THE TRANSCRIPTION OF THE DNA MOLECULE*

By PETER FONG

PHYSICS DEPARTMENT, EMORY UNIVERSITY, ATLANTA, GEORGIA

Communicated by Robert R. Wilson, June 2, 1967

In the transcription process a messenger RNA is made using one strand of the DNA double helix as the template. A mechanism of transcription is proposed here and the related unwinding problem is discussed.

As far as the synthesis of a new nucleotide chain with an old chain as the template is concerned, the transcription process is similar to the replication process. It is natural to assume that the transcription mechanism is somewhat similar to the replication mechanism. A mechanism of DNA replication has been proposed by the author.¹ In this mechanism, hydrogen-bonded base pairs are assumed to drift into the big groove of the DNA molecule side by side with the parent pairs. In the plane of replication (Fig. 1 of ref. 1) two new base pairs are formed by transferring hydrogen bonds from the two old pairs. The two new pairs are then incorporated into the two daughter DNA's by covalent bond formation along the two new chains. The transcription mechanism proposed here is closely related to it.

There are two possibilities concerning the physical configuration of the DNA molecule during transcription. Either the bases of the two strands separate or they remain bonded. Since the base pairing is assumed to provide specificity in biosynthesis, it is natural to assume that the bases of the two strands of DNA separate during transcription so that specificity of hydrogen bonding of the bases will be made available for the synthesis of the messenger RNA.

Once we assume the double helix to be opened, the next question is how the two strands separate and how much of the DNA double helix is opened in transcription. The separation of the two strands involves the problem of unwinding, which has been discussed by the author;² the theory of unwinding previously developed may be applied here. According to this theory, unwinding results from the fluctuation of the random Brownian rotation of the two end parts of the linear DNA molecule, thus unravelling the turns in the middle part of DNA. The question of the extent of opening also occurs in the replication problem, and some of the previous considerations are also valid here.

As in the replication process, the most naïve mechanism one may think of is that the two strands of DNA completely separate before transcription begins. While the complete separation may be achieved by the mechanism of unwinding discussed before,² the time involved in unwinding may be very long (3 days for *E. coli* DNA), and the mechanism is too inefficient. It is also inefficient in the sense that in the transcription of one operon only a small section of the DNA is used, and so it is not necessary to open up the whole DNA molecule. Nevertheless, as in the replication process, this could be the mechanism in the early stages of biological evolution, which was later superseded by more efficient mechanisms developed in the history of evolution.

The consideration of the unwinding time forces us to take the position that if there is any separation of the two strands of DNA in transcription, it must be limited to a small section while the rest remains bonded. Assume the DNA to be a linear

structure, disregarding the complication of circularity, which will be discussed elsewhere in connection with the replication process.³ We need a mechanism to open up a part of the DNA molecule, say, an operon, in the middle of the linear structure without opening up the two ends. The mechanism of unwinding previously proposed which assumes the unwinding to start from the middle rather than from the ends provides just such a mechanism, which is an additional desirable feature of the unwinding mechanism. Take the histidine operon of *Escherichia coli* as an example; the messenger RNA transcribes the whole operon containing ten genes for ten enzymes with a total length of 11,000 nucleotides or 4 microns. The time for unwinding 4 microns of the *E. coli* DNA, according to the previous paper,² is ten seconds. The time for replicating 4 microns of the *E. coli* DNA, according to Cairns, is about eight seconds; the corresponding time for transcription probably is about the same. Thus the unwinding time for one operon is comparable to the transcription time of the operon; the assumption of opening up a section of DNA as much as the length of one operon seems reasonable.

Yet, it is not necessary to assume the whole operon to be completely opened before transcription. Just as in the process of replication, the mechanism will be more efficient the less amount of DNA section is opened. The previously proposed mechanism of replication¹ is the most efficient possible (the ultimate) with the least possible amount of unwinding at one time, that is, 36° for the angular displacement of one base pair. Analogously we may propose the most efficient possible mechanism of transcription on the basis of infinitesimal unwinding, that is, the DNA is opened up only in a length of a few base pairs and transcription takes place in this short section. The region of transcription then moves down the DNA double helix, eventually covering up a whole operon.

The breaking up of the few base pairs involved presumably is due to chemical action associated to the starting of the transcription process. Once broken, according to the mechanism of unwinding previously proposed, the DNA will immediately unwind in this region, resulting in the appearance of a loop in the middle of the double helix. The transcription enzyme presumably will thread its way through the loop, preparing to transcribe. In order to transcribe a section of the DNA, the loop and the enzyme must move over that section, opening up base pairs in front of them and closing up base pairs behind. The enzyme will thus work like a double zipper moving down the chains, one zipper opening the closed chains in the front and the other closing the chains left behind. The mechanical problem associated to this movement can easily be solved. First, the number of hydrogen bonds is conserved (on the coarse-grained average) and so no energy supply is required from outside for opening the base pairs. Second, the mechanism for the forward and backward movement of the loop is provided for by the Brownian rotation of the two parts of the DNA separated by the loop, in the same way as discussed in the previous paper on unwinding. Third, the mechanism to prevent backward movement, and so to make the movement unidirectional, is provided for by the ratchet mechanism of the newly synthesized nucleotide chain in the same way as discussed in the previous paper on replication.

There is one major difference between the transcription mechanism here proposed and the replication mechanism previously discussed. In replication the enzyme more or less stands still while the parent DNA threads its way through the enzyme

by rotation, and the two daughter DNA's are separate. In transcription the enzyme threads its way through the parent DNA spiral in relative rotation, and the messenger RNA produced is wrapped around the DNA in a three-stranded structure. The histidine operon will produce a messenger RNA wrapped around the DNA 1100 turns. A new question thus arises: how to unravel the 1100 turns to separate the RNA from the DNA.

To remove the messenger RNA from the DNA seems to require the turning around of the long RNA chain 1100 turns about the DNA in a viscous medium; the amount of viscous drag on the RNA chain in this process seems tremendous. However, we can bypass this difficulty by the following proposed mechanism of peeling, which is a natural outgrowth of the theory of unwinding previously proposed.

Separation of RNA and DNA may be achieved by letting the viscous medium hold the loose part of RNA still while letting the DNA turn 1100 turns in one direction to unreel the RNA. We can always expect a few turns of the RNA at the end to spring loose from the DNA and to float around in the viscous medium. When the length of this loose part becomes sufficiently long, the viscous drag will become sufficiently large to hold it still in the medium. The DNA molecule executes Brownian rotation all the time, reeling and unreeling the RNA as it rotates. What is required to achieve separation is a fluctuation of the random process of Brownian rotation in which the DNA turns a total of 1100 turns in one direction (a very high-order fluctuation indeed, but still not as high as in the complete unwinding process), thus unreeling the RNA completely. The time required for the *E. coli* DNA to turn 1100 turns in one direction, calculated according to the previous theory of unwinding, is 20 seconds, which is about twice the transcription time of the operon. Since unreeling proceeds simultaneously with transcription, it does not increase the total time of making messenger RNA substantially, and the mechanism of peeling seems reasonable. Thus separation of RNA from DNA may be realized by merely the fluctuation of thermal agitation, and the peeling problem presents no difficulty to the proposed mechanism of transcription. It may also be mentioned that this mechanism of peeling works just the same if the DNA is a closed circle.

We next consider the kinematics of transcription. Many possibilities may be proposed; some of them may have happened in the past evolutionary history and may have been superseded by more efficient mechanisms developed later. Therefore, in the same spirit as in the replication paper, we propose the most efficient possible (the ultimate) mechanism as follows. The configuration and kinematics are basically the same as in the replication process previously proposed (see ref. 1 for the details) except for the following changes: (1) The enzyme recognizes uracil instead of thymine, thus producing RNA instead of DNA. (2) Only one new nucleotide chain is formed. In replication, the enzyme makes two covalent bonds for the two chains; we need to assume that in transcription the process is just the same except that covalent bonds are made only along one chain but fail to be made along the other. (The mononucleotides of the unformed chain are thrown away as waste.) Thus the transcription enzyme leaves behind one double helix of DNA-RNA hybrid and one single DNA chain instead of two double helices, as in replication. (3) The double helix and the single chain then form a three-stranded mole-

cule in which the hydrogen bonds between bases are later transferred from DNA-RNA back to DNA-DNA because of configurational stability. (In replication, the two double helices are separate and cannot form a four-stranded molecule.) The RNA chain, now unbonded, is ready to be peeled off.

The last point is related to the question of why only one strand of RNA is made in transcription when the template DNA has two strands. If instead of (2) we assume that two RNA chains are formed just as in replication, then the products of transcription would be two separate DNA-RNA hybrid helices and there will be no free, unbonded RNA strand. Assuming that two RNA strands are formed, we can still construct a proper mechanism, but all mechanisms will lead to bonded RNA strands. Only when we introduce the asymmetry of synthesizing one strand can we obtain an unbonded RNA strand.

The proposed mechanism carries over the idea of the hydrogen-bonded base pair as the working unit first proposed in replication. This idea has been used by Kubitschek to explain the behavior of mutation without segregation.⁴ It may also provide the explanation of the low rate of spontaneous mutation in reproduction (10^{-8}) due to error of replication in contradistinction with the high rate of error in recognition along a single chain ($\lesssim 10^{-3}$).^{3, 5} By incorporating this idea into the transcription mechanism, the same advantage of low error rate in information transfer is carried over, which is a most desirable feature to possess from the evolutionary point of view.

The mechanism of transcription proposed here is so closely related to that of replication that the transcription enzyme may be regarded and may even arise as a "mutant" of the replication enzyme. One may conceive a slight modification of the replication enzyme to change its recognition of thymine to uracil. Many simple mechanisms are available for the enzyme to lose the ability to make covalent bonds on one chain while retaining the ability on the other. The replication enzyme may be an oligomer with subunits responsible for the formation of covalent bonds of the two chains.⁶ The transcription enzyme could be one that has lost one of the subunits, or one that has one subunit not functioning properly because of deleterious mutation, or one that does not have that subunit to start with. A "mutant" of the replication enzyme with the above two properties (1) and (2) is expected to carry out the transcription process described above as a modified form of replication when it threads itself through a loop opening in the middle of a DNA molecule. The disparity between the molecular weights of the two enzymes does not pose any difficulty in this argument. It may well be that the essential functioning part of the transcription enzyme has such an origin and the rest has a different origin serving other functions not discussed here.

The relation between replication and transcription discussed above leads to an interesting speculation on the origin of life. Current ideas on the origin of life assume the formation of simple compounds from inorganic material, then the formation of complex compounds such as proteins from simple ones, and then the formation of simple organisms from the compounds.⁷ If this is so, it would be very difficult for the organism to develop the ability to reproduce itself. Reproduction by simple direct physical-chemical processes is ruled out by quantum mechanics.⁸ The development of the genetic system and machinery is such a complex phenomenon that the probability of its occurrence at once is practically nil; moreover,

it cannot appear step by step because each step taken out of the whole has no survival value, and the Darwinian mechanism will not select and retain it.

The problem disappears if we take the view that reproduction goes before the organism instead of the view that the organism goes before reproduction. In other words, the very earliest appearance of "life" is just the replicating molecule with no genetic information whatsoever. The replication may be mediated by some other molecule (primitive enzyme, which is certainly not the present form of replication enzyme). Based on the above discussion on transcription, we may conceive that the primitive replication enzyme suffers deleterious changes such that the progeny of replication is not exactly the same as the parent (for example, DNA produces RNA instead of DNA as it should). Many such mutations may take place by which the replicating molecules produce irrelevant products (trash) instead of replicas. As the trash builds up from all the nonsense replicating molecules, the probability increases that one of the trash products may become useful to help replication and thus has a survival value in the evolution process. For example, a protective molecule may help preserve the replicating molecule and has a survival value. The replicating molecule that produces this useful trash will survive better than others; furthermore, it is no longer completely nonsense because it has the information to produce another useful molecule. One may conceive of a step-by-step process in the evolutionary history that many such useful trash products come into being together with their genetic information coded on the replicating molecules. Eventually the useful trash and the replicating molecules associate themselves together in a space with clear demarcation, and this is the beginning of the individual organism which already has the ability to reproduce. Along this line of thought the appearance of life may be understood in terms of step-by-step evolution, which would not be possible in the other approach. The transcription mechanism of today is the result of many evolutionary changes, but it is basically a deleterious mutation of the replication mechanism that eventually becomes useful by chance. Therefore, the transcription mechanism still bears close relation to the replication mechanism.

A general theory of life along this line has been reported,⁹ and will be published in a more detailed sketch soon.

* This work is supported by the National Science Foundation (grant GB-4060) and, in part, by the U.S. Atomic Energy Commission.

¹ Fong, P., these PROCEEDINGS, 52, 641 (1964).

² *Ibid.*, 52, 239 (1964).

³ To be published.

⁴ Kubitschek, H. E., these PROCEEDINGS, 52, 1374 (1964); 55, 269 (1966).

⁵ Kubitschek, H. E., and T. R. Henderson, these PROCEEDINGS, 55, 512 (1966).

⁶ Lee-Huang, S., and L. F. Cavalieri, *Science*, 148, 1474 (1965).

⁷ See, for example, Keosian, J., *The Origin of Life* (Reinhold Publishing Corp., 1964).

⁸ Wigner, E. P., *The Logic of Personal Knowledge. Essays Presented to Michael Polanyi* (London: Routledge and Kegan Paul, 1961); Landsberg, P. T., *Nature*, 203, 928 (1964); Fong, P., to be published.

⁹ Fong, P., *Bull. Am. Phys. Soc.*, 11, 373 (1966); *Chem. Eng. News*, 44, no. 37, 62 (1966).