4 Bové, J. M., Carle, P., Garnier, M., Laigret, F., Renaudin, J., and Saillard, C., Molecular and cellular biology of spiroplasmas, in: The Mycoplasmas, vol. 5, pp. 243–364. Eds R. F. Whitcomb and J. G. Tully. Publishers, New York 1989.
5 Cabrera, M., Nghiem, Y., and Miller, J. H., *Mut*M, a second mutator locus in *Escherichia coli* that generates G.C → T.A transversions. J. Bact. *170* (1988) 5405–5407.
6 Cox, E. C., and Yanofsky, C., Altered base ratios in the DNA of an *Escherichia coli* mutator strain. Proc. natl Acad. Sci. USA *58* (1967) 1895–1902.
7 Crick, F. H. C., Codon-anticodon pairing: the wobble hypothesis. J. molec. Biol. *19* (1966) 548–555.
8 Heckman, J. E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S., and Raj Bhandary, U. L., Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. Proc. natl Acad. Sci. USA *77* (1980) 3159–3163.
9 Hori, H., and Osawa, S., Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Molec. Biol. Evol. *4* (1987) 445–472.
10 Ikemura, T., Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. molec. Biol. *146* (1981) 1–21.
11 Inamine, J. M., Ho, K., Loechel, S., and Hu, P., Evidence that UGA is read as tryptophan rather than stop by *Mycoplasma pneumoniae*, *Mycoplasma genitalium* and *Mycoplasma gallisepticum*. J. Bact. *172* (1990) 504–506.
12 Ishikura, H., Murao, K., and Yamada, Y., EMBO-FEBS meeting, Strasbourg, July 1980.
13 Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T., Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. natl Acad. Sci. USA *86* (1989) 9355–9359.
14 Jukes, T. H., The genetic code, II. Am. Sci. *53* (1965) 477–487.
15 Jukes, T. H., A change in the genetic code in *Mycoplasma capricolum*. J. molec. Evol. *22* (1985) 361–362.
16 Jukes, T. H., and Bhushan, V., Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. J. molec. Evol. *24* (1986) 39–44.
17 Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., and Oshima, T., High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. J. biol. Chem. *259* (1984) 2956–2960.
18 Kimura, M., The Neutral Theory of Molecular Evolution. Cambridge University Press, 1983.
19 Komine, Y., Adachi, T., Inokuchi, H., and Ozeki, H., Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. J. molec. Biol. *212* (1990) 579–598.
20 Lee, C. C., Timms, K. M., Trotman, C. N. A., and Tate, W. P., Isolation of a rat mitochondrial release factor. Accommondation of the changed genetic code for termination. J. biol. Chem. *262* (1987) 3548–3552.
21 Muramatsu, T., Yokoyama, S., Horie, N., Matsuda, A., Ueda, T., Yamaizumi, Z., Kuchino, Y., Nishimura, S., and Miyazawa, T., A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. J. biol. Chem. *263* (1988) 9261–9267.
22 Muto, A., and Osawa, S., The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. natl Acad. Sci. USA *84* (1987) 166–169.
23 Nghiem, Y., Cabrera, M., Cupples, C. G., and Miller, J. H., The mutY gene: A mutator locus in *Escherichia coli* that generates G:C → T:A transversions. Proc. natl Acad. Sci. USA *85* (1988) 2709–2713.
24 Ohama, T., Muto, A., and Osawa, S., Spectinomycin operon of *Micrococcus luteus*: Evolutionary implications of organization and novel codon usage. J. molec. Evol. *29* (1989) 381–395.
25 Ohama, T., Muto, A., and Osawa, S., Role of GC-biased mutation pressure on synonymous codon choice in *Micrococcus luteus*, a bacterium with a high genomic GC-content. Nucleic Acids Res. *18* (1990) 1565–1569.
26 Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F., and Osawa, S., The ribosomal protein gene cluster of *Mycoplasma capricolum*. Molec. gen. Genet. *210* (1987) 314–322.
27 Osawa, S., and Jukes, T. H., Evolution of the genetic code as affected by anticodon content. Trends Genet. *4* (1988) 191–198.
28 Sueoka, N., Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. Proc. natl Acad. Sci. USA *47* (1961) 1141–1149.
29 Sueoka, N., On the genetic basis of variation and heterogeneity of DNA base composition. Proc. natl Acad. Sci. USA *48* (1962) 166–169.
30 Tanaka, R., Muto, A., and Osawa, S., Nucleotide sequence of tryptophan tRNA gene in *Acholeplasma laidlawii*. Nucleic Acids Res. *17* (1989) 5842.
31 Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., and Osawa, S., UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. natl Acad. Sci. USA *82* (1985) 2306–2309.
32 Yamao, F., Iwagami, S., Azumi, Y., Muto, A., Osawa, S., Fujita, N., and Ishihama, A., Evolutionary dynamics of tryptophan tRNAs in *Mycoplasma capricolum*. Molec. gen. Genet. *212* (1988) 364–369.
33 Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., and Miyazawa, T., Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. Proc. natl Acad. Sci. USA *82* (1985) 4905–4909.

0014-4754/90/11-12/1097-10$1.50 + 0.20/0

Eucaryotic codes

F. Caron

Laboratoire de Génétique Moléculaire, Ecole Normale Supérieure, 46 rue d'Ulm, F-75230 Paris Cedex 05 (France)

Summary. This article is a review of the rules used by eucaryotic cells to translate a nuclear messenger RNA into a polypeptide chain. The recent observation that these rules are not identical in two species of a same phylum indicates that they have changed during the course of evolution. Possible scenarios for such changes are presented.
Key words. Genetic code; eucaryotic cell; evolution; code ambiguity; code universality; convergence.

Introduction

The genetic program of a cell is entirely contained in its DNA. Execution of this program involves many different steps, but each of them proceeds invariably in two phases which, for a eucaryotic cell, take place first in its nucleus and thereafter in the cytoplasm: in the nucleus, a copy of a part of the DNA is made and is afterwards modified by capping, polyadenylation, splicing and editing; in the cytoplasm, this modified copy (the messenger RNA) is

deciphered by the ribosomal machinery to make a protein. This chapter is a review of the current status of knowledge about the second phase. In the first part of this review, I have adopted a plan which follows the chronology of the deciphering process when a ribosome first encounters a mRNA molecule: Where does the message start? How is it read? Where does it stop? In the second and last part, I have made a survey of the known species which have changed some of the rules in the deciphering process, and I have tried to analyze the implications of these variations for the evolution of eucaryotic organisms.

Finding the beginning of the message: Initiation

Where does the message start? Ten years ago, the two canonical rules about the initiation of translation in eucaryotes were that the initiation codon is an AUG and that among the many AUG triplets present on a message, the first one is selected. This led to the proposal of the scanning model[66,71]: the ribosome binds to the mRNA via the cap and slides along the mRNA until the first AUG is encountered; from this point, the message is read in triplets. These rules (and the model) are still valid in most cases, but recent exceptions have been quoted in the literature that I shall review here.

1) The nature of the initiation codon

Of the several hundred eucaryotic mRNA sequences which have been published in the past years, AUG appears to be the almost exclusive initiation codon[63]. However, there are reports of in vivo initiation on CUG in the protooncogene c-myc[39] and in the human basic fibroblast growth factor[106], on ACG in one of the capsid proteins of an adeno-associated virus[5] and in a mouse transcription factor[82], on UUG in infectious plant plasmids[2], and even on UAG in the *Plasmodium falciparum* aldolase gene[34]. Departures from AUG always lead to a decrease in the efficiency of messenger RNA translation. This has been demonstrated by several in vivo and in vitro studies[1,17,100]. Among the various non-AUG initiation codons which have been tested, GUG seems to be the most efficient in yeast[17], and in mammalian in vitro translation systems, substitution of AUG by CUG leads only to a 50–60% reduction in translational efficiencies[23]. The explanation of the decrease is that the molecular recognition of the initiator codon by the ribosome is mediated by the initiator tRNA (tRNAi) whose anticodon is 3'- UAC 5'. A very illustrative example of this property has been given recently by Cigan et al.[16] who have mutated the anticodon of one of the 4 initiator tRNAs of *Saccharomyces cerevisiae* from 3' UAC 5' to 3' UCC 5' and shown that this specifically promotes the initiation of translation at an AGG codon. It is likely that natural variations in the initiation codon are used by the eucaryotic cell to regulate negatively the translation of a particular message.

2) Codon context

The nature of the nucleotides in the vicinity of the initiation codon influences the efficiency of translation of the message. An optimum sequence PuNNAUGG (Pu : purine; N : any base) has been determined, and any departure from this leads to a decrease in the efficiency of translation of the message[64,69]. At variance with initiation in procaryotic cells, in eucaryotic cells tRNAi binds invariably to the ribosome prior to its interaction with the message. Therefore, both are responsible for the recognition of the initiation sequence: the tRNAi for the initiation codon and the ribosome for the surrounding nucleotides. An ACG initiator codon in an optimal context could be as efficient as an AUG in a suboptimal context. The contributions of both interactions seem to vary from one species to another: yeast, which is less sensitive to codon context, will discriminate much more between AUG and non-AUG codons than will higher eucaryotes[17].

Sequences further downstream from the initiation codon might influence the efficiency of recognition. Recently, it has been shown that the presence of a downstream sequence, which may potentially adopt a hairpin structure, increases the recognition of an upstream AUG initiation codon[67].

3) First AUG

Initiation on the first AUG is the general rule[66]. However, this rule does not stand in at least three different cases: alternative initiation, multiple open reading frames (ORF) on a single mRNA, and internal initiation. Analysis of the in vivo products of some eucaryotic mRNAs have shown that initiation may occur at alternative initiation codons[65,70]. The simplest mechanism to explain this phenomenon is that a fraction of the ribosomes do not recognize the first initiation codon very efficiently but slide to the second (leaky scanning). As a consequence, the nature of the two initiation codons and their contexts will determine the relative efficiency of the initiation of translation at both sites (i.e., the relative amounts of the two polypeptide species). This has been verified with in vitro constructions tested in vivo[69,72,83].

Eucaryotic mRNA is monocistronic in nature: it contains a unique translated open reading frame (ORF) bounded on either side by an initiation codon and a termination codon. Artificial introduction of another initiation codon upstream of the true initiator codon leads to initiation at the new site. This is consistent with the scanning model. Subsequent insertion of a termination codon in between the two initiation codons restores the correct translation of the message, albeit with a reduced efficiency[83]. The interpretation of such a phenomenon is that ribosomes do not dissociate at the first termination codon, but move along the mRNA molecule and resume reading at the next initiation codon. Such constructed mRNAs are formally bicistronic although the ORF introduced is short (minicistron). The efficiency of reinitia-

tion has been measured as a function of the intercistronic distance and, at variance with the procaryotic rule observed in polycistronic bacterial mRNA, the efficiency is very low for short distances but increases with intercistronic length and eventually reaches the value obtained in the absence of the minicistron[68]. Short upstream ORFs have been found in different mRNAs of different species, where they play a role in regulation of the expression of their products (for instance, the yeast positive regulator GCN4[15,49,50,93]).

Highly efficient translation of uncapped viral mRNA has recently been observed in picornaviruses[57,102] although this type of mRNA, which lacks the cap recognition signal used for the entry of the ribosomes, is usually very inefficiently translated. In the case of encephalomyocarditis virus (EMC)[57] and poliovirus[102], the sole presence of a long (574 nucleotides for EMC and 736 for poliovirus) untranslated region upstream of the initiation sequence has been shown to cause an internal initiation. A possible mechanism would be that this region has a secondary structure which, by interacting with the ribosome, directs it to the internal initiation site[56]. It has also been reported that a cellular protein could be involved in this process[87]. In the case of the hepatitis B virus (HBV), the 3.3 kb mRNA codes for two proteins are the core (or nucleocapsid) protein C and the polymerase P (reverse transcriptase). The two ORFs for C and P are not in the same frame, and they overlap by about 300 nucleotides. Different experiments indicate that the downstream P gene is translated at an AUG by internal initiation[14,112]. An analogous phenomenon is observed in translation of the X protein of Sendai virus[22].

Reading the message: The 3 by 3 rule

The translational machinery is remarkable in its high reading speed and in its phasing accuracy. At the molecular level, mRNA must be translocated rapidly and by steps of 3 nucleotides. Inaccuracies in this process reveal how it proceeds. Frameshift is one example.

1) Frameshift

Frameshifting occurs at particular sites where the tRNA-ribosome complex moves by 1 or 2 nucleotides with respect to the mRNA[20,44] (longer shifts, called tRNA hopping[53], have not yet been observed in eucaryotes). It is almost exclusively observed in several retroviruses (for instance, the Rous sarcoma virus (RSV)[55], the human T-cell leukemia virus (HLTV)[85], the human immunodeficiency virus (HIV)[128], and in retrovirus-like sequences such as the yeast Ty transposon[89,129]. The site of frameshift is located in the overlapping portion of two out-of-phase ORFs. Frameshift will cause the expression of the downstream ORF (pol gene in the case of HIV[128] and RSV[55]) as a fusion product with the upstream gene (gag gene). In a very economical way, the virus can express two genes at two different levels, which are determined by the frequency of frameshift (5–10%).

2) Sites of frameshift

The positions on the mRNA where the two anticodons of the two tRNAs at the A and P ribosomal sites shift with respect to the mRNA sequence, have been determined precisely in the case of RSV[55] and HIV[128] where the shift is -1. It is obvious (and in fact true) that the tRNAs must in some way bind to the mRNA sequence in the shifted position (at least by the first two positions of the two codons). Such a constraint necessitates the presence of homopolymeric runs at the site of shifting, with runs of A or U being more 'slippery' than G or C because of the weaker interaction of the AU pair. And indeed, the 7-nucleotide frameshift sites of RSV and HIV are respectively $5'A_3U_3A3'$ and $5'U_6A3'$. Mutations at these positions will affect the efficiency of frameshift differently: the first six nucleotides on the 5′ side, which correspond to the first 2 positions of both codon-anticodon pairs at the initial and final position (before and after shifting), will affect it drastically. But the last one, the wobble base for the tRNA in the A site before shifting, is indifferent since it is not directly implicated in any pairing with the tRNAs after the shift. Moreover, in the case of RSV, a change of A to U at this last position will involve $tRNA^{phe}$ instead of $tRNA^{leu}$ at the A site, indicating that frameshift can be mediated with comparable efficiency by different tRNAs. Lastly, it is worth mentioning that the presence of homopolymeric runs at a known position of frameshift has also been recently observed in procaryotes (γ subunit of *E. coli* DNA polymerase III[28,122]).

3) Context effect in frameshift

Is the 7-nucleotide sequence solely responsible for the frameshift? The answer is no and, as in initiation, context plays a role. Mutagenesis of a sequence located about 80 nucleotides downstream of the RSV shifting sequence impairs the efficiency of the frameshift[55]. Examination of the potential RNA secondary structure reveals that this sequence perfectly complements another sequence immediately downstream of the frameshift sequence. The stem structure formed in this way seems to be effective, since any destabilizing mutation reduces frameshift and a second compensatory mutation restores it. The interpretation given by the authors[55] is that the hairpin structure either slows down the translational process at the site of frameshifting, forcing the ribosome to mark a pause in front of the secondary structure and give it time to frameshift, or even plays a more active part in pushing it backwards. In the case of RSV[55], the sequence containing the frameshift site and the hairpin (about 150 nucleotides) can be transposed to another genetic context and gives rise to frameshifts in mammalian systems as well as in yeast. In the case of HIV[128], no downstream hairpin can be found, and a short sequence of 26 nucle-

otides containing the site of frameshift is sufficient (but possibly not necessary) to give rise to the frameshift in other genetic contexts and systems. The fact that a relatively short sequence can cause a frameshift and severely affect the regular translocation of the mRNA led to a computer search for homopolymeric runs of 6 or more T's in the GenBank data base. This motif is rare in coding sequences (6 found), which suggests that living organisms selectively avoid it.

Instead of a hairpin structure, a termination codon in phase with the upstream ORF and immediately following the frameshift site has been found to contribute to frameshifting[18–20]. However, in the case of RSV, replacement of the UAG termination codon by mutagenesis with another non termination codon does not change the frameshift efficiency[55]. Moreover, HIV[128] and other retroviruses (sequences reviewed in Jacks et al.[55]) do not possess a termination codon at this position.

From statistical studies of coding sequences, Trifonov[121] discovered that any coding sequence contains an underlying message in the form of a repeated pattern of three bases (G-nonG-N)$_n$ and he proposed that this periodicity is used in some way by the translation machinery as a guide mark for the accurate translocation of mRNA (frame monitoring). In this type of analysis, frameshift sites are readily detected as points where this periodicity is shifted. 3 surface-exposed sites of the *E. coli* 16S RNA show this periodic complementary sequence, which suggests that frame monitoring is mediated by some kind of ribosome-mRNA interactions and that their local disruption leads to a frameshift. Therefore, context effects would lie, as in the initiation process, mainly in the ribosome-mRNA interaction.

Deciphering the message: The genetic code

25 years ago, the rules of correspondence between a nucleotide triplet (codon) and either an amino acid (sense codon) or a termination signal (nonsense codon) were established. 61 codons directed the incorporation of 20 amino acids and 3 the release of the polypeptide chain from the ribosome. Two statements about the genetic code, which were considered during a long period of time to be valid, have recently been shown to be false: the first one is its accuracy and non-ambiguity, and the second its universality.

1) Non-ambiguity of the genetic code

The non-ambiguity of the genetic code means, at the molecular level, that 3 codons are never recognized by tRNAs, but only by releasing factors, and 61 are accurately recognized by the aminoacylated tRNAs. And indeed, living organisms have developed sophisticated kinetic devices to improve the accuracy of the recognition and ligation steps at the expense of translational speed[61, 96]. I shall not review these questions here, but instead show examples of ambiguity. Ambiguity exists in the eucaryotic genetic code only because of the ability of some tRNAs to read either specifically or by mistake nonsense codons. Therefore, in competition with releasing factors, they introduce ambiguity. We shall describe briefly a few examples of such natural suppressions which will be relevant to the discussion of the evolution of eucaryotic genetic codes. Excellent reviews on this subject have already been published[42, 124, 125], and a complete survey is to be found in Dr Murgola's review (this issue).

1a) Nonsense suppression via misreading

In all the eucaryotic species studied, a few normal major cytoplasmic tRNAs recognize not only their complementary codons but also, to a lesser extent, a termination codon. These ambiguous recognitions are used by viruses as a strategy to express their genes (reviewed in refs 124, 125). They have also been observed in artificial situations, by introducing a nonsense mutation in a gene and then searching for the tRNA which, once overexpressed via gene cloning on a multicopy plasmid, will suppress the mutation[109, 126].

Some viruses possess a termination codon (UAG or UGA) in their gene coding sequences and use the property of a host tRNA that can misread it to synthesize 2 polypeptide chains which will differ by their length (the long one differing from the short one by a carboxy terminal extension) and by their respective amount, with the ratio being given by the frequency of readthrough (10–20%). This process is a convenient and economical way used by the virus to synthesize two related proteins at 2 different levels – and it has also greatly helped the molecular biologist to understand more deeply a variety of codon-anticodon interactions (tRNA-mRNA interactions in the ribosome).

In the tobacco mosaic virus (TMV), 2 proteins of 126 kDa and 183 kDa are produced, the latter by a translational readthrough of the UAG termination codon of the former[8, 35, 101]. 2 tRNA^{tyr} extracted from tobacco leaves with anticodon 5′GΨA3′ are presumably responsible for the in vivo suppression[7]. Post-transcriptional modification of the G base into Q (queuosine), which occurs to various extents with these tRNAs according to the type of tissue, completely abolished the suppression[6, 10]. These results demonstrate that an unorthodox G-G base pair may occur in the wobble position and that the suppression will depend on the type of modification of the suppressor tRNA.

In the Moloney murine leukemia virus (MoMuLV), an UAG codon is inserted between the 2 in-frame gag and pol genes and the pol gene is synthesized as the fused gag-pol protein by readthrough of the amber codon[133]. The suppressor tRNA is a minor tRNA^{gln} with an anticodon of 5′UmUG3′ (m is 2′-O-methyl)[74]. Divergent results have been reported concerning an increase in the amount of this tRNA upon infection[27, 74]. The suppressive activity of this tRNA shows that a wobble may occur

in the first position of the codon (G-U base pairing). The same type of pairing has been obtained in yeast using the following procedure[109]: an ochre mutation has been characterized in the rad10 gene conferring UV-sensitivity to the corresponding yeast strain. By screening a yeast genomic library made with a multicopy plasmid vector and the UV-sensitive strain, partial suppression of the phenotype was obtained when the recombinant plasmid contained a normal $tRNA^{gln}$ gene (anticodon 5'UUG3'). An analogous experiment done with an amber mutant[126] in the rad4 gene has led to the selection of a second $tRNA^{gln}$ gene with an anticodon of 5'CUG3'. Both codon-anticodon pairs involved, as in the case of MoMuLV, a G-U base pair in the first position of the codon.

It is important to realize that the previously-mentioned cases of orthodox or unorthodox base pairing cannot be considered as general rules. Although non-Watson-Crick pairing in the anti-anti configuration has been observed in crystallographic studies of the B form of nucleic acids[54,108], G in the wobble position usually pairs with U or C rather than G, and G in the third position of the anticodon pairs with C but not with U except in the special case of yeast $tRNA^{leu}$[127]. Moreover, *Tetrahymena* $tRNA^{gln}$ with an anticodon of 5'UmUG3' never reads UAG or UAA[40], whereas the above-mentioned mouse $tRNA^{gln}$ can read it. Therefore, these suppressions of nonsense codons in the translational process cannot merely be reduced to unorthodox codon-anticodon interactions. Before giving any tentative explanation, let us examine the second case of nonsense suppression.

1b) Nonsense suppression via specific recognition

Both selenium and sulfur belong to the same column of the periodic table of the elements and so, despite its larger atomic radius, selenium may replace sulfur randomly in a protein in the form of selenomethionine[118,119]. By taking advantage of the low redox potential of selenium compared to sulfur, selenium is incorporated specifically as selenocysteine in the catalytic sites of enzymes involved in cell protection from damage by oxygen derivatives (membrane, DNA,...). Examples of such enzymes include glycine reductase and formate dehydrogenase in procaryotes and glutathione peroxidase in higher eucaryotes.

Let us first briefly summarize the procaryotic pathway for incorporating selenocysteine into a polypeptide chain. In *E. coli*, 4 genes are responsible for this incorporation (selA, selB, selC, selD). The unique Se-cyst of formate dehydrogenase is encoded by an in-frame UGA codon[134] which is specifically read by a tRNA[81,115,117] (product of gene selC). This tRNA has a 5'UCA3' anticodon which specifically recognizes UGA in an in vitro test. Besides having a very unusual structure, it is ligated to serine by ser tRNA ligase and the seryl-tRNA is converted to selenocysteyl-tRNA by the action of the products of genes selA and selD[79,80]. Recently, the product of the selB gene has been cloned and sequenced: it encodes a novel translation factor which has similarities with the initiation factor 2 (IF2) and the elongation factor EF-Tu[29]. Its interaction with the Se-cyst-tRNA (product of gene selC) is extremely selective: the novel factor does not bind to other elongator tRNAs[29] and the Se-cyst-tRNA interacts specifically with the novel factor, but 100-fold less than the other tRNAs with the usual EF-Tu factor[30].

In eucaryotes, the gene for mouse glutathione peroxidase has been cloned and sequenced[13]. As in procaryotes, the unique Se-cyst residue is encoded by a UGA codon. In bovine liver, two tRNAs with anticodons of 5'CmCA3' and 5'NCA3' (N being a modified nucleoside) almost exclusively recognize UGA in ribosome-binding assays[24,43]. Despite their very unusual structures and the fact that their anticodons resemble a Trp anticodon, they are aminoacylated by serine. Once aminoacylated, the serine moiety becomes the substrate of a very specific phosphorylation reaction, so that these tRNAs are mainly isolated in the form of phosphoseryl-tRNAs. Recently, they have also been purified as Se-cyst-tRNAs, which suggests that phosphoseryl-tRNAs are a metabolical intermediate between the seryl-tRNA form and the Se-cyst-tRNA form[77]. Therefore, it appears that the procaryotic and eucaryotic systems for incorporating Se-cyst proceed through the suppression of a UGA codon by a very special tRNA[76]. It seems useful here to remind the reader that the UGA codon in the coding sequence is also a termination codon (UGA is not 100% suppressed), and as long as we do not know whether the truncated form has a role, it is impossible to rationalize the choice of UGA to code for the 21st amino acid[117]. In any case, there is no essential difference between nonsense suppression via misreading or specific recognition, since they both lead to ambiguity in the genetic code.

1c) Kinetic theory of nonsense suppression

The incorporation of any amino acid into a polypeptide chain is a kinetic process involving two elements (reviewed in refs 61, 96): an mRNA-ribosome with a tRNA in the P site and a ternary complex tRNA-EF1-GTP which can interact at the A site. (This kinetic process is a theoretical model developed for procaryotes, which has been shown to be correct in many instances. We shall transpose it to eucaryotes, assuming that the two translation systems are not too different). In our case, among all the ternary complexes made with all the possible tRNAs, the one with the suppressor tRNA is supposed to have the best interaction with the A site but it must compete for this site with the releasing factor RF. If the ternary complex interacts with the A site, then the formation of the peptide bond will proceed in two steps, and at each of these steps the ternary complex can dissociate, and let the RF interact and release the poylpeptide chain. Therefore, the efficiency of suppression (the degree of ambigu-

ity of the genetic code) will depend upon a balance between the actions of these two mutually exclusive interacting elements: the RF and the ternary complex.

The rabbit peptide chain release factor has been recently cloned and sequenced [78] and although it displays a striking sequence similarity with tryptophanyl-tRNA synthetases, no data are available on the functioning of eucaryotic RF. However, two arguments tend to support the idea that the efficiency with which RF releases the chain might depend upon the context: 1) Injection of efficient yeast tRNA suppressors into *Xenopus* oocytes leads to only a 10% increase in the number of longer polypeptide chains synthesized [9]. Artificial introduction of a tRNA suppressor gene into a eucaryotic cell may have variable consequences but does not seem to be lethal: in yeast, it does not affect the cell viability except when it is expressed at a high level [48]. In *Drosophila*, cases of sterility have been reported suggesting an interference of the expression of such a gene with the cell differentiation program [25]. In some way, the eucaryotic cell is buffered against the fortuitous emergence of a new suppressor tRNA. 2) Compilation of sequences around termination codons display a significant bias in base composition [62]. Therefore, it is possible that the mRNA context around the suppressible nonsense codon deviates in such a way from the 'optimal' context of a true termination codon that it markedly decreases the efficiency of RF.

The stronger the interaction of the ternary complex with the A site, the longer the 'sticking' time and the more likely it is that the peptide bond will be made and the nonsense codon suppressed. Therefore, two parameters will be important: the amount of ternary complex, and the strength of the interaction of the ternary complex as a whole at the A site (and not just the codon-anticodon interaction, which is only part of it). The role of the amount of ternary complex has been illustrated in the case of the mouse and yeast $tRNAs^{gln}$ [74, 104, 126]. It is obvious that no simple answer can be given at the moment as to the nature of the interaction of the ternary complex with the A site owing to its extreme complexity. However, the contribution of EF1 to this interaction could be important if, as in procaryotes [29], another EF1 is synthesized in the eucaryotic cell to direct the incorporation of Se-cys. Also, the structure of the suppressor tRNA has been shown to play a part: 1) In the case of Se-cyst, the tRNA structure does not possess some of the well-conserved features common to all tRNAs [24, 76, 81] (length of the acceptor stem, for instance). 2) The base modifications at the first position of the anticodon (position 34) change the in vivo specificity of codon recognition [11, 73] (for instance, as shown above in the case of TMV suppression, Q cannot replace G [6, 88]). 3) Future work might reveal the importance of base modification at position 37 in stabilizing the G-U interaction at the first position of the codon. Moreover, it is worth remembering that tRNA modification is never a complete chemical reaction [11, 60, 97] and therefore, the extent of the modification (the ratio of modified to unmodified forms) may vary according to external conditions and in multicellular organisms, according to cell type and age [52]. This point is extremely interesting because it raises the possibility of a flexible genetic code [26] whose level of ambiguity would be adapted to cell requirements (for example, possible variation of the suppressive activity of Se-cyst-tRNA between aerobic and anaerobic conditions). Finally, the poorly characterized mRNA context may play a role in the ternary complex interaction with the A site. For instance, an endogeneous tRNA in rabbit reticulocytes suppresses in vitro a UAG codon embedded in the TMV context, but not in the MoMuLV context [123]. In viral suppression, nucleotide motifs around the nonsense codon are conserved. Also, in the case of the incorporation of Se-cys, special organization of the sequences surrounding the UGA codon has been observed [134].

2) Universality of the genetic code

20 years ago, the allocations of the 64 codons were established in *E. coli*. Examination of other organisms has shown that they have the same codon allocations. To explain this uniformity, a two-stage theory has been put forward: the stereochemical [130] and frozen accident theory [21]. The reasoning was the following: during the early course of evolution, codon X was chosen to code for amino acid Y because in some way, the structures of X and Y were related (stereochemical theory). Afterwards, when the stereochemical interaction between X and Y was no longer direct but via an adaptor molecule, and therefore changes could in principle occur, reallocation of a codon would have been lethal because of the complexity of the genomes (frozen accident theory). Whatever the validity of this theory, it can no longer be used as an argument in favor of the universality of the genetic code, since three families of eucaryotes (complex organisms in the sense of the frozen accident theory) have adopted genetic codes which display deviations from universality. The first example of deviation (ciliates) was discovered in 1984 [12, 46, 51, 107] and the two others in 1989 [59, 113]. It seems obvious that the number of cases of deviations observed will increase rapidly in the future.

2a) The genetic codes of ciliates

Ciliates form a large phylum of more than 7000 species which has been divided on morphogenetic criteria into three sub-phyla: the holotrichs, the hypotrichs and the heterotrichs [94]. Genes of two representatives of holotrichous ciliates, *Tetrahymena* and *Paramecium*, have been cloned and sequenced. UAG has invariably been found as the unique termination codon. Apart from one tubulin gene [3], all these genes contain UAA and (or) UAG codons scattered in their coding sequences. By different types of arguments, it has been shown that these codons are not termination codons but instead code for glutamine [12, 40, 46, 51, 75, 105, 107]. In vitro translation of

Paramecium mRNAs[90] in heterologous systems is blocked for almost all the polypeptide chains. This shows that the use of UAA and UAG as glutamine codons is a property of the whole *Paramecium* genome. Searches for other deviations from the universal genetic code, although not exhaustive, were unfruitful, which suggests that only UAA and UAG allocation have been modified[51,91,104,105]. Three $tRNA^{gln}$s from *Tetrahymena* have been isolated and their structures determined[40,75]. Two of them have 5'CUA3' and 5'UmUA3' anticodons. $tRNA_{UmUA}$ has a strong readthrough activity in rabbit reticulocyte lysate over UAA and UAG codons of α-globin and TMV mRNA, but $tRNA_{CUA}$ only reads through UAG. The third $tRNA^{gln}$ has a 5'UmUG3' anticodon and recognizes the 2 other glutamine codons CAA and CAG[40].

In the few genes sequenced in hypotrichous ciliates, *Stylonychia* uses UAA as a glutamine codon[46], *Oxytricha* uses UAA and UAG as glutamine codons[47], whereas in *Euplotes* UAA is a termination codon and the 34 glutamine residues present in the 3 proteins whose genes have been sequenced are encoded by CAA and CAG[41,92]. Full length polypeptide synthesis has been obtained from in vitro translation of *Euplotes* mRNAs in heterologous systems[41]. No data has yet been published on the genetic code of heterotrichs.

Although these results do not allow the determination of the complete allocations of the 64 codons for each above mentioned species, it is nevertheless clear that UAA and UAG are used as glutamine codons by some ciliates and as termination codons by others.

2b) The genetic codes of algae

Eucaryotic algae are traditionally divided into three subphyla: rhodophyte, chromophyte, and chlorophyte. Genes from 2 species of the chlorophyte subphylum, *Chlamydomonas* and *Acetabularia*, have been sequenced and as in the hypotrichous ciliates, it appears that UAA and UAG code for glutamine in *Acetabularia*[113] whereas they are definitely termination codons in *Chlamydomonas*[36]. Again, in vitro translation of Acetabularia mRNAs in heterologous systems results in the synthesis of mostly small polypeptides[114].

2c) The genetic codes of fungi

Candida cylindracea[131] (*Candida rugosa* ATCC14830) is an asporogenic yeast. Direct comparison of the experimental amino acid composition of lipase I and that deduced from the cDNA nucleotide sequence reveal discrepancies only in the leucine and serine content[59]. Comparison of the amino acid sequence deduced from the nucleotide sequence and the partial amino acid sequence of some peptides obtained from protease-digested lipase I demonstrates that CUG codes for serine instead of leucine[59]. With this unique change, an excellent agreement is obtained between the two amino acid compositions. In in vitro translation experiments, a synthetic capped mRNA containing several repeated CUG codons immediately after an initiator AUG directs the specific incorporation of serine and not leucine, if the cell free translation system is prepared from *Candida cylindracea*, and the reverse, if it is prepared from *Saccharomyces cerevisiae*[59]. Therefore, the *C. cylindracea* genetic code displays a novel kind of deviation from universality with at least one change in the allocation of a sense codon.

Evolution of genetic codes

1) Time scale of evolution

The genetic code is the keystone for the coordinated expression of thousands of genes, and all cell functions are extremely sensitive even to very subtle changes in this code. If evolution allows variations in genetic codes, it cannot occur abruptly but must take place gradually. The frozen accident theory can be restated in this way: for complex organisms, the more frequently a codon is used, the longer it will take to change its allocation or its level of ambiguity. I shall take two extreme examples.

As discussed above, nonsense suppression via misreading probably occurs almost exclusively in the expression of viral mRNA, since up to now no cellular mRNA has been reported to use this strategy for the differential expression of 2 genes. Therefore, variations in the level of ambiguity of these codons might occur on a short time scale without any dramatic consequence except perhaps for the virus. Indeed, unmodified $tRNA^{tyr}$ with G in the first position of the anticodon, which suppresses the UAG codon in TMV mRNA, is the major form in tobacco leaves, whereas the modified form with Q, which does not suppress, is present at 85% in the germ line[6]. Since the two forms of tRNA do not show any marked specificity for one of the two tyr codons, these variations must be triggered by some unknown cellular functions not based on the ambiguity of the termination codon.

In contrast, reallocation of a codon is undoubtedly a process taking place on a long time scale, since it concerns thousands of genes. If, inside a subphylum, two species have different genetic codes, the variation must have taken place between now and the time of divergence of the two species. The most recent and trustworthy phylogenetic tree of ciliates (A. Adoutte, personal communication and refs 4, 84, 110), based on ribosomal RNA sequences, establishes that the second earliest diverging branch corresponds to the heterotrich subphylum and the next one to *Euplotes* (which forms the euhypotriches class); the rest of the ciliates (holotriches and pseudohypotriches) branch off from this point. Ciliates emerged approximatively 10^9 years ago, at a period of intense evolutionary radiation (late precambrian) and therefore, *Euplotes* appeared somewhat later. Fungi and algae emerged at about the same time as ciliates[38,103]. Therefore, branchings within the chlorophyte subphylum (between *Chlamydomonas* and *Acetabularia*) or within the

fungi (between *Sacharomyces* and *Candida*) occurred later. The above-mentioned deviations in the genetic code took place according to current estimates, in the last $5 \cdot 10^8 - 10^9$ years. More elaborate phylogenetic trees will give a more precise estimate of these numbers. A possible discovery of different codes among closer orders, families or groups will show that codon rellocation may occur on even shorter time scales.

2) Convergent versus non-convergent evolution

Since algae and ciliates form two well-separated clusters, which branched off 10^9 years ago, and since within these two phyla species adopted either the universal code (TAA and TAG as termination codons) or the divergent code (TAA and TAG as glutamine codons), then whatever the nature of the unknown code of their common ancestor, one can conclude that this must be a case of convergent evolution. It is hardly possible for the time being to imagine what has 'pushed' some living organisms to adopt the divergent code, but in the face of the same problems, they must have found the same (perhaps unique) solution.

3) Scenarios for the evolution of codes[31-33, 37, 40, 41, 51, 58, 92, 95, 99, 120]

If we observe nowadays two species with two different genetic codes, then either their common ancestor had an ancestral code different from both of the two codes, or the ancestral code was identical to one of them and the other has diverged. Although we are far from knowing the extent of variation of the codes, many different species which diverged very early in evolution have adopted the universal code. Therefore, unless convergence has occurred throughout evolution, which seems unlikely, then the universal code must have been fixed in its present state before the branching off of ciliates, algae and so on.... The scenarios have to answer the question: how, with our current knowledge of molecular mechanisms, can we imagine a termination codon becoming a glutamine codon or a leucine codon, a serine codon?

We have shown that the deciphering rules in their present state are much more complex than previously expected. As in any language, some codons (words) are nonambiguous, while others are ambiguous and their sense depends upon the context. The comparison with languages can be pushed further. Languages evolve, although on a shorter time scale than genetic codes: words which share the same root may have two different meanings in two different languages ('eventual', 'promiscuity' in the English language; 'éventuel', 'promiscuité' in the French language), while others within the same language can acquire a new meaning by losing the first one (often because the object or action they represent has disappeared, e.g. 'silly'). Whatever the mechanisms by which such a phenomenon happens, time is required for it to occur, because abrupt changes in the meaning of commonly used words would result in people misunderstanding each other. Therefore, the scenarios we have to imagine are gradual changes in the usage of a codon in two phases: loss of a first sense and acquisition of a second.

3a) Termination codon into glutamine codon

Due to our ignorance of eucaryotic RF, and in particular those of ciliates and algae, it is hard to speculate on how 2 codons can lose their specificity for termination. The acquisition of UAA and UAG as glutamine codons is conceivable because a tRNA gene could have been duplicated and the anticodon modified, as was previously suggested[58, 99]. The close structural relationship between the 3 $tRNA^{gln}$ of *Tetrahymena* support a filiation mechanism among them[40]. The fact that glutamine in two different species has captured these 2 codons could be related 1) to the close proximity of the glutamine codons CAA and CAG, 2) to the ability of $tRNA^{gln}$ to act as a natural suppressor in different eucaryotic species[33, 74, 109, 126], 3) to a property which has been demonstrated in procaryotes but could also exist in eucaryotes, namely that the anticodon is not an identity element of tRNAgln in the recognition by the synthetase[98] (even more, glntRNA synthetase misacylates many different amber suppressors).

A drift in the base composition of the genome has often been proposed as an explanation for these changes[40, 58, 99, 132]. Ciliates have AT-rich genomes: the AT percentage reaches 80–85% in noncoding regions, and a significant bias for AT-rich codons is observed in codon usage[86]. It is possible to imagine that AT pressure has, in some unknown way, led these organisms to use AT-rich termination codons as sense codons, and glutamine would have been the most likely candidate to 'capture' UAA and UAG. This type of argument has to be considered with some caution for the following reasons: first, some organisms (*Dictyostelium*, for instance), which diverged before ciliates and algae, have a high AT content and a strong bias in codon usage, but still use the universal code. Second, genomes of ciliates have a high AT content but the codes they have adopted are different (see above). Third, UAA and UAG usage as glutamine codons varies significantly between *Paramecium* and *Tetrahymena*, which have the same AT content. From recent compilations of codon usages (P. Dupuis, personal communication and refs 3, 86), the ratio of the frequency of codon usage (CAA/G)/(UAA/G) is 0.83 for *Tetrahymena*[8] and 0.37 for *Paramecium* (P. Dupuis, personal communication). Comparison of the same ratio for the same family of genes (α and β tubulins) gives 12.2 for *Tetrahymena* and 0.83 for *Paramecium* (the sequenced α-tubulin gene of *Tetrahymena* does not even contain one UAA or UAG but 13 CAA/G). Variation in this codon usage is reflected in the sizes of the polypeptides synthesized in heterologous in vitro translation systems; the molecular weights of the vast majority of these polypeptides are under 10 kDa for *Paramecium*[90], whereas they go up to 25 kDa for *Tetrahymena*[109]. In conclusion, although it is conceivable that a drift in the genomic base

composition can cause a change in codon usage and even a change in codon allocation, it is much more likely that this is due to an evolution of the tRNA pool. I shall develop this argument now.

3b) Leucine codons to serine codons

Considerable evidence for a fit between codon usage on the one hand and tRNA population and specificity on the other have been reported during the recent years (Hatfield and Rice[45] and refs therein). Therefore, a change in codon usage can be seen as a readaptation of a genome to a modification in the tRNA pool. Changes in the specificity of tRNA molecules may occur at the level of their nucleotide sequence or of their post-transcriptional modifications. Change in the tRNA population is possible through the modulation of the expression of their genes or through their modifying enzymes[11,60]. Modifications in tRNA specificities and population are more likely to occur gradually through changes in tRNA post-transcriptional modifications since they are numerous, diverse and not always complete (equilibrium between the modified and the unmodified forms). One can imagine that any change in the amount of a modifying enzyme (mutations in the promoter) or in its specificity (missense mutations) could lead to a small variation in the tRNA population and (or) specificity that can be corrected later on by a readjustment of the pattern of codon usage. Series of steps of this kind will lead to such a modification in the codon usage that, for instance, 2 codons like CUA and CUG will no longer be used as leucine codons. Studies of the leucine tRNAs of *Candida* will give us clues to confirm or refute such mechanisms. Acquisition of the two codons by serine can proceed through a mechanism similar to the one described previously.

Conclusion

The process of translation at the molecular level involves a set of complex interactions among numerous macromolecules. As a consequence, the deciphering rules cannot be reduced to the codon allocation table; they look more and more like grammatical rules of a language on which the effects of context are superimposed. With our present limited knowledge, once given a nucleotide sequence, it is often impossible to predict where the message starts or whether or not it starts at different positions. It is also presumptuous to believe that the deciphering process will proceed all the way by 3-nucleotide steps, since shifting may occur, and that the first termination codon will signal the end of the message, since it can sometimes signify the incorporation of a special amino acid. In all these steps, the meaning of a triplet or of several triplets will depend on the surrounding nucleotides (the context). However, despite the large number of messages which are translated according to this complex set of deciphering rules, changes in the codon allocation table have occurred during the course of evolution more recently than the time when the different kingdoms are thought to have diverged (less than 10^9 years ago): allocations of triplets to amino acids may vary from one species to another within the same phylum. Although the described molecular mechanisms which lead to such changes can be considered, it is doubtful that we will have in the future an even partial knowledge of the causes.

Acknowledgments. I thank Jacques Ninio who generously helped me by stimulating discussions during the preparation of this review. The literature cited was arbitrarily limited and I apologize for all omissions. I thank also R. Valle, A. Adoutte, E. Kubli and M. Dreyfus for useful suggestions, P. Dupuis and A. Adoutte for the communication of experimental results prior to publication and L. Sperling and K. Tanner for a critical reading of the manuscript.

1 Anderson, C. W., and Buzash-Pollert, E., Can ACG serve as an initiation codon for protein synthesis in eucaryotic cells? Molec. cell. Biol. *5* (1985) 3621–3624.
2 Aoyama, T., Hirayama, T., Tamamoto, S., and Oka, A., Putative start codon TTG for the regulatory protein virG of the hairy-root-inducing plasmid pRiA4. Gene *78* (1989) 173–178.
3 Barahona, I., Soares, H., Cyrne, L., Penque, D., Denoulet, P., and Rodrigues-Pousada, C., Sequence of one α- and two β-tubulin genes of *Tetrahymena pyriformis*. J. molec. Biol. *202* (1988) 365–382.
4 Baroin, A., Perasso, R., Qu, L., Brugerolle, G., Bachellerie, J. P., and Adoutte, A., Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. Proc. natl Acad. Sci. USA *85* (1988) 3474–3478.
5 Becerra, S. P., Rose, J. A., Hardy, M., Baroudy, B. M., and Anderson, C. W., Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. Proc. natl Acad. Sci. USA *82* (1985) 7919–7923.
6 Beier, H., Barciszewska, M., and Sickinger, H., The molecular basis for the differential translation of TMV RNA in tobacco protoplasts and wheat germ extracts. EMBO J. *3* (1984) 1091–1096.
7 Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R., and Gross, H. J., UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs tyr with suppressor activity from tobacco plants. EMBO J. *3* (1984) 351–356.
8 Beier, H., Mundry, K. W., and Issinger, O. G., In vivo and in vitro translation of the RNAs of four tobamoviruses. Intervirology *14* (1980) 292–299.
9 Bienz, M., Kubli, E., Kohli, J., de Henau, S., Huez, G., Marbaix, G., and Grosjean, H., Usage of three termination codons in a single eucaryotic cell, the *Xenopus laevis* oocyte. Nucl. Acids Res. *9* (1981) 3835–3850.
10 Bienz, M., and Kubli, E., Wild type Tyr tRNAG reads the TMV RNA stop codon but Q base modified tyr tRNAQ does not. Nature *294* (1981) 188–190.
11 Bjork, G. R., Ericson, J. U., Gustafson, C. E. D., Hagerval, T. G., Jonsson, Y. H., and Wikstrom, P. M., Transfer RNA modification. A. Rev. Biochem. *56* (1987) 263–287.
12 Caron, F., and Meyer, E., Does *Paramecium primaurelia* use a different genetic code in its macronucleus? Nature *314* (1985) 185–188.
13 Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P. R., The structure of the mouse glutathione peroxydase gene: the selenocysteine in the active site is encoded by the termination codon, TGA. EMBO J. *5* (1986) 1221–1227.
14 Chang, L. J., Pryciak, P., Ganem, D., and Varmus, H. E., Biosynthesis of the reverse transcriptase of hepatitis B viruses involves de novo translational initiation not ribosomal frameshifting. Nature *337* (1989) 364–368.
15 Cigan, A. M., and Donahue, T. F., Sequence and structural features associated with translational initiator regions in yeast: a review. Gene *59* (1987) 1–18.
16 Cigan, A. M., Feng, L., and Donahue, T. F., tRNAi meth functions in directing the scanning ribosome to the start site of translation. Science *242* (1988) 93–97.
17 Clements, J. M., Laz, T. M., and Sherman, F., Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. Molec. cell. Biol. *8* (1988) 4533–4536.

18 Craigen, W., and Caskey, C. T., Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc. natl Acad. Sci. USA *82* (1985) 3616–3620.
19 Craigen, W., and Caskey, C. T., Expression of peptide chain release factor 2 requires high efficiency frameshift. Nature *322* (1986) 273–275.
20 Craigen, W., and Caskey, C. T., Translational frameshifting: where will it stop? Cell *50* (1987) 1–2.
21 Crick, F. H. C., The origin of the genetic code. J. molec. Biol. *38* (1968) 367–379.
22 Curran, J., and Kolakofsky, D., Scanning independent ribosomal initiation of the Sendai virus X protein. EMBO J. *7* (1988) 2869–2874.
23 Dasso, M. C., and Jackson, R. J., Efficient initiation of mammalian translation at a CUC codon. Nucl. Acids Res. *17* (1989) 6485–6497.
24 Diamond, A., Dudock, B., and Hatfield, D., Structure and properties of a bovine liver UGA suppressor serine tRNA with a tryptophan anticodon. Cell *25* (1981) 497–506.
25 Doerig, R. E., Suter, B., Gray, M., and Kubli, E., Identification of an amber nonsense mutation in the rosy^{516} gene by germline transformation of an amber suppressor tRNA gene. EMBO J. *7* (1988) 2579–2584.
26 Engelberg-Kulka, H., and Schoulaker-Schwarz, R. A., flexible genetic code, or why does selenocysteine have no unique codon? TIBS *13* (1988) 419–421.
27 Feng, Y., Hatfield, D. L., Rein, A., and Levin, J. G., Translational readthrough of the murine leukemia virus gag gene amber codon does not require virus-induced alteration of tRNA. J. Virol. *63* (1989) 2405–2410.
28 Flower, A. M., and Mchenry, C. S., The γ-subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. Proc. natl Acad. Sci. USA *87* (1990) 3713–3717.
29 Forchlammer, K., Leinfelder, W., and Bock, A., Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature *342* (1989) 453–456.
30 Förster, C., Ott, G., Forchhammer, K., and Sprinzl, M., Interaction of a selenocysteine-incorporating tRNA with elongation factor Tu from *E. coli*. Nucl. Acids Res. *18* (1990) 487–491.
31 Fox, T. D., Diverged genetic codes in protozoans and a bacterium. Nature *314* (1985) 132–133.
32 Fox, T. D., Natural variation in the genetic code. A. Rev. Genet. *21* (1987) 67–91.
33 Friedberg, E. C., and Weiss, W. A., Divergent genetic codes. Nature *325* (1987) 306.
34 Ghersa, P., Shrivastava, I. K., Perrin, L. H., Döbeli, H., Becherer, D. J., Matile, H., Meyer, B., and Certa, U., Initiation of translation at a UAG stop codon in the aldolase gene of *Plasmodium falciparum*. EMBO J. *9* (1990) 1645–1649.
35 Goelet, P., Lomonossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J., and Karn, J., Nucleotide sequence of tobacco mosaic virus RNA. Proc. natl Acad. Sci. USA *79* (1982) 5818–5822.
36 Goldschmidt-Clermont, M., and Rahire, M., Sequence, evolution and differential expression of two genes encoding variant small subunits of ribulose biphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. J. molec. Biol. *191* (1986) 421–432.
37 Grivell, L. A., Deciphering divergent codes. Nature *324* (1986) 109–110.
38 Gunderson, J. H., Elwood, H., Ingold, A., Kindle, K., and Sogin, M. L., Phylogenetic relationship between chlorophytes, chrysophites and oomycetes. Proc. natl Acad. Sci. USA *84* (1987) 5823–5827.
39 Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W., and Eisenman, R. N., A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell *52* (1988) 185–195.
40 Hanyu, N., Kuchino, Y., and Nishimura, S., Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two *Tetrahymena* tRNAsGln. EMBO J. *5* (1986) 1307–1311.
41 Harper, D. S., and Jahn, C. L., Differential use of termination codons in higher eucaryotes. Proc. natl Acad. Sci. USA *86* (1989) 3252–3256.
42 Hatfield, D., Suppression of termination codons in higher eucaryotes. TIBS *10* (1985) 201–204.
43 Hatfield, D., Diamond, A., and Dudock, B., Ipal suppressor serine tRNAs from bovine liver form phosphoseryl-tRNA. Proc. natl Acad. Sci. USA *79* (1982) 6215–6219.
44 Hatfield, D., and Oroszlan, S., The where, what and how of ribosomal frameshifting in retroviral protein synthesis. TIBS *15* (1990) 186–190.
45 Hatfield, D., and Rice, M., Aminoacyl-tRNA (anticodon): codon adaptation in human and rabbit reticulocytes. Biochem. Int. *13* (1986) 835–842.
46 Helftenbein, E., Nucleotide sequence of a macronuclear DNA molecule for α-tubulin from the ciliate *Stylonichia lemnae*. Nucl. Acids Res. *13* (1985) 415–433.
47 Herrick, G., Hunter, D., Williams, K., and Kotter, K., Alternate processing during development of a macronuclear chromosome family in *Oxytricha fallax*. Genes Develop. *1* (1987) 1047–1058.
48 Hieter, P., Pridmore, D., Hegemann, J. H., Thomas, M., Davis, R. W., and Philippsen, P., Functional selection and analysis of yeast centromeric DNA. Cell *42* (1985) 913–921.
49 Hinnebusch, A. G., Involvement of an initiation factor and protein phosphorylation in translational control of GCN4 mRNA. TIBS *15* (1990) 148–152.
50 Hinnebusch, A. G., Jackson, B. M., and Mueller, P. P., Evidence for regulation of reinitiation in translational control of GCN4 mRNA. Proc. natl Acad. Sci. USA *85* (1988) 7279–7283.
51 Horowitz, S., and Gorovsky, M. A., An unusual genetic code in nuclear genes of *Tetrahymena*. Proc. natl Acad. Sci. USA *82* (1985) 2452–2455.
52 Hosbach, H. A., and Kubli, E., Transfer RNA in aging *Drosophila*: II. Isoacceptor patterns. Mech. Aging Dev. *10* (1979) 141–149.
53 Huang, W. M., Ao, S., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R., Winge, D., and Fang, M., A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. Science *239* (1988) 1005–1012.
54 Hunter, W. N., Brown, T., Anand, N. N., and Kennard, O., Structure of an adenosine-cytosine base pair in DNA and its implications for mismatch repair. Nature *320* (1986) 552–555.
55 Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E., Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell *55* (1988) 447–458.
56 Jackson, R. J., Picornaviruses break the rules. Nature *334* (1988) 292–293.
57 Jang, S. K., Krausslich, H. G., Nicklin, M. J. H., Duke, G. M., Palmenberg, A. C., and Wimmer, E., A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J. Virol. *62* (1988) 2636–2643.
58 Jukes, T. H., Osawa, S., Muto, A., and Lehman, N., Evolution of anticodons: variations in the genetic code. Cold Spring Harbor Symp. *52* (1987) 769–776.
59 Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J., and Iwasaki, S., The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. Nature *341* (1989) 164–166.
60 Kersten, H., On the biological significance of modified nucleosides in tRNA. Prog. nucl. Acid Res. *31* (1984) 59–114.
61 Kirkwood, T. B. L., Rosenberger, R. F., and Galas, D. J., Accuracy in Molecular Processes. Its Control and Relevance to Living Organisms. Chapman and Hall, London 1986.
62 Kohli, J., and Grosjean, H., Use of three termination codons: compilation and analysis of the known eukaryotic and prokaryotic translation termination sequences. Molec. gen. Genet. *182* (1981) 430–439.
63 Kozak, M., An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucl. Acids Res. *15* (1987) 8125–8148.
64 Kozak, M., At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. molec. Biol. *196* (1987) 947–950.
65 Kozak, M., Bifunctional messenger RNAs in eukaryotes. Cell *47* (1986) 481–483.
66 Kozak, M., Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. Microbiol. Rev. *47* (1983) 1–45.
67 Kozak, M., Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Molec. cell. Biol. *9* (1989) 5073–5080.
68 Kozak, M., Effect of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Molec. cell. Biol. *7* (1987) 3438–3445.
69 Kozak, M., Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44* (1986) 283–292.
70 Kozak, M., Regulation of protein synthesis in virus-infected animal cells. Adv. Virus Res. *31* (1986) 229–292.
71 Kozak, M., The scanning model for translation: an update. J. Cell Biol. *108* (1989) 229–241.
72 Kozak, M., Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucl. Acids Res. *12* (1984) 3873–3893.

73 Kubli, E., Codon Usage and Q-Base Modification in *Drosophila melanogaster*. Chromatography and Modifications of Nucleosides. Part B. Elsevier, 1990.

74 Kuchino, Y., Beier, H., Akita, N., and Nishimura, S., Natural UAG suppressor glutamine tRNA is elevated in mouse cells infected with Moloney murine leukemia virus. Proc. natl Acad. Sci. USA *84* (1987) 2668–2672.

75 Kuchino, Y., Hanyu, N., Tashiro, F., and Nishimura, S., *Tetrahymena thermophila* tRNA and its gene that corresponds to UAA termination codon. Proc. natl Acad. Sci. USA *82* (1985) 4758–4762.

76 Lee, B. J., Rajagopalan, M., Kim, Y. S., Jacobson, K. B., and Hatfield, D., Selenocysteine tRNA(Ser)Sec gene is ubiquitous within the animal kingdom. Molec. cell. Biol. *10* (1990) 1940–1949.

77 Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T., and Hatfield, D. L., Identification of a selenocysteyl-tRNAser in mammalian cells that recognizes the nonsense codon, UGA. J. biol. Chem. *264* (1989) 9724–9727.

78 Lee, C. C., Craigen, W. J., Muzni, D. M., Harlow, E., and Caskey, C. T., Cloning and expression of a mammalian peptide chain release factor with sequence similarity to tryptophanyl-tRNA synthetases. Proc. natl Acad. Sci. USA *87* (1990) 3508–3512.

79 Leinfelder, W., Forchhammer, K., Veprek, B., Zehelein, E., and Böck, A., In vitro synthesis of selenocysteinyl-$tRNA_{UCA}$ from seryl-$tRNA_{UCA}$: involvement and characterization of the selD gene product. Proc. natl Acad. Sci. USA *87* (1990) 543–547.

80 Leinfelder, W., Stadtman, T. C., and Bock, A., Occurrence in vivo of selenocysteyl-tRNAser in *Escherichia coli*. J. biol. Chem. *264* (1989) 9720–9723.

81 Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A., and Bock, A., Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature *331* (1988) 723–725.

82 Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R., and Charnay, P., The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptional activator. Molec. cell. Biol., in press.

83 Liu, C. C., Simonsen, C., and Levinson, A. D., Initiation of translational internal AUG codons in mammalian cells. Nature *309* (1984) 82–85.

84 Lynn, D. H., and Sogin, M. L., Assessment of phylogenetic relationship among ciliated protists using partial ribosomal RNA sequences derived from reverse transcripts. BioSystems *21* (1988) 249, 254.

85 Mador, N., Panet, A., and Honigman, A., Translation of gag, pro and pol gene products of human T-cell leukemia virus type 2. J. Virol. *63* (1989) 2400–2404.

86 Martindale, D. W., Codon usage in *Tetrahymena* and other ciliates. J. Protozool. *36* (1989) 29–34.

87 Meerovitch, K., Pelletier, J., and Sonenberg, N., A cellular protein that binds to the 5′-noncoding region of poliovirus RNA: implications for internal translation initiation. Genes Develop. *3* (1989) 1026–1034.

88 Meier, F., Suter, B., Grosjean, H., Keith, G., and Kubli, E., Queuosine modification of the wobble base in tRNAHis influences in vivo decoding properties. EMBO J. *4* (1985) 823–827.

89 Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., and Kingsman, A. J., A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon. Nature *313* (1985) 243–246.

90 Meyer, E., Caron, F., and Guiard, B., Blocking of in vitro translation of *Paramecium messenger* RNAs is due to messenger RNA primary structure. Biochimie *66* (1984) 403–412.

91 Meyer, E., Caron, F., and Baroin, A., Macronuclear structure of the G surface antigen gene of *Paramecium primaurelia* and direct expression of its repeated epitopes in *Escherichia coli*. Molec. cell. Biol. *5* (1985) 2414–2422.

92 Miceli, C., La Terza, A., and Melli, M., Isolation and structural characterization of cDNA clones encoding the mating pheromone Er-1 secreted by the ciliate *Euplotes raikovi*. Proc. natl Acad. Sci. USA *86* (1989) 3016–3020.

93 Mueller, P. P., Jackson, B. M., Miller, P. F., and Hinnebusch, A. G., The first and fourth upstream open reading frame in GCN4 mRNA have similar initiation efficiencies but respond differently in translational control to changes in length and sequence. Molec. cell. Biol. *8* (1988) 5439–5447.

94 Nanney, D. L., Experimental ciliatology. Wiley, New York 1980.

95 Ninio, J., Divergence in the genetic code. Biochem. System. Ecol. *14* (1986) 455–457.

96 Ninio, J., Kinetic devices in protein synthesis, DNA replication and mismatch repair. Cold Spring Harbor Symp. *52* (1987) 639–645.

97 Nishimura, S., Structure, biosynthesis and function of queuosine in tRNA. Prog. nucl. Acid Res. *28* (1983) 49–73.

98 Normanly, J., and Abelson, J., tRNA identity. A. Rev. Biochem. *58* (1989) 1029–1049.

99 Osawa, S., and Jukes, T. H., Evolution of the genetic code as affected by anticodon content. TIG *4* (1988) 191–198.

100 Peabody, D. S., Translation initiation at an ACG triplet in mammalian cells. J. biol. Chem. *262* (1987) 11 847–11 851.

101 Pelham, H. R. B., Leaky UAG termination codon in tobacco mosaic virus. Nature *272* (1978) 469–471.

102 Pelletier, J., and Sonenberg, N., Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus. Nature *334* (1988) 320–325.

103 Perasso, R., Baroin, A., Qu, L., Bachellerie, J. P., and Adoutte, A., Origin of the algae. Nature *339* (1989) 142–144.

104 Prat, A., Conserved sequences flank variable tandem repeats in two alleles of the G surface protein of *Paramecium primaurelia*. J. molec. Biol. *211* (1990) 521–535.

105 Prat, A., Katinka, M., Caron, F., and Meyer, E., Nucleotide sequence of the *Paramecium primaurelia* G surface protein. A huge protein with a highly periodic structure. J. molec. Biol. *189* (1986) 47–60.

106 Prats, H., Kaghad, M., Prats, A. C., Klagsbrun, M., Lélias, J. M., Liauzun, P., Chalon, P., Tauber, J. P., Amalric, F., Smith, J. A., and Caput, D., High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. Proc. natl Acad. Sci. USA *86* (1989) 1836–1840.

107 Preer, J. R., Preer, L. B., Rudman, B. M., and Barnett, A. J., Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. Nature *314* (1985) 188–190.

108 Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L., and Dickerson, R. E., Helix geometry, hydration, and G-A mismatch in a B-DNA decamer. Science *238* (1987) 498–504.

109 Pure, G. A., Robinson, G. W., Naumovski, L., and Friedberg, E. C., Partial suppression of an ochre mutation in *Saccharomyces cerevisiae* by multicopy plasmids containing a normal yeast tRNAgln gene. J. molec. Biol. *183* (1985) 31–42.

110 Qu, L., Perasso, R., Baroin, A., Brugerolle, G., Bachellerie, J. P., and Adoutte, A., Molecular evolution of the 5′-terminal domain of large-subunit rRNA from lower eucaryotes. A broad phylogeny covering photosynthetic and non-photosynthetic protists. BioSystems *21* (1988) 203–208.

111 Rinaldy, A. R., Westhoff, P., Jauker, F., Seyfert, H. M., and Cleffmann, G., Properties of total and poly(A) + RNA from exponentially growing and from resting cultures of *Tetrahymena thermophila*. Exp. Cell Res. *134* (1981) 417–423.

112 Schlicht, H. J., Radziwill, G., and Schaller, H., Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core-polymerase fusion proteins. Cell *56* (1989) 85–92.

113 Schneider, S. U., Leible, M. B., and Yang, X., Strong homology between the small subunit of ribulose-1, 5-biphosphate carboxylase/oxygenase of two species of *Acetabularia*. Molec. gen. Genet. *218* (1989) 445–452.

114 Schoeman, R. L., and Schweiger, H. G., Gene expression in *Acetabularia*. II. Analysis of in vitro translation products. J. Cell Sci. *58* (1982) 35–48.

115 Schon, A., Bock, A., Ott, G., Sprinzl, M., and Söll, D., The selenocysteine-inserting opal suppressor tRNA from *E. coli* is highly unusual in structure and modification. Nucl. Acids Res. *17* (1989) 7159–7165.

116 Sedivy, J. M., Capone, J. P., RajBhandary, U. L., and Sharp, P. A., An inducible mammalian amber suppressor: propagation of a poliovirus mutant. Cell *50* (1987) 379–389.

117 Söll, D., Enter a new amino acid. Nature *331* (1988) 662–663.

118 Stadtman, T. C., Selenium-dependent enzymes. A. Rev. Biochem. *49* (1980) 93–110.

119 Stadtman, T. C., Specific occurrence of selenium in enzymes and amino acid tRNAs. FASEB J. *1* (1987) 375–379.

120 Syvanen, M., Cross-species gene transfer: a major factor in evolution? TIG *2* (1986) 63–66.

121 Trifonov, E. N., Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16S rRNA nucleotide sequences. J. molec. Biol. *194* (1987) 643–652.

122 Tsuchihashi, Z., and Kornberg, A., Translational frameshifting generates the γ subunit of DNA polymerase III holoenzyme. Proc. natl Acad. Sci. USA *87* (1990) 2516–2520.

123 Valle, R. P. C., Etude de la suppression naturelle de codons de terminaison chez les eucaryotes supérieurs : modèles viraux et synthé-

tiques, tRNA suppresseurs et mécanismes de régulation. Thèse de Doctorat d'Université Paris 1989.
124 Valle, R. P. C., and Haenni, A. L., Regulation of peptide chain termination in higher eucaryotes; in: Translation in Eucaryotes. Ed. Hans Trachsel. Tedford Press, Caldwell, N.J. 1990 in press.
125 Valle, R. P. C., and Morch, M. D., Stop making sense. FEBS Letters *235* (1988) 1–15.
126 Weiss, W. A., and Friedberg, E. C., Normal yeast $tRNA^{gln}$ can suppress amber codons and is encoded by an essential gene. J. molec. Biol. *192* (1986) 725–735.
127 Weissenbach, J., Dirheimer, G., Falcoff, R., Sanceau, J., and Falcoff, E., Yeast $tRNA^{leu}$ (anticodon UAG) translates all six leucine codons in extracts from interferon treated cells. FEBS Letters *82* (1977) 71–76.
128 Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J., HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell *55* (1988) 1159–1169.
129 Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J., and Kingsman, S. M., Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. Nucl. Acids Res. *14* (1986) 7001–7016.
130 Woese, C. R., The Genetic Code. Harper & Row, New York 1967.
131 Yamada, K., and Machida, H., Nippon Nogeikagaku Kaishi *36* (1962) 858–860.
132 Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., and Osawa, S., UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. natl Acad. Sci. USA *82* (1985) 2306–2309.
133 Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S., Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. natl Acad. Sci. USA *82* (1985) 1618–1622.
134 Zinoni, F., Birkman, A., Stadtman, T. C., and Bock, A., Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase from *Escherichia coli*. Proc. natl Acad. Sci. USA *83* (1986) 4650–4654.

0014-4754/90/11-12/1106-12$1.50 + 0.20/0

The genetic code in mitochondria and chloroplasts

T. H. Jukes and S. Osawa

Space Sciences Laboratory, University of California/Berkeley, 6701 San Pablo Avenue, Oakland (California 94608, USA) and Nagoya University, Laboratory of Molecular Genetics, Department of Biology, Nagoya 464 (Japan)

Summary. The universal genetic code is used without changes in chloroplasts and in mitochondria of green plants. Non-plant mitochondria use codes that include changes from the universal code. Chloroplasts use 31 anticodons in translating the code; a number smaller than that used by bacteria, because chloroplasts have eliminated 10 CNN anticodons that are found in bacteria. Green plant mitochondria (mt) obtain some tRNAs from the cytosol, and genes for some other tRNAs have been acquired from chloroplast DNA. The code in non-plant mt differs from the universal code in the following usages found in various organisms: UGA for Trp, AUA for Met, AGR for Ser and stop, AAA for Asn, CUN for Thr, and possibly UAA for Tyr. CGN codons are not used by *Torulopsis* yeast mt. Non-plant mt, e.g. in vertebrates, may use a minimum of 22 anticodons for complete translation of mRNA sequences. The following possible causes are regarded as contributing to changes in the non-plant mt: directional mutation pressure, genomic economization, changes in charging specificity of tRNAs, loss of release factor RF2, changes in RF1, changes in anticodons, loss of lysidine-forming enzyme system, and disappearance of codons from coding sequences.
Key words. Genetic code; mitochondria; evolution; organelles.

Introduction

Several differences from the universal genetic code have been discovered in mitochondria (mt) of organisms other than green plants. Chloroplasts and green plant mt have retained the universal code, but have undergone genomic economization so that some anticodons in chloroplasts have been discarded, and, in green plant mt, some tRNAs are imported from the cytosol, and some tRNA genes have been acquired from chloroplast DNA.

We have proposed[25] that evolutionary changes in the universal code must have taken place after a codon disappeared from coding sequences, because an abrupt change in assignment of a codon would be disruptive and therefore lethal. Disappearance of a codon, according to our proposal, is accompanied by a change in or loss of the corresponding anticodon. Such changes in the codon or anticodon may sometimes result from directional mutation pressure towards either AT or GC. The deleted codon reappears with a different assignment at locations in the coding sequence corresponding to the amino acid to which it has been reassigned. Examples of these proposed events will be described in this review, occurring in non-plant mt. Similar changes in nuclear codes are described by Jukes elsewhere in this issue of Experientia.

Evolution of mitochondria

The origin of mitochondria (mt) is inferred to be from endosymbiotic bacteria (reviewed by Küntzel and Köchel[15]). Comparison of 16S ribosomal RNA sequences led Yang et al.[37] to conclude that the endosym-