cloned and their sequences determined. DNA fragments thus obtained could be classified into five types, A, A', B, C and D, according to the similarity in their flanking sequences (Fig. 6a). All the types except A' contain identical 122-nucleotide segments, including the tRNA region, as shown in Fig 6a. In addition, types A and B possess the same sequences in the 3'-flanking region and should thus belong to the same subclass; types C and D can be put in the same subclass as well based on the existence of the same 3'-flanking region. We also obtained a DNA fragment comprising both type A and type C. Contrary to the other fragment types, type A' has completely different 5'- and 3'-flanking sequences.

Another intriguing point revealed by the sequence of the tRNA^{Ser}CAG gene is that the tRNA genes possess the CCA sequence in their 3' termini (Fig. 6a). Although more than 30 tRNA genes in various yeasts have been sequenced [11], no tRNA genes have ever been found that have the CCA sequence. Furthermore, it has been believed that the CCA sequence is absent in all eukaryotic tRNA genes [11]. This unique feature of the tRNA^{Ser}CAG gene suggests that the gene might have been generated by the reverse transcription of a tRNA molecule. We do not know whether the other tRNA genes of *C. cylindracea* sequenced here also possess the CCA sequence because they were sequenced using the PCR products. It is necessary to characterize other tRNA genes to clarify whether CCA-encoding genes are peculiar to the tRNA^{Ser}CAG gene or common in other tRNA genes of *C. cylindracea*.

Table 2. Aminoacylation of tRNA^{Ser}CAG by seryl- and leucyl-tRNA synthetases of *C. cylindracea* and *S. cerevisiae*

	Aminoacyl-tRNA synthetases <i>S. cerevisiae</i>	<i>C. cylindracea</i>
Serine acceptance	609 pmole/A ₂₆₀	1068 pmole/A ₂₆₀
Leucine acceptance	18 pmole/A ₂₆₀	25 pmole/A ₂₆₀

The reaction mixtures containing 100 mM Tris-HCl (pH 7.5), 10mM KCl, 15mM MgCl₂, 10mM dithiothreitol, 2mM ATP, 50mM NaCl, 10μM [¹⁴C]-amino acid (6.33GBq/m mole for serine and 11.8GBq/m mole for Leucine), 0.5μg of purified tRNA^{Ser}CAG and 0.1A₂₈₀unit/ml of partially purified aminoacyl-tRNA synthetase were incubated at 30°C for 15 minutes. The acid-insoluble radioactivities were then measured.

Aminoacyl-tRNA synthetases were partially purified from *C. cylindracea* and *S. cerevisiae*. Low acceptance of tRNA^{Ser}CAG may be due to incomplete purification of the enzymes

DISCUSSION

We have presented the structures of all the serine and leucine tRNAs from *C. cylindracea* which could be purified in sufficient amounts for sequence determination, and related the individual tRNAs to codons for serine and leucine. As described previously and also in this paper, CUG was identified not as a leucine codon but as a serine codon to be translated by tRNA^{Ser}CAG [3]. However, CUA has not been assigned either as a leucine or as a serine codon. Munz *et al.* proposed, based on the result of their genetic analysis, that inosine at the first letter of the anticodon of tRNA^{Ser}IGA cannot recognize adenosine at the third letter of the serine codon UCA in a fission yeast, *Schizosaccharomyces pombe* [19], which is inconsistent with the wobble hypothesis[2].

Figure 4. Assignment of the codons CUA and CUG by an *in vitro* translation system using synthetic mRNAs. (a) Sequences of two mRNAs containing either the CUA or CUG codon. (b) Incorporation of [¹⁴C]leucine (left) or [¹⁴C]serine (right) in an *in vitro* translation system using the S30 fraction of *C. cylindracea* and CUA- (○) or CUG-containing mRNA (●). (c) Incorporation of [¹⁴C]leucine (left) or [¹⁴C]serine (right) started from [¹⁴C]leucyl-tRNA^{Leu}IAG or [¹⁴C]seryl-tRNA^{Ser}CAG in an *in vitro* translation system using the S30 fraction of *C. cylindracea* and CUA- (○) or CUG-containing mRNA (●).

Moreover, in many organisms, including yeast, *E. coli*, archaeobacteria and mammals, a tRNA^{Leu} with a UAG anticodon (or its gene) which would be able to read the CUA leucine codon has been identified [11]. However, no such tRNA^{Leu} was detected in this study.

Using the *in vitro* translation system, we demonstrated that the codon CUA is translated as leucine by a tRNA^{Leu} having the anticodon IAG, which suggests that even in eukaryotes, inosine at the wobble position of tRNAs can base-pair with adenosine at the third position of codons, which means that the eukaryotic genetic system also conforms to the wobble rule [2]. It is possible that the result obtained by Munz *et al.* [19] may have been due to some secondary effect caused by disruption of the tRNA^{Ser}UGA gene *in vivo*.

In the above-mentioned *in vitro* translation system using similar synthetic mRNAs we also confirmed that the codons CUU and CUC are translated as leucine by tRNA^{Leu}IAG (data not shown). Taking both this finding and the wobble rule together,

Figure 5. (a) Nucleotide sequence of tRNA^{Ser}CAG (Ser1) gene. The solid line around the clover-leaf structure of the tRNA^{Ser}CAG gene indicates the region complementary to the oligonucleotide probe for hybridization. (b) Southern hybridization analysis of genomic DNA of *C. cylindracea*. *C. cylindracea* genomic DNA was digested with *Pst*I or *Sac*I, followed by electrophoresis on agarose-gel, and detected by hybridization with the probe. Although the bands for types A and A' appear to have similar lengths at 4.3 kbp, two distinct bands were detected.

we could deduce the codon-anticodon pairings with tRNA^{Ser}s (Ser1–5) and tRNA^{Leu}s (Leu1–3) in *C. cylindracea*, as shown in Table 1. The only ambiguous codon is UUA. Although we have not examined the translational capability of tRNA^{Leu}CmAA towards it, UUA seems to be an absent or unassigned codon [5], or at least very rare, in *C. cylindracea*, judging from the finding that the codon does not appear in the lipase I gene, which is one of the most highly expressed genes in *C. cylindracea*, probably because of the high GC content (63%) of its genome [1]. The tRNA^{Leu} corresponding to the UAA codon may thus be absent or so small in amount that it was not isolated in the present study.

Thus, CUG belongs to a single codon box for serine in *C. cylindracea*, like AUG (methionine) and UGG (tryptophan) in most living organisms. We have already found that the CUG codon is also used as serine in several other *Candida* species, judging from the results with an *in vitro* translation assay using a synthetic mRNA containing CUG codons and the presence of tRNA^{Ser}CAG corresponding to the CUG codon [12].

Since tRNA^{Ser}CAG is aminoacylated with serine by either of the seryl-tRNA synthetases from *C. cylindracea* and *S. cerevisiae*, it is concluded that the codon change of CUG from leucine to serine was not caused by a mutational change that occurred in these synthetase genes. Rather, it is postulated that a mutational change that occurred in a tRNA gene resulted in a change in its anticodon so as to make a base-pair with the CUG codon. It is still unknown whether the original tRNA concerned in this change could have been tRNA^{Ser} or tRNA^{Leu}. In a previous paper [3], we suggested that it would have derived from tRNA^{Ser} for UCN codons. The sequence homology of tRNA^{Ser}CAG is slightly higher with other serine tRNAs than with leucine tRNAs, but this is not sufficient to explain the origin of tRNA^{Ser}CAG, because tRNA^{Ser}CAG has acquired recognition sites for seryl-tRNA synthetase.

A unique feature of tRNA^{Ser}CAG is that G is present at the 5'-adjacent position of the anticodon (position 33), where U33 is present in the usual tRNAs and serves to make a U-turn in

Figure 6. (a) Nucleotide sequences of the genes for tRNA^{Ser}CAG and its flanking regions. Five different DNA fragments containing tRNA^{Ser}CAG genes were cloned into λEMBL3 phages and sequenced after subcloning. Nucleotides identical to the upper sequence are shown by bars. CCA sequences at the 3' termini of tRNA genes are indicated by a black background. **(b)** A possible evolutionary process of gene multiplication of Ser1 tRNA inferred on the basis of the homology of their flanking sequences. The shaded parts indicate that the DNA fragments are actually present in the *C. cylindracea* genome. Type ABCD indicates a putative ancestral gene of types A, B, C, D and type A+C, all of which are found in the genome.

the anticodon loop. G33 of tRNA^{Ser}CAG may play a role in its unusual translation capacity toward the CUG codon or in the prevention of misaminoacylation with leucyl-tRNA synthetase [21].

The structural analysis of serine and leucine tRNA genes revealed that the presence of an intron is unique only in the tRNA^{Ser}CAG gene among the serine and leucine tRNA genes. The sequence similarities between tRNA^{Ser}CAG and other serine tRNAs are lower than the similarities among other serine tRNAs themselves, suggesting that the tRNA^{Ser}CAG gene might be derived from a minor serine tRNA gene with an intron, whose RNA product cannot be identified, probably because its quantity in cells is very small or the gene itself has already been changed to a pseudogene. In contrast to the case of *C. cylindracea*, we have recently isolated an intron-containing tRNA^{Ser} gene from *Candida zeylanoides* having the CGA anticodon, in which CUG is also used as a serine codon (unpublished result), and found

that the tRNA^{Ser}CGA gene has a high sequence homology with that of tRNA^{Ser}CAG from *C. zeylanoides*. Thus, it is possible that tRNA^{Ser}CAG is derived from tRNA^{Ser}CGA corresponding to codon UCN (probably UCG). We have sequenced tRNA^{Ser}CAGs and their genes of 5 *Candida* species including *C. zeylanoides*, closely related to *C. cylindracea*, *C. parapsilosis*, *C. albicans*, *C. rugosa* and *C. melibiosica* [12, 20] and found that these tRNAs have a unique feature of having G33 instead of the conserved U33 in the usual tRNAs, and have rather high sequence homologies to each other (more than 70%), which implies that these genes are derived from a single origin [20, 21].

The evolutionary process of gene multiplication of tRNA^{Ser}CAG in *C. cylindracea* is suggested by the presence of similar flanking regions. It has been reported that the copy number of tRNA genes correlates with the frequency of the codon usage in yeasts [22]. As only the lipase I gene has been sequenced in the genome of *C. cylindracea*, the codon usage of this yeast genome has not been sufficiently clarified. However, considering that the lipase I gene is highly expressed in this yeast [1, 23], its codon usage may reflect those of the abundant proteins in the cells. The CUG codon occurs most frequently among serine codons in the lipase I gene of *C. cylindracea* (19 times out of 45 times for all the serine codons) [1] and, thus it is reasonable to believe that tRNA^{Ser}CAG is one of the major serine isoacceptor tRNAs in *C. cylindracea*, as shown in Fig. 1. The copy number of this gene on the genome may correlate with the abundance of this tRNA.

Judging from the sequence homologies in the DNA fragments containing tRNA^{Ser}CAG genes shown in Fig. 6a, a possible evolutionary process of the gene duplication may be drawn as follows; a putative ancestral gene (single copy, as shown in Fig. 6b) was duplicated to give ancestral genes of type A–D (denoted as type ABCD) and type A'. The tRNA^{Ser}CAG gene of type ABCD was duplicated to produce a DNA fragment in which two tRNA genes were tandemly repeated in an opposite orientation, as shown by type A+C [this type of DNA fragment has actually been detected in the *C. cylindracea* genome. (data not shown)], from which 4 different tRNA^{Ser}CAG gene units (types A–D) would have been generated. During these processes, the tRNA^{Ser}CAG gene with the 3'-flanking region would have been dispersed through the genome. This gene amplification resulted in an abundant amount of tRNA^{Ser}CAG molecules within the cells of *C. cylindracea*, so that the codon CUG would have appeared frequently in the protein genes. Indeed, codon CUG is used most frequently among serine codons, as described above.

It is noteworthy that CCA, the 3'-terminal sequence of tRNA, is present in all the genes of tRNA^{Ser}CAG. This unique feature, which has never been found in other eukaryotic tRNAs [11], suggests the possibility that tRNA^{Ser}CAG could have been generated by the reverse transcription of a tRNA molecule. On the other hand, it cannot explain why only tRNA^{Ser}CAG genes have an intron, which remains to be answered by further study. If our speculation is correct, the next interesting problem is whether there is any relationship between the generation of tRNA genes mediated by RNA and the mechanism of codon change in *C. cylindracea*.

ACKNOWLEDGEMENTS

We are grateful to Drs Syozo Osawa (at present, Biohistory Research Hall, Takatsuki, Japan) and Takashi Ohama (at present,

National Institute of Health, USA.), formerly of Nagoya University, for valuable discussions. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture of Japan and by a JSPS Fellowships for Japanese Junior Scientists (to T.S.).

REFERENCES

1. Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S. (1989) *Nature* 341, 164–166.
2. Crick, F. H. C. (1966) *J. Mol. Biol.* 19, 548–555.
3. Yokogawa, T., Suzuki, T., Ueda, T., Mori, M., Ohama, T., Kuchino, Y., Yoshinari, S., Motoki, I., Nishikawa, K., Osawa, S. and Watanabe, K. (1992) *Proc. Nat. Acad. Sci. U.S.A.* 89, 7408–7411.
4. Barrell, B. G., Bankier, A. T. and Drouin, J. (1979) *Nature* 282, 189–194.
5. Osawa, S., Jukes, T. H., Watanabe K. and Muto, A. (1992) *Microbiol. Rev.* 56, 229–264.
6. Gillam, I. C. and Tener, G. M. (1971) *Methods Enzymol.* 20, 55–70.
7. Saiki, R. K., Gelfand D. H., Stoffel, S., Schorf, S.J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) *Science* 239, 487–491.
8. Holley, R. W. (1963) *Biochem. Biophys. Res. Commun.* 10, 186–188.
9. Andachi, Y., Yamao, F., Muto, A. and Osawa, S. (1989) *J. Mol. Biol.* 209, 37–54.
10. Pearson, R. L. Weiss, J. F. and Kelmers, A. D. (1971) *Biochim. Biophys. Acta* 228, 770–774.
11. Sprinzl, M., Hartmann, T., Weber, J., Blank, J. and Zeidler, R. (1989) *Nucleic Acids Res.* 17, Supplement r1–r172.
12. Ohama, T., Suzuki, T., Mori, M., Osawa, S. Ueda, T., Watanabe, K. and Nakase, T. (1993) *Nucleic Acids Res.*, 21, 4039–4045.
13. Rozijn, T. H. and Tonino, G. J. M. (1964) *Biochim. Biophys. Acta* 91, 105–112.
14. Rozijn, T. H., Tonino, G. J. M., Frens-van der Bilt, E. M. B., Bloemer, H. P. S. and Koningsberger, V. V. (1964) *Biochim. Biophys. Acta* 80, 675–677.
15. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
16. Donis-Keller, H. (1980) *Nucleic Acids Res.* 8, 3133–3142.
17. Kuchino, Y., Hanyu, N. and Nishimura, S. (1987) *Methods Enzymol.* 155, 379–396.
18. Messing, J. (1983) *Methods Enzymol.* 101, 20–78.
19. Munz, P., Leupold, U., Agris, P. and Kohli, J. (1981) *Nature*, 294, 187–188.
20. Suzuki, T., Ueda, T., Ohama, T., Osawa, S. and Watanabe, K. (1993) *Nucleic Acids Res.* 21, 356.
21. Watanabe, K., Ueda, T., Yokogawa, T., Suzuki, T., Nishikawa, K., Mori, M., Ohama, T., Nakabayashi, H., Nakase, T. and Osawa, S. (1993) in *Proceedings of the Translational Apparatus* (Nierhaus, K. H. ed), Plenum London, in press.
22. Ikemura, T. (1982) *J. Mol. Biol.* 158, 572–597.
23. Yamada, K. and Machida, H. (1962) *J. Agr. Chem. Soc. Japan*, 36, 858–860.
24. Keith, G., Desgres, J., Pochart, P., Heyman, T., Kuo, K. C., and Gehrke, C. W. (1990) *Biochim. Biophys. Acta* 1049, 255–260.