On the Possibility of a Stage in the Evolution of the Genetic Message in which Replication was Imprecise[1]

Arthur M. Lesk[2]

Program in Biochemical Sciences
Princeton University
Princeton, New Jersey

Received January 7, 1970

The capacity for precise replication is an essential property of the genetic message as it is known today, and a striking feature of the structure of DNA is its natural mechanism of self-replication derived from complementary base pairing. In recent discussions of the origin of the genetic code, it has been assumed, explicitly or implicitly, that replication was precise at all stages of the evolution of the genetic message (1-3). We have questioned the need for this assumption, and have considered alternatively the possibility of an imperfectly replicating polynucleotide which directed polypeptide synthesis. Two such possibilities would be polynucleotides composed of the bases[3] I, A and U or G, A and U; the imprecision of replication would result from the same ambiguity in base pairing postulated by Crick to occur in the interaction between codon and anticodon during protein synthesis (4).

We shall first consider circumstances under which imperfect replication might not have been lethal, and then describe the nature of the genetic coding associated with it.

The advantage of the precise replication of contemporary DNA depends

---

1. Third in a series of papers on Molecular Aspects of Evolution. Predecessor: J. Theor. Biol., in press.
2. Supported by grants to Prof. J.R. Fresco from the National Institutes of Health (GM-07654) and the National Science Foundation (GB-6664).
3. The bases adenine, cytosine, guanine, inosine, thymine and uracil are abbreviated to their initials. C.t. stands for "chain termination".

on the facts that: (a) the nucleotide sequence is "decoded" into polypeptide with high accuracy, and (b) the protein molecules synthesized are highly evolved, so that many arbitrary amino acid alterations are detrimental. But in the early stages of evolution, neither of these need have been true.

Woese has postulated that during the development of the code, the correspondence between triplets and amino acids was somewhat ambiguous: that groups of codons corresponded to groups of amino acids (5). This would have been the case before the tRNA + synthetase adaptor system became precise. But, imperfect replication would then result in no impairment of "protein" synthesis, provided that the replication errors changed each codon only to others associated with the same class of amino acids. Moreover, while the amino acid sequences of proteins were evolving, a mechanism which introduced modest amounts of variation might even have been useful -- or at least, not disadvantageous. On the other hand, it would clearly be efficient to have precise replication and adaptor systems, after the "proteins" produced achieved reasonable catalytic competence.

Orgel and Crick have suggested that the original nucleic acid might have been composed of the two bases A and I (2). Consider the following hypothetical stages in the evolution of the base composition:[4]

A + I → A + I + U → A + G + U → A + G + C + U

In these different stages, the following base pairings are possible (+ means a possible pairing, - an unlikely one):

AI:	A	I
A	-	+

AIU:	A	I	U
A	-		
I	+	-	
U	+	+	-

AGU:	A	G	U
A	-		
G	-	-	
U	+	+	-

Nonstandard base pairings are designated as possible if model building suggests that they are stereochemically feasible (4). Studies of synthetic

4. At some point, U became T, but we consider here only the base pairing, which is the same for both.[6]

polynucleotides provide some direct evidence of their existence (7). Pairs are designated as unlikely if they are stereochemically incompatible in the same helix with those considered possible.

Suppose that during replication or transcription, each base has a certain chance of pairing with the other possible partners. In the A-I-U case, each base might pair with any different base; any triplet might pair with any of eight of the twenty-seven possible triplets. In reproduction of the original strand, each triplet might be replaced by any other triplet, according to some distribution of probability. At equilibrium, the fractions of different trinucleotides would be constant.

This case is perhaps too general to be informative in any detailed way, unless environmental conditions are discovered to bias the pairing strongly in favor of less ambiguity. However, it does suggest the general notion that imprecision in replication provided a means for experimenting with amino acid sequences, whereas imprecision in the adaptor system provided a means for experimenting with different amino acids at a particular site.

In the case of a genetic message composed of A, G and U, the ambiguity is reduced by the disfavor of the A-G pair. The codons are not all interconvertible, and the possible amino acid substitutions can be enumerated precisely.

The portion of the genetic code from E. coli, restricted to codons not containing C, is as follows:

UUU phe	UAU tyr	UGU cys
UUA leu	UAA c.t.	UGA c.t.
UUG leu	UAG c.t.	UGG trp
AUU ile	AAU asn	AGU ser
AUA ile	AAA lys	AGA arg
AUG met	AAG lys	AGG arg
GUU val	GAU asp	GGU gly
GUA val	GAA glu	GGA gly
GUG val	GAG glu	GGG gly

The ambiguity of pairing -- either G or A with U -- leads to

the exact replication of U, but to the interconversion of A with G. The resulting ambiguities of amino acids can be classified into three groups, shown below, according to the number of purines in the codons. We assume, for lack of alternative information, that the correspondence between codons and amino acids is the one observed today.

1. No purines. There is one codon, UUU = phe, and no possible substitution.

2. One purine. There are twelve codons, interconvertible within three sets of four:

AAU asn, AGU ser, GAU asp, GGU gly

AUA ile, GUA val, AUG met, GUG val

UAA c.t., UAG c.t., UGA c.t., UGG trp

3. Three purines. Eight interconvertible codons:

AAA lys, AAG lys, AGA arg, GAA glu, AGG arg, GAG glu, GGA gly, GGG gly

Many, but not all, of the amino acid substitutions occur between structurally related amino acids; such substitutions might correspond to non-lethal "mutations". In addition, the chain termination codons are in the same class.

There is, of course, no evidence that the code did develop in the way formulated. We desire to conclude from these considerations only that there is nothing inherently impossible about an imperfectly replicating genetic message, at an early stage of molecular evolution.

References

1. T.H. Jukes, Molecules and Evolution, (Columbia University Press, New York, 1966). Chapter 2.
2. F.H.C. Crick, J. Mol. Biol. 38, 367 (1968).
3. L.E. Orgel, J. Mol. Biol. 38, 381 (1968).
4. F.H.C. Crick, J. Mol. Biol. 19, 548 (1966).
5. C.R. Woese, The Genetic Code, (Harper and Row, New York, 1967).
6. A.M. Lesk, J. Theor. Biol. 22, 537 (1969).
7. G. Felsenfeld and H.T. Miles, Ann. Revs. Biochem. 36, 407 (1967).