RECENT ADVANCES IN STUDIES OF EVOLUTIONARY RELATIONSHIPS BETWEEN PROTEINS AND NUCLEIC ACIDS*

THOMAS H. JUKES

University of California, Space Sciences Laboratory, Berkeley, Calif., U.S.A.

(Received 15 May, 1968)

Abstract. Evolution depends upon the occurrence of occasional changes, large or small, in hereditary characteristics. Molecular genetics gave rise to the new field of molecular evolution, which is currently exploring the changes that take place in proteins and nucleic acids over long periods of time. The following are some of the fundamental assumptions:

(1) The phenotypic characteristics of organisms depend directly on proteins.

(2) Proteins are synthesized in accordance with information carried in molecules of DNA as sequences of the four bases, adenine, guanine, cytosine, and thymine. The information is transcribed into molecules of messenger RNA and is translated into proteins by the intervention of the genetic code.

(3) Changes in the composition of the base sequences in DNA can take place in living organisms, and these changes can affect the phenotypic characteristics of the next generation.

(4) The process of natural selection favors the perpetuation of organisms which compete successfully in the struggle for existence. This process leads to the elimination of all but a small fraction of the astronomical number of possible protein molecules that could result from genetic translation of the possible variants of DNA. Furthermore, the number of protein molecules was originally much smaller than it is to-day, and it has increased by hereditary processes rather than by the chance appearance of entirely new proteins.

(5) The DNA present in any single cell contains the complete information for all the hereditary characteristics of the organism. The amount of DNA per cell may increase during evolution and this increase has produced modern organisms that are 'higher', more specialized, and more complex, from earlier and simpler forms.

(6) Protein molecules are slowly and steadily differentiated during evolution if their genes are physically separated from each other, by allopatric speciation or even by duplication and translocation, whether or not the functions of the proteins are changed.

(7) Mutations, together with recombination, contribute to changes in the genetic pool which provide the variability within populations that is necessary for evolution of species.

The field of molecular evolution should include a theory of the chemical events leading to the formation of the first living organism from molecules of non-living origin.

The genetic code may have evolved through multiplication of transfer RNA molecules by gene duplication followed by differentiation. This proposal is supported by the similarities between all tRNA molecules of known structures.

The DNA of higher organisms contains families of repetitive sequences. The families may contain thousands or hundreds of thousands of individual members. The 'family resemblance' within each group grows less with the passage of time because this leads to differentiation resulting from the accumulation of point mutations.

1. Introduction

The diversity of living organisms has been a subject of interest to human beings for many years. At one time it was customary to ascribe the origin of the so-called lower forms of life to spontaneous generation. This theory was destroyed by the experiments

* Supported by NASA Grant NsG 479 to the University of California, Berkeley.

of Spallanzani and Pasteur, but the theme of a special creation continued in many circles to be reserved for man himself. About one hundred years ago, the publication of the *Origin of Species* led to wide acceptance of the idea that all life had a common origin; this was mentioned even in 'The Mikado' by Gilbert and Sullivan. The concept of divergent evolution has been reinforced by an increasing body of evidence, including taxonomy and systematics, ecology, paleontology, biochemistry, Mendelian genetics, and mutations. It leads back to a single organism. We shall explore this concept in terms of deoxyribonucleic acid (DNA).

The formula for DNA proposed by WATSON and CRICK (1953) showed that the molecular basis of heredity rests on the sequences of four bases; adenine, cytosine, guanine and thymine (A, C, G and T) in DNA molecules. This provides a direct key to the molecular mechanism of evolution. Since all hereditary information is carried as linear sequences of four variables, evolution can take place only through changes in such sequences, consisting of repetitive changes, molecular shortening, and replacement of one base with another in DNA. These alterations produce phenotypic changes in the organism. Most of these changes occur in proteins. The net result is change in the fitness of the organism for its environment. This leads to natural selection: the emergence of some species, and the extinction of others.

The new field of molecular evolution measures the evolutionary changes in DNA molecules by comparing their base sequences. One method for comparing them is by the annealing procedure, in which the ability of single strands of DNA from two different species to form hybrids is quantitatively measured. Another is the indirect method of determining and comparing the amino acid sequences in proteins. In a very few cases, it has been possible to analyze and directly compare the base sequences in molecules of RNA.

It is fortunately possible, although the procedures are laborious, to determine the amino acid sequences of proteins. The genetic code enables the sequences to be translated back into the base sequences in DNA that code for the proteins. There are some limitations to this procedure because of certain ambiguities in the code. However, the method is sufficiently accurate to provide a lot of information on the evolution of proteins. Such information is deduced from comparisons of proteins that have identical or analogous functions, and similar but not identical sequences. Such proteins are obtained in most cases from different species of organisms.

To re-emphasize the preceding, the information for heredity and hence for evolution is carried as a linear sequence of four variables; the bases in the two strands of DNA molecules.

Changes in the length and ordering of the sequence are responsible for all inherited changes, and hence for the *genotypic* component of evolution.

The other two controlling forces in evolution are *phenotypic expression*, which is the direct result of translation of the information in the sequence, and *natural selection*, which is exerted on the phenotype.

So far, the DNA model has been completely adequate for the study of genetics

and evolution. There is no need to propose any other model until the DNA system has been found wanting. The chances of this are vanishingly small.

DNA is essentially inert. It functions solely as a repository for information. Its physical and physico-chemical properties are of the greatest interest and importance, because they have a bearing on the manner in which it is replicated and transcribed, but they are not properties that we associate with living organisms. The *transcription* of DNA is defined as the production of complementary RNA copies of one of the two DNA strands. All transcription is carried out by the action of an enzyme, RNA polymerase. The length of the single-stranded RNA molecules that are produced by transcription is controlled by *start* and *stop* signals on the DNA molecules. Presumably these signals, like all other information in DNA, are linear permutations with specific sequences. Recent evidence is that the *start* signal consists of a short sequence of C's or, at least, a C-rich cluster (SUMMERS and SZYBALSKI, 1968).

There is no evidence that modified bases in DNA, such as methyl cytosine in plant DNA, are needed for coding purposes.

2. RNA and Its Functions

The coding properties of DNA are studied by analyzing the structure and function of RNA. Three main functional classes of RNA are recognized: messenger, transfer and ribosomal. Messenger RNA consist of long single strands of RNA, containing unmodified bases. Their function is to specify amino acid sequences in polypeptides. The code in messenger RNA molecules (Table I) consists entirely, or almost entirely, of 64 codons, each consisting of one of the 64 three-letter permutations of A, C, G and U (uracil).

TABLE I

The genetic code

UUU Phenylalanine	CUU Leucine	AUU Isoleucine	GUU Valine
UUC Phenylalanine	CUC Leucine	AUC Isoleucine	GUC Valine
UUA Leucine	CUA Leucine	AUA Isoleucine	GUA Valine
UUG Leucine	CUG Leucine	AUG Methionine	GUG Valine
UCU Serine	CCU Proline	ACU Threonine	GCU Alanine
UCC Serine	CCC Proline	ACC Threonine	GCC Alanine
UCA Serine	CCA Proline	ACA Threonine	GCA Alanine
UCG Serine	CCG Proline	ACG Threonine	GCG Alanine
UAU Tyrosine	CAU Histidine	AAU Asparagine	GAU Aspartic acid
UAC Tyrosine	CAC Histidine	AAC Asparagine	GAC Aspartic acid
UAA Chain Termn.	CAA Glutamine	AAA Lysine	GAA Glutamic acid
UAG Chain Termn.	CAG Glutamine	AAG Lysine	GAG Glutamic acid
UGU Cysteine	CGU Arginine	AGU Serine	GGU Glycine
UGC Cysteine	CGC Arginine	AGC Serine	GGC Glycine
UGA Chain Termn.	CGA Arginine	AGA Arginine	GGA Glycine
UGG Tryptophan	CGG Arginine	AGG Arginine	GGG Glycine

The transfer RNA (tRNA) molecules are more complex than those of messenger RNA. They double back on themselves to form helical regions and loops. One of the loops contains an anticodon of 3 consecutive bases, which pairs with a messenger RNA codon. It is possible that there are not more than 55 anticodons because of ambiguous pairing in the third base position of the codons. Eight anticodons have been identified by analyzing tRNA molecules.

There are about 75 to 85 nucleotides in a tRNA molecule. The molecule is first formed as a sequence of A, C, G and U, following which, some of these bases are changed by enzymes into modified bases, such as dimethyl guanine, etc. TRNA molecules have a distinctive and complex three-dimensional structure. Each tRNA molecule contains at least four regions with specific functions: first, the anticodon; next, the sequence TΨCG or TΨCA, which may bind tRNA to a ribosome; third, an unidentified sequence that is specific for the enzyme which attaches one, and only one, of the 20 amino acids to the end of the molecule; and, fourth, the terminal – CCA for the attachment of the amino acid, which is the first step in translation of the code.

We must conclude that the short length of the tRNA molecule and its base composition enable the cell and its enzyme systems to distinguish it from other RNA molecules, such as those of messenger RNA.

The third class of RNA molecules are the ribosomal RNAs. These molecules probably do not contain information for protein synthesis. The two larger ribosomal RNA molecules contain some modified bases, but the smaller, third one, does not. The function of ribosomal RNA is to combine with several specific proteins to form a ribosome, which is about 20 to 30 mμ in diameter. Proteins are synthesized on the surface of ribosomes, and ribosomes participate in the pairing reaction between codons in messenger RNA and anticodons in transfer RNA. This pairing leads to the selection of a specific sequence of amino acids by a messenger RNA molecule. This sequence is coupled together by peptide linkages to form a protein such as insulin, hemoglobin or casein. There are thousands of different proteins in each of the higher species of living organisms.

3. Hydrogen Bonding

Hydrogen bonding in the nucleic acids was discovered by WATSON and CRICK (1953). They proposed that two hydrogen bonds between each AT and GC pair held the two strands of DNA molecules together. Later, PAULING and COREY (1956) pointed out that the GC pair had 3 hydrogen bonds. Hydrogen-bonded strands of DNA and RNA in helices always run in opposite directions as defined by the sugar-phosphate linkage in the nucleic acids.

Hydrogen bonding in RNA is usually by means of AU and GC pairs. As a result of the nature of the RNA polymerase reaction, RNA occurs in single strands and these can double back on each other to form two-stranded helical regions, as in transfer RNA.

Hydrogen bonding takes place between the codon of messenger RNA and the anticodon of transfer RNA in the selection of amino acids in protein synthesis. The results of this hydrogen bonding reaction are included in a group of relationships which is termed the amino acid code.

It was proposed by CRICK (1966) that codon-anticodon pairing includes a special type of hydrogen bonding, as follows: the first two bases of the codon and the second two bases of the anticodon pair according to the Watson-Crick rule, but there is a certain amount of play or 'wobble' between the third base of the codon and the first base of the anticodon. The pairing in this scheme is shown in Table II.

TABLE II

Codon-Anticodon pairing: the wobble hypothesis*

Anticodon –I–G–C– → ϟ ||| ||| –C–C–G– ← –U– –A– Codon	Anticodon –G–U–A– → ϟ || || –C–A–U– ← –U– Codon
Anticodon –C–A–A– → ||| || || –G–U–U– ← Codon	Anticodon –U–U–U– → ϟ || || –A–A–A– ← –G– Codon

* I = inosine.

This complex pattern is at the root of the translation of the genetic code. Apparently, it is Nature's device for apportioning the codes for 20 amino acids among the 64 possible permutations of the 4 bases in messenger RNA and for providing two anticodons to translate each codon. It is not known whether transfer RNAs carrying all of the possible anticodons are present in all organisms; in fact, it may well be that some are missing from some species. The 'wobble' hypothesis implies that anticodons with A in the first position do not occur, and so are not used in the translation of the genetic message. Presumably, the A occurring in this position is deaminated to hypoxanthine by a specific enzyme during the modification of transfer RNAs subsequently to their transcription. Hypoxanthine riboside is termed inosine, and hypoxanthine pairs preferentially with C. However, in the 'wobble' position it pairs with U or A as well as C. Table III shows the known, indicated and predicted anticodons.

4. Protein Homology and Evolution

In studying evolution, the genetic code is used to compare the amino acid sequences in two homologous proteins. An example would be a comparison between the hemo-

TABLE III

The known (bold face), indicated (italics) and predicted anticodons

	IAG Leu	**IAU** Ile	**IAC** Val
GmAA Phe	GAG Leu	*GAU* Ile	*GAC* Val
UAA Leu	UAG Leu		*UAC* Val
CAA Leu	*CAG* Leu	**CAU** Met	CAC Val
IGA Ser	IGG Pro	*IGU* Thr	**IGC** Ala
GGA Ser	*GGG* Pro	GGU Thr	GGC Ala
UGA Ser	UGG Pro	UGU Thr	*UGC* Ala
CGA Ser	CGG Pro	CGU Thr	CGC Ala
GΨA Tyr	GUG His	GUU Asn	GUC Asp
UUA End	UUG Gln	*UUU* Lys	UUC Glu
CUA End	CUG Gln	*CUU* Lys	CUC Glu
	ICG Arg		ICC Gly
GCA Cys	GCG Arg	*GCU* Ser	*GCC* Gly
	UCG Arg	*UCU* Arg	*UCC* Gly
CCA Trp	*CCG* Arg	*CCU* Arg	*CCC* Gly

The known anticodons have been identified in tRNA molecules; the indicated anticodons have been identified by binding of tRNA on ribosomes with polynucleotides of known sequence and the remainder of the anticodons are predicted by the wobble hypothesis.

globins of a fish and a mammal. A hemoglobin molecule contains two different polypeptide chains, each with about 140 to 150 amino acids. The α chains of human and carp hemoglobin are identical in about half of the amino acids. It is also possible to compare the hemoglobins in a single species, for there are several different hemoglobins in each vertebrate species, such as human beings, and the chains may be compared with each other to study the manner in which the globins (hemoglobins and myoglobins) have diverged from a common origin, which is termed an archetypal gene.

TABLE IV

Amino acid sequences in portions of globin chains

	79														94
M	–Lys–	Lys–	Gly–	His–	His–	Glu–	Ile	–Glu–	Leu–	Lys–	Pro–	Leu–	Ala–	Gln–	Ser –His–
α	–His–	Val–	Asp–	Asp–	Met–	Pro–	Asn–	Ala–	Leu–	Ser	–Ala–	Leu–	Ser–	Asp–	Leu–His–
β	–His–	Leu–	Asp–	Asn–	Leu–	Lys–	Gly–	Thr–	Phe–	Ala	–Thr–	Leu–	Ser–	Glu–	Leu–His–
γ	–His–	Leu–	Asp–	Asp–	Leu–	Lys–	Gly–	Thr–	Phe–	Ala	–Gln–	Leu–	Ser–	Gly–	Leu–His–

The numbering is from the beginning of all four chains, including gaps.
M = myoglobin; α = α hemoglobin; β = β hemoglobin; γ = γ hemoglobin.

An example of this procedure is shown in Table IV, in which segments of four different globin peptide chains, each corresponding to about 11% of a hemoglobin chain, are compared with each other. The first segment is taken from 'muscle hemoglobin', or myoglobin, the next from the α chain of hemoglobin; the third from the

TABLE V

The sequences in the globin chains in Table IV written in the form of their genes

	79	94
M	A–A–R–A–A–R–G–G–N–C–A–Y–C–A–Y–G–A–R–A–T–Y–G–A–R–T–T–R–A–A–R–C–C–N–C–T–N–G–C–N–C–A–R–T–C–N–C–A–Y	
α	C–A–Y–G–T–N–G–A–Y–G–A–Y–A–T–G–C–C–N–A–A–Y–G–C–N–T–T–R–T–C–N–G–C–N–C–T–N–T–C–N–G–A–Y–T–T–R–C–A–Y	
β	C–A–Y–T–T–R–G–A–Y–A–A–Y–T–T–R–A–A–R–G–G–N–A–C–N–T–T–Y–G–C–N–A–C–N–C–T–N–T–C–N–G–A–R–T–T–R–C–A–Y	
γ	C–A–Y–T–T–R–G–A–Y–G–A–Y–T–T–R–A–A–R–G–G–N–A–C–N–T–T–Y–G–C–N–C–A–R–C–T–N–T–C–N–G–G–N–T–T–R–C–A–Y	

Minimum Differences: M:α 20 α:β 12 β:γ 4 N = A, C, G or T;
M:β 20 α:γ 13 R = A or G;
M:γ 22 Y = C or T.

β chain and the fourth from the γ chain. All these chains are present in every human being; the γ chain being present prior to birth.

All four of the globin chains are structurally and chemically related; they are of almost equal lengths, and they coil themselves into convoluted globular molecules of almost exactly the same shape, as shown by crystallography. Each of them holds in its folds a unit of heme, the oxygen-carrying pigment. A striking difference is found, however, in the sequences of amino acids. Myoglobin differs from each of the other three at about 75% of the amino acid sites. The α chain differs from the β and γ chains by about 55%. Finally, the β and γ chains resemble each other more closely than they do either of the other two chains, and differ by only 27%. This state of affairs is just what would be expected if all four of the polypeptide chains had a common ancestor from which they have diverged by a series of widely spaced events of separation followed by differentiation. The mechanism for this is straightforward in terms of DNA and the genetic code. In Table V the same sequences of amino acids are shown in terms of possible codons representing them, just as they would be found in the regions of the DNA molecule that are responsible for carrying the genetic information for the globins. The minimum base differences are also shown. Table VI

TABLE VI

Minimum differences per codon in the relationship between certain globin chains

Comparison	Sites Compared	Minimum Difference				
		None	1	2	3	Average
Myoglobin: α human	140	35	57	47	1	1.10
Myoglobin: β human	144	36	53	54	1	1.14
Myoglobin: γ human	144	36	57	50	1	1.11
*Lamprey: α human	100	38	40	22	0	0.84
α human: γ human	139	59	55	25	0	0.76
α human: β human	139	64	53	22	0	0.70
α human: α carp	140	73	42	25	0	0.66
β human: γ human	146	106	29	10	0	0.34
β human: β horse	146	120	18	8	0	0.23
α human: α horse	142	126	12	5	0	0.16
β human: δ human	146	136	9	1	0	0.08
β human: β *M. mulatta*	146	139	6	1	0	0.05

* The complete sequence of lamprey hemoglobin is not known.

shows that the β and γ chains separated from each other subsequently to the time when human beings and fishes had a common ancestor, but before the lines of descent separated that led to human beings and horses, in which the β chains differ by 0.23 minimum base differences per codon.

All the homologous proteins that have been so far studied show relationships that are quite analogous to the one set forth for the hemoglobins. These include, among other, the cytochromes *c*, the bacterial and plant ferredoxins, the insulins, the fibrinopeptides, and pancreatic ribonuclease. The complete amino acid sequences of

spinach and *Scenedesmus* ferredoxins (MATSUBARA *et al.*, 1967; SUGENO and MATSUBARA, 1968) are in Table VII. These differ by 0.41 minimum base differences per codon. The differences have accumulated during a period of about 350 million years of divergent evolution in the plant kingdom. A comparison of three bacterial ferredoxins is in Table VIII.

Some sets of homologous proteins have diverged from each other more rapidly than other sets during evolution. A possible explanation for this is that some proteins are subject to greater constraints than others in the relation of their structures to their functions. Such constraints will lead to rigorous natural selection and the rejection of numerous deleterious mutants by lethality. The net result would be that the protein would appear to evolve more slowly. An alternative explanation for the same result might be that some regions of the chromosomes are better protected than others against mutations.

It is often necessary to postulate the presence of 'gaps' in aligning homologous sequences. The criteria for this are explained by CANTOR (1968).

5. Hybridization of DNA as a Measure of Evolution

MARMUR and LANE (1960) showed that after separating two strands of DNA, obtained from *Diplococcus pneumoniae*, by heating, it was possible to reunite them by slowly cooling the mixture. This necessitated the pieces of the strands aligning in exactly the same manner as that in which they were originally held together; for the randomization of the sequence of bases in a bacterial DNA molecule is such that, with few exceptions, each sequence is unique when measured over distance of about 15 or more base pairs. Each region must therefore 'find' its original partner before it can reunite. It was also found that a single strand of DNA from one organism could form a double strand with fragments of single-stranded DNA of a closely related organism (SCHILDKRAUT *et al.*, 1962). The hybrid double strand did not bind together as firmly as did a pair of single strands from the same organism. Many refinements of this procedure were subsequently introduced by Bolton, McCarthy, Hoyer, Britten and Kohne (BOLTON and MCCARTHY, 1962; HOYER *et al.*, 1963, 1964; MCCARTHY and BOLTON, 1963, 1964; MCCARTHY and HOYER, 1964; BRITTEN and KOHNE, 1965–66). An extensive literature has accumulated in which the quantitative nature of binding between single strands of DNA from different organisms has been used as a measure of taxonomy and evolution. The results are quite concordant with those obtained by other procedures, including comparisons of amino acid sequences in homologous proteins, and also comparisons made by classical methods of taxonomy and systematics. Two new findings have emerged: one is that the DNA of higher organisms contains regions of highly repetitive sequences, some of which are found to have been duplicated 100000 or more times, and the other is that some portions of DNA, such as the portion that carries the information for ribosomal RNA, have differentiated much more slowly during evolution than have other regions that presumably code for proteins. The studies with DNA lead to the same conclusion

TABLE VII

Sequences of Two Type 1 (Chloroplast Type) Ferredoxins

Sp NH_2Ala–Ala–Tyr–Lys–Val–Thr–Leu–Val–Thr–Pro–Thr–Gly–Asn–Val–Glu–Phe–Gln–Cys–Pro–Asp–Asp–Val–Tyr–Ile–Leu–Asp–Ala–Ala–
1 2 1 1 2 2 1 1 2

Sc NH_2Ala–Thr–Tyr–Lys–Val–Thr–Leu–Lys–Thr–Pro–Ser–Gly–Asp–Gln–Thr–Ile–Glu–Cys–Pro–Asp–Asp–Thr–Tyr–Ile–Leu–Asp–Ala–Ala–

Sp –Glu–Glu–Glu–Gly–Ile–Asp–Leu–Pro–Tyr–Ser–Cys–Arg–Ala–Gly–Ser–Cys–Ser–Ser–Cys–Ala–Gly–Lys–Leu–Lys–Thr–Gly–Ser–Leu–
1 1 1 1 1 1 1 1 1

Sc –Glu–Glu–Ala–Gly–Leu–Asp–Leu–Pro–Tyr–Ser–Cys–Arg–Ala–Gly–Ala–Cys–Ser–Ser–Cys–Ala–Gly–Lys–Val–Glu–Ala–Gly–Thr–Val–

Sp –Asn–Gln–Asp–Asp–Gln–Ser–Phe–Leu–Asp–Asp–Asp–Gln–Ile–Asp–Glu–Gly–Trp–Val–Leu–Thr–Cys–Ala–Ala–Tyr–Pro–Val–Ser–Asp–
1 2 2 1 1 2 1 2

Sc –Asp–Gln–Ser (Asp, Gln) Ser–Phe–Leu–Asp–Asp–Ser–Gln–Met–Asp–Gly–Gly–Phe–Val–Leu–Thr–Cys–Val–Ala–Tyr–Pro–Thr–Ser–Asp–

Sp –Val–Thr–Ile–Glu–Thr–His–Lys–Glu–Glu–Glu–Leu–Thr–AlaCOOH
2 1 1 2

Sc –Cys–Thr–Ile–Ala–Thr–His–Lys–Glu–Glu–Asp–Leu–Phe–COOH

Identical sites 67; Single-base differences 19; Two-base differences 10; Minimum base differences per codon (MBDC) 0.41.

Sp = Spinach (MATSUBARA *et al.*, 1967)
Sc = Scenedesmus (SUGENO and MATSUBARA, 1968)

TABLE VIII

Bacterial Ferredoxins: Primary Structures and Base Differences per Codon in Internal Duplication

(i) *Clostridium pasteurianum* (a):

```
                                                          *
 1       *          *        *  10      *   *      *          *  20   *      *             *  29
 Ala–Tyr–Lys–Ile– – –Ala–Asp–Ser–Cys–Val–Ser–Cys–Gly–Ala–Cys–Ala–Ser–Glu–Cys–Pro–Val–Asn–Ala–Ile–Ser–Gln–Gly–Asp–Ser–
 30                             40                                   50                               58
–Ile–Phe–Val–Ile–Asp–Ala–Asp–Thr–Cys–Ile–Asp–Cys–Gly–Asn–Cys–Ala–Asn–Val–Cys–Pro–Val–Gly–Ala–Pro–Val–Gln–Glu        Ave.
 2   1   2   0       0   0   1   0   1   2   0   0   2   0   0   1   1   0   0   0   2   0   2   2   0   1          0.77
```

(ii) *Clostridium butyricum* (b):

```
 Ala–Phe–Val–Ile– – –Asn–Asp–Ser–Cys–Val–Ser–Cys–Gly–Ala–Cys–Ala–Gly–Glu–Cys–Pro–Val–Ser–Ala–Ile–Thr–Gln–Gly–Asp–Thr
–Gln–Phe–Val–Ile–Asp–Ala–Asp–Thr–Cys–Ile–Asp–Cys–Gly–Asn–Cys–Ala–Asn–Val–Cys–Pro–Val–Gly–Ala–Pro–Asn–Gln–Glu
 2   0   0   0       2   0   1   0   1   2   0   0   2   0   0   2   1   0   0   0   1   0   2   1   0   1          0.69
```

(iii) *Micrococcus aerogenes* (c):

```
 Ala–Tyr–Val–Ile– – –Asn–Asp–Ser–Cys–Ile–Ala–Cys–Gly–Ala–Cys (Lys, Pro, Glu, Cys, Pro, Val, Asn, – , Ile, Gln, Gln,Gly) – Ser
–Ile–Tyr–Ala–Ile–Asp–Ala–Asp–Ser–Cys–Ile–Asp–Cys–Gly–Ser–Cys–Ala–Ser–Val–Cys–Pro–Val–Gly–Ala–Pro–Asn–Pro–Glu–Asp
 2   0   1   0       2   0   0   0   0   1   0   0   1   0   2   1   1   0   0   0   2   0   2   2   1   1          0.77
```

Prototype:

```
1                               10                                  20                               29
Ala–Phe–Val–Ile–Asp–Ala–Asp–Ser–Cys–Ile–Asp–Cys–Gly–Ala–Cys–Ala–Ser–Val–Cys–Pro–Val–Gly–Ala–Ile–Asn–Gln–Gly–Asp–Ser
    Tyr†                                                            Glu†                    Pro†        Glu†
```

Comparison	Base Differences			
	1 base	2 base	total	Per codon
Cl. past./Cl. but.	6	3	12	0.22
Cl. past./Mic. aer.	9	5	19	0.35
Cl. but./Mic. aer.	11	4	19	0.35

* Invariant sites in both halves of molecule. The alignment is that proposed by TSUNODA *et al.* (1967).

† Alternate possibilities in prototype.

References: (i) TANAKA *et al.* (1964); (ii) BENSON *et al.* (1966); (iii) TSUNODA *et al.* (1967).

as the studies with proteins in estimating the rate of evolution. This rate for an entire species corresponds to 1 to 2 base-pair substitutions per year as taking place during evolution in the total genetic DNA of a higher organism containing about 3 billion base pairs. These changes are beneficial or neutral rather than deleterious. The rate of evolutionary change is a phenomenon separate from the rate of mutations. Evolution cannot proceed unless mutations take place, but evolution proceeds as a result of natural selection acting upon the pool of genetic variations which is distributed throughout the millions of individuals in a species.

Let us now consider further aspects of DNA and its role in evolution. A present viewpoint is that randomly distributed base replacements or point mutations occur incessantly in DNA whenever it is replicated, which takes place whenever cells divide during the life of living organisms. This theory predicts that if two groups of organisms of the same species are separated by geographic barriers, the genetic DNA of one group will slowly and inexorably differentiate from that of the other group until a point is reached at which the descendants of the two original lines of inheritance are so different from each other that they can no longer interbreed. Such a barrier will be produced when a base difference of about 2% is reached (McCARTHY, 1968).

Proteins derive their amino acid sequences directly from the base sequences in messenger RNA, and therefore changes in amino acid sequences will be produced by the changes in the base sequences of DNA. It follows that homologous proteins in two different species will have different amino acid sequences and that the quantitative differences will be proportional to the time of evolutionary separation of the two species and to the biochemical characteristics of the protein; for certain proteins are inherently more flexible in their composition than others.

The theory predicts that any molecule which is directly transcribed from DNA, such as transfer RNA or ribosomal RNA, will undergo differentiation in different species of organisms, and the experimental results so far obtained are in agreement with this.

6. Evolution of the Code

It is possible to apply this model to the genetic code itself. In so doing, we must take consideration of the various mechanisms involved in the translation of the code. Messenger RNA is an exact complementary copy of the specifications in DNA for amino acid sequences in polypeptides, and, as such, has a comparatively passive role to play. The other components of the translation mechanism, such as transfer RNA, are likely to play a more active role in evolution.

An examination of the code (Table I) shows that certain amino acids have more codons than others. Obviously, the fewer codons per amino acid, the more amino acids will be concerned in protein synthesis and the more complex will be the biological system. It therefore seems indicated that, if complex systems evolved from simple ones, there must have been a time when fewer than 20 amino acids where involved in protein synthesis (JUKES, 1966b).

The most likely candidates for the amino acids used in this earlier era are those

which have four or more codons apiece. These are alanine, glycine, valine, threonine, leucine, proline, arginine and serine. We have therefore proposed that at an earlier stage of evolution no amino acid had fewer than four codons. This proposal leads to the concept that a new amino acid is introduced or was introduced into protein synthesis by capturing one or two codons from a group of four that coded for an 'old' amino acid. It is necessary to assume that such changes in the code took place before evolutionary divergence started; in other words, that they took place in the single organism that was, according to current evolutionary hypothesis, the ancestor of all the millions of living forms now present on earth, because in all these forms the code is thought to be universal.

The four codons starting with AU perhaps give a clue as to how an increase took place in the number of amino acids used in protein synthesis. Let us suggest that at one time AUG was a codon for isoleucine. It is next necessary also to assume that the anticodon UAU disappeared. This could be brought about by a mutation that destroyed the function of the tRNA carrying this anticodon. The non-functional gene might then vanish. The codon AUG could change its assignment to methionine if the transfer RNA carrying the anticodon CAU had its properties altered by mutation so that it combined with methionine rather than isoleucine. Methionine would then replace isoleucine throughout all proteins in the organism when the codon AUG was translated (JUKES, 1967).

The next step is that one of the codons loses its function in peptide synthesis. This is illustrated by the example of the four codons starting with UG. Two of these, UGC and UGU, are codes for cysteine; UGG is the codon for tryptophan; and UGA is unassigned to an amino acid but is an interval or chain-terminating codon. The stage is now set for the capture of the unassigned codon by the amino acid that has only one codon. This could happen by duplication of the gene for the tRNA which has the anticodon CCA, followed by a point mutation of CCA to UCA. Such a mutation, GUA to CUA, is known to exist in the anticodon of *E. coli* tyrosine suppressor RNA.

The result of such a postulated event is illustrated in the next example; that of lysine and asparagine. In this, and in similar examples, a 'quartet' of codons is divided equally between two amino acids. Notice that the division is always in the same way; there are no cases where one amino acid is coded by XYA and XYU and the other by XYC and XYG. This pattern fits the wobble hypothesis.

If the code evolved by duplication and differentiation of the transfer RNAs, it should be possible to find evidence of similarities between their sequences.

All the tRNA molecules can be arranged in the 'clover leaf' pattern of secondary structure that was proposed by HOLLEY *et al.* (1965). The general shape of this is shown in Figure 1. It contains the following regions:

(i) Four helical regions, characteristically with seven, four, five and five Watson-Crick base pairs respectively. The first helical region starts with the left-hand terminal nucleotide which has a phosphate group esterified to the 5′-position. Unpaired bases often occur in the helical regions, especially in the first one. It was suggested that

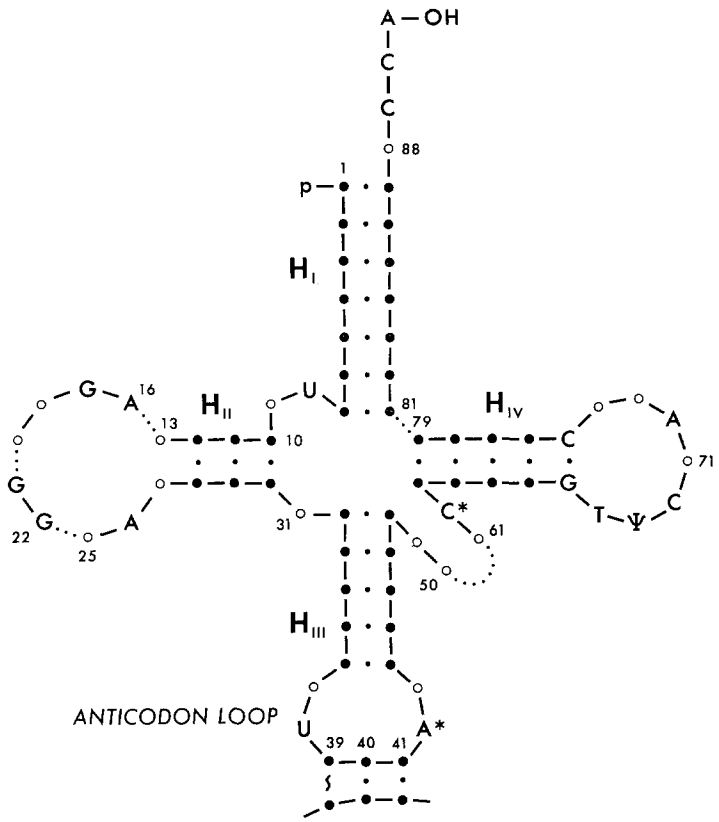


Fig. 1. General formula for transfer RNA molecules, showing helices, loops, and connecting regions. Invariant bases are lettered. ● = hydrogen-bonded bases; ○ = unbonded bases; * designates a base that is sometimes modified; H_I–H_{IV} helical regions. A messenger codon is shown paired with the anticodon; ~, wobble pairing.

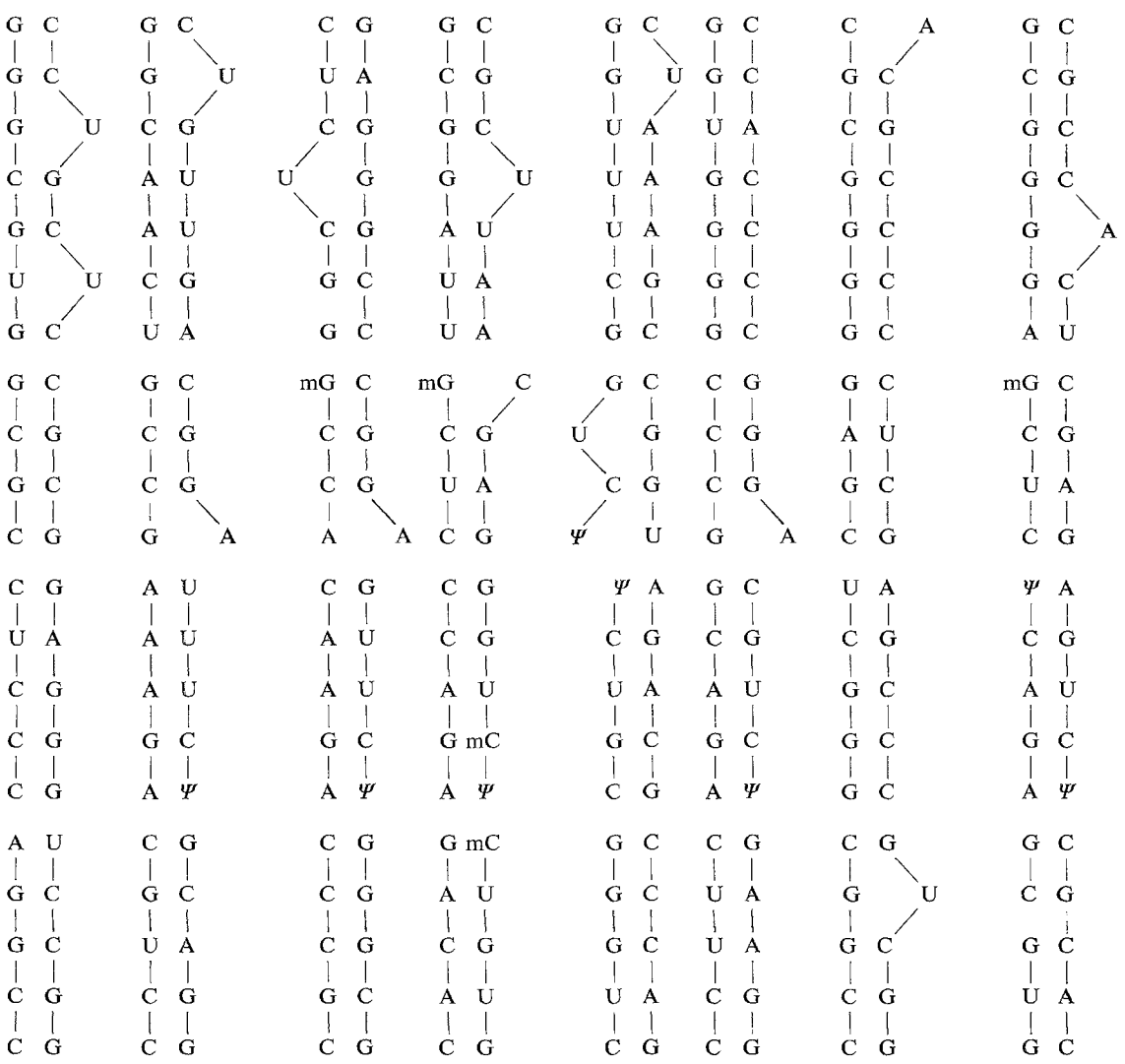
these may result from single-base changes occurring during evolution and that they represent the intermediate stage of a change from one base pair to another, e.g., G≡C→G~U→A=U (JUKES, 1966a). All the helical regions are shown in Table IX.

It was suggested that the helicity of these regions is their only essential property, because none of the bases in these regions are identical at corresponding loci in all the known tRNAs except the GC pair which ends the fourth helix and adjoins the TΨC loop. However, the top three pairs in H_I may recognize the activating enzyme (CHAMBERS, 1968).

(ii) A fifth helical region connects the third and fourth helices in yeast serine and *E. coli* phenylalanine tRNAs. This may have resulted from crossing-over and recombination. There is evidence of this when the 'extra' sections present in yeast serine tRNA and *E. coli* tyrosine tRNA are compared. The comparison is shown in Figure 2. A polynucleotide sequence of 11 residues has apparently been duplicated at a preceding sequence in these two tRNAs, followed by differentiation. The repe-

TABLE IX

Helical regions of tRNAs of known sequence



tition is best seen when yeast tyrosine tRNA, which does not contain the 'extra' segment, is compared with *E. coli* tyrosine tRNA.

(iii) The 'dihydrouridine loop' is variable in size. It contains –A–G–; –G–G–, and –A– in homologous locations.

(iv) The 'anticodon loop' contains seven bases. The anticodon is preceded by U and followed by A or modified A.

(v) The loop containing the TΨC sequence evidently has no amino acid specificity, for two pairs of tRNAs, valine and phenylalanine yeast tRNAs; and formyl methionine *E. coli* and yeast serine-2 tRNAs, are identical with each other in this sequence of seven nucleotides.

(vi) The initial pG is complemented by the terminal base of the fourth helix except in the case of formyl methionine tRNA and it was suggested that this exception might have a functional significance (DUBE *et al.*, 1968). The dihydrouridine loop of tRNA has been suggested as carrying the recognition site for the amino acid

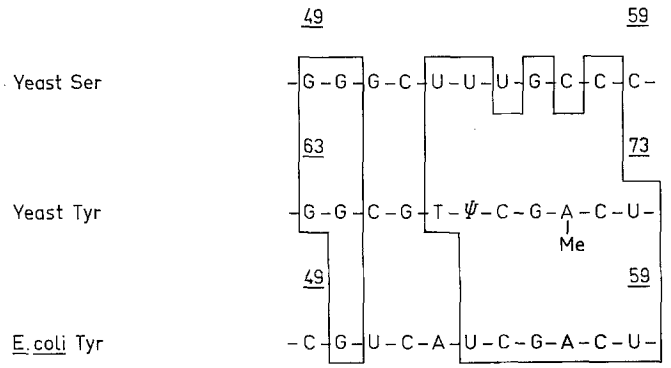


Fig. 2. Repetitive sequence in Ser and Tyr tRNAs.

activating enzyme (DUDOCK *et al.*, 1968f. The dihydrouridine loops are very different for yeast and *E. coli* tyrosine tRNAs as follows:

Yeast –A–G–hU–hU–Gm–G–hU–hU–hU–A–
E. coli –A–G– – –C –Gm–G–C –C –A –A–

but the two anticodon loops are quite similar, as follows:

Yeast –C–U–G–Ψ–A–iA–A–
E. coli –C–U–G–U–A–A†–A–

Yeast and *E. coli* tyrosine tRNAs are therefore potentially useful for comparing the effects of the tyrosine tRNA activating enzyme from one of the respective organisms on the tRNA from the other organism. This has been done by DOCTOR *et al.* (1966). They found that there was absolute species-specificity. The *E. coli* enzyme would not charge yeast tyrosine tRNA I and, vice versa, the yeast enzyme would not charge either of two tyrosine tRNAs obtained from *E. coli* as follows:

Source of tRNA	Source of enzymes	
	E. coli	Yeast
Tyr I Yeast	–	1.37
Tyr I *E. coli*	1.08	–
Tyr II *E. coli*	1.33	–

The figures denote the mμmoles of tyrosine incorporated per unit of tRNA calculated from ultraviolet absorption.

In contrast, when a similar experiment was carried out with yeast and wheat germ phenylalanine tRNAs by DUDOCK *et al.* (1968), they found that either enzyme would charge either tRNA. The dihydrouridine loops of these two tRNA are identical, as follows:

–A–G–hU–hU–G–G–G–A–

but the other regions of the two molecules differ from each other. These comparisons led to the conclusion by DUDOCK *et al.* (1968) that the dihydrouridine loop identifies the tRNA to the activating enzyme. An alternative proposal was made by SCHULMAN and CHAMBERS (1968). They studied the inactivation of alanine tRNA by ultraviolet light, and they concluded that the target for inactivation was located in the first three nucleotide pairs of the top helix (Table IX). These are identical in yeast and wheat phenylalanine tRNAs. Perhaps the activating enzyme binds with *both* the dihydrouridine loop and the first three base pairs of the top helix. The comparisons also raise questions of the manner of differentiation of the tyrosine tRNA of yeast from that of *E. coli*. If these two tRNAs had a common origin, how is it possible that they have diverged so much? Each evolutionary change in the recognition site of each tRNA would have to be accompanied by a corresponding adjustment in the primary structure of the activating enzyme. An alternative mechanism would be convergent evolution. The two possibilities are considered below.

The sequences of eight tRNAs, aligned for homology, are in Table X. Fifteen loci are genetically identical in all eight molecules, assuming that the unidentified nucleosides in position 42 are adenine derivatives and are in the unmodified form at the time of transcription. The random probability for identity of any two sites in two RNA sequences is about 25%, so that there will be an expectation of about 20 identical sites in two tRNA sequences each containing 80 nucleotides, 5 in three sequences, one in four sequences, but less than one in more than four sequences.

There is a marked homology between the primary structures of the phenylalanine tRNAs of yeast and wheat (Table X). These two tRNAs share with the other six tRNAs in Table X the identity at the DNA level of the fifteen bases marked by *; some of these bases are modified subsequently to transcription by methylation, etc. There are 58 other bases in the two phenylalanine tRNAs in addition to the 15 that are identical in all eight tRNAs. Of these 58, 43 are identical and 15 are different in the yeast:wheat comparison; a difference of 26%. This is about the same as is the case with the cytochromes *c* of baker's yeast and wheat, which have 43 amino acid residue differences in 107 sites compared, corresponding to 61 minimum base differences in 107 codons. If it is assumed that 23 of the amino acids are invariant, the 61 minimum base differences are equivalent to 0.72 MBDC or 24 base substitutions in 100 base pairs. The 61 minimum base differences presumably correspond to a larger number of actual base differences because of ambiguity in the third position of codons. The conclusion may be drawn that the cistrons for the phenylalanine tRNAs of yeast and wheat have diverged during evolution at about the same rate or perhaps somewhat more slowly than have the cistrons for the cytochromes *c* of these two organisms.

TABLE X

Base sequences of yeast (Y) alanine, *E. coli* (EC) formylmethionine, yeast and *Torulopsis* (TU) valine, yeast serine, yeast and *E. coli* tyrosine, and yeast and wheat (W) phenylalanine transfer RNAs

Ala	F Met	Val		Ser	Tyr		Phe	
Y	EC	Y	TU	Y	Y	EC	Y	W
pG	pC	pG		pG	pC	pG	pG	pG
G	G	G		G	U	G	C	C
G	C	U		C	C	U	G	G
C	G	U		A	U	G	G	G
G	G	U		A	C	G	A	G
U	G	C		C	G	G	U	G
G	G	G		U	G	G	U	A
*U	U†	U		U	U	Ut	U	U
mG	G	mG		G	A	Ut	A	A
G	G	G		G	mG	C	mG	mG
C	A	U		C	C	C	C	C
G	G	C		acC	C	C	U	U
C	C	Ψ		G	A	G	C	C
G	–	–		–	–	–	–	–
U	–	–		–	–	–	–	–
*A	A	A		A	A	A	A	A
*G	G	G		G	G	G	G	G
hU	C	hU		hU	hU	–	hU	hU
C	C	C	hU	–	hU	C	hU	hU
–	U	–		–	–	–	–	–
*G	G	G		Gm	Gm	Gm	G	G
*G	G	G		G	G	G	G	G
–	–	hU	C	hU	hU	C	–	–
–	–	–		–	hU	C	–	–
hU	hU	hU		hU(C)	hU	A	G	G
*A	A	A		A	A	A	A	A

Ala	F Met	Val		Ser	Tyr		Phe	
Y	EC	Y	TU	Y	Y	EC	Y	W
G	G	U		A	A	A	G	G
C	C	G		G	G	G	A	A
G	U	G		G	G	G	G	G
C	C	C		C	C	G	C	C
m_2G	G	A		m_2G	m_2G	A	m_2G	m_2G
C	U	Ψ		A	C	G	C	Ψ
U	C	C		A	A	C	C	C
C	G	U		A	A	A	A	A
C	G	G		G	G	G	G	G
C	G	C		A	A	A	A	A
U	Cm	Ψ		Ψ	C	C	Cm	Cm
*U	U	U		U	U	U	U	U
I	C	I		I	G	G†(C)	Gm	Gm
G	A	A		G	Ψ	U	A	A
C	U	C		A	A	A	A	A
*mI	A	A†(A)		iA	iA	A†	N	N
Ψ	A	C		A	A	A	A	A
G	C	G		Ψ	Ψ	Ψ	Ψ	Ψ
G	C	C		C	C	C	mC	C
G	C	A		U	U	U	U	U
A	G	G		U	U	G	G	G
G	A	A		U	G	C	G	A
A	A	A		Um	A	–	A	A
G	G	C		G	G	C	G	G
	mG(A)	–		G	A	G	mG	N

Table X (continued)

Ala	F Met	Val		Ser	Tyr		Phe	
Y	EC	Y	TU	Y	Y	EC	Y	W
–		–		G	–	U	–	–
–		–		C	–	C	–	–
–		–		U	–	A	–	–
–		–		U(C)	–	C(U)	–	–
–		–		U	–	A(C)	–	–
–		–		G	–	G	–	–
–		–		C	–	A	–	–
–		–		C	–	C	–	–
–		–		C	–	U	–	–
U†	U	hU	–	G	hU	U	U	C
*C	C	mC		mC	mC	C	C	C
U	G	C		G	G	G	mC	G
C	U	C		C	G	A	U	C
C	C	C		A	G	A	G	G
G	G	A		G	C	G	U	U
*G	G	G		G	G	G	G	
*T	T	T		T	T	T	T	T
*Ψ	Ψ	Ψ		Ψ	Ψ	Ψ	Ψ	Ψ
*C	C	C		C	C	C	C	C
G	A	G		G(A)	G	G	G	G
*A	A	mA		A	mA	A	mA	mA
U	A	U		G(A)	C	A	U	U
U	U	C		U	U	U	C	C

Ala	F Met	Val		Ser	Tyr		Phe	
Y	EC	Y	TU	Y	Y	EC	Y	W
*C	C	C		C	C	C	C	C
C	C	U		C	G	C	A	A
G	G	G		U	C	U	C	C
G	G	G		G	C	U	A	G
A	C	G		C	C	C	G	C
–	–	G	–	–	–	–	–	–
C	C	C		A	C	C	A	U
U	C	G		G	C	C	A	C
C	C	A		U	G	C	U	A
G	C	A		U	G	C	U	C
U	G	A		G	G	A	C	C
C	C	U		U	A	C	G	G
C	A	C		C	G	C	C	C
A	A	A	–	G	A	A(C)	A	A

C

C

A–OH

Abbreviations: N, unspecified nucleoside; hU, dihydrouridine; Cm, O-methylcytidine; Gm, O-methyl guanosine; mG, methyl guanosine; mA, methyl-adenosine; Ψ, pseudouridine; m_2G, dimethylguanosine; mI, methylinosine; m_2A, dimethyl adenosine; m_2C, dimethyl cytidine; iA, isopentenyl adenine; acC, acetylcytidine; mC, methylcytidine; †, modified nucleoside; * designates nucleosides that are identical in all the tRNAs listed; lines separate the helical regions; dashes indicate evolutionary 'gaps' (CANTOR, 1968). Letters in parentheses are residues in variants of the tRNAs beside which they appear; those beside the yeast valine sequence are differences that are present in TU valine tRNA. Ut = 4-thiouridine.

Sources: HOLLEY *et al.* (1965); ZACHAU *et al.* (1966a, b); MADISON *et al.* (1966); RAJBHANDARY *et al.* (1967); BAYEV *et al.* (1967); TAKEMURA *et al.* (1968); GOODMAN *et al.* (1968); DUBE *et al.* (1968); DUDOCK (1968); HISHIMURA *et. al.* (1968).

The clear-cut nature of the evolutionary divergence of these two phenylalanine tRNAs encourages an attempt to compare the two tyrosine tRNAs (Table X). Without the availability of the yeast:wheat comparison, one might conclude that it was difficult to represent the large difference between the two tyrosine tRNAs in terms of divergent evolution. This difference is 32 bases in 58 comparisons, or 55%, which is about twice that of the yeast:wheat comparison. This infers that the evolutionary separation of bacteria from plants and yeast took place about twice as long ago as the yeast; higher plant separation if it is assumed that the rate of differentiation in the tRNAs is approximately linear with time.

The next comparison that can logically be made is that of a tRNA for one amino acid with a tRNA of a different amino acid. This should have a bearing on the origin of the amino acid code. The results are in Table XI. The average of all such com-

TABLE XI

Percent differences between homologous sites of eight transfer RNA molecules, excluding sites common to all eight

	Y Ala	EC F Met	Y Val	Y Ser	Y Tyr	EC Tyr	Y Phe	W Phe
Y Ala		56	54	63	69	70	65	67
EC F Met	56		65	68	64	56	69	62
Y Val	54	65		65	69	63	68	63
Y Ser	63	68	65		52	57	58	62
Y Tyr	69	64	69	52		(55)	53	52
EC Tyr	70	56	63	57	(55)		64	55
Y Phe	65	69	68	58	53	64		(26)
W Phe	67	62	63	62	52	55	(26)	
Average of comparisons between pairs of different amino acids	63.4	62.9	63.9	60.7	59.7	60.7	63.0	58.1

parisons, excluding, of course, the yeast: *E. coli* tyrosine, and yeast:wheat phenylalanine comparisons, is 61.8%. The difference is not much greater than the yeast: *E. coli* tyrosine difference of 55%, a finding which raises the interesting possibility that the genetic code originated not long before the evolutionary separation of the lines of descent leading to yeast and *E. coli* took place. There is a considerable amount of variation in the comparisons between tRNAs for two different amino acids; the greatest difference is between yeast alanine and the two tyrosine tRNAs. Most of the smaller differences are in the serine, tyrosine and phenylalanine group. The corresponding codons for these three amino acids are related by single-base minimum differences.

Another possibility to be considered is that the tRNAs for different amino acids have been separated so long that they have reached an equilibrium. This would represent a random difference of about 75%, for there is about one chance in four for two of the four bases to be identical at random. The values in Table XI are for the

most part substantially less than 75%. This fact, together with the observation that the yeast:*E. coli* tyrosine tRNA difference is about twice as great as the yeast: wheat phenylalanine tRNA difference, suggests that evolution in the tRNA series has not reached an equilibrium with respect to the differences between pairs of tRNAs for different amino acids.

We assume that the evolution of the code ceased at the stage shown in Table I. If the evolution had continued to increase the number of amino acids in protein synthesis, a biological system could have developed in which 30 or 31 amino acids participated in protein synthesis. To proceed even further into diversification requires that the wobble pairing be replaced by something more specific. A system would then be possible where 61 amino acids, each with one codon, participate in protein synthesis. For this, it would be necessary to have further differentiation of the transfer RNAs and to have a biological system sufficiently flexible so that new amino acids could be introduced into proteins.

But this did not happen. Instead the code froze in its present form as soon as divergent evolution started. From this point on, the complexity of the biochemistry of terrestrial life was too great to permit changes throughout the proteins of any organism. The complexity increased with time, so that introduction of a new amino acid into protein synthesis would undoubtedly be lethal.

The model of evolution that is presented leads back to a single organism which used the present code. This organism had competitive advantages so great that its descendants have crowded out all other forms of life. They have done this by divergent evolution. Prior to this ancestral organism, we conjecture there was a long period of parallel or convergent evolution in which more primitive codes and proteins existed.

Acknowledgment

The author thanks Dr. C. R. Cantor for numerous discussions and suggestions.

References

BAYEV, A.A., VENKSTERN, T.V., MIRZABEKOW, A.D., KRUTILINA, A.I., LI, L., and AXELROD, V.D.: 1967, *Mol. Biol.* (U.S.S.R.) **1,** 754.

BENSON, A.M., MOWER, H.F., and YASUNOBU, K.T.: 1966, *Proc. Natl. Acad. Sci. U.S.* **55,** 1532.

BOLTON, E.T. and MCCARTHY, B.J.: 1962, *Proc. Natl. Acad. Sci. U.S.* **48,** 1390.

BRITTEN, R.J. and KOHNE, D.E.: 1965–66, *Carnegie Inst. Washington, Yearbook 1965,* p. 78.

CANTOR, C. R.: 1968, *Biochem. Biophys. Res. Commun.* (in press).

CRICK, F.H.C.: 1966, *J. Mol. Biol.* **19,** 548.

DOCTOR, B. P., LOEBEL, J. E., and KELLOGG, D. A.: 1966, *Cold Spring Harbor Symp. Quant. Biol.* **31,** 543.

DUBE, S.K., MARCKER, K.A., CLARK, B.F.C., and CORY, S.: 1968, *Nature* **218,** 232.

DUDOCK, B. A.: 1968, presented at Symposium on Transfer RNA, New York, June 7–8.

DUDOCK, B.A., KATZ, G., TAYLOR, E.K., and HOLLEY, R.W.: 1968, *Federation Proc.* **27,** 342.

GOODMAN, H.M., ABELSON, J., LANDY, A., BRENNER, S., and SMITH, J.D.: 1968, *Nature* **217,** 1019.

HOLLEY, R. W., APGAR, J., EVERETT, G. A., MADISON, J. T., MARQUISEE, M., MERRILL, S. H., PENSWICK, J. R., and ZAMIR, A.: 1965, *Science* **147,** 1462.

HOYER, B.H., MCCARTHY, B.J., and BOLTON, E.T.: 1963, *Science* **140,** 1408.
HOYER, B.H., MCCARTHY, B.J., and BOLTON, E.T.: 1964, *Science* **144,** 959.
JUKES, T.H.: 1966a, *Molecules and Evolution*, Columbia Univ. Press, New York.
JUKES, T.H.: 1966b, *Biochem. Biophys. Res. Commun.* **24,** 744.
JUKES, T.H.: 1967, *Biochem. Biophys. Res. Commun.* **27,** 573.
MCCARTHY, B.J.: 1968, Presented at the Western Experiment Station Collaborators Conference; Albany, California, March 13, 1968.
MCCARTHY, B.J. and BOLTON, E.T.: 1963, *Proc. Natl. Acad. Sci. U.S.* **50,** 156.
MCCARTHY, B.J. and BOLTON, E.T.: 1964, *J. Mol. Biol.* **8,** 184.
MCCARTHY, B.J. and HOYER, B.H.: 1964, *Proc. Natl. Acad. Sci. U.S.* **52,** 915.
MADISON, J.T., EVERETT, G.A., and KUNG, H.K.: 1966, *Science* **153,** 531.
MARMUR, J. and LANE, D.: 1960, *Proc. Natl. Acad. Sci. U.S.* **46,** 453.
MATSUBARA, H., SASAKI, R.M., and CHAIN, R.K.: 1967, *Proc. Natl. Acad. Sci. U.S.* **57,** 439.
NISHIMURA, S., HARADA, F., NARUSHIMA, U., and SENO, T.: 1967, *Biochim. Biophys. Acta* **142,** 133.
NISHIMURA, S. and co-workers: 1968, personal communication.
PAULING, L. and COREY, R.B.: 1956, *Arch. Biochem. Biophys.* **65,** 164.
RAJBHANDARY, U.L., CHANG, S.H., STUART, A., FAULKNER, R.D., HOSKINSON, R.M., and KHORANA, H.G.: 1967, *Proc. Natl. Acad. Sci. U.S.* **57,** 751.
SCHILDKRAUT, C.L., WIERZCHSOWSKI, K.L., MARMUR, J., GREEN, D.M., and DOTY, P.: 1962, *Virology* **18,** 43.
SCHULMAN, L.H. and CHAMBERS, R.W.: 1968, presented at Symposium on Transfer RNA, New York, June 7–8.
SUGENO, K. and MATSUBARA, H.: 1968, *Biochem. Biophys. Res. Commun.* **32,** 951.
SUMMERS, W.C. and SZYBALSKI, W.: 1968, *Virology* **34,** 9.
TAKEMURA, S., MIZUTANI, T. and MIYAZAKI, M.: 1968, *J. of Biochem. (Tokyo)* **63,** 277.
TANAKA, M., NAKASHIMA, T., BENSON, A., MOWER, H.F., and YASUNOBU, K.T., 1964, *Biochem. Biophys. Res. Commun.* **16,** 422.
TSUNODA, J., WHITELEY, H., and YASUNOBU, K.T.: 1967, *Pacific Slope Biochemical Conference*, abstracts, p. 103.
WATSON, J.D. and CRICK, F.H.C.: 1953, *Nature* **171**, 737.
ZACHAU, H.G., DÜTTING, D., and FELDMANN, H.: 1966a, *Angew. Chem.* **5,** 122.
ZACHAU, H.G., DÜTTING, D., and FELDMANN, H.: 1966b, *Z. Physiol. Chem.* **347,** 212.