A CYBERNETIC APPROACH TO THE ORIGIN OF THE GENETIC CODING MECHANISM

I. *Methodological Principles*

V. A. RATNER and A. G. BATCHINSKY
Institute of Cytology & Genetics of the USSR Academy of Sciences, Siberian Department, Novosibirsk 90, U.S.S.R

(Received 8 August 1975; in revised form 23 January 1976)

Abstract. It is postulated that some quasi-deterministic code features (universality, connectedness, systematic degeneracy, symmetry, regularity and so on) resulted from *unique* (and therefore universal) *realization* of a stochastic evolutionary process. The evolution of real genetic systems should satisfy the *principle of succession*; that is, loss of a feature that is necessary for a genetic system means death to its carrier. The hypothesis of *unique key coincidence* is proposed which indicates the mechanisms of arising of the primary correspondence between the linear structures of polynucleotides and polypeptides. If the collinear coincidence was to appear in the key positions of pra-protein with, at least, some of the primitive properties of the pra-amino-acyl-t-RNA-synthetase required for the accelerated recognition of the key positions of pra-template, the *positive feed-back mechanism* in the system would be short-circuited so that the repetitive reproduction of pra-synthetase would be much accelerated.

Attempts to reveal specific amino acid–nucleotide interactions, which would help to clarify the structural features of the genetic code, have not met with appreciable success. But the absence of a chemical correspondence between monomers seems very likely to be the basic difference between the process of nucleic acid translation and nucleic acid transcription and replication, because all codons have been assigned to definite amino acids.

There is reason to believe that the genetic code was originally determined by chance events. However, the genetic code possesses a set of quasi-deterministic features such as universality, connectedness, systematic degeneracy, symmetry and regularity (Ratner, 1974, 1975; Batchinsky and Ratner, 1976) which could not have emerged by chance, all at a stroke. Based on this concept of the genetic code, some methodological principles implied in the analytic treatment of a number of relevant problems will be discussed.

The general heuristic rules for the simulation of the origin and evolution of the genetic code may be formulated as follows:

(a) All models should contain a minimal number of events with negligibly small occurrence probability and operate only with events and processes established in evolving genetic systems (models should not explain the unclear using the unknown);

(b) Models should incorporate a minimal number of components;

(c) The introduction of a new element of property should increase the probability or rate of self-reproduction;

(d) Models should satisfy the *principle of succession: loss of a feature that is necessary for a genetic system means death to its carrier.*

In this approach the process suggested for the origin of the genetic code is treated as a stochastic process. It is postulated that the quasi-deterministic features of the code are the necessary consequences of its stochastic evolution and unique

Origin of Life 7 (1976) 225–228.

(and, consequently, universal) realization of this process. This become a reality only when there is a limited choice of stable variants of the process whose occurrence probability is negligibly small.

Since the principle of succession holds true, the evolution of the code should be reflected in some of its structural features. Thus, the triplet nature of the code should be considered as its initial property (altered codon length makes all the previous genetic templates senseless); connectedness and systematic degeneracy of code series, symmetry and regularity indicate that similar mechanisms and processes underlie the origin of code series; universality is evidence of the monophyletic origin of all existing forms.

All schemes of the origin of the genetic code are based on the hypothetical appearance of the primary structures of pra-cistrons, their products, pra-proteins, pra-t-RNAs and pra-aminoacyl-t-RNA-syntheses (pra-ARSs), and of a system of their interrelations. We may minimize this variety by suggesting that pra-ARS might have been a product of a pra-cistron, and that pra-t-RNA might have been a double-stranded, self-complementary pra-cistron variant closed at one end.

Now, omitting the chemical and energetic aspects of the problem, let us consider only the cybernetic one. There was once a pool or reservoir where all the conditions necessary for the origin of life were partially fulfilled (Ponnamperuma and Gabel, 1968; Fox and Dose, 1972; Rutten, 1971). The conditions, among others, include:

(a) a set of monomers, incorporated into irregular biopolymers of nucleic acids and proteins;

(b) nucleotide-like substances, capable of polymerization using oligonucleotides as templates for the synthesis of complementary sequences;

(c) amino acids, also capable of polymerization, which gave rise to protein-like substances with stochastic structures (these structures, though incapable of reproducing, presumably possessed some catalytic properties); and

(d) polynucleotides which possibly served as sites of amino acid condensation and promoted unspecific non-template synthesis of polypeptides with stochastic sequences.

Of these substances, only the polynucleotides were capable of self-reproduction; the system itself was not. In such systems, amino acids sequences in protein-like substances were not fixed but determined by the relative concentrations of monomers.

At some point of time there arose a primary correspondence between the triplets of nucleotides and amino acids. How this occurred is open to speculation. It is by no means impossible that weak, barely specific interactions between amino acids and nucleotides might have promoted the occurrence of this correspondence. One can also imagine that this correspondence might have been the accidental result of *unique key coincidence* (Ratner, 1966, 1975). The way in which this coincidence was produced may be described in general outlines as follows. Through the course of time a poly-nucleotide appeared, among innumerable others, that might have had a definite cluster (B) of nucleotides which were arranged in some positions collinearly with respect to the positions of definite amino acids (K), which were crucial in the formation of the functional centre of primitive pra-ARS for this set of

amino-acids (K-group) and nucleotide triplets (clusters B) or triplet-antitriplet pairs. Let us suppose that this protein with properties of this pra-ARS has conceivably arisen at some time when these polynucleotides existed. This key coincidence reduced the variety of amino acids in the determined positions of a protein. Since this protein was a pra-ARS, there emerged the possibility of reproducing with relatively high probability some patterns of protein structure based on the pra-template, i.e., copying the sequential positioning of bases in nucleic acids.

It is not so hard to imagine that a protein with a definite function arose by chance. The now existing (and, probably, the primitive) ARS as well as other protein families require not more than several dozen amino acids in invariant positions to ensure their functions. (To be more precise, these amino acids have to be incorporated into their functional centres.) For this reason, there is no absolute necessity in the formation of a perfect primary structure with a given function; it is sufficient to determine the number of key positions. Availability of key positions would make the accidental formation of proteins much more probable. If collinear coincidence was to appear in the key positions of pra-protein with, at least, some of the primitive properties of the pra-ARS required for the accelerated recognition of the key positions of pra-template, the *positive feed-back mechanism* in the system would be short-circuited so that the repetitive reproductions of pra-proteins representing the same ARS family would be much accelerated.

Whatever may have been the cause of a faintest correspondence between amino acids and nucleotides, this event has drastically affected the evolutionary course of primary genetic systems. Had there been no amino acid–nucleotide correspondence,

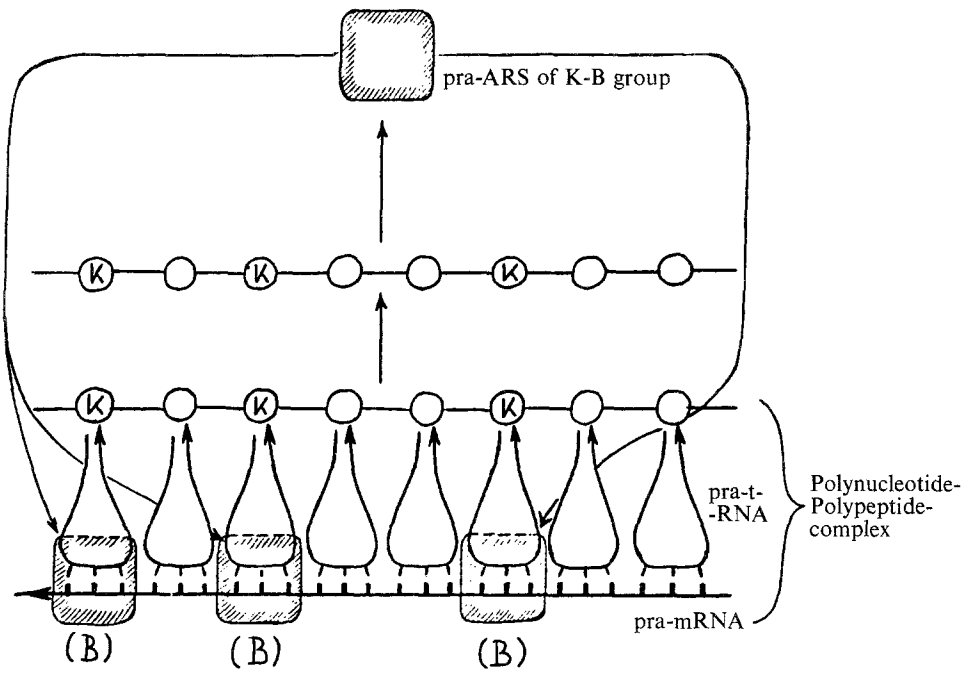

Fig. 1. Hypothetical scheme for the origin of a polynucleotide – polypeptide complex with a collinear coincidence of a polynucleotide with polypeptide, and a positive feed-back mechanism through an ARS.

only self-replicating polynucleotides could have evolved in the successive way, whereas amino acid sequences in proteins would have had to be synthesized each time anew with no template as guide. Amino acid-nucleotide correspondence not only made the self-reproduction of the structure of definite macromolecules possible, it also opened up possibilities of reproducing the system of interrelations between them, to select systems and subsystems for efficiency, i.e. for evolution in the Darwinian sense.

Cybernetic relations between the elements of primitive coding mechanisms, of course, had their physico-chemical basis. Our hypothesis is based on the idea of Thomas (1970) about weak specific interactions between triplet-antitriplet pairs and amino acids in primitive polynucleotide-polypeptide complexes, where the main features are polarity-nonpolarity and the dimension of amino acids. Thomas noted that small amino acids could not cover the length of a 3rd nucleotide pairs of triplet-antitriplet complex. This feature may be the reason why the 3rd nucleotides of codons are not used in the coding of these amino acids. Some physico-chemical details of our hypothesis were published elsewhere (Ratner, 1975).

References

Batchinsky, A. G. and Ratner, V. A.: 1976, *Biometrische Z.* **18,** 1, 53–67.

Fox, S. W. and Dose, K.: 1972, *Molecular Evolution and the Origin of Life*, Freeman & Co., San Francisco.

Ponnamperuma, C. and Gable, N.: 1968, *Space Life Sci.* **1,** 65.

Ratner, V. A.: 1974, 'The Genetic Language', in R. Rosen and F. Snell (ed.), *Progress in Theoretical Biology*, Academic Press, New York, V. 3, 143–228.

Ratner, V. A.: 1975, *Molecular Genetic Regulatory Systems*, "Nauka", Novosibirsk. (In Russian).

Rutten, M.: 1971, *The Origins of Life by Natural Causes*, Elsevier Publ. Co., Amsterdam-London-New York.

Thomas, B. R.: 1970, *Biochem. Biophys. Res. Coms.*, **40,** 6, 1289–1296.