*A THEORY ON THE MECHANISM OF MESSENGER-RNA SYNTHESIS**

By Geoffrey Zubay

BIOLOGY DEPARTMENT, BROOKHAVEN NATIONAL LABORATORY, UPTON, NEW YORK

Communicated by Milislav Demerec, January 12, 1962

A few years ago, I proposed a template mechanism whereby the genetic message is transmitted from DNA to nuclear RNA in the cells of higher organisms in the form of a specified nucleotide sequence.[1, 2] Since then, further information has accumulated which suggests that a similar type of mechanism, for what has been called the synthesis of messenger-RNA, exists in bacterial cells. The mechanism initially proposed for messenger-RNA synthesis involves the formation of a triple helical intermediate containing a double helical DNA molecule as the template and a single polynucleotide chain of newly formed RNA. The transient formation of the triple helical intermediate is visualized (Fig. 1) as taking place during the polymerization of the RNA from mononucleotide triphosphates on the DNA surface.

The advantages in using the DNA as a template in its two-chain helical form rather than in the single chain form which would lead to a double helical intermediate like DNA (as suggested by Rich[3]) are manifold. (1) Most of the DNA in the cell nucleus is probably in the double-helical form. (2) The formation of the three-chain intermediate necessitates some modification of the DNA structure but no uncoiling. It would be difficult, if not impossible, to attain the speed and selectivity with which DNA responds to the action of inducers and repressors[4] if the DNA first had to uncoil into single chains before it could manufacture messenger-RNA. (3) The three-chain intermediate involves RNA in a relatively loosely bound state with the DNA, so that the RNA could readily dissociate after synthesis. (4) The nucleotide base ratios observed for DNA and nuclear RNA lead to the following relation:

$$\left(\frac{A + T}{G + C}\right)_{DNA} = \left(\frac{G + C}{A + U}\right)_{RNA}. \quad [1]$$

(A = adenylic acid, T = thymidylic acid, G = guanylic acid, C = cytidylic acid, U = uridylic acid; the subscript DNA means the base ratios refer to DNA and the subscript RNA means the base ratios refer to RNA.) A semiregular triple-helical intermediate with reasonable hydrogen bonding, bond lengths, and bond angles has been built that would account for this relationship.[2] (5) If the native double-helical DNA molecule had to unwind into single chains to serve in RNA manufacture, then presumably only one of the two chains of RNA so produced could act as a template. This would be wasteful. On the other hand, with a three-chain intermediate, once the direction of synthesis along the DNA molecule is specified (this would probably be determined by the orientation of the DNA in the chromosome), with the appropriate polymerizing enzyme a unique template RNA should be made.

When the three-chain intermediate was proposed as the mechanism for nuclear RNA synthesis, I found that there was a certain ambiguity resulting from the fact that it was possible to build two different kinds of triple-helical intermediate complexes with molecular models.[2] The base triplets involved in these two triple

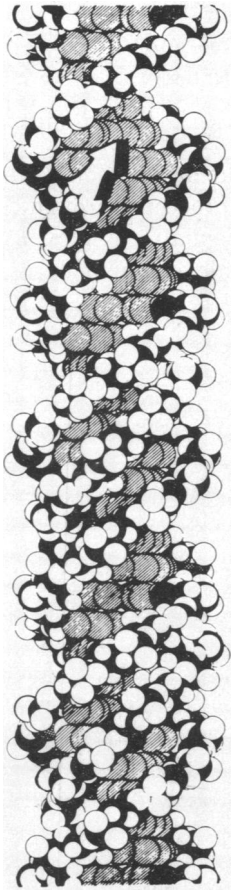

FIG. 1.—Synthesis of messenger RNA on a double-helix DNA template. This model is not exact.

helical complexes are reproduced in Figures 2 and 3 and shall be referred to as template model A and template model B, respectively. It can be seen that only the three-chain intermediate resulting from model A would explain the base ratio relationship observed between DNA and nuclear RNA in animal cells. In contrast, model B would give a base ratio relationship where $(A + T/G + C)_{DNA} = (A + U/G + C)_{RNA}$ and, furthermore, the RNA chain which would be produced by model B as an intermediate would reflect the sequence of nucleotides in one of the DNA chains and be complementary to the other. A three-chain model giving the same base ratios as model B was simultaneously proposed by Stent.[5] The structural features of this model, however, make it less acceptable (Zubay, to be published). It has not yet been possible to find structural criteria for believing

model A should be appreciably more stable than model B. Therefore, the task of choosing which of the two types of RNA chain is fabricated is most likely relegated to the RNA polymerizing enzyme.

In contrast to the base ratio relationship which exists between DNA and messenger-RNA in animal cells, Volkin and Astrachan[6] and others[7] have found that in some bacteriophages and bacteria the purine and pyrimidine base ratios of the messenger-RNA produced *in vivo* and *in vitro* are similar to those of the DNA. It has also been demonstrated that RNA can be induced to form a specific complex with denatured single-chain DNA[8, 9] from which it was made, indicating that the nucleotide sequence of the RNA is complementary to one of the DNA chains, hence, of the same as that of the other. These facts have led others[3] to speculate that messenger-RNA synthesis proceeds through the uncoiling of the DNA and the formation of a two-chain intermediate complex involving one polynucleotide chain of DNA and one polynucleotide chain of RNA with a hydrogen-bonded base

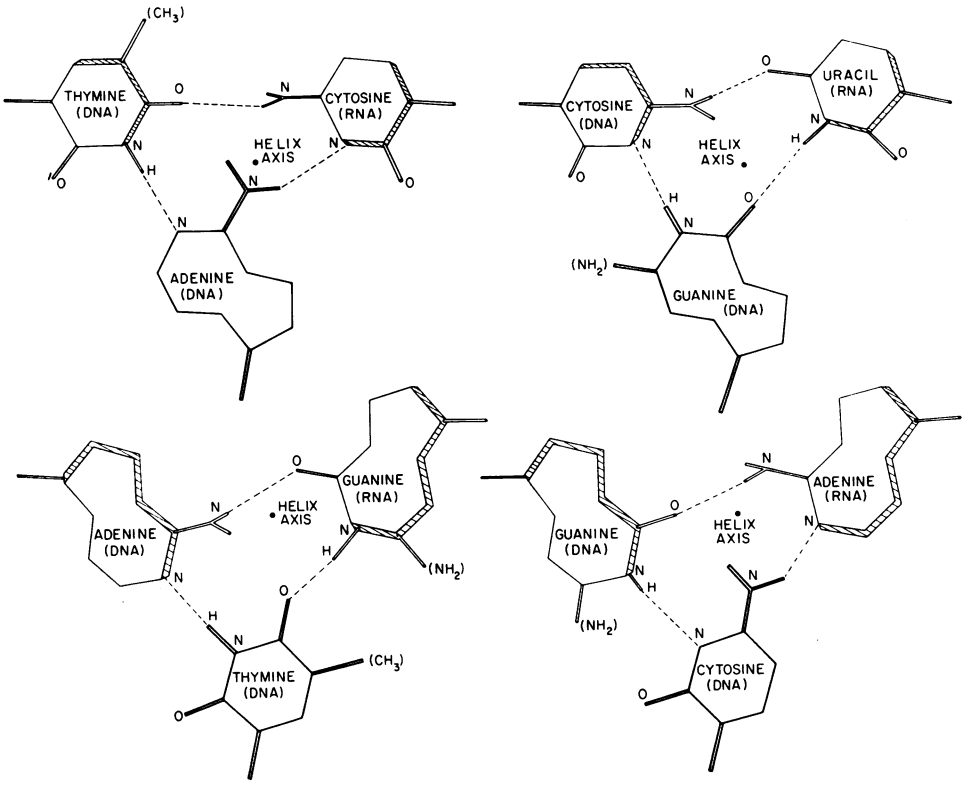

FIG. 2.—Nucleotides triplets involved in template model A for animal cells.

paired structure similar to that of native DNA. Apart from the general reasons already given for favoring a three-chain intermediate, new biochemical facts have come to light which argue strongly in favor of a three-chain intermediate in bacteria also. Firstly, no messenger-RNA synthesized *in vitro* is found in strong combination with single-chain DNA. Only in intricately devised annealing experi-

ments is a strong complex formed.[9] Secondly, single-chain DNA produced by heat denaturation does not prime effectively for messenger-RNA synthesis. If RNA synthesis involved a single-chain DNA template, then heat-denatured DNA should make a better template for RNA synthesis. The reverse is observed,[10] indicating that intact double-helical DNA is necessary for RNA synthesis.

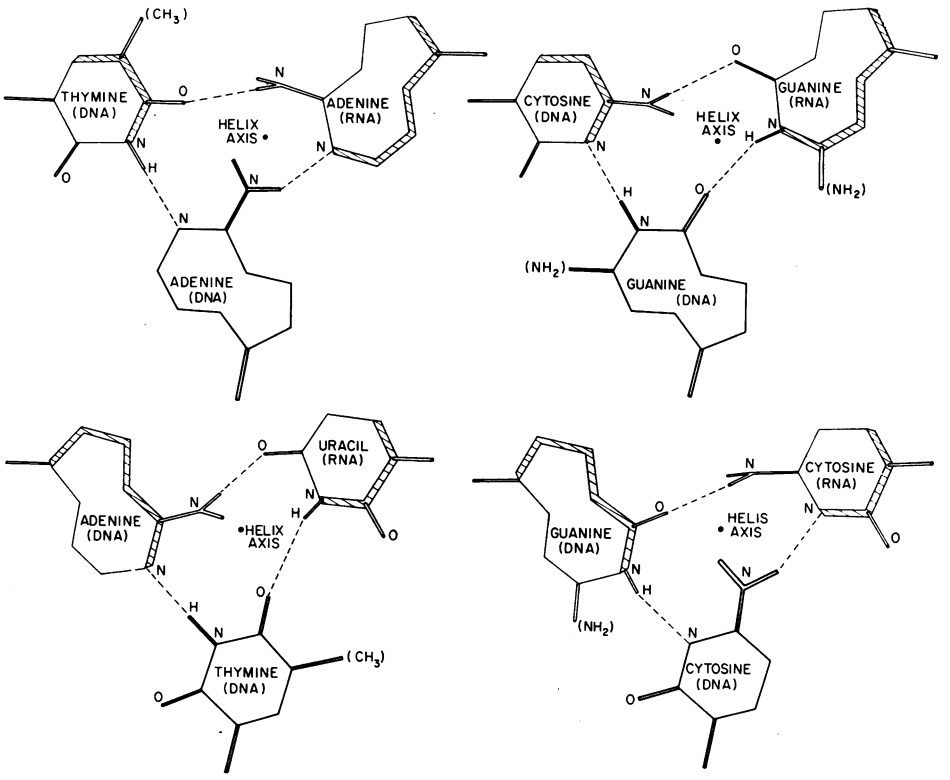

FIG. 3.—Nucleotide triplets involved in template model B for bacterial cells.

How are we to resolve this biochemical evidence suggesting that a three-chain intermediate is involved in messenger-RNA synthesis in bacteria and some bacteriophages with the observations that "complementary RNA" is in fact produced? These apparent incompatabilities can be resolved if we adopt the three-chain intermediate complex represented here by model B as the pathway for messenger-RNA synthesis in these microorganisms. Model B gives the observed base ratios and sequence of bases for messenger-RNA in bacteria[11] and is consistent with the biochemical data that suggests that in bacteria, as in animal cells, a three-chain intermediate is involved in the synthesis of the messenger. I am thus inclined to believe that in most cells messenger-RNA is synthesized on a double-helical DNA template involving the formation of a triple chain intermediate complex between the two deoxypolynucleotide strands of DNA and a single ribopolynucleotide strand of RNA. In animal cells, this appears to lead to a triple-helical intermediate represented by model A (Fig. 2). In bacteria, a different triple-helical intermediate

complex represented by model B (Fig. 3) is suggested. The job of determining the type of RNA synthesized becomes the joint responsibility of the DNA and the RNA polymerizing enzyme. The latter presumably dictates by which model the template will operate and the DNA then dictates the detailed sequence of nucleotides in the RNA.

Evolutionary Consequences of These Two Distinct Modes of Messenger-RNA Synthesis.—Messenger-RNA most likely serves as a template for protein synthesis.[12] Thus, each amino acid becomes chemically linked to an adaptor RNA molecule which has a specific sequence of nucleotides related to the amino acid. The amino acid-adaptor RNA complexes become ordered on a messenger-RNA template and finally the amino acids are polymerized into a polypeptide chain. In this general way, the sequence of amino acids in a protein is believed to be determined by the sequence of nucleotides in a messenger-RNA. The possibility of there being two different ways of producing messenger-RNA from DNA, the one observed in animal cells and the one observed in bacteria, might lead one to believe that there are two different coding schemes for protein synthesis. However, it has been demonstrated[13] that the amino acid code carried by the messenger-RNA is universal by showing that amino acid-adaptor RNA from a bacterium is interchangeable with amino acid-adaptor RNA from an animal cell. In view of this, it is clear that the higher and lower forms of life considered here must have had a common origin, involving basically similar protein-synthesizing systems and RNA. However, it is difficult, if not impossible, to imagine that animal cells evolved from bacteria at that stage when DNA served as their basic genetic substance as we know them today. Because of these facts, we think it likely that RNA functioned as the major genetic apparatus before DNA and that the separation between animal cells and bacteria came at this early stage. As DNA gradually evolved as the gene-bearing system, it apparently did so in the two different ways discussed here.

Note added in proof: A recent paper by M. Chamberlain and P. Berg (these PROCEEDINGS, **48**, 81, 1962) indicates that RNA polymerase from *Escherichia coli* can use single-stranded DNA primer. These results would appear to be contradictory to earlier observations cited here[10] and to invalidate the proposed template mechanism. Alternatively, it may mean that there is more than one kind of RNA polymerase.

* Research carried out at Brookhaven National Laboratory under the auspices of the U.S. Atomic Energy Commission.

[1] Zubay, G., *Nature*, **182**, 112 (1958).

[2] *Ibid.*, 1290 (1958).

[3] Rich, A., *Ann. N. Y. Acad. Sci.*, **81**, 709 (1959).

[4] e.g., Jacob, F., and J. Monod, *J. Mol. Biol.*, **3**, 318 (1961).

[5] Stent, G., *Advances in Virus Research*, **5**, 98 (1958).

[6] Volkin, E., and L. Astrachan, *Virology*, **2**, 149 (1956); Volkin, E., L. Astrachan, and J. L. Countryman, *ibid.*, **6**, 545 (1958).

[7] Weiss, S. B., and T. Nakamoto, these PROCEEDINGS, **47**, 694 (1961).

[8] Spiegelman, S., and B. D. Hall, these PROCEEDINGS, **47**, 1135 (1961); Hall, B. D., and S. Spiegelman, *ibid.*, **47**, 137 (1961); Hayashi, M., and S. Spiegelman, *ibid.*, **47**, 1564 (1961).

[9] Geiduschek, P., T. Nakamoto, and S. B. Weiss, these PROCEEDINGS, **47**, 1405 (1961).

[10] Burma, D. P., H. Kröger, S. Ochoa, R. C. Warner, and J. D. Weill, these PROCEEDINGS, **47**, 749 (1961).

[11] The Stent 3-chain intermediate would also predict the observed base ratios in bacterial messenger-RNA. The chain direction of the RNA would be the opposite of that produced by template model B. Only the RNA made by template model B would be capable of forming a

complementary DNA-RNA double helix on annealing. Therefore, the Stent 3-chain intermediate can be definitely excluded.

[12] Crick, F. H. C., in *The Structure of Nucleic Acids and Their Role in Protein Synthesis* (Cambridge University Press, 1957), p. 25.

[13] Von Ehrenstein, G., and F. Lipmann, these PROCEEDINGS, **47**, 941 (1961).

---

*AN RNA-PROTEIN CODE BASED ON REPLACEMENT DATA**

BY G. ZUBAY AND H. QUASTLER†

BIOLOGY DEPARTMENT, BROOKHAVEN NATIONAL LABORATORY, UPTON, NEW YORK

Communicated by Milislav Demerec, January 18, 1962

The primary structures of proteins are polypeptide chains with amino acids arranged in well-defined sequences. The most acceptable theory of how these sequences are specified is mainly attributable to Crick.[1] According to this theory, an amino acid first becomes linked to an adapter RNA which somewhere contains a sequence of nucleotides characteristic of that particular amino acid. The complex of amino acid and adapter RNA then becomes attached to the messenger-RNA, a molecule considerably larger than the first. The messenger-RNA contains a linear arrangement of sites which are complementary to different adapter molecules. The particular arrangement of sites on the messenger-RNA determines the order in which adapter RNA molecules with their linked amino acids are arranged on the messenger-RNA and, subsequently, the sequence of amino acids in the polypeptide chain. Each site for attachment on the messenger-RNA must be sufficiently specific to admit only adapter RNA with a particular amino acid. In this sense, each site *represents* a particular amino acid, and vice versa. The mutual representation is implicit in the specificity of two processes, that of attaching an amino acid to an adapter RNA and that of attaching an adapter RNA to the messenger-RNA. Hence, the mutual representation of amino acids and sites on the messenger-RNA does not have to involve any physical similarity or complementarity between the two. It could even be that sites on the messenger-RNA may be overlapping, that the nucleotides forming a particular site need not be contiguous, and that sites associated with the same kind of amino acid do not all have to be identical.

An explicit statement of the mapping relations between individual amino acids and individual sites on the messenger-RNA, regardless of the mechanisms which impose the mapping, constitutes the *code* of the mutual representation. Once the detailed mechanisms involved in the mapping process are known, the code follows automatically; in the meantime, attempts to "crack the code" have been made by making logical deductions from partial knowledge (real or assumed) of the mapping process and comparing the results with experimental data. This approach, pioneered by Gamow,[2] has led to a number of theories, some published, some unpublished, some very ingenious, and none completely successful.

There must be at least 18 different sites on the messenger-RNA. It is commonly