THE FUNDAMENTAL NATURE OF THE GENETIC CODE: PREBIOTIC INTERACTIONS BETWEEN POLYNUCLEOTIDES AND POLYAMINO ACIDS OR THEIR DERIVATIVES*

BY CARL R. WOESE

DEPARTMENT OF MICROBIOLOGY, UNIVERSITY OF ILLINOIS, URBANA

Communicated by T. M. Sonneborn, October 12, 1967

The essence of the cell lies in the processes of replication and translation of nucleic acid—the former a process in which exact copies are made of pre-existing genetic material, the latter a process in which nucleic acid primary structure is mapped into protein primary structure. The essence of replication lies in the base-pairing interactions; thus we comprehend its fundamental nature, although we have yet to elucidate all its details. Translation is another matter. It is known to involve a "genetic code"; to each of the 64 possible trinucleotides, or "codons," there corresponds a single amino acid.[1, 2] The polynucleotide, taken then as a linear array of codons, is "translatable" into a unique, corresponding, and colinear array of amino acids, the polypeptide. However, we at present have no idea of why there exists this particular, unique, precise correspondence between amino acids and codons. Consequently, essentially nothing is understood about the fundamental nature of the translation process.

While we can hope for an ultimate understanding of nucleic acid replication in simple terms, this is unlikely in the case of translation; the former is a simple process, the latter a complex one. Such complexity demands a complex evolution.[3] It is even conceivable that the translation of today is so many evolutionary steps removed from its origins that the two have little or nothing in common. (I emphasize this point in that preconceptions as to the nature of translation derived from knowledge of the process as it is found in cells today could well prejudice and/or narrow our thinking regarding translation's evolution.) No wonder, then, that we have yet to find evidence for any sort of "codon relationship," i.e., any sort of unique or preferential association of certain oligonucleotides with certain amino acids in a system simple enough to be considered prebiotic.[4] To compound this paucity, at present we can not even point to one single reasonable *hypothesis* which accounts for the origins of the components of the translation apparatus.

Given the complexity of the problem and the paucity in both conception and fact, it seems most profitable at this time to approach the fundamental nature of translation in essentially a blind way—by first attempting to determine what sorts of prebiotic interactions between polynucleotides and polyamino acids or their derivatives were plausible, and then hope that this knowledge will in turn lead to the development of a concept of what a workable, primitive form of "translation" could have been like, and perhaps how this form evolved into the present complex machine.

Our starting point is, of course, in the very probable assumption that biopolymers were an important part of a prebiotic milieu. Syntheses performed under "primitive earth conditions" readily yield all kinds of amino acids, bases,

sugars, etc.[5] Thus an early presence of monomer units from which polyamino acids and polynucleotides could be constructed is nearly certain. The actual polymerizations under prebiotic conditions, though not nearly so well substantiated, have factual support, and we can expect a good deal more in the near future.[6]

For a polymer to have any direct evolutionary significance, the least we must require is that it be a part of an autocatalytic cycle which catalyzes the synthesis of the polymers involved therein. Over and above this the cycle should eventually be capable of propagating certain types of changes (randomly) produced in its complexion.

In discussing various possible primitive biopolymer autocatalysis cycles, it is best to begin with a clear definition of the different stages or steps in the synthesis of a biopolymer. As a first step we can envision monomer unit *accumulation*. The cell's building up of pools of amino acids is one example of this. Prebiotically it might occur through adsorption onto surfaces, and in particular, in the physical binding of monomers to some pre-existing polymer, as happens in the binding of nucleosides to polynucleotides.[7] A second *positioning* step should be distinguished from a third, *ordering* one. *Positioning*, by definition, is the aligning of monomer units relative to one another in an orientation conducive to their polymerization. *Ordering* is used to describe any processes which place constraints on the primary structure of the polymer being synthesized. Very often positioning and ordering go together, as would occur again, for example, in the alignment of polynucleotide monomer units by a pre-existing polynucleotide chain. (This example is, more or less, one of *precise* ordering; various degrees of imprecise or "statistical" ordering must be recognized as well.) For the biologically important monomer units, a fourth, *activation*, stage is necessary—e.g., in the triphosphorylation of nucleic acid monomer units. Finally we need to recognize the actual *condensation* step, by which monomers are linked together to form the polymer. *Condensation* and *activation* could be simultaneous reactions, or they could, as is generally the case in the cell, be separate stages. Obviously, in a primitive biopolymer autocatalysis cycle a pre-existing polymer might function catalytically in any of the above five steps.

The main purpose of the present communication is to examine some simple biopolymer autocatalysis cycles, with the view of determining which is the most plausible in a prebiotic world and/or which could contain relationships that might be termed "primitive translation." Arguing from the few facts now available, it is concluded that the most plausible cycle is one involving polynucleotides of certain compositions and polyamino acids of certain compositions, the synthesis of the one being catalyzed by the other and vice versa. Such a cycle, too, automatically gives what might be "primitive translation." What shall not be discussed at this time, however, is the rather involved matter of how this type of "translation" could have evolved into what we now see—a true, precise translation.

Materials and Methods.—The nucleoside mono- and diphosphates, polynucleotides, amino acid derivatives, and polyamino acids used in this study were purchased from one or more of the following suppliers: Pabst Laboratories, Mann

Laboratories, Miles Chemical Co., Calbiochem, and Pilot Chemical Co. For equilibrium dialysis studies, a two-chambered cell was used. In one chamber the polymer dissolved in buffer was placed, in the other the monomer, in the same buffer, the volume in each chamber being 2 ml. The cells were kept at 4°C with constant rotation until equilibrium was reached, about three days.

Results and Specific Discussion.—Let us assume that the initial forms of life that evolved into modern cells were themselves based only upon polymers similar to those we now term "informational biopolymers"—i.e., upon polynucleotides and polyamino acids. In this case, we have to consider only four general types of situations where a pre-existing polymer molecule might somehow catalyze the formation of another polymer molecule. And from these we can form only three simple biopolymer autocatalysis cycles.

First, take the synthesis of a polynucleotide catalyzed by a pre-existing polynucleotide. Although this synthesis has not been accomplished in the test tube as yet (without the help of some enzyme), all the evidence suggests that it should be possible to observe it. Not only do the strong base-base stacking interactions exhibited by the purine nucleosides and nucleotides provide a ready rationale for the initial, spontaneous formation of polynucleotides, but this property plus the base-pairing interactions and the double-stranded helix↔ single-stranded coil transition suggest mechanisms whereby a pre-existing polynucleotide might catalyze further polynucleotide synthesis.[7, 8] The observed binding of purine mononucleosides and nucleotides by the complementary polypyrimidines gives clear experimental support for *accumulation, positioning*, and strict *ordering* roles for the polymer.[7] It is important to note that in this case monomer binding to polymer is a cooperative phenomenon: At low monomer concentrations, a negligible fraction of the monomer is bound; but at concentrations above roughly 10^{-3} *M*, the fraction of polymer-bound monomer can become very appreciable.[7] Were primitive monomer unit concentrations to achieve this critical level, then a primitive biopolymer autocatalysis cycle based solely upon polynucleotides is certainly a feasible one.

The above example makes it clear that two sorts of interactions are probably involved in any "optimal" nucleotide, etc., binding: (1) an attractive interaction between the individual monomers and the binding macromolecule, and (2) a base-base stacking interaction among the bound monomers themselves. In this example the first interaction is, of course, the result of base pairing. There is no reason to suppose, however, that this interaction could not just as wel occur through the phosphate groups of the monomer if they were attracted by properly spaced positive charges—which brings us to the second system to be considered, the synthesis of polynucleotides catalyzed by pre-existing polyamino acids.

It has long been known that the interresidue spacing in a polypeptide closely approximates the interresidue spacing for a polynucleotide in a base-stacked configuration.[9] Thus a poly*basic* amino acid should present, then, the correct charges in the correct geometry to *accumulate* and *position* polynucleotide monomer units. Figure 1 shows that this is probably the case. Here we see that as concentration of monomer is increased, a critical point is reached above which the fraction of nucleotide monomer bound to polyarginine rises (befoer

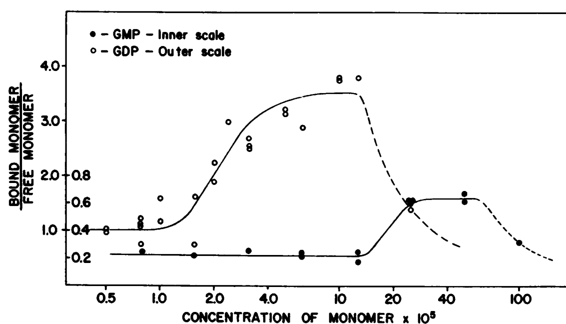

Fig. 1.—Equilibrium dialysis of guanosine mono- and diphosphate vs. polyarginine. Polyarginine concentration 1×10^{-3} mono-molar. Medium is 0.1 *M* Tris buffer, pH 7.5. *Abscissa:* concentration of nucleoside mono- or diphosphate. *Ordinate:* ratio of bound monomer to free monomer.

dropping eventually at polymer "saturation," which, in this case, is precipitation of monomer-polymer complexes). This apparently cooperative behavior in monomer binding is most simply understood in terms of base-base stacking among the bound monomer units.

A second prediction of such stacking would be that the cooperative aspect of the binding of nucleotides to polybasic amino acids should be a function of the base component of the nucleotide, the strength of the stacking interactions being a strong function of the nature of the base. Stacking interactions order the bases (from strongest to weakest) G > A > C > U.[10] Figure 2 shows the verification of this prediction. Here the concentration of mononucleotide at which polymer "saturation" occurs is measured by the formation of insoluble monomer-polymer complexes (i.e., by appearance of turbidity). More than a two-decade difference in concentration exists between the best monomer, GMP, and the worst one, UMP, by this criterion. The same base ranking holds for other polybasic amino acids, polylysine or polyornithine, as well. At lower ionic strengths formation of these complexes occurs at monomer concentrations as low as 10^{-5} *M*. The details of these and the related experiments about to be discussed are to be reported elsewhere.[11]

Certainly the composition and properties of polybasic amino acids make it impossible that these could have arisen merely by unguided polymerization of amino acids from a general amino acid pool. A reasonable mechanism for producing polybasic amino acids, however, is essentially the reverse of the above scheme—through the use of pre-existing polynucleotide chains to catalyze the polymerization of polybasic amino acid monomer units. The negatively charged backbone in this case would serve to bind the positively charged amino acids, or their activated derivatives—i.e., would select these from the general amino acid pool—and when the polynucleotide is in a stacked configuration, spacings would again be correct for positioning the polyamino acid monomer units.

Factually, the situation here is almost completely analogous to the mononucleotide-polybasic amino acid interactions. Insoluble complexes form upon

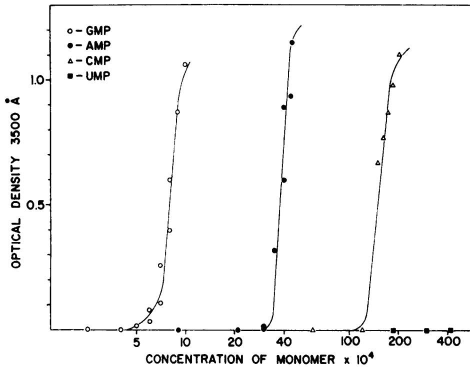

FIG. 2.—Formation of insoluble complexes between polyarginine and nucleoside monophosphates.

Abscissa: concentration of nucleoside monophosphate.

Ordinate: turbidity as measured by optical density at 3500 Å, 1-cm light path.

Polyarginine concentration is 5×10^{-4} monomolar. Medium is 0.1 M Tris buffer, pH 7.5.

addition of the monomer to a solution of the polymer when monomer is in excess. And again, "activated" monomer units bind better than free amino acid. Thus, for example, a complex between arginine methyl ester (a derivative resembling in some ways a truly "activated" amino acid) and poly G will form at all ester concentrations above 1×10^{-3} M, and in salt concentrations up to about 2×10^{-2} M.[11] As might be expected, formation of such complexes occurs almost solely with basic amino acids.

Thus we see that under what could be considered fairly reasonable primitive earth conditions—i.e., monomer unit concentrations in the range 10^{-3} M (or in some cases lower)—both the polybasic amino acid plus nucleoside mono- or diphosphates system, and the polynucleotide plus amino acid (or "activated" amino acid) system, form monomer-polymer complexes. These two systems together then constitute a two-component biopolymer autocatalysis cycle.

The question now arises as to whether this type of two-polymer-component autocatalysis cycle would actually be as "good" as (i.e., have as much selective advantage as) the previously discussed single-polymer-component polynucleotide cycle. On comparing the two types of autocatalysis cycles, they both appear roughly equivalent with regard to an *accumulation* function. The formation of complexes in all cases occurs at roughly the same minimum monomer concentration—i.e., $\sim 10^{-3}$ M. Of course, in the two-polymer-component cycle an appreciable fraction of monomer is bound no matter how low its concentration, which is not true for the pure polynucleotide cycle.[7] And depending upon what the prevailing primitive mechanisms for activation and/or condensation are, it may also be advantageous that this double cycle preferentially binds *activated* monomer units—nucleoside diphosphates or amino acid esters. Both cycles are probably capable of *positioning* monomers. This is a certainty for the pure polynucleotide cycle; the evidence presented above makes this very likely for the polybasic amino acid–mononucleotide interaction; and what little evidence there is, is consistent with *positioning* in a polynucleotide-amino acid ester interaction. (The order of effectiveness of forming complexes with basic amino acid derivatives is again the base-stacking ranking: poly G > poly A > poly C > poly U.)

The major difference between the two cycles being contrasted is in the kind of *ordering* they produce. The pure polynucleotide cycle seems capable of precise

ordering—i.e. "exact" (complementary) copying of a pre-existing primary structure. The alternative cycle manifests a very imprecise, "statistical" *ordering*. The crucial consideration, then, is what advantages would precise ordering confer upon the pure polynucleotide cycle. Put another way, would two polynucleotides of different primary structures (but similar gross composition) be *functionally* differentiable in an environment devoid of proteins (on any other basis than their capacity to form a double helix with a complementary strand)? We at present have no reason to suspect that oligonucleotide sequences of moderate size could be so distinguished from one another in the environment so defined. Thus the pure polynucleotide cycle should not be any less effective were the base-pairing interactions inexact, ambiguous ones. (Actually they are ambiguous if we include the "noncoding" bases.) Therefore, there is no reason to suppose that capacity to *order* monomers would give a polynucleotide cycle a selective advantage over the alternative two-polymer-component cycle.

The remaining biopolymer autocatalysis cycle yet to be considered has polyamino acid synthesis catalyzed by pre-existing polyamino acids. We have no precedent for assuming that one amino acid chain can serve to *accumulate*, *position*, or *order* the monomers for any new amino acid chain—with the possible exception of a relationship similar to the above sorts between the basic amino acids and the acidic ones. Although such a system may be feasible, it appears to offer the least evolutionary potential of all the systems considered, and so it will be somewhat arbitrarily discarded here.

General Discussion and Summary.—The hypothesis put forth herein is that evolution of the cell began with two very general kinds of polymers, polynucleotides that were purine-rich and polyamino acids composed largely if not solely of basic amino acids. The one type of polymer is viewed as catalyzing, in particular ways, the synthesis of the other, and vice versa. Although somewhat more complex variants of this basic type of biopolymer autocatalysis cycle are imaginable, and might offer certain hypothetical advantages over the simplest type, there is little point in discussing them at this time. It suffices to recognize at this point that the simple cycle invoked here may eventually need some modification to include polynucleotide replication, etc.

The plausibility of the present scheme rests in large part upon the availability in a primitive environment of appreciable quantities of basic amino acids. Although lysine has been reported to be synthesized under "primitive earth conditions," the yield is low.[12] However, in dealing with primitive polyamino acids one is not restricted to those basic amino acids now found "encoded" in cells—i.e., lysine, arginine, and histidine. Therefore, it seems reasonable to take as the prevalent primitive basic amino acid a simpler one, 2,4 diaminobutyric acid. This compound might well have been relatively abundant on the primitive earth, for it is probably synthesized from compounds themselves considered to have been very abundant and/or reactive—i.e., β-amino propionitrile or derivatives of aspartic acid.[5]

The primitive cycle considered here is radically different from the cell we see today, and it is pictured as arising and evolving in a way that is, according to present concepts, atypical. Take this last matter first. Partly from experience

with the effects of mutations on the cells of today (and perhaps on a deeper level due to survival of an ethos that life is somehow a "miraculous" happening), early evolution is very often pictured as occurring through the sudden appearance of some novel protein sequence that performs a function which in turn somehow leads to its own propagation. How rare such events are is beyond the realm of useful conjecture, but one tends to think they might be vanishingly so. In any case this prevalent view of evolution assumes a system in a steady state waiting upon the very rare events at the far reaches of probability distributions to slowly nudge it along some evolutionary course. Any of the systems discussed here can by no means be considered as evolving through highly improbable events. Although I shall not discuss the matter of the present system's further evolution, it seems reasonable to consider that the early aspects of its evolution did not progress through events far *less* probable than the mutations operating in evolution today, but that initially the significant evolutionary events were relatively *more* probable, and only as evolution progressed did they become less and less so.

This last, somewhat paradoxical, point is intimately associated with another point of difference between the primitive autocatalysis cycle and the modern cell. The latter is capable of extremely accurate information storage, retrieval, and transfer, and the various informational states of the system are readily distinguishable from one another. Thus the information storage capacity of the modern cell is very high, as is the potential number of states of the system. On the other hand, a very primitive system such as we envision would have very inaccurate information handling processes, and poor capacity to distinguish various informational states of the system from one another. The number of (distinguishable) states the primitive system can potentially assume is relatively small and the system thus has a very low information storage capacity. In this situation, perhaps, the *fraction* of (distinguishable) potential states of the primitive system that are viable may be much, much greater than is the case for the modern cell.

Finally let me recapitulate the predictions of the present model regarding the main characteristics of the primitive system. Initially there was only one kind of polyamino acid produced biologically, a "homopolymeric" polybasic amino acid (not a strict homopolymer in the sense that the term is usually employed, but a "group" homopolymer). An analogous situation held for the polynucleotides, they were also group homopolymers—polypurines (high in G), and if typical nucleic acid replication is permitted, polypyrimidines as well. The dominant relationship between polynucleotides and polyamino acids was initially a colinear 1:1 physical binding, probably not unlike that occurring in DNA:histone complexes today (which raises the possibility that one function of modern histones might be in *accumulation* of nucleic acid monomer units). This polynucleotide:polyamino acid relationship is simpler than most of its modern counterparts. Consider nucleic acid tape reading, or enzymatic production of activated mononucleotides, etc.

According to the present model, translation began as a "direct templating" and in addition, "translation" was initially a reciprocal matter, not unidirectional as it is now. The coding ratio for primitive translation was unity, com-

pared to the contemporary value of three. Probably the most important difference between "translation" now and then was the extreme inaccuracy, the ambiguity of the primitive version. In fact, there is doubt that such an imprecise ordering could be termed "translation." This point will be discussed further at some later time.

It is quite clear that a number of conceptual hurdles remain in the way of getting any such system as the one considered here to evolve into a system that translates in the sense of the modern cell. Such considerations far exceed the limits and space given to the present context, however.

I am pleased to acknowledge many delightful discussions on this general topic with Drs. S. Spiegelman and K. Atwood, and in particular with Dr. K. vanHolde. The work reported here was supported by NSF grant GB-6023.

* Earlier articles in this series are entitled: "Evolution of the genetic code," these PROCEEDINGS, **54,** 1546 (1965), and "The molecular basis to the genetic code," these PROCEEDINGS, **55,** 966 (1966), respectively.

[1] Perhaps the best summary to date on the codon assignments and their determination appears in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 31 (1966).

[2] With the exception of three codons that are "assigned" to no amino acid and probably function as "punctuation."

[3] Woese, C., these PROCEEDINGS, **54,** 1546 (1965).

[4] In addition to the published works here—see, for example, Zubay, G., and P. Doty, *Biochim. Biophys. Acta*, **29,** 47 (1958)—there are many more unpublished negative studies—for example, that of R. Britten and C. Woese, who employed a poly U column in an unsuccessful attempt to distinguish among amino acids.

[5] Choughuley, A. S., and R. M. Lemmon, *Nature*, **210,** 628 (1966); Fox, S. W., and K. Harada, *Science*, **133,** 1923 (1961); Fox, S. W., J. E. Johnson, and M. Middlebrook, *J. Am. Chem. Soc.*, **77,** 1048 (1955); Miller, S. L., *Science*, **117,** 528 (1953); Oro, J., *Biochem. Biophys. Res. Commun.*, **2,** 407 (1960); Ponnamperuma, C., and R. Mack, *Science*, **148,** 1221 (1965); Ponnamperuma, C., C. Sagan, and R. Mariner, *Nature*, **199,** 222 (1963); Sanchez, R., J. Ferris, and L. E. Orgel, *Science*, **153,** 72 (1966).

[6] Fox, S. W., and K. Harada, *Science*, **128,** 1214 (1958); Matthews, C. N., and R. E. Moser, these PROCEEDINGS, **56,** 1087 (1966); Schwartz, A. W., and S. W. Fox, *Biochim. Biophys. Acta*, **134,** 9 (1967); Steimann, G., and M. N. Cole, these PROCEEDINGS, **58,** 735 (1967).

[7] Howard, F. B., J. Frazier, M. F. Singer, and H. T. Miles, *J. Mol. Biol.*, **16,** 415 (1966); Huang, W. M., and P. O. P. Tso, *J. Mol. Biol.*, **16,** 523 (1966).

[8] Schweizer, M. P., S. J. Chan, and P. O. P. Tso, *J. Am. Chem. Soc.*, **87,** 5241 (1965); van Holde, K., *Biochem. Biophys. Res. Commun.*, **26,** 717 (1967).

[9] Gamow, G., *Nature*, **173,** 318 (1954).

[10] Pullman, B. J., *Chem. Phys.*, **43,** 5233 (1965).

[11] Woese, C., manuscript in preparation.

[12] Grossenbacher, K. A., and C. A. Knight, in *The Origins of Prebiological Systems and of Their Molecular Matrices* (New York: Academic Press, 1965), p. 173.