Genetic code 1990. Outlook

T. H. Jukes

Space Sciences Laboratory, University of California/Berkeley, 6701 San Pablo Avenue, Oakland (California 94608, USA)

Summary. The genetic code is evolving as shown by 9 departures from the universal code: 6 of them are in mitochondria and 3 are in nuclear codes. We propose that these changes are preceded by disappearance of a codon from coding sequences in mRNA of an organism or organelle. The function of the codon that disappears is taken by other, synonymous codons, so that there is no change in amino acid sequences of proteins. The deleted codon then reappears with a new function.
Wobble pairing between anticodons and codons has evolved, starting with a single UNN anticodon pairing with 4 codons. Directional mutation pressure affects codon usage and may produce codon reassignments, especially of stop codons. Selenocysteine is coded by UGA, which is also a stop codon, and this anomaly is discussed.
The outlook for discovery of more changes in the code is favorable, and open reading frames should be compared with actual sequential analyses of protein molecules in this search.
Key words. Genetic code; amino acid code; evolution of code; mitochondria; codon capture; directional mutation pressure; selenocysteine.

Introduction

A notable development has taken place in the genetic code within the past few years: the discovery of exceptions to the universal code. These were first found in mitochondria, and more recently in the nuclear codes of bacteria, a yeast, ciliated protozoa and algae[29–33]. In addition, selenocysteine is incorporated into a few proteins in vertebrates and bacteria by 'universal' stop codon UGA[26]. Changes in the universal code are shown in table 1.
Some of the changes have occurred twice in different lines of descent, and one change has been reversed. Codon AUA has been reconverted to isoleucine in mitochondria of echinoderms[4,17], although AUA codes for methionine in the mitochondria of other metazoa as a change from isoleucine in the universal code.
It is now evident that the code is not 'frozen', as was proposed in the frozen accident theory, which stated that the code had reached a point where no further changes could be tolerated[7]. Another theory had stated that the universal code represented a stereochemical relationship between an amino acid and its anticodon[22]. This idea is no longer tenable in view of the number of examples in which an amino acid has acquired a different anticodon during evolution.

Table 1. Changes in universal code

Codon in universal code	Changes to	System		Change caused by
		Nuclear	Mitochondrial	
UGA, Stop	Trp	Mycoplasma	All except plant	Anticodon CCA to UCA
AUA, Ile	Met		Yeast, metazoa, except echinoderms	tRNA (CAU) changed from pairing with AUG to AUR
	Met to Ile		Echinoderms	tRNA (CAU) changed from pairing with AUR to AUG anticodon GAU to IAU
AGR, Arg	Ser		Metazoa, except vertebrates	tRNA (GCU) changed from pairing with AGY to AGN
	Stop		Vertebrates	tRNA (GCU) changed from pairing with AGN to AGY
AAA, Lys	Asn		Flatworms, echinoderms	Anticodon GUU to IUU
				Anticodon UUU disappeared
UAR, Stop	Gln	*Acetabularia*		Anticodon UUG to UUA
		Ciliated protozoa except *Euplotes*		Gln codons CAR to UAR
CUN, Leu	Thr		Yeast	tRNA (UAG) changed from amino-acylating Leu to Thr
CUG, Leu	Ser	*Candida cylindracea*		tRNA (CAG) changed from amino-acylating Leu to Ser

The idea that changes in codon assignments in mitochondria were tolerated solely because of their small genome [15] was challenged by the finding of departures from the universal code in intact organisms (table 1).

There are at least two approaches to evolution of the code: First, that new assignments of codons to amino acids can take place abruptly. The second is that codon capture can result from disappearance of a codon from coding sequences, followed by its reappearance in a new role [30, 32].

Wobble rules

Codon-anticodon pairing was shown by Crick [6] to follow rules of wobble pairing between the first anticodon base and the third base of codons: G in the first anticodon position paired with C and U, U with A and G, C only with G. I (inosine) paired with U, C and A. A in the first anticodon position was rare, and paired with U.

The universal code table conformed to these rules, because G paired with both U and C, so that GNN anticodons paired with both NNU and NNC codons and therefore all NNU and NNC codon pairs are synonymous. Most NNA and NNG codon pairs are also synonymous. The exceptions are AUA (isoleucine) and AUG (methionine), and UGA (stop) and UGG (tryptophan). The wobble rules provide for this, because anticodons CNN pair only with codons NNG, so that anticodon CUA (methionine) pairs with AUG and anticodon CCA (tryptophan) pairs with UGG.

The discovery of the mitochondrial code led to reexamination of the wobble rules [15], because, in family boxes, a single UNN anticodon, U unmodified, paired with four codons by 'four-way' wobble. On the other hand, modification of U was necessary to restrict its pairing in mitochondrial two-codon sets in boxes where GNN anticodons exist. These pair with NNY two-codon sets. This modification of U in mitochondria is still being investigated and has not been identified. The wobble rules may have evolved by a series of steps as shown in table 2. These steps would provide for expansion of the code from 15 to 20 amino acids.This expansion would include emergence of anticodon GUA capturing codons UAU and UAC for tyrosine (if these were formerly stop codons). The anticodons were increased in number by the appearance of CNN anticodons, pairing strongly with NNG codons. Inosine (I) was introduced into anticodons, pairing with U, C and A in family boxes. Seven INN anticodons are in eukaryotes and one (ICG) in bacteria.

Most of the additions to the 1966 wobble rules have resulted from studies by Nishimura and his colleagues. Yokoyama and co-workers [48] divided modified uridines into two groups. The first group has uridines in the first anticodon position modified to 5-methyl-2-thiouridine derivatives (2-thiolated U, table 2). These primarily recognize A as the third codon base and the recognition of G is much less efficient. Anticodons for glutamine, lysine and glutamic acid contain this modification of U. Interestingly, Stadtman [43] has found that when selenium is substituted for sulfur in this modification of U, the recognition of G is facilitated, in the case of lysine UUU and glutamic acid UUC anticodons. The other group of modified uridines are 5-hydroxyuridine derivatives, found in *E. coli* and *B. subtilis* (table 2), in family boxes for valine, serine, threonine and alanine, and pairing with U, A and G.

Lysidine, an unusual modification of C, pairing with A in isoleucine codon AUA, is found in eubacterial anticodon *CAU (C = 2-lysyl C, lysidine) [27].

Codon-anticodon pairing in mitochondrial protein synthesis is discussed in the article by Dr R. P. Martin. As he points out, some mt tRNAs have short anticodon loops, and 'unconventional features are found in mt tRNAs'.

Codon capture

Codon capture can account for the changes in the code (table 1) according to the present hypothesis, which postulates that codon capture can take place in at least three different ways. These postulations are:

1) *Mutation in an anticodon.* Anticodon CCA in the universal code mutated to UCA by duplication of the tRNA gene for tryptophan, and CCA to UCA mutation in one of the duplicates. UCA pairs with both codons UGG and UGA. UGA in stop codon sites mutated to UAA, and UGA was captured by tryptophan, pairing with anticodon UCA, which also pairs with UGG [30].

   Anticodon UUU for lysine codon AAA disappeared in echinoderm mt, together with codon AAA, because all lysine AAA codons mutated to AAG pairing with anticodon CUU. Later, anticodon GUU for asparagine codons AAY changed to (probably) IUU, pairing with AAY and AAA. Asparagine was then coded by AAA, reappearing by mutation of AAY codons [29].

   Anticodon UUG in ciliates [21] (except *Euplotes*) and *Acetabularia* [36] duplicated [32], and one duplicate mutated to UUA. Universal stop codons UAA and UAG are not used by these organisms, which use UGA as their only stop codon. Some CAA and CAG gluta-

Table 2. Evolution of anticodon-codon pairing

1) UNN paired with 4 codons, NNU, NNC, NNA, NNG in 15 family boxes, except UAN. This would provide for 15 amino acids and 4 stop codons UAN.

2) Two-codon sets became possible when UNN duplicated to GNN, pairing with NNU and NNC, and *UNN pairing with NNA and NNG. *U = modified U. Code could expand to 20 amino acids, including tyrosine UAY.

3) *UNN duplicated to form **UNN, pairing mostly with NNA, and CNN, pairing only with NNG. **U is thiolated.

4) INN formed by GNN→ANN→INN mainly in eukaryotes. INN pairs with NNU, NNC, NNA.

mine codons mutated to UAA and UAG, pairing with 'new' glutamine anticodon UUA. Later, some UUA anticodons mutated to CUA, pairing strongly with UAG, glutamine, in *Tetrahymena*[13,21].

It is important to note that high AT in DNA does not by itself bring about codon reassignments such as have occurred in *Tetrahymena, Stylonychia, Paramecium* and *Acetabularia*. There must also be disappearance of stop codons UAA and UAG, and acquisition of sole chain termination function by UGA. This makes UAA and UAG available for capture. This turn of events has not occurred in *Euplotes*. Similarly, in *Mycoplasma*, reassignment of UGA to tryptophan could not take place until UGA stop codons mutated to UAA, and anticodon UCA appeared in tRNA Trp.

2) *Change to a different amino acid in aminoacylation of a tRNA molecule.* In yeast mt, tRNA with anticodon UAG (leucine) changed to aminoacylation by threonine, so that CUN codons became codons for threonine instead of leucine[33].

   In the yeast *Candida cylindracea*, tRNA with anticodon CAG (leucine) changed to aminoacylation by serine, so that CUG codes for serine instead of leucine[23].

3) *Changes in codon-anticodon pairing in mitochondria.* This has taken place in mitochondrial tRNAs for 2 different amino acids: tRNA serine (GCU) changed from pairing only with AGY codons to pairing with AGN codons (table 1). UCU anticodon disappeared. Later, anticodon GCU reverted to pairing only with AGY codons and AGR became stop codons in vertebrate mt[31].

   Anticodon CUA (methionine) in yeast and in metazoa (except echinoderms) pairs with AUG and, anomalously, also with AUA (methionine). In echinoderms, this pairing has reverted to pairing solely between CAU and AUG, methionine, and AUA has become an isoleucine codon, pairing with an isoleucine anticodon, probably IAU, which also pairs with AUU and AUC[4].

   These hypothesized changes all provide for disappearance of a codon followed by its reappearance with a different assignment. In almost every case, the amino acid sequences of the proteins involved are unchanged, so the reassignments are non-disruptive, but sometimes sequences are changed temporarily (yeast mt leucine to threonine and *Candida cylindracea* leucine to serine, see below).

Directional mutation pressure and codon capture

AT pressure has evidently been effective in capture of UGA by tryptophan in *Mycoplasma* through the changes CCA to UCA in the anticodon and UGA stop codon to UAA. It is possible that a reversal of these changes by GC pressure took place during evolution of the code, so that anticodon UCA in an earlier code was converted to CCA, UGA tryptophan codons to UGG, and some UAA stop codons to UGA, thus introducing the first use of UGA for chain termination. Apparently, AT pressure, by changing glutamine anticodon UUG to UUA in ciliates and *Acetabularia*, enabled UAA and UAG, unused as stop codons, to be captured by glutamine. The change of AAA from lysine to asparagine is well explained by GC pressure converting all AAA lysine codons to AAG, followed by disappearance of anticodon UUU (see above). Acquisition of CUN by threonine and CUG by serine could have been made possible by AT pressure, converting CUN leucine codons to UUR, and CUG to CUA. *Candida cylindracea* and its related species show wide differences in genomic GC content, helping to explain changes in directional mutation pressure.

Most arginine residues coded by AGR in yeast or *Trypanosoma* mt are replaced by CGN in metazoan mt. A few of those AGR residues were replaced by codons for other amino acids, not including serine. AGR became unassigned because of deletion of tRNA Arg (UCU). Changes in serine tRNA GCU then enabled its anticodon to pair with AGR, and some AGR codons reappeared as serine codons, by mutations of AGY and other codons to AGR[31]. Later, however, in vertebrate mt, these codons apparently mutated back to AGY, and AGR codons once again disappeared. A few reappeared in terminal sites as stop codons[31]. The alternation of AUA between isoleucine and methionine has resulted from changes in anticodon-codon pairing. In the universal code, isoleucine codon AUA pairs with *CAU (*C = lysidine) in bacteria and with IAU in eukaryotes. If *CAU is changed to CAU, it pairs only with AUG, methionine, and AUA would disappear, together with its tRNA. AUA isoleucine codons would mutate to AUY. AUA could reappear as a methionine codon if the anticodon CAU changed its pairing properties to pair with both AUG and AUA, as in metazoan mitochondria, except echinoderms (see above).

RNA editing

This subject is reviewed in detail in the chapter by Roberto Cattaneo, who defines RNA editing as 'a process leading to pre-determined modifications of the coding region of a primary gene transcript'. Some comments are added here on the type of RNA editing found in plant mitochondria. The discovery of RNA editing in plant mitochondria, of both monocots and dicots, by several groups of investigators[5,12,16] has abruptly revised an earlier conclusion that CGG was a tryptophan codon in these organelles. The finding also explains why a tryptophan anticodon CCG could not be found in plant mt; it presumably does not exist. Its absence would likely prevent translation of unedited plant mt mRNA.

The findings are summarized in table 3, which shows that all but one of the editing changes so far reported consist of C to U. Whether these are by deamination or replace-

Table 3. Codons of certain sites in gene for COXIII in mitochondria (mt) of various species compared with edited sites in mRNA of wheat and evening primrose mt

Site	26	56	57	87	129	184	188	196	207	213	228	235	250
DNA NEU	TTT	GGA	TGA	TGA	TGA	GCA	ATA	TTA	TTA	TTT	TGT	ATG	TGA
DNA YSC	TTT	TCT	TGA	TGA	TGA	GCT	ATT	TTA	TTA	TTA	TGT	ATG	TGA
DNA BOV	TTC	–	–	TGA	TGA	TCT	TTA	CTA	CTA	ATA	TGC	ATA	TGA
DNA HUM	CTA	–	–	TGA	TGA	TCA	TTG	TTA	CTA	ACG	TGT	ATG	GGG
DNA WHT	TCT	TCA	CGG	CGG	CGG	CCC	CCT	TCA	TCA	TCG	CGT	ACG	TGG
RNA WHT	UUU	UUA	UGG	UGG	UUG	UCC	CUU	UUA	UUA	UUG	UGU	AUG	UGG
'Consensus'	Phe	–	Trp	Trp	Trp	–	Leu, Ile	Leu	Leu	–	Cys	Met	Trp
DNA OBE	TCT	TCA	CGG	CGG	CGG	CCT	CCT	TCA	TTA	TCG	TGT	ATG	CGG
RNA ONE						UCU	CUU	UUA		UUG			UGG

Abbreviations: NEU = *Neurospora*, YSC = *Saccharomyces cerevisiae*, BOV = bovine; HUM = human; WHT = wheat; OBE = evening primrose. Underlined codons have been edited to identify with consensus.

ment has not been reported. Hiesel and co-workers[16] found six C to U changes in the non-coding sequence 3' to the gene for COXIII in *Oenothera*. They also found[39] a coding U to C change, isoleucine to threonine in the cytochrome *b* gene. The editing found in plant mt RNA includes arginine to tryptophan codons, serine to leucine and phenylalanine, leucine to phenylalanine, proline to serine, leucine and phenylalanine, histidine to tyrosine, arginine to cysteine, and threonine to isoleucine and methionine.

Additional findings on editing in genes of *Oenothera* mt have been reported by Schuster et al.[37–39] as follows: (all, except one, are C to U changes) *nad 3* gene, Ser to Leu (twice), Ser to Phe (three times), Pro to Leu (three times), Pro to Ser, Pro to Phe (2-base change), Leu to Phe, Arg to Trp and 2 silent changes: ACC to ACU and TCC to UCU. In the cytochrome *b* gene, Pro to Leu, Pro to Ser, Leu to Phe, His to Tyr (five times) Thr to Ile, Arg to TRP and one silent change, CTC to CUU. The exception from C to U was Ile to Thr, T to C[39].

There were two Pro to Leu changes in an unidentified reading frame. Of interest was the finding that no editing was found in a 34-codon sequence of an *rps* pseudogene. Gualberto et al.[12a] examined gene COXIII in wheat mt and found 12 C to U conversions in the coding region, including Leu to Phe, Ser to Phe (five times) Pro to Leu (four times) Ser to Leu, Arg to Trp and one U to C (silent) Thr to Thr.

An obvious question is whether RNA editing occurs in 'non-plant' mitochondria. In cases where a transfer RNA and anticodon have been identified, corresponding to the code change, the changes cannot be ascribed to RNA editing. This is the case with anticodon UCA (tryptophan), pairing with UGA in mammalian mt. Yeast mt have threonine tRNA with anticodon UAG, pairing with CUN, and also the 'regular' threonine tRNA with anticodon UGU pairing with ACN. Echinoderm mt have an anticodon for asparagine that pairs with AAA, so in none of these cases can editing be expected to occur.

Cattaneo (this issue) discusses several other types of RNA editing, especially in trypanosome mitochondria. RNA editing is a field that is actively being developed.

Will the real twenty-first amino acid please stand up?

Selenomethionine (SeMet) can replace methionine in protein synthesis. It is introduced into polypeptide formation by internal methionine tRNA, but it has no codon in the universal code, and hence it is not coded in DNA. Its presence in polypeptide chains as a partial replacement for methionine is randomized, apparently depending on the relative populations of tRNA Met and tRNA SeMet, and for these reasons it does not qualify as the 21st amino acid.

Selenocysteine (SeCys) has been nominated[41] as the 21st amino acid. It does not become attached to a tRNA molecule, but instead is formed by selenolation of an unusual seryl-tRNA. The failure of SeCys to charge a tRNA molecule may be due to the high reactivity of the HSe-group, in contrast to that of the CH_3Se-group. SeCys is coded by UGA in specific sites of a few enzymes, but UGA is a stop codon in the universal code and is a tryptophan codon in all mitochondrial codes except those of plant mt in which it is a stop codon. UGA also codes for tryptophan in *Mycoplasma*. The claims for SeCys as the 21st amino acid are therefore shaky. Only a very few proteins contain SeCys. We give SeCys a 20.5 rating.

Organisms that are exposed to selenium, especially plants in seleniferous soils, contain various compounds of selenium, in some of which selenium has replaced sulfur. Many of the compounds are quite labile. A review, 82 pages, gave descriptions of these earlier findings[28]. As investigations continued, SeMet was found incorporated into proteins by the same steps as in the case of 'internal' methionine, including aminoacylation of methionyl-tRNA by SeMet, and peptide synthesis in which the seleno amino acid analog replaces methionine in the completed protein molecule (Stadtman[43]). Hartmanis and Stadtman[14] isolated a selenium-containing protein, thiolase,

from *Clostridium kluyveri*. The selenium in the protein was identified as selenomethionine. It was shown that selenomethionine residues were distributed randomly throughout the polypeptide subunits. This non-specific replacement of methionine by small amounts of selenomethionine had no observable effect on the catalytic activity of the enzyme. Stadtman[43] pointed out that '... in this enzyme the nonspecific substitution of a sulfur amino acid by its selenium analog is limited to methionine. No selenocysteine was detected in numerous preparations of thiolase from *C. kluyveri* in spite of the fact that there are 16 cysteine residues in the 160-kDa native protein'. This absence of selenocysteine in thiolase was no doubt because SeCys is incorporated in proteins only by the special procedure explained above.

The requirement for SeCys is not necessarily absolute, because Zinoni et al.[49] found that SeCys in formate dehydrogenase of *E. coli* could be replaced by cysteine, and the enzyme retained about 10% of its activity. Even more notable was their finding that this enzyme in *Methanobacterium formicicum* contained cysteine at the position (residue 132) corresponding to SeCys in the *E. coli* enzyme (residue 140) when the sequences were aligned.

Selenocysteine and anaerobic life

Selenocysteine (SeCys) occurs in specific positions in the polypeptide chains of several enzymes, including glycine reductase, formate dehydrogenase and hydrogenase in bacteria and glutathione reductase of mammals and chickens[43]. SeCys is inserted by a tRNA that pairs with codon UGA through its anticodon UCA[26]. Only the UGA codons at the specific sites in mRNAs of these enzymes are subject to the insertions. The tRNA initially is charged with phosphoserine and this becomes pretranslationally converted to SeCys by an enzymatic modification that exchanges the hydroxy group of serine for a selenol group[26].

Leinfelder et al.[26] have suggested that 'UGA was originally a serine codon' (presumably for SeCys) and 'after introduction of oxygen into the biosphere this highly oxidizable amino acid could be maintained only in anaerobic organisms or in aerobic systems which evolved special protective mechanisms'. In the aerobic world codon UGA could have 'acquired other functions such as its more familiar role in termination'.

UGA could not have changed abruptly from coding for SeCys to a stop codon, because this would have caused widespread disruption in amino acid sequences of proteins. We have proposed[30] that UGA was a tryptophan codon that became a stop codon when GC pressure changed anticodon UCA to CAA, pairing only with codon UGG. This left codon UGA unassigned, and it became a stop codon as a result of mutations of UAA to UGA. The following series of events is possible: In the anaerobic world, the UGN family box in the code was assigned to both Cys and SeCys, anticodon UCA. UGN coded indiscriminately for both SeCys and Cys, depending on their relative abundance. This is analogous to the randomization of selenomethionine in place of methionine residues in proteins of various organs, depending on the levels of selenium in the growth media and diets of the organisms when methionine is being synthesized. SeCys, however, under such circumstances, does not enter proteins randomly, and occurs only at low levels as a component of a few selenium-dependent enzymes. When oxygen entered the biosphere, nearly all the SeCys was oxidized to Cys, which retained UGN codons and anticodon UCA. This anticodon duplicated, and one duplicate mutated to GCA, the present cysteine anticodon, pairing with UGY. The other UCA anticodon was captured by a 'new' amino acid, tryptophan, with codons UGR. GC pressure changed anticodon UCA to CCA, pairing only with UGG, the present tryptophan codon, and UGA disappeared, except for rare use in coding for SeCys. It reappeared as a mutant of stop codon UAA in chain-termination locations.

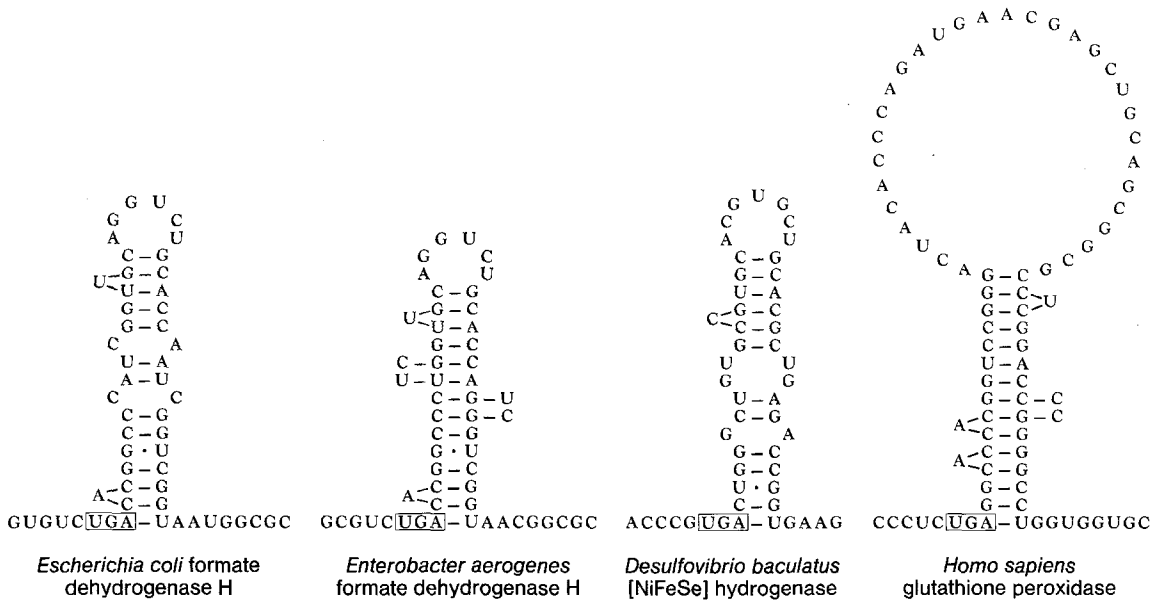
SeCys in enzymes in perhaps a vestigial reminder of the anaerobic world of 2 or 3 billion years ago. The enzymes containing SeCys all function at very low redox potentials.

A rule is usually followed that no codon can have 2 different meanings simultaneously, or the amino acid sequences of proteins would became hopelessly scrambled. Every rule may have its exception. UGA is used simultaneously as a codon for selenocysteine and as a stop codon in bacteria and in vertebrates. The reason for this anomaly has been discovered by Zinoni et al.[50]. The *fdHF* gene for the selenopolypeptide subunit of formate dehydrogenase (*E. coli*) has a TGA codon which codes for selenocysteine at amino acid position 140. Reading of this codon depends on the presence and identity of at least 40 bases of *fdHF* mRNA downstream from the UGA codon. Apparently these bases form a hairpin loop next to the UGA codon (fig.). Deletions from this loop drastically reduced UGA translation, and deletions that destroyed the hairpin structure completely prevented this translation, in which a special protein, the product of the Sel beta gene, participates[10].

In addition, selenocysteine incorporation also was found when the UGA codon at position 140 was replaced by cysteine codons UGU or UGC. Therefore selenocysteinyl tRNA could compete with cysteinyl tRNAs for pairing at this position, in addition to competing for UGA with chain termination (release factor 2). The authors point out that the mRNA context around UGA-140 forces selenocysteine incorporation into this position.

Role of nucleotide 37 in codon-anticodon pairing

Nucleotide 37, adjoining the third anticodon position, is often extensively modified, and the modifications may be



Potential secondary structures in the vicinity of UGA codons in mRNAs coding for selenocysteine in proteins. From Zinoni et al.[49].

concerned with strengthening A · U and U · A pairing between the third anticodon nucleotide (position 36) and the first codon nucleotide[20]. In contrast, the tRNA nucleotide at position 33 is an invariant unmodified U[42]. Presumably an unmodified U would be the nucleotide least apt to interfere with wobble pairing by the adjacent first anticodon nucleotide (position 34).

In the case of pairing between U at position 36 (anticodon position 3) and A in codon position 1 (3-1 pairing), the modification of nucleotide 37 is usually a bulky threonyl-containing side chain on position 6 of adenine. Such side chains may prevent, by steric hindrance, wobble-pairing between U in the third anticodon position and G in the first codon position. Such prevention of U · G pairing would be essential to fidelity of translation of all codons starting with A except initiator methionine, which is sometimes coded in bacteria by GUG, and it is significant that initiator methionine tRNA in bacteria has unmodified A at position 37 in all 14 sequences listed in the *Nucleic Acids Research* compilation[42] plus 3 in chloroplasts. Unmodified A at 37 would presumably facilitate U · G pairing at 36.

In 3-1 pairing between A and U, the most common modification at position 37 is a large isopentenyl side chain attached to adenine. An even more bulky group, the 'Y-base' occupies position 37 in eukaryotic phenylalanine tRNAs for 12 of the 17 species in which this nucleotide has been identified; the other 5 have 1-methylguanosine at this position.

Codons starting with U must not pair with anticodons ending with G. These are anticodons for leucine (CUN), proline, histidine, glutamine and arginine (CGN). Most nucleotides in position 37 in tRNAs for these amino acids are methyl-G. Evidently there is no 3-1 pairing between G and U but only between G and C. Position 37 is often unmodified following anticodons ending in C, but in other cases it is methylated.

UAG as an initiation codon: Varshney and RajBhandary[44] mutated the initiator codon of *E. coli* chloramphenicol acetyltransferase gene (CAcT) from AUG to UAG. They found that mRNA derived from the mutant gene was translated in the presence of a mutant *E. coli* initiator tRNA with its anticodon changed from CAU to CUA. They added glutaminyl-tRNA synthetase. The initial amino acid in the polypeptide was likely glutamine rather than methionine. Evidently the change in the anticodon to CUA caused the tRNA to be aminoacylated with glutamine instead of methionine. Possibly anticodon CUA is the recognition site for glutamine in this tRNA. Anticodon CUA is not present in the universal code, but it occurs in *Tetrahymena* in tRNA Gln (UAG). *Tetrahymena* and some other ciliated protozoa use codons UAG and UAA, (in addition to CAA and CAG) for glutamine[13,21]. The findings by Varshney and Raj Bhandary indicate that the initiator tRNA in *E. coli* that is bound to the peptide site on the 30S ribosomal unit decides the process of initiation even when its anticodon is CUA rather than CAU, and even when the amino acid is glutamine rather than methionine. The authors also found the same pathway was used efficiently in vivo: transformants of *E. coli* initiated protein synthesis with the mutated genes for CAcT and for the tRNA (CUA). They record observations that the mutant tRNA, aminoacylated with glutamine, can be formylated enzymatically to formylglutamine. It would be of interest to determine whether aminoacylation of tRNA glutamine (UAG) of *Tetrahymena*[13] was dependent on the nucleotide sequence in the tRNA with anticodon CUA. Gher-

sa et al.[11] found that the gene and mRNA in *Plasmodium falciparum* for fructose-1,6-diphosphate aldolase lacked a functional AUG start codon, and, instead, a UAG codon was recognized as the start signal of protein synthesis in vivo and in vitro. UAG was also used for chain termination.

Context of stop codons in prokaryotes

Brown et al.[3] examined sequences around stop codons in 862 *E. coli* genes. They noted a preference for NAR codons immediately preceding UAA, and for uridine following all three stop codons. They suggest that for efficient chain termination in *E. coli* the stop signal may be a tetranucleotide, especially UAAU, recognized by release factor 2.

Future possibilities

It has been assumed by many investigators that open reading frames in DNA may be used to predict the amino acid sequences of proteins that in many cases are not yet known. This method of 'sequencing proteins' will fail to detect changes in the code, as shown strikingly by Kawaguchi et al.[23]. They found that a known protein, lipase I of *Candida cylindracea*, differed in leucine and serine content from the levels predicted by the sequence of its gene as found in cDNA, constructed from a mutation that produced a large amount of lipase 1. By this means, Kawaguchi et al.[23] showed CUG coded for serine in this gene and, presumably, in all proteins in *C. cylindracea*.

Which codons can not undergo changes in meaning? Probably NNU or NNC codons cannot, unless they are members of a family box that has only one anticodon (UNN). With this exception, NNY codons are translated by a single GNN anticodon, so that if either NNU or NNC disappears, GNN will be retained to translate the remaining NNY codon, and this will ensure that the deleted codon, if it reappears, will be translated to the same amino acid as before. Also, codons AUG and UGG cannot disappear unless a second codon first acquires the ability to translate methionine or tryptophan. This has actually happened in *Mycoplasma*, in which anticodon UCA pairs with both UGA and UGG, so that UGG is apparently disappearing[47].

Several NNA codons seem to have disappeared from use in *Micrococcus luteus*, as discussed by S. Osawa in Chapter 4. UUA, CUA, AUA, GUA, CAA and AGA were not found in a total of 6814 codons, and tRNAs with anticodons (*UNN) pairing with these codons were not detected. In effect, *M. luteus* uses only 58 codons, and 10 of these codons are almost completely unused. If *M. luteus* changed from high GC to low GC usage, it is conceivable that many of the undetected codons could be captured by amino acids.

In contrast *Mycoplasma capricolum* had no CUC or CGG, and Osawa et al. point out that anticodon CCG has become unnecessary and has disappeared, so that CGG has become an unassigned codon.

Evolution of the genetic code history

The idea that the code has evolved attracts many authors, as shown by numerous publications during the past 28 years. The concept of a doublet code interested Roberts[35] and E. L. Smith[40]. The first studies of amino acid incorporation in vitro were made with templates of polyuridine containing low levels of other nucleotides. It seemed from these results that all codons contained at least one uridine, and therefore that there was always one silent U per codon. Soon it was shown that synthetic polyribonucleotides not containing uridine could also code for polypeptide formation. Eck[9] proposed assignments for all 64 of the triplet codons, and the 'doublet code' with a silent U was no longer considered to be feasible. In any case, a doublet code, as pointed out by Dounce[8], could furnish only 16 pairs, insufficient for 20 amino acids, and an evolutionary change from a doublet to a triplet code would abolish all the previous genetic information, as emphasized by Crick[7].

Another possibility was that a doublet code could have existed with the third base, N, having no specificity. This was attractive because, in the 'family boxes' (4-codon sets) of the universal code, 8 amino acids are coded on this basis. Perhaps, therefore, an earlier code could have consisted entirely of family boxes, with a maximum of 15 amino acids, assuming four stop codons in one box. Introduction of additional 'new' amino acids could take place by dividing a family box into two pairs of codons, one for each of 2 amino acids. This was proposed by Jukes[18,19]. Such a step could not take place until a procedure had evolved that would distinguish between UNN anticodons with four-way wobble pairing, and *UNN (*U = modified U) anticodons that paired only with NNA and NNG codons. The NNY codons for a different amino acid in the same box would be translated by anticodon GNN[15].

The caveat in discussing expansions of the code is not to speculate as to which were the 'new' amino acids, because nature is inscrutable in such matters. Nevertheless, various authors have hypothesized extensively as to which were the primitive amino acids that were in early codes and which are the newcomers. Wong has drawn on metabolic pathways in making such proposals. He[45] postulates that the genetic code initially contained only the small number of amino acids that were formed most readily in the primordial soup. (This assumes the existence of such a soup, containing 'paleokaryotes'). He then says 'these amino acids later served as precursors for the formation of other amino acids along prehistoric pathways which became enzymatized with the amino acid biosynthesis pathways of present day organisms'. This proposal is esthetically and biochemically pleasing. It cannot be verified, and leaves us with the problem of

disruption of proteins resulting from wholesale introduction of 'new' amino acids, so it must be assumed that these introductions were acceptable. In support of such introductions, Wong[46] notes that 4-fluoro tryptophan could enter a strain of *Bacillus subtilis* by serial mutation, carried out by 'feeding' the tryptophan analog to the organism at a high level. It is known from other examples that an amino acid can be slightly modified, as in the case of selenomethionine, so that it will still charge the cognate tRNA, and be used in protein synthesis. Other examples of 'non-protein' amino acids that can be acylated by tRNAs and can be incorporated into proteins in vivo are norleucine, which can charge methionyl tRNA, and canavanine, which can charge arginyl-tRNA[1,2,25,34]. Norleucine and canavanine are not normally found in proteins, but human epidermal growth factor containing norleucine has been prepared from *E. coli*[25]. An antibiotic, furanomycin, is produced by *Streptomyces* L-803. Its side chain has a 5-membered heterocyclic ring with a double bond. Furanomycin competes with isoleucine in charging isoleucine-tRNA in *E. coli* and is incorporated into proteins[24]. The perturbation in such proteins caused by furanomycin incorporation is probably responsible for their cytotoxicity. These examples of occasional introduction of a 'non-protein' amino acid differ from the drastic introduction of an amino acid (produced by metabolism of another amino acid) into the genetic code throughout all proteins by capturing an existing codon.

Summary. The outlook for discovery of further changes in the genetic code is good. The codon capture hypothesis explains how such changes can take place, without altering the amino acid sequences of proteins, by disappearance of a codon from mRNA and its reappearance with a different assignment.

Acknowledgments. I acknowledge the advice of Syozo Osawa and David Collins, the collaboration of Roxanne Short and Carol Fegté in preparation of the manuscript and support by NIH grant GM 40903 01 to the University of California, Berkeley.

1 Anfinsen, C. B., and Corley, L. G., An active variant of staphylococcal nuclease containing norleucine in place of methionine. J. biol. Chem. *244* (1969) 5149–5152.
2 Attias, J., Schlesinger, J. J., and Schlesinger, S., The effect of amino acid analogues on alkaline phosphatase formation in *Escherichia coli* K-12. J. biol. Chem. *244* (1969) 3810–3817.
3 Brown, C., Stockwell, P., Trotman, C., and Tate, W., The signal for the termination of protein synthesis in procaryotes. Nucleic Acids Res. *18* (1990) 2079.
4 Cantatore, P., Roberti, M., Rainaldi, G., Gadaleta, M. N., and Saccone, C., The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. biol. Chem. *264* (1989) 10965–10975.
5 Covello, P. S., and Gray, M. W., RNA editing in plant mitochondria. Nature *341* (1989) 662–666.
6 Crick, F. H. C., Codon-anticodon pairing: The wobble hypothesis. J. molec. Biol. *19* (1966) 548–555.
7 Crick, F. H. C., The origin of the genetic code. J. molec. Biol. *38* (1968) 367–379.
8 Dounce, A. L., Enzymologia *15* (1952) 251.
9 Eck, R. V., Science *140* (1963) 477.
10 Forchhammer, K., Leinfelder, W., and Böck, A., Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature *342* (1989) 453–456.
11 Ghersa, P., Shrivastava, I., Perrin, L., Dobeli, H., Becherer, D., Matile, H., Meier, 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.
12 Gualberto, J. M., Lamattina, L., Bonnard, G., Weil, J-H., and Grienenberger, J-M., RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature *341* (1989) 660-662.
12a Gualberto, J. M., Weil, J., and Grienenberger, J., Editing of the wheat COXIII transcript: evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. Nucleic Acids Res. *18* (1990) 3771–3773.
13 Hanyu, N., Kuchino, Y., Nishimura, S., and Beier, H., Dramatic events in ciliate evolution: Alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two *Tetrahymea* tRNAs Gln. EMBO J. *5* (1986) 1307–1311.
14 Hartmanis, M. G. N., and Stadtman, T. C., Isolation of a selenium-containing thiolase from *Clostridium kluyveri*: identification of the selenium moiety as selenomethionine. Proc. natl Acad. Sci. USA *79* (1982) 4912–4916.
15 Heckman, J. E., Sarnoff, J., Alzner-de Weerd, 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.
16 Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A., RNA editing in plant mitochondria. Science *246* (1989) 1632–1634.
17 Jacobs, H. T., Elliot, D., Math, V. V., and Farquharson, A., Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. molec. Biol. *202* (1988) 185–217.
18 Jukes, T. H., Coding triplets and their possible evolutionary implications. Biochem. biophys. Res. Commun. *19* (1965) 391–396.
19 Jukes, T. H., Molecules and Evolution, p. 284. Columbia University Press, New York 1966.
20 Jukes, T. H., Possibilities for the evolution of the genetic code from a preceding form. Nature *246* (1973) 22–26.
21 Jukes, T. H., Osawa, S., and Muto, A., Divergence and directional mutation pressures. Nature *325* (1987) 668.
22 Jungck, J. R., The genetic code as a periodic table. J. molec. Evol. *11* (1978) 211–224.
23 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.
24 Kohno, T., Kohda, D., Haruki, M., Yokoyama, S., and Miyazawa, T., Nonprotein amino acid furanomycin, unlike isoleucine in chemical structure, is charged to isoleucine tRNA by isoleucyl-tRNA synthetase and incorporated into protein. J. biol. Chem. *265* (1990) 6931–6935.
25 Koide, H., Yokoyama, S., Kawai, G., Ha, J-M., Oka, T., Kawai, S., Miyake, T., Fuwa, T., and Miyazawa, T., Biosynthesis of a protein containing a nonprotein amino acid by *Escherichia coli*: L-2-aminohexanoic acid at position 21 in human epidermal growth factor. Proc. natl Acad. Sci. USA *85* (1988) 6237–6241.
26 Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M-A., and Böck, A., Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature *331* (1988) 723–725.
27 Muramatsu, T., Yokoyama, S., Hirose, N., Matsuda, A., Ueda, T., Yamaizumi, Z., Kuchino, Y., Nishimura, S., and Miyazawa, T., A novel lysine-substituted nucleotide in the first position of the anticodon of amino isoleucine tRNA from *Escherichia coli*. J. biol. Chem. *263* (1988) 9261–9267.
28 National Research Council, Chapt. 5, Biological Effects, in: Selenium, pp. 51–133. National Academy of Sciences, 1976.
29 Ohama, T., Osawa, S., Watanabe, K., and Jukes, T. H., Evolution of the mitochondrial genetic code IV. AAA as an asparagine codon in some animal mitochondria. J. molec. Evol. *30* (1990) 329–332.
30 Osawa, S., and Jukes, T. H., Evolution of the genetic code as affected by anticodon content. Trends Genet. *4* (1988) 191–198.
31 Osawa, S., Ohama, T., Jukes, T. H., and Watanabe, K., Evolution of the mitochondrial genetic code I. Origin of AGR serine and stop codons in metazoan mitochondria. J. molec. Evol. *29* (1989) 202–207.
32 Osawa, S., and Jukes, T. H., Codon reassignment (codon capture) in evolution. J. molec. Evol. *28* (1989) 271–278.
33 Osawa, S., Collins, D., Ohama, T., Jukes, T. H., and Watanabe, K., Evolution of the mitochondrial genetic code III. Reassignment of

CUN codons from leucine to threonine during evolution of yeast mitochondria. J. molec. Evol. *30* (1990) 322–328.
34 Pines, M., Rosenthal, G. A., and Applebaum, S. W., In vitro incorporation of L-canavanine into vitellogenin of the fat body of the migratory locust *Locusta migratoria migratorioides*. Proc. natl Acad. Sci. USA *78* (1981) 5480–5483.
35 Roberts, R. B., Further implications of the doublet code. Biochemistry *48* (1962) 1245–1250.
36 Schneider, S. U., Leible, M. B., and Yang, X-P., Strong homology between the small subunit of ribulose-1,5-biphosphate carboxylase/oxygenase of two species of *Acetabularia* and the occurrence of unusual codon usage. Molec. gen. Genet. *218* (1989) 445–452.
37 Schuster, W., Unseld, M., Wissinger, B., and Brennicke, A., Ribosomal protein S14 transcripts are edited in *Oenothera* mitochondria. Nucl. Acids Res. *18* (1990).
38 Schuster, W., Wissinger, B., Unseld, M., and Brennicke, A., Transcripts of the NADH-dehydrogenase subunit 3 gene are differentially edited in *Oenothera* mitochondria. EMBO J. *9* (1990) 263–269.
39 Schuster, W., Hiesel, R., Wissinger, B., and Brennicke, A., RNA editing in the cytochrome *b* locus of the higher plant *Oenothera berteriana* includes a U-to-C transition. Molec. cell. Biol. *10* (1990) 2428–2431.
40 Smith, E. L., Nucleotide base coding and amino acid replacements in proteins. Biochemistry *48* (1962) 677–684.
41 Söll, D., Enter a new amino acid. Nature *331* (1988) 662–663.
42 Sprinzl, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R., Compilation of tRNA sequences and sequences of tRNA genes. Nucl. Acids Res., Suppl. to vol. *17* (1989) 1–173.
43 Stadtman, T. C., Specific occurrence of selenium in enzymes and amino acid tRNAs. FASEB J. *1* (1987) 375–379.
44 Varshney, U., and RajBhandary, U., Initiation of protein synthesis from a termination codon. Proc. natl Acad. Sci. USA *87* (1990) 1586–1590.
45 Wong, J. T-F., The evolution of a universal genetic code. Proc. natl Acad. Sci. USA *73* (1976) 2336–2340.
46 Wong, J. T-F., Evolution of the genetic code. Microbiol. Sci. *5* (1988) 174–181.
47 Yamao, F., Iwagami, S., Azumi, Y, Muto, A., and Osawa, S., Evolutionary dynamics of tryptophan tRNAs in *Mycoplasma capricolum*. Molec. gen. Genet. *212* (1988) 364–369.
48 Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., and Miyazawa, T., Molecular mechanism of codon recognition by tRNAs with modified uridine in the first position of the anticodon. Proc. natl Acad. Sci. USA *82* (1985) 4905–4909.
49 Zinoni, F., Birkmann, A., Leinfelder, W., and Böck, A., Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. Proc. natl Acad. Sci. USA *84* (1987) 3156–3160.
50 Zinoni, F., Heider, J., and Böck, A., Features of the formate-dehydrogenase mRNA necessary for decoding of the UGA codon as cysteine. Proc. natl Acad. Sci. USA *87* (1990) 4660–4664.

0014-4754/90/11-12/1149-09$1.50 + 0.20/0

Hypoxia: On the borderline between physiology and pathophysiology

The Editors wish to thank Prof. Christian Bauer for coordinating this multi-author review.

Hypoxia: On the borderline between physiology and pathophysiology. A foreword

C. Bauer

Physiologisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich (Switzerland)

Molecular oxygen was not present when life began on Earth. However, within about 1 billion years, photosynthetic organisms appeared, and within a relatively short time, oxygen, a by-product of photosynthesis, became a dominant chemical entity in the atmosphere. Oxygen was then used to produce chemical energy in the form of adenosine triphosphate (ATP) in specific cellular compartments, the mitochondria. A *lack of oxygen* (hypoxia) is not very easy to define, because hypoxia may mean different things to different organs and organisms. Quite a number of physiological processes are actually geared to maintaining a *constant flow of oxygen*, for example to the heart, the skeletal muscles or the brain. In other words, biological systems have a number of protective devices that actually *prevent* the occurrence of harmful levels of hypoxia. These devices function as regulated and regulatory elements in the chain of transporters that carry oxygen from the environment to the mitochondria. The papers that comprise the present multi-author review are based on three symposia that were held in conjunction with the annual meeting of the Swiss Union of Biological Sciences that took place in Fribourg in 1989. The subjects that were dealt with embraced a broad spectrum of hypoxia-sensitive systems in man and other mammalian species. It was hoped that by putting these interdisciplinary contributions together, a scientific fermentation would occur which might not otherwise have arisen.

The transport of oxygen from the environment to the final consumer, the mitochondrion, involves the following forces: a) convectional transport of O_2 in the lung and in the blood; b) diffusional transport from the lung alveoli into the blood capillaries and out of the tissue capillaries to the mitochondria. It is obvious that a shortage of oxygen, i.e. a mismatch between oxygen supply and oxygen demand, can occur at all levels of this oxygen transport chain. One of the very first cellular structures that would sense such a diminished oxygen supply is, of course, the mitochondrion itself. Wilson discusses and reviews in his contribution the possible role of the *mitochondria* as *'oxygen sensors'* and concludes that in the *physiological range* of *oxygen pressures*, the oxygen-de-