© 1995 Oxford University Press

Nucleic Acids Research, 1995, Vol. 23, No. 9 1481–1486

# The CUG codon is decoded *in vivo* as serine and not leucine in *Candida albicans*

**Manuel A.S. Santos and Mick F. Tuite***

Research School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK

Received February 10, 1995; Accepted March 20, 1995

## ABSTRACT

**Previous studies have shown that the yeast *Candida albicans* encodes a unique seryl-tRNA$_{CAG}$ that should decode the leucine codon CUG as serine. However, *in vitro* translation of several different CUG-containing mRNAs in the presence of this unusual seryl-tRNA$_{CAG}$ result in an apparent increase in the molecular weight of the encoded polypeptides as judged by SDS–PAGE even though the molecular weight of serine is lower than that of leucine. A possible explanation for this altered electrophoretic mobility is that the CUG codon is decoded as modified serine *in vitro*. To elucidate the nature of CUG decoding *in vivo*, a reporter system based on the *C.albicans* gene (*RBP1*) encoding rapamycin-binding protein (RBP), coupled to the promoter of the *C.albicans TEF3* gene, was utilized. Sequencing and mass-spectrometry analysis of the recombinant *RBP* expressed in *C.albicans* demonstrated that the CUG codon was decoded exclusively as serine while the related CUU codon was translated as leucine. A database search revealed that 32 out of the 65 *C.albicans* gene sequences available have CUG codons in their open reading frames. The CUG-containing genes do not belong to any particular gene family. Thus the amino acid specified by the CUG codon has been reassigned within the mRNAs of *C.albicans*. We argue here that this unique genetic code change in cellular mRNAs cannot be explained by the 'Codon Reassignment Theory'.**

## INTRODUCTION

The genetic code was established over a quarter a century ago from *in vitro* studies carried out with *Escherichia coli* cell-free translation systems. These studies showed that the mRNA decoding processes were common to viruses, *E.coli* and vertebrates and lead to the suggestion that the genetic code was universal (1). This has been largely confirmed since then, although several exceptions have been reported. For example, in some non-plant mitochondrial genomes, decoding of the UGA, AUA, AAA, AGR and CUN (R is purine, N is any nucleotide) codons does not follow the universal genetic code. Furthermore, the UGA stop codon codes for tryptophan in the eubacteria *Spiroplasma* and *Mycoplasma,* while the UAA and UAG stop codons code for glutamine in the eukaryotic cytoplasmic mRNAs of *Acetabularia*, *Tetrahymena* and *Paramecium.* In *Euplots octocarinatus* the UGA stop codon codes for cysteine (reviewed in reference 2). These exceptions show that different organisms can evolve non-universal genetic codes and question the validity of the 'Frozen Accident Theory' which postulates that 'the code is universal because any change would be lethal, or at least strongly selected against' (1).

To explain the evolutionary mechanisms that allow for alternative genetic codes to evolve, Osawa *et al.* (2) proposed the 'Codon Reassignment Theory' which postulates that under strong selective GC- or AT-biased pressure, certain codons can disappear allowing for subsequent loss of the corresponding tRNAs. At a later stage these unassigned codons reappear by mutation and can be captured by mutant tRNAs from different isoacceptor families thereby introducing genetic code deviations. By implication, the termination codons are the main source of genetic code changes since cells can survive with a single stop signal, leaving two of the three termination codons free for capture by mutant tRNAs. In this case the release factors must lose their affinity for the captured termination codons to allow for efficient decoding. In non-mitochondrial genetic codes this is precisely what happens.

An alternative view of the molecular mechanisms required for the evolution of genetic code deviations was recently proposed by Schultz and Yarus (3). This theory postulates that tRNA structural change is the key element in any genetic code deviation, that is, mutations that allow tRNAs to recognize near-cognate codons will permit a single codon to be assigned to two amino acids at a certain point in time. This first step, involving codon ambiguity, is the critical element in the evolution of alternative genetic codes since it permits reassignment of the codon via tRNA selection without disrupting protein function. The decoding of the UGA stop codon as selenocysteine is a good example of a codon having two different meanings. This genetic code change requires an mRNA 'recoding signal', a unique tRNA and, in *E.coli* at least, a special elongation factor (selB) (reviewed in reference 4), but is restricted to the selenoprotein-encoding mRNAs (e.g. formate dehydrogenase and glutathione peroxidase) and therefore represents a specialized class of codon reassignment. Other known genetic code deviations are not restricted to any specific class of mRNAs.

In eukaryotic cytoplasmic mRNAs the termination codons are usually the only codons involved in genetic code changes. An

* To whom correspondence should be addressed

exception to this rule appears to be the decoding of the leucine-CUG codon as serine in the unicellular fungi *Candida cylindracea* (5) and *Candida maltosa* (6). This genetic code change is not a universal feature of the genetic codes of all *Candida* species, but *in vitro* studies have indicated that it might also occur in the cellular mRNAs of *C.albicans, C.parapsilosis, C.zeylanoides, C.rugosa* and *C.melibiosica* (7). In each of these species, the CUG codon appears to be decoded by a unique $tRNA_{CAG}$ which can be aminoacylated *in vitro* with serine even though it translates the leucine-CUG codon *in vitro* (7,8). However, translation of several natural CUG-encoding mRNAs *in vitro* in the presence of this tRNA purified from *C.albicans, C.parapsilosis, C.guilliermondii* and *C.tropicalis* results in an apparent increase in the molecular weight of the encoded polypeptides (9), even though the molecular weight of serine is lower than that of leucine. This brings into question the validity of the assumption that the CUG codon is decoded simply as serine *in vivo*. To elucidate the nature of the decoding of the CUG codon *in vivo* in *C.albicans,* and thus to identify the nature of the non-standard translational event mediated by this novel seryl-$tRNA_{CAG}$, the *C.albicans* rapamycin-binding protein gene, *RBP1* (10) was modified to contain two CUG codons and expressed in *C.albicans*. By sequencing the engineered RBP we show that the CUG and the CUU codons are decoded as serine and leucine respectively in *C.albicans*. Furthermore, mass-spectrometry analysis of peptides obtained from the recombinant RBP protein confirmed the insertion of serine and not a modified amino acid in response to the CUG codon *in vivo*.

## MATERIALS AND METHODS

### Strains and growth conditions

*E.coli* strain JM109 [*recA1 SupE44 endA1 hsdR17 gyrA96 relA1 thi*$\Delta$(*Lac-proAB*) F′(*traD36 proAB*$^+$ *lacI lacZ* $\Delta$*M15*)] was used as a host for all DNA manipulations. *C.albicans* strain 1006 (*arg57 ser57 lysl ura3 MPA1*; 11) was grown at 30°C in YEPD (2% glucose, 1% yeast extract, 1% peptone) or minimal medium lacking uridine (2% glucose, 0.67% yeast nitrogen base without amino acids, 100 µg/ml of each of the required amino acids; 11).

### Plasmid transformation

Transformation of *E.coli* was carried out by electroporation using a BioRad Gene Pulser according to the manufacturer's instructions. Transformation of the *C.albicans* strain 1006 was carried out using the sphaeroplast method developed by Goshorn *et al.* (11).

### Construction of the *C.albicans RBP1* expression vector

The *C.albicans* vector pCARS/AcZcyc (12) was used to construct an expression vector based on the *C.albicans TEF3* gene promoter (13) and the rapamycin-binding protein gene, *RBP1* (10). For this, the *E.coli lacZ* gene and the *C.albicans* actin promoter were removed from the pCARS/AcZcyc plasmid by digestion with *Bam*HI. The *TEF3* promoter was amplified by PCR from *C.albicans* genomic DNA using the primer pair TEF3-5′ (CCGCAGATCTAAGCACTTCAACCAATCTTG) and TEF3-3′ (GCAACCATGGCCAGAACCACGTGGAACTAAGACCATT-TTCCAAAATGTTGTTGATA). Similarly, the *RBP1* gene was PCR amplified with the primer pair RBP5′ (GCAAC-CATGGTCTGAAGCTGGAAGAACTTCCACAAATTGA) and

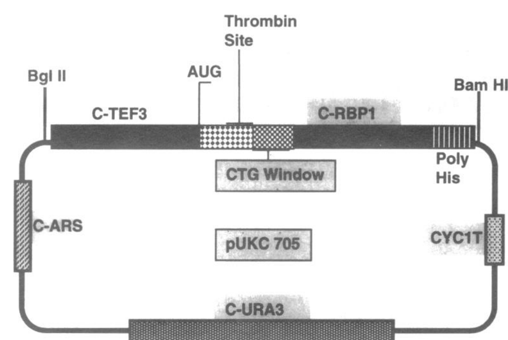


**Figure 1.** Diagram of the vector used to express recombinant rapamycin-binding protein (RBP) in *C.albicans*. C-TEF3, promoter of the *C.albicans TEF3* gene. C-RBP1, recombinant *C.albicans RBP1* gene; PolyHis, tail of six histidines used to facilitate the purification of the protein by nickel affinity chromatography; C-ARS, *C.albicans* autonomous replicating sequence; CYC1T, transcription terminator of the *S.cerevisiae* iso-1-cytochrome-c gene; C-URA3, *C.albicans URA3* gene (12). The sequence that specifies thrombin cleavage is Leu–Val–Pro–Arg–Gly–Ser (14).

RBP3′ (GCCCGGATCCTTAATGGTGATGGTGATGGTGTT-GACCATTAACACCAAGTA). *Bgl*II and *Nco*I restriction sites were introduced at the 5′ and 3′ ends of the *TEF3* gene promoter sequence and *Nco*I and *Bam*HI sites were introduced at the 5′ and 3′ ends of the *RBP1* gene respectively, using tailed oligonucleotides for DNA amplification. Tailed oligonucleotides were also used to introduce CUG codons and a thrombin cleavage site (Leu–Val–Pro–Arg–Gly–Ser; 14) at the N-terminus of the *RBP1* gene and six histidines at its C-terminus (Fig. 1). PCR amplified DNA was cloned into the plasmid pUC18 and sequenced by the dideoxy chain termination method (15) using the Sequenase II kit (US Biochemicals). The *TEF3/RBP1* expression cassette was assembled into the plasmid pPOLYIII (16) as a *Bgl*II–*Bam*HI fragment, prior to cloning into the *C.albicans* vector. The resulting *C.albicans* expression vector (Fig. 1) was designated pUKC705. All DNA amplifications were carried out in a Techne PHC-3 thermocycler using Ultma™ DNA polymerase (Perkin Elmer). The oligonucleotides were purchased from Oswell (UK). Restriction enzymes were purchased from BCL, Promega or Life Sciences and used according to the manufacturer's instructions. All DNA manipulations were carried out essentially as described by Sambrook *et al.* (17).

### Purification of the recombinant RBP expressed in *E.coli* and in *C.albicans*

For overexpression of the engineered *RBP1* gene in *E.coli*, a 0.4 kb *Nco*I–*Bam*HI DNA fragment carrying the *RBP1* gene was cloned into the expression vector pJLA504 (18) under the control of the $\lambda$ $P_L/P_R$ promoters to give plasmid pUKC710. Induction of expression was achieved by a heat shock at 42°C for 4 h. Synthesis of RBP in the *C.albicans* strain 1006 transformed with plasmid pUKC705 was monitored by Western blotting (19,20) using a mouse anti-RBP polyclonal antibody prepared with purified RBP protein over-expressed in *E.coli* as the antigen. The RBP expressed in *E.coli* was purified by nickel-affinity chromatography using the conditions recommended by the 'QIAexpressionist kit' (Qiagen Inc.). The modified RBP expressed in *C.albicans* was purified using the same method, but with the following alterations: 85 g of cells harvested from 8 l culture

grown to a $OD_{600}$ of 1.2 in YEPD were resuspended in 100 ml of 50 mM sodium phosphate pH 7.8, 300 mM sodium chloride, 1% Triton X-100, 1 mM phenylmethyl sulfonyl fluoride (PMSF) and 1 mM benzamidine, and disrupted in a Bead-Beater (Biospec Products, Bartlesville, OK, USA). The lysate was cleared by ultracentifugation at 40 000 *g* for 2 h, then diluted with 2 vol of cell disruption buffer (50 mM sodium phosphate, 300 mM sodium chloride, 1 mM PMSF, 1 mM benzamidine, pH 7.8) and passed through a 8 ml nickel affinity column (Qiagen Inc.). Bound RBP was washed with 50 mM sodium phosphate pH 6.0, 500 mM sodium chloride, 10% v/v glycerol until the $OD_{280}$ of the eluate was 0.008. RBP was then eluted from the matrix in the washing buffer with a gradient of decreasing pH from 6.0 to 4.0.

### Protein sequencing and mass-spectrometry analysis

Purified RBP was sequenced after blotting onto PVDF membrane (21) using an Applied Biosystems 473A Automatic Sequencer (Krebs Institute, University of Sheffield, UK). Cleavage of the RBP N-terminus by thrombin was carried out in a 40 µl reaction of 10 mM Tris–Cl pH 7.0, 1.5 mM $CaCl_2$ containing 3–4.5 µg RBP and 0.1 µg thrombin (Sigma). For V8-protease (Sigma) cleavage, 3 µg RBP was resuspended in 50 mM ammonium bicarbonate pH 7.8 and cleaved overnight at 37°C with a RBP:V8 protease ratio of 20:1 (22). AspN-protease (BCL) cleavage was carried out essentially as described for V8-cleavage but the RBP was resuspended in 50 mM sodium phosphate buffer pH 8.0 (23). Mass spectrometry analysis was carried out using a laser-desorption time of flight mass-spectrometer (VG Tofspec). Partially pure RBP from the nickel affinity chromatography (see above) was purified to homogeneity by reverse phase chromatography using a Pharmacia 'Smart System'. Elution buffers used were, A: 5% acetonitrile in 0.13% TFA (trifluoracetic acid), B: 95% acetonitrile in 0.1% TFA.

## RESULTS

### The CUG codon is translated as serine *in vivo* in *C.albicans*

To elucidate the nature of the decoding of the CUG codon *in vivo* in *C.albicans*, the gene coding for rapamycin-binding protein (*RBP1*) was engineered to contain two CUG codons at positions 12 and 14 of the open reading frame (Figs 1 and 2). In addition, a sequence encoding a thrombin cleavage site was introduced immediately upstream of the CUG codons to ensure that a free N-terminus could be obtained to allow for sequencing the N-terminal region of the RBP protein. To facilitate the purification of the modified RBP by nickel affinity chromatography (24) a sequence encoding six histidine residues was inserted in-frame at the 3′ end of the gene's coding sequence (Fig. 1). The modified *RBP1* gene was expressed in *C.albicans* using the promoter of the *C.albicans TEF3* gene, encoding the translation elongation factor EF3 (13), as an efficient constitutive promoter.

Nickel affinity chromatography proved an effective method for purification of the modified RBP protein from *C.albicans* (1006) transformed with the *RBP1* expression plasmid pUKC705, in a single step (Fig. 3). However, several minor contaminating proteins eluted with the RBP protein indicating that a number of endogenous *C.albicans* proteins are rich in histidines and also bind to the nickel affinity matrix. Following cleavage of the RBP with thrombin the partially pure RBP was further purified by

| | |
|---|---|
| DNA Seq | ATG GTC TTA GTT CCA CGT GGT TCT GGC CAT GGT CTG AAG |
| ddProt-Seq | MET VAL LEU VAL PRO ARG GLY SER GLY HIS GLY LEU LYS |
| dtProt-SeqCa | ..........................................GLY SER GLY HIS GLY SER LYS |
| dtProt-SeqEc | ..............................................................................GLY LEU LYS |
| DNA Seq | CTG GAA GAA CTT CCA CAA ATT GAA ATT GTT CAA GAA GGT |
| ddProt-Seq | LEU GLU GLU LEU PRO GLN ILE GLU ILE VAL GLN GLU GLY |
| dtProt-SeqCa | SER GLU GLU LEU PRO.................................................................. |
| dtProt-SeqEc | LEU GLU GLU LEU PRO GLN ILE GLU ILE VAL GLN GLU GLY |

**Figure 2.** The DNA and corresponding protein sequences of the N-terminus of the RBP expressed in *C.albicans* and in *E.coli*. DNA Seq, DNA sequence; ddProt-Seq, RBP sequence deduced from the DNA sequence; dtProt-SeqCa, sequence of the RBP expressed in *C.albicans* and determined experimentally; dtProt-SeqEc, sequence of the RBP expressed in *E.coli* and determined experimentally. The cleavages sites for thrombin, V8- and ApsN proteases are indicated by the triangles, squares and circles, respectively.

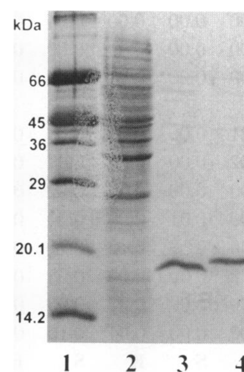


**Figure 3.** SDS–PAGE analysis of the recombinant RBP expressed in *C.albicans* before and after thrombin cleavage. Lane 1, molecular mass markers; lane 2, 30 µg of total *C.albicans* protein extracted from *C.albicans* strain 1006 [pUKC705]; lane 3, 2 µg of RBP partially purified from the protein sample shown in lane 2, by nickel affinity chromatography; lane 4, as for lane 3 but after thrombin cleavage. Note that the thrombin cleavage removes six amino acids from the RBP N-terminus, yet the cleaved protein shows a slower electrophoretic mobility than the uncleaved RBP.

SDS–PAGE and blotted onto PVDF membrane for protein sequencing. Ironically, the thrombin cleaved RBP protein which is six amino acids shorter than the uncleaved form, migrates more slowly on SDS–PAGE than the latter (Fig. 3). This emphasizes the unreliability of the SDS–PAGE gels in determining the correct molecular weight of proteins.

The protein sequencing data for the purified, thrombin-cleaved RBP (Table 1), demonstrated that serine was inserted at the two introduced CUG codons and that leucine was inserted in response to the CUU codon at position 17 (Fig. 2). The protein sequencing data also indicated that no unusual amino acid or non-standard translational event (e.g. frameshifting) occurred during the decoding of either of the two CUG codons. Protein sequencing analysis of the modified RBP without thrombin cleavage demonstrated that the N-terminal methionine residue was cleaved and that the second residue (valine) was not modified thus allowing for sequencing of the RBP N-terminus (data not shown). This suggests that the processing of the N-terminus by methionine aminopeptidase and the subsequent acetylation (or not) of the next residue in *C.albicans*

may follow the same rules as determined for *S.cerevisiae* polypeptides (25). Protein sequencing analysis of the non-cleaved RBP expressed in *E.coli* gave rise to leucine at the two CUG-codons (Fig. 2) confirming that this bacterium does not deviate from the universal meaning of the CUG codon.

**Table 1.** Protein sequencing data for amino acid residues 5–12 of the thrombin-cleaved RBP expressed in *C.albicans*

| | pmol of PTH-amino acid in cycle number: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Amino Acid | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| G | **0.81** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.27 | 0.64 |
| S | 0.00 | **0.19** | 0.00 | **0.62** | 0.00 | 0.17 | 0.14 | 0.00 |
| K | 0.00 | 0.00 | **2.10** | 1.20 | 0.49 | 0.22 | 0.01 | 0.06 |
| E | 0.00 | 0.00 | 0.00 | 0.00 | **1.52** | **2.37** | 1.06 | 0.18 |
| L | 0.12 | 0.00 | 0.00 | 0.17 | 0.00 | 0.07 | **1.88** | 1.18 |
| P | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | **1.32** |
| H | 0.86 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 | 0.11 |
| A | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.02 | 0.09 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.03 | 0.00 | 0.00 |
| N | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.03 | 0.02 | 0.00 |
| D | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 |
| C | | | | | | | | |
| Q | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | 0.02 | 0.00 | 0.00 | 0.07 | 0.08 | 0.00 | 0.00 | 0.03 |
| M | 0.00 | 0.00 | 0.00 | 0.22 | 0.18 | 0.14 | 0.19 | 0.00 |
| F | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 |
| T | 0.00 | 0.00 | 0.12 | 0.11 | 0.00 | 0.00 | 0.41 | 0.12 |
| W | 0.28 | 0.00 | 0.00 | 0.02 | 0.00 | 0.10 | 0.00 | 0.13 |
| Y | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.03 | 0.10 | 0.00 |
| V | 0.00 | 0.00 | 0.04 | 0.10 | 0.02 | 0.00 | 0.00 | 0.00 |
| Deduced Sequence | **G** | **S** | **K** | **S** | **E** | **E** | **L** | **P** |

Cycles 6 and 8 show incorporation of serine in response to the CUG codons. Cycle 11 shows incorporation of leucine in response to the CUU codon (see Fig. 2). The variation in PTH-amino acid concentration between cycles is due to different recoveries of each of the amino acids and to the fact that the efficiency of cleavage at the N-terminus is partially dependent upon the protein sequence (R. Hlodan, personal communication).

A major difficulty encountered during expression of the recombinant RBP in *C.albicans* from plasmid pUKC705 was that the levels of the protein were significantly different between independently obtained *C.albicans* transformants. This might be related to the fact that the pUKC705 vector has no centromere and therefore the copy number of the vector may be significantly different between mother and daughter cells due to unequal partitioning of the plasmid during mitosis (26). Both large and small colonies were found on the primary transformation plates which is consistent with this assumption. A similar result was previously described by Leuker *et al.* (12) when expressing the *Kluyveromyces lactis LAC4* reporter gene in the *C.albicans* strain SGY-243. To detect pUKC705 transformants that expressed high levels of the modified RBP, 10 independent *C.albicans* transformants were screened by Western blotting using a mouse anti-RBP antibody. Of these, two showed levels of expression of the modified RBP similar to the endogenous levels of the wild type RBP while the remaining eight transformants expressed significantly lower levels of the modified RBP (data not shown).

To confirm that serine and not modified serine was inserted during decoding of the CUG codon, the recombinant RBP, partially purified by nickel affinity chromatography, was purified to homogeneity by reverse phase chromatography. The pure RBP was then cleaved with V8- and AspN-proteases which specifically cleave on the carboxyl side of glutamic acid residues at pH 7.8 in ammonium bicarbonate and on the amino side of asparagine, respectively (22,23). The peptides obtained were analyzed by laser desorption time of flight mass-spectrometry (Fig. 2 and Table 2). The experimentally determined molecular mass of fragments V8-F1 and AspN-F1 obtained from V8 and AspN-protease cleavage of the recombinant RBP, which contained the two CUG-encoded serines, deviated by only 2.4 and 11 Da from the predicted mass for those polypeptides respectively (Table 2) thereby confirming that the serine residues inserted by the seryl-$tRNA_{CAG}$ are unlikely to be modified.

**Table 2.** Molecular mass of uncleaved and of specific proteolytic fragments of recombinant *C.albicans* RBP expressed in *E.coli* and *C.albicans*, as determined by laser-desorption time of flight mass-spectrometry

| Polypetides[a] | Molecular mass (Da) | | |
|---|---|---|---|
| | Predicted | Determined | Deviation |
| Coli-RBP | 14 398 | 14 349 | 49 |
| Cand-RBP | 15 117 | 15 082 | 35 |
| V8-F1 | 2117 | 2114.6 | 2.4 |
| ASPN-F1 | 2643 | 2654 | 11 |

[a]Coli-RBP, *C.albicans* RBP1 expressed in *E.coli*; CAN-RBP, RBP1 expressed in *C.albicans* V8-F1 and Asp N-F1, peptides produced by proteolytic cleavage of the RBP expressed in *C.albicans* (see Fig. 2). The mass of the different polypeptides was predicted from the respective DNA sequences using the UWGCG software package. Note: The N-terminus of the RBP expressed in *E.coli* is shorter by nine amino acids than that of the RBP expressed in *C.albicans*. This is due to the cloning procedure used.

## The use of the CUG codon in *C.albicans* genes

The *C.albicans* seryl-$tRNA_{CAG}$ and analogous tRNAs isolated from a number of other *Candida* species all have guanosine at position 33 within the anticodon loop which is unique among elongator tRNAs; U is usually found at this position (27). It is not yet clear what the functional significance of the G33 nucleotide is (if any). Since G is a much larger residue than U one would expect that the accommodation of this residue in the anticodon-loop might induce an alteration in the stacking of the anticodon bases and therefore decoding of the CUG codon would be inefficient or perhaps restricted to certain codon-contexts present in a specific class of genes.

We have examined the distribution of leucine codons in the open reading frames of 65 genes of *C.albicans* (35 212 codons). This analysis shows that the UUA/G are the most frequently used codons while the CUN codon family is rarely used (Table 3; see also reference 28 for an earlier summary). Among the CUN codon family, the CUU codon is the most frequently used followed by the CUA codon and then the CUG and CUC codons. The low frequency use of the CUG codon relates well with the low abundance of the seryl-$tRNA_{CAG}$ (8) and to the copy number of the seryl-$tRNA_{CAG}$ gene in *C.albicans*; which is present at one copy per haploid genome (8; V. Perreau, M. A. S. Santos and M. F. Tuite, unpublished data). This low frequency use of the CUG codon in *C.albicans* contrasts sharply with the use of the CUG codon in *C.cylindracea* where the CUG codon appears to be the most frequently used CUN codon (5). An analysis of the distribution of the CUG codon in the 65 *C.albicans* mRNAs

revealed that 32 (49.2%; Table 3) contained CUG codons in their deduced open reading frames and that the distribution of these codons was not restricted to any specific gene family. However, consistent with its designation as a rare codon, the CUG codon appears to be preferentially found in genes that are expressed at low level in *C.albicans* (28).

**Table 3.** Usage of CUN codons in 65 *C.albicans* gene sequences

| | Codons | | | |
|---|---|---|---|---|
| | CUG | CUA | CUC | CUU |
| No. of codons | 62 | 73 | 58 | 312 |
| Frequency (per 1000) | 1.7 | 2.07 | 1.65 | 8.68 |
| % of genes | 49.2 | 32.3 | 49.2 | 92.3 |

Sixty-five *C.albicans* gene sequences containing 35 312 codons were analysed. The *C.albicans* gene sequences were accessed via the World Wide Web at the locus Http://alces.med.umn.edu/candida/gb-candida. The frequencies of each of the CUN codons was determined manually.

## DISCUSSION

### CUG is decoded as serine and not modified serine in *C.albicans*

*In vitro* translation studies have indicated that a variety of polypeptides synthesized in the presence of the *C.albicans* seryl-tRNA$_{CAG}$ have a reduced electrophoretic mobility in denaturing SDS–PAGE gels (8,29). Since serine has a smaller molecular weight than leucine (87.03 and 113.08 respectively) the simple replacement of leucines by serines encoded by the CUG codons cannot be used to explain the slower electrophoretic mobility of the polypeptides. Therefore, one might speculate that a modified serine or some other amino acid is inserted in response to CUG codons in *C.albicans*. Alternatively, a non-standard translational event such as frameshifting could result in an extension in the length of the encoded polypeptides, although in the cases of the mRNAs whose translation we have examined, (rabbit α- and β-globins, Brome Mosaic Virus RNA-4 and *S.cerevisiae HSP26* mRNA) the ribosomes would soon encounter stop codons in the +1 and +2 frames.

The protein sequencing data we present here conclusively shows that the CUG codon is decoded *in vivo* in *C.albicans* as serine in contravention of the 'Universal' genetic code. This suggests that the leucine for serine replacement is responsible for the altered electrophoretic mobility of the polypeptides synthesized *in vitro* (8,29). However, one cannot rule out the possibility that some of the serines inserted in response to the CUG codons are modified post-translationally or that a non-standard translational event occurs *in vitro*. Assuming, however, that the *C.albicans* seryl-tRNA$_{CAG}$ is solely responsible for CUG-decoding *in vivo,* the insertion of a modified amino acid can be ruled out. Furthermore, given that the most likely form of modification would be phosphorylation, our finding that treating the *in vitro* translation products with alkaline phosphatase does not change their electrophoretic mobility (M. A. S. Santos and M. F. Tuite, unpublished data), further argues against CUG being used as a signal for the insertion of a modified amino acid in *C.albicans*.

That simple amino acid substitutions alter the electrophoretic mobility of proteins on SDS–PAGE is not a new phenomenon. For example, single amino acid substitutions cause aberrant SDS–PAGE migration of mutant forms of the human H-*ras* protein (30) and the *Salmonella typhimurium* histidine-transport protein (31). That the altered electrophoretic mobility of the rabbit α- and β-globins synthesized in presence of the *C.albicans* seryl-tRNA$_{CAG}$ is independent of an altered protein structure, has been shown previously by determining their electrophoretic mobility on urea–tricine gels (8). Therefore, if the altered electrophoretic mobility of the proteins synthesised *in vitro* in presence of the seryl-tRNA$_{CAG}$ is only due to the altered decoding of the CUG codon, the replacement of the hydrophobic leucines with polar serines might alter the SDS-binding capacity of the mutant proteins which would result in a slow electrophoretic migration.

### Evolution of the reassignment of the CUG codon in *Candida* species

The pattern of genetic code deviations described to date indicates that in mitochondrial mRNAs both sense and nonsense codons can be 'captured' by mutant tRNAs from different isoacceptor families (2). In eubacteria the termination codons are also available for capture, but some sense codons are unassigned. For example, in *Mycoplasma capricolum* (2,32), an organism whose genome has a very low GC content (25%), the arginine-CGG codon does not code for any amino acid while in *Micrococcus luteus,* which has a genome with 74% GC, some codons ending with A, namely Arg-AGA, Ile-AUA and possibly Leu-UUA, Leu-CUA, Val-GUA and Gln-CAA codons, are not found in the deduced coding sequences (33). The unassignment of these codons can be explained by the 'Codon Reassignment Theory' (2).

In eukaryotic cytoplasmic mRNAs, termination codons are the main protagonists of genetic code deviations (2) with the only exception being the decoding of the leucine CUG codon as serine in *C.cylindracea*, *C.albicans*, *C.maltosa* and possibly a number of other *Candida* species (5–9). This major deviation to the pattern of genetic code changes found in eukaryotic cytoplasmic mRNAs cannot be explained by the 'Codon Reassignment Theory' for two main reasons. First, the sequencing of the *S.cerevisiae* genome has shown that in a complex eukaryotic genome the relative GC/AT content is not evenly distributed within single chromosomes (34). Therefore it is very unlikely that codons disappear from the entire set of mRNAs due to GC or AT pressure. Secondly, if one assumes that the *C.albicans* seryl-tRNA$_{CAG}$ arose via insertion of a C in the intron of a seryl-tRNA$_{IGA}$ gene as previously suggested (2,35), two mutations are required to mutate a serine codon (UCN) to the CUG codon. Therefore, an intermediary step is required in which the mutant anticodon decodes either the most frequently used leucine codon UUG or the proline-CCG codon. Since ambiguous decoding is not allowed by the 'Codon Reassignment Theory' (2), these other codons in addition to CUG should also disappear from the mRNAs of *C.albicans*; an evolutionary event which would appear to be highly unlikely.

The evolution of serine-CUG decoding in *Candida* can be better explained by a model that does not preclude ambiguous decoding, such as the model recently proposed by Schultz and Yarus (3). Such a model could also explain the presence of G at position 33 in the seryl-tRNA$_{CAG}$ sequence since studies carried out with A/G33 mutant tRNAs in *E.coli* have shown that the mutant tRNAs have a higher dissociation constant from the ribosome than the wild type tRNA, i.e., they decode their cognate codons with lower efficiency (36–38).

In the early stages of the evolution of the genetic code change in *C.albicans*, inefficient decoding would have been required to counteract the deleterious effect of CUG recoding. At later stages, tRNAs carrying a U33 to G33 mutation could have evolved for efficient decoding by means of compensatory mutations in the anticodon-stem or other parts of the tRNA as predicted by the 'Extended Anticodon Theory' which states that there is a correlation between the nucleotides on the 3′ side of the anticodon (nucleotide 36 in particular) and the nucleotide sequence in the anticodon stem, that is, mutations in the anticodon-loop, that lower the efficiency of tRNA decoding, can be compensated by mutations in the anticodon stem (3[illegible] Ambiguous decoding of the CUG codon as leucine and serine [illegible]s not exclude the possibility that the seryl-tRNA$_{CAG}$ evolved f[illegible] a leucyl-tRNA$_{CAG}$. Intriguingly, in *C.cylindracea* the seryl-tR[illegible]A$_{CAG}$ is the only seryl-tRNA that has a m$^2$G at position 10 in the D-stem and this modified nucleoside is also present in the leucyl-tRNA$_{CmAA}$ of *C.cylindracea* (40). This clearly shows that the 2G methylase recognizes a structural element present in both the seryl-tRNA$_{CAG}$ and the leucyl-tRNA$_{CmAA}$. Furthermore, the seryl-tRNA$_{CAG}$ is the only *C.cylindracea* tRNA that has U in the discriminator position (N73) at the top of the acceptor stem. All *C.cylindracea*, seryl-tRNAs have G at position 73 and this nucleotide has been shown to be important for aminoacylation of mammalian seryl-tRNAs (41). If the seryl-tRNA$_{CAG}$ evolved from a leucyl-tRNA$_{CAG}$ one is left with the question of knowing what physiological advantage could the Leu to Ser identity switch have brought about in these yeasts that allowed for the selection of the mutant tRNA. Expression of the *C.albicans* seryl-tRNA$_{CAG}$ in *S.cerevisiae* using a single copy vector is compatible with viability, but significantly reduces the growth rate (M. A. S. Santos, and M. F. Tuite, unpublished data) which suggests that the seryl-tRNA$_{CAG}$ would have been eliminated by negative selection unless it created a new selective advantage for these yeasts.

The data we present here on the decoding of the CUG codon in *C.albicans*, together with the *in vivo* data previously obtained for *C.cylindracea* (5) and *C.maltosa* (6) are clearly insufficient to provide a convincing model to account for the evolution of this unique codon reassignment. Nevertheless, this will become more feasible once more sequences of seryl-tRNA$_{CAG}$s and their genes from other *Candida* species become available.

## ACKNOWLEDGEMENTS

The *C.albicans* vector pCARS/AcZcyc was a kind gift of J. Ernst (University of Dusseldorf). The *C.albicans* strain 1006 was a kind gift of S. Scherer (University of Minnesota). We are indebted to John Atkins, Ray Gestland and Norma Wills (Howard Hughes Medical Institute, Salt Lake City) for helpful comments, support and guidance during the course of this work. We also thank Arthur Moir (University of Sheffield) for protein sequencing and Roman Hlodan (University of Kent) for mass spectrometry analysis. The mass spectrometry work was carried out in the Wellcome Trust-supported Protein Analytical Facility at the University of Kent. This work was supported by a project grant (Ref: M/91/2902) from the Wellcome Trust.

## REFERENCES

1 Crick,F.H.C. (1968) *J. Mol. Biol.*, **38**, 367–379.
2 Osawa,S., Jukes,T.H., Watanabe,K. and Muto,A. (1992) *Microbiol. Rev.*, **5**, 229–264.
3 Schultz,D.W. and Yarus,M. (1994) *J. Mol. Biol.*, **235**, 1377–1380.
4 Böck,A., Forchhammer,K., Heider,J. and Baron,C. (1991) *Mol. Microbiol.*, **5**, 515–520.
5 Kawaguchi,Y., Honda,H., Morimura-Taniguchi,J. and Iwasaki,S. (1989) *Nature*, **341**, 164–166.
6 Sugiyama,H., Ohkuma,M., Masuda,Y., Park,S.M., Ohta,A. and Takagi,M (1995) *Yeast*, **11**, 43–52.
7 Ohama,T., Susuki,T., Mori,M., Osawa,S., Ueda,T., Watanabe,K. and Nakase,T. (1993) *Nucleic Acids Res.*, **21**, 4039–4045.
8 .Santos,M.A.S., Keith,G. and Tuite,M.F. (1993) *EMBO J.*, **12**, 607–616.
9 Santos,M.A.S., Colthurst,D.R., Wills,N., McLaughlin,C.S. and Tuite,M.F. (1990) *Curr. Genetics*, **17**, 487–491.
10 Ferrara,A., Cafferkey,R. and Livi,G.P. (1992) *Gene*, **113**, 125–127
11 Goshorn,A.K., Grindle,S.M. and Scherer,S. (1992) *Inf. Immuni*, **60**, 876–884.
12 Leuker,C.E., Hahn,A.M. and Ernst,J. (1992) *Mol. Gen. Genet.*, **235**, 235–241.
13 Domenico,B.J.D., Lupisella,J., Sandbaken,M. and Chakraburtty,K. (1992) *Yeast*, **8**, 337–352.
14 Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
15 Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5464.
16 Lathe,R., Vilotte,J.L. and Clark,A.J. (1987) *Gene*, **57**, 193–201.
17 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
18 Schauder,B., Blöcker,H., Frank,R. and McCarthy,J.E.G. (1987) *Gene*, **52**, 279–283.
19 Towbin,H., Stahelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354.
20 Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21 LeGendre,N. and Matsudaira,P. (1989) In Matsudaira, P.T. (Ed.), *A Pratical Guide to Protein and Peptide Purification for Microsequencing*. Academic Press, Inc. London. pp 49–70.
22 Houmard,J. and Drapear,G.R. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 3506–3512.
23 Drapeau,G.R. (1980) *J. Biol. Chem.*, **255**, 839–840.
24 Hochuli,E., Döbeli,H. and Schacher,A. (1987) *J. Chromatography*, **411**, 177–184.
25 Arfin,S.M. and Bradshaw,R.A. (1988) *Biochemistry*, **27**, 7979–7984.
26 Broach,J.R. and Volkert,F.C. (1991). In Broach,J.R., Pringle,J.R. and Jones,E.W (Eds), *The Molecular Biology of the Yeast Saccharomyces*. Genome Dynamics, Protein Synthesis, and Energetics. Cold Spring Harbor Laboratory press, Cold Spring Harbor, pp. 297–331.
27 Steinberg,S., Mish,A. and Sprinzl,M. (1993) *Nucleic Acids Res.*, **21**, 3011–3015.
28 Lloyd,A.T. and Sharp,P.M. (1992) *Nucleic Acids Res.*, **20**, 5289–5295.
29 Santos,M.A.S., Keith,G. and Tuite,M.F. (1993) In Brown,A.J.P., Tuite,M.F. and McCarthy,J.E.G., (Eds). *Protein Synthesis and Targeting in Yeasts*. Springer Verlag, Berlin. pp. 111–121.
30 Fasano,O., Aldrich,T., Tamanoi,F., Taparowsky,E., Furth,M. and Wigler,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4008–4012.
31 Noel,D., Nikaido,K. and Ames,G.F.L. (1979) *Bichemistry*, **18**, 4159–4165.
32 Andachi,Y., Yamao,F., Muto,A. and Osawa,S. (1989) *J. Mol. Biol.*, **209**, 37–54.
33 Ohama,T., Yamao,F., Muto,A. and Osawa,S. (1987) *J. Bacteriol.*, **169**, 4770–4777.
34 Sharp,P.M. and Matassi,G. (1994) *Curr. Opin. Genet. Dev.*, **4**, 851–860.
35 Yokogawa,T., Susuki,T., Ueda,T., Mori,M., Ohama,T., Kuchino,Y., Yoshinari,S., Motoki,I., Nishikawa,K., Osawa,S. and Watanabe, K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7408–7411.
36 Smith,D. and Yarus,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4397–4401.
37 Curran,J.F. and Yarus, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6538–6542.
38 Dix,D.B., Wittenberg,L.W., Uhlenbeck,O.C. and Thompson,R.C. (1986) *J. Biol. Chem.*, **261**, 10112–10118.
39 Yarus,M. (1982) *Science*, **218**, 646–652.
40 Susuki,T., Ueda,T., Yokogawa,T., Nishikawa,K. and Watanabe,K. (1994) *Nucleic Acids Res.*, **22**, 115–123.
41 Breitschopf,K. and Gross,H.J. (1994) *EMBO J.*, **13**, 3166–3169.