# A HYPOTHESIS AND AN EXPERIMENT ON THE ORIGIN OF THE GENETIC CODING PROCESS

Simon Black

Laboratory of Biochemical Pharmacology
National Institute of Arthritis and Metabolic Diseases
National Institutes of Health, Bethesda, Maryland 20014

Received February 24, 1971

From thermodynamic considerations it is argued that the earliest templates for protein synthesis were very unlike present nucleic acids. It is suggested that they were composed of linked uracil derivatives bearing hypothetical side chains that could specifically recognize and interact with individual amino acid side chains (Fig. 1). An experiment supporting the possibility of such a specific interaction is described. The hypothesis accounts for the grouping of amino acids in the present codon table according to their relative hydrophobic character.

In efforts to discern how life originated from inanimate substances (1-9), two paramount questions are encountered: (1) How was a process that now utilizes chemically generated energy initiated? (2) What molecular interactions led to the template function of nucleic acids in protein synthesis? These questions may be intimately related, as illustrated by the following concept of the circumstances that brought the first self-replicating structures into being.

## *I. THE GENERAL HYPOTHESIS*

*Why Chemically Generated Energy Came Into Use.* Basic assumptions in this discussion are that insoluble prebiotic polypeptides formed spontaneously from hydrophobic amino acids; that the polymerization was extremely slow; and that the earliest polypeptides thus formed had a very weak catalytic capability susceptible of evolutionary improvement. With the driving force supplied by the difference in energy levels between soluble amino acids and their insoluble polymers, the infinite slowness of the spontaneous polymerization would favor development of a molecular apparatus that could autocatalytically accelerate polymer accumulation (10). Once evolution of a self-replicating apparatus of this kind was underway, however, a crucially important conflict of direction would develop. The driving force would be in the direction of the more insoluble or thermodynamically most stable structures, but selection for catalytic effectiveness would favor those structures that were more soluble and flexible, and thus less stable. As catalytic function improved, the difference between the energy levels of the prebiotic amino acids and their most effective polymeric products would diminish, and the driving force for formation of more improved catalysts would approach zero. In this circumstance an enormous selective advantage would result from a mechanism to overcome the thermodynamic limitation on evolution of a faster-multiplying, though less durable, primitive organism. Such a mechanism appeared in the form of coupled reactions, through which degradative processes generate chemical energy for polymer synthesis.

*Relation of the Availability of Chemical Energy to the Coding Mechanism.* The most plausible beginning for the coding process would involve a direct interaction of amino acid side chains with a template. Such an interaction would necessarily result in a relatively stable complex of template and protein product. The interaction could thus provide a driving force for the formation of peptide bonds from non-activated carboxyl and amino groups (11), in addition to specifying the sequence of amino acid residues. The stability of the complex, however, would seriously hamper further function of its members as primitive enzymes and templates. For this reason, when chemical energy for polymer synthesis eliminated the need for the driving force of a template-peptide interaction, the original coding mechanism would become obsolete; it would be selectively forced to give way to one involving an indirect or more readily reversible association of amino acid side chains and template. A most important corollary of this idea is that *the original direct coding mechanism would not utilize nucleic acids of the present kind,* which have evolved for highly specialized indirect coding. For the direct coding envisioned, *a template must closely complement and specifically bind the prominent groups of a peptide molecule.*

## *II. A SPECIFIC HYPOTHESIS AND AN EXPERIMENT*

*Hypothetical Primitive Templates and How They May Have Self-Assembled Together with the First Proteins.* Molecular models indicate that the nucleotide bases should be well-suited to form hydrogen bonds with the carbonyl and imide groups of a peptide, and hypothetical side chains extending from the base rings could be positioned to recognize and specifically bind the peptide side chains (Fig. 1). Additional substituent groups must be postulated on the bases to link the units together in a nucleic acid-like polymer. Formation of the first templates and proteins would begin through interaction of such side chain-bearing nucleotide bases with small prebiotic peptides. Stabilization of formed peptide bonds by this interaction would allow the peptide chains to grow in length and number beyond the low equilibrium limits of the reversed hydrolytic reaction through which they originally arose (11). Interbase bonds would likewise form and the template chain similarly grow. With strands of substantial length, the intramolecular forces responsible for secondary and

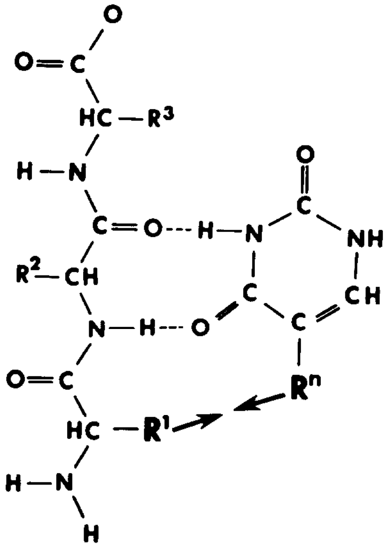

Fig. 1. *Diagrammatic illustration of the binding of a hypothetical uracil derivative to a tripeptide.* Arrows indicate hydrophobic attraction of peptide side chain, $R^1$, to base side chain, $R^n$. Mutual recognition of $R^1$ and $R^n$ forms the basis of the hypothetical primitive code.

tertiary protein structure would become stronger than the inter-chain forces, causing an eventual slow separation of the template from the essentially insoluble primitive protein. The template could then serve as a pattern for additional protein synthesis and for its own replication, and the stable protein particle could function as a very weak catalyst, as previously outlined (10).

*The Rationale for an Experiment.* The severe restriction of position imposed on $R^n$ by the hydrogen bonds in the scheme of Fig. 1 should allow one particular hydrophobic $R^n$ group to bind (12) more firmly than any other to a particular hydrophobic $R^1$ side chain. Thus, of the three conceivable $R^n$ groups, phenyl, phenylmethyl, and phenylethyl, one might bind better than the others to the phenylalanyl side chain of phenylalanylglycylglycine (Phe-Gly-Gly). The formation of a complex would be expected to cause an apparent increase in the solubility of the $R^n$-bearing uracil derivative, because Phe-Gly-Gly is several hundred-fold more soluble than any of the $R^n$ derivatives considered.

*An Experimental Test.* A sterile buffered solution of Phe-Gly-Gly was agitated with solid samples of each of three uracil derivatives in sealed tubes for five weeks at room temperature. With 5-phenylmethyl and 5-phenylethyl uracils, the resulting solutions were indistinguishable by their UV spectra from control solutions containing no Phe-Gly-Gly. With 5-phenyluracil, however, UV absorption did indicate the amount dissolved to be somewhat greater in the presence of the tripeptide. A second experiment with 5-phenyluracil gave results (Fig. 2) closely similar to those of the first. In the latter experiment a trial was also made with Val-Gly-Gly, which produced a much smaller effect on 5-phenyluracil solubility than was observed with Phe-Gly-Gly (Fig. 2). Thus, there is some evidence of specificity for the side chains of both peptide and base ($R^1$ and $R^n$ of Fig. 1). These preliminary observations encourage the undertaking of more extensive experiments, when appropriate nucleotide base derivatives become available in greater quantity and variety.

## *III. THE PRESENT CODE AS A VESTIGIAL RECORD OF ITS OWN EVOLUTION*

The present association of the five most hydrophobic amino acids with trip-

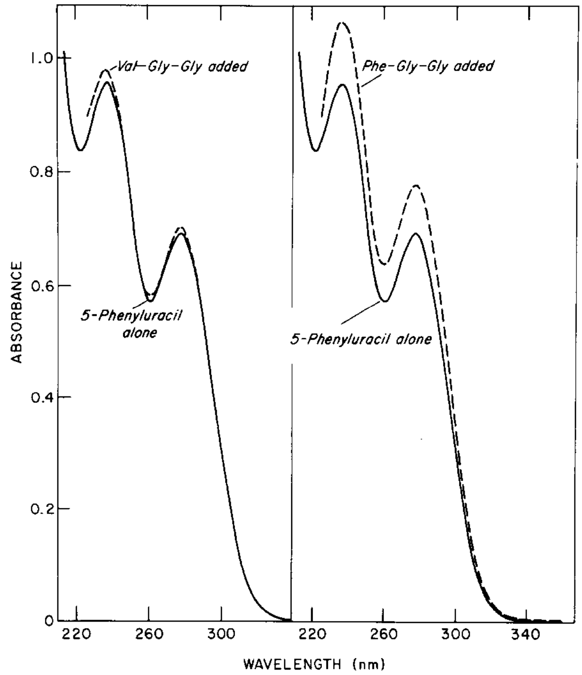

Fig. 2. *UV absorption spectra of solutions of 5-phenyluracil.* 5-Phenyluracil in the amounts of 0.919, 0.905, and 1.025 mg were placed in three pyrex tubes, 10 mm in diameter. Three ml of 5 mM sodium arsenate buffer, adjusted to pH 7.0 with HCl, were added to the first tube; three ml of the same buffer containing 1 mM Phe-Gly-Gly (Cyclo) to the second; and three ml of the buffer containing 1 mM Val-Gly-Gly (Cyclo) to the third. The tubes were sealed in a flame to a length of about 15 cm, and then placed in a boiling water bath for 6 minutes. Upon cooling they were rotated end over end 60 times per minute at about 22° for 5 weeks with a mechanical device. The tubes were then opened and the clear solutions scanned for UV absorption in quartz cells of 1 cm light path in a Cary 14 spectrophotometer, using as references the respective original solutions, without 5-phenyluracil, which had been similarly sterilized and agitated. The 5-phenyluracil was a gift from Professor B. R. Baker.

---

let codons having uracil as the central base (13) can be accounted for if the first proteins consisted largely of hydrophobic amino acid residues, the primitive codons were all derivatives of uracil, and the association of hydrophobic residues with uracil persisted during evolution of the tRNA coding mechanism. Uracil does appear to be the most suitable base for the purpose assigned in Fig. 1, on the basis of positions available for substitution, and a high proportion of hydrophobic side chains would be essential for the relatively insoluble, long-lived primitive proteins. If, then, the later introduction into proteins of the amino acids with the least hydrophobic side chains accompanied introduction of purines into the new code, the presence of aspartic and glutamic acids, lysine, arginine, serine, and glycine in the lower right quadrant of the codon table could be accounted for as a resultant of this coincidental relationship. There would be no need for assuming molecular complementation between codons and amino acids.

## ACKNOWLEDGMENT

I am grateful to Professor B. R. Baker for the uracil derivatives (14).

## REFERENCES

1. Oparin, A. I., *Genesis and Evolutionary Development of Life*, Academic Press, New York, 1968; Moscow, 1966.
2. Haldane, J. B. S., *Origin of Life*, Penguin Books, London, 1954.
3. Calvin, M., *Chemical Evolution*, Oxford University Press, New York and Oxford, 1969.
4. Kenyon, D. H., and Steinman, G., *Biochemical Predestination*, McGraw-Hill, New York, 1969.
5. Bernal, J. D., *The Origin of Life*, World, Cleveland, 1967.
6. Orgel, L. E., *J. Mol. Biol.*, 38, 381 (1968).
7. Crick, F. H. C., *J. Mol. Biol.*, 38, 367 (1968).
8. Woese, C. R., *The Genetic Code*, Harper and Rowe, New York, 1967.
9. Blum, H. F., *Time's Arrow and Evolution*, Princeton University Press, Princeton, New Jersey, 1951.
10. Black, S., *Nature*, 226, 754 (1970).
11. Dobry, A., Fruton, J. S., and Sturtevant, J. M., *J. Biol. Chem.*, 195, 149 (1952).
12. Scheraga, H. A., in H. Neurath (Editor), *The Proteins*, Volume I, Academic Press, New York, 1963, p. 527.
13. Nirenberg, M., Caskey, T., Marshall, R., Brimacombe, R., Kellogg, D., Doctor, B., Hatfield, D., Levin, J., Rottman, F., Pestka, S., Wilcox, M., and Anderson, F., *Cold Spring Harbor Symp. Quant. Biol.*, 31, 11 (1966).
14. Baker, B. R., and Kelley, J. L., *J. Med. Chem.*, 11, 686 (1968).