The biologic synthesis of deoxyribonucleic acid

Nobel Lecture, December 11, 1959

The knowledge drawn in recent years from studies of bacterial transformation and viral infection of bacterial cells combined with other evidences, has just about convinced most of us that deoxyribonucleic acid (DNA) is the genetic substance. We shall assume then that it is DNA which not only directs the synthesis of the proteins and the development of the cell but that it must also be the substance which is copied so as to provide for a similar development of the progeny of that cell for many generations. DNA, like a tape recording, carries a message in which there are specific instructions for a job to be done. Also like a tape recording, exact copies can be made from it so that this information can be used again and elsewhere in time and space.

Are these two functions, the expression of the code (protein synthesis) and the copying of the code (preservation of the race) closely integrated or are they separable? What we have learned from our studies over the past five years and what I shall present is that the replication of DNA can be examined and at least partially understood at the enzymatic level even though the secret of how DNA directs protein synthesis is still locked in the cell.

**DNA structure**

First I should like to review very briefly some aspects of DNA structure which are essential for this discussion. Analysis of the composition of samples of DNA from a great variety of sources and by many investigators revealed the remarkable fact that the purine content always equals the pyrimidine content. Among the purines, the adenine content may differ considerably from the guanine, and among the pyrimidines, the thymine from the cytosine. However, there is an equivalence of the bases with an amino group in the 6-position of the ring, to the bases with a keto group in the 6-position. These facts were interpreted by Watson and Cricks in their masterful hypothesis on the structure of DNA. As shown in Fig. 1, they proposed in
connection with their double-stranded model for DNA, to be discussed presently, that the 6-amino group of adenine is linked by hydrogen bonds to the 6-keto group of thymine and in a like manner guanine is hydrogen-bonded to cytosine, thus accounting for the equivalence of the purines to the pyri-
midines. On the basis of these considerations and the results of X-ray crystallographic measurements by Wilkins and associates, Watson and Crick proposed a structure for DNA in which two long strands are wound about each other in a helical manner. Fig. 2 is diagrammatic representation of a fragment of a DNA chain about ten nucleotide units long. According to physical measurements, DNA chains are on the average 10,000 units long. We see here the deoxypentose rings linked by phosphate residues to form the backbone of the chain; the purine and pyrimidine rings are the planar structures emerging at right angles from the main axis of the chain. Fig. 3 is a more detailed molecular model and gives a better idea of the packing of the atoms in the structure. The purine and pyrimidine bases of one chain are bonded to the pyrimidine and purine bases of the complementary chain by the hydrogen bonds described in Fig. 1. The X-ray measurements have indicated that the space between the opposing chains in the model agrees with

![Molecular model of DNA](image)

Fig. 3. Molecular model of DNA (After M. Feughelman, et al.).
the calculated value for the hydrogen-bond linkage of a purine to a pyrimidine; it is too small for two purines and too large for two pyrimidines. Most rewarding from the biological point of view, the structure provides a useful model to explain how cellular replication of DNA may come about. For, if you imagine that these two chains separate and that a new chain is formed complementary to each of them, the result will be two pairs of strands, each pair identical to the original parent duplex and identical to each other.

Enzymatic approach to the problem of DNA replication

Although we have in the Watson and Crick proposal a mechanical model of replication, we may at this point pose the question: "What is the chemical mechanism by which this super molecule is built up in the cell?" Some sixty years ago the alcoholic fermentation of sugar by a yeast cell was a "vital" process inseparable from the living cell, but through the Buchner discovery of fermentation in extracts and the march of enzymology during the first half of this century, we understand fermentation by yeast as a, now familiar, sequence of integrated chemical reactions. Five years ago the synthesis of DNA was also regarded as a "vital" process. Some people considered it useful for biochemists to examine the combustion chambers of the cell, but tampering with the very genetic apparatus itself would surely produce nothing but disorder. These gloomy predictions were not justified then, nor are similar pessimistic attitudes justified now with regard to the problems of cellular structure and specialized function which face us. High adventures in enzymology lie ahead and many of the explorers will come from the training fields of carbohydrate, fat, amino acid and nucleic acid enzymology.

I feel now, as we did then, that for an effective approach to the problem of nucleic acid biosynthesis it was essential to understand the biosynthesis of the simple nucleotides and the coenzymes and to have these concepts and methodology well in hand. It was from these studies that we developed the conviction that an activated nucleoside 5'-phosphate is the basic biosynthetic building block of the nucleic acids*. You will recall that the main pathways of purine and pyrimidine biosynthesis all lead to the nucleoside 5'-phosphate; they do not, except as salvage mechanisms, usually include the free bases or nucleosides. While the 2' and 3' isomers of the nucleotides are known, they probably arise mainly from certain types of enzymatic degradation of the nucleic acids. You will also recall from the biosynthesis of coen-
zymes, the simplest of the nucleotide condensation products, that it is ATP which condenses with nicotinamide mononucleotide to form diphosphopyridine nucleotide, with riboflavin phosphate to form FAD, with pantetheine phosphate to form the precursor of coenzyme A and so forth. This pattern has been amplified by the discovery of identical mechanisms for the activation of fatty acids and amino acids and it has been demonstrated further that uridine, cytidine and guanosine coenzymes are likewise formed from the respective triphosphates of these nucleosides.

This mechanism (Fig. 4), in which a nucleophilic attack on the pyrophosphate-activated adenyl group by a nucleoside monophosphate leads to the formation of a coenzyme, was adopted as a working hypothesis for studying the synthesis of a DNA chain. As illustrated in Fig. 5, it was postulated that the basic building block is a deoxynucleoside 5'-triphosphate which is attacked by the 3'-hydroxyl group at the growing end of a polydeoxynucleotide chain; inorganic pyrophosphate is eliminated and the chain is lengthened by one unit. The results of our studies on DNA synthesis, as will be mentioned presently, are in keeping with this type of reaction.

Properties of the DNA-synthesizing enzyme

First let us consider the enzyme and comment on its discovery. Mixing the triphosphates of the four deoxynucleosides which commonly occur in DNA with an extract of thymus or bone-marrow or of Escherichin coli would not be expected to lead to the net synthesis of DNA. Instead, as might be expected, the destruction of DNA by the extracts of such cells and tissues was by far the predominant process and one had to resort to the use of more subtle devices for detection of such a biosynthetic reaction. We used a 14C-
labeled substrate of high specific radioactivity and incubated it with ATP and extracts of Escherichia coli, an organism which reproduces itself every 20 minutes. The first positive results represented the conversion of only a very small fraction of the acid-soluble substrate into an acid-insoluble fraction (50 or so counts out of a million added). While this represented only a few \( \mu \text{mole} \)s of reaction, it was something. Through this tiny crack we tried to drive a wedge, and the hammer was enzyme purification\(^1\). This has been and still is a major preoccupation. Our best preparations are several thousand-fold enriched with respect to protein over the crude extracts, but there are still contaminating quantities of one or more of the many varieties of nuclease and diesterase present in the coli cell. The occurrence of what appears to be a similar DNA-synthesizing system in animal cells as well as in other bacterial species has been observed\(^2\). We must wait for purification of the enzymes from these sources in order to make valid comparisons with the coli system.
The requirements for net synthesis of DNA with the purified coli enzyme are shown in the equation in Fig. 6. All four of the deoxynucleotides which form the adenine-thymine and guanine-cytosine couples must be present. The substrates must be the tri- and not the diphosphates and only the deoxy sugar compounds are active. DNA which must be present may be obtained from animal, plant, bacterial or viral sources and the best indications are that all these DNA samples serve equally well in DNA synthesis provided their molecular weight is high. The product, which we will discuss in further detail, accumulates until one of the substrates is exhausted and may be 20 or more times greater in amount than the DNA added and thus is composed to the extent of 95% or more of the substrates added to the reaction mixture. Inorganic pyrophosphate is released in quantities equimolar to the deoxynucleotides converted to DNA.

Should one of these substrates be omitted, the extent of reaction is diminished by a factor of greater than 10^4 and special methods are now required for its detection. It turns out that when one of the deoxynucleotide substrates is lacking, an extremely small but yet significant quantity of nucleotide is linked to the DNA primer. We have described this so-called “limited reaction”, and have shown that under these circumstances a few deoxynucleotides are added to the nucleoside ends of some of the DNA chains but that further synthesis is blocked for lack of the missing nucleotide. Current studies suggest to us that this limited reaction represents the repair of the shorter strand of a double helix in which the strands are of unequal length, and that the reaction is governed by the hydrogen bonding of adenine to thymine and of guanine to cytosine.

When all four triphosphates are present, but when DNA is omitted, no reaction at all takes place. What is the basis for this requirement? Does the DNA function as a primer in the manner of glycogen or does it function as a template in directing the synthesis of exact copies of itself? We have good reason to believe that it is the latter and as the central and restricted theme of
this lecture I would like to emphasize that it is the capacity for base-pairing by hydrogen bonding between the preexisting DNA and the nucleotides added as substrates that accounts for the requirement for DNA.

The enzyme we are studying is thus unique in present experience in taking directions from a template - it adds the particular purine or pyrimidine sub-

![Fig. 7. Mechanism for enzymatic DNA replication.](image)

strate which will form a hydrogen-bonded pair with a base on the template (Fig. 7). There are five major lines of evidence that I would like to present to support this thesis.

**Physical properties of enzymatically synthesized DNA**

The first line of evidence is derived from studies of the physical nature of the DNA produced by the enzyme. It may be mentioned again that in these descriptions as in those of the chemical nature of DNA, to be discussed shortly, 90-95% of the DNA sample comes from the substrates used in the reaction. From collaborative studies with Dr. Howard K. Schachman, to whom we are greatly indebted, it can be said that the enzymatic product is
indistinguishable from high-molecular weight, double-stranded DNA isolated from nature. It has sedimentation coefficients in the neighbourhood of 25, reduced viscosities of 40 deciliters per gram and, on the basis of these measurements, it is believed to be a long, stiff rod with a molecular weight of about 6 million. Upon heating the DNA, the rod collapses and the molecule becomes a compact, randomly coiled structure; it may be inferred that the hydrogen bonds holding the strands together have melted and this is borne out by characteristic changes in the viscometric and optical properties of the molecule. Similar results are found upon cleavage of the molecule by pancreatic deoxyribonuclease. In all these respects the enzymatically synthesized DNA is indistinguishable from the material isolated from nature, and may thus be presumed to have a hydrogen-bonded structure similar to that possessed by natural DNA.

Would one imagine that the collapsed jumbled strands of heated DNA would serve as a primer for DNA synthesis? Very likely one would think not. Guided by intuition derived from everyday experience with a jumbled strand of twine one might regard this as a hopeless template for replication. It turns out that the collapsed DNA is an excellent primer and the non-viscous, randomly coiled, single-stranded DNA leads to the synthesis of highly viscous, double-stranded DNA. Sinsheimer has isolated from the tiny ØX 174 virus a DNA which appears to be single-stranded. Like heated DNA it has proved to be an excellent primer and a favorable material in current studies for demonstrating in density gradient sedimentations that it is progressively converted to a double-stranded condition during the course of enzymatic synthesis.

While a detailed discussion of the physical aspects of replication is not feasible in this lecture, it should be mentioned that the DNA in the single-stranded condition is not only a suitable primer but is the only active form when the most purified enzyme preparations are used. With such coli preparations, the native, double-stranded DNA is inert unless it is heated or pretreated very slightly with deoxyribonuclease. Bollum has made similar observations with the enzyme that he has purified from calf thymus.

Substitution of analogues in DNA synthesis

The second line of evidence is derived from studies of the activity of the substrates when substitutions are made in the purine and pyrimidine bases
From the many interesting reports on the incorporation of bromouracil, azaguanine and other analogues into bacterial and viral DNA, it might be surmised that some latitude in the structure of the bases can be tolerated provided there is no interference with their hydrogen bondings. When experiments were carried out with deoxyuridine triphosphate or 5-bromodeoxyuridine triphosphate, it was found that they supported DNA synthesis when used in place of thymidine triphosphate but not when substituted for the triphosphates of deoxyadenosine, deoxyguanosine or deoxycytidine. As already described, S-methyl- and 5-bromocytosine specifically replaced cytosine; hypoxanthine substituted only for guanine; and, as just mentioned, uracil and 5-bromouracil specifically replaced thymine. These findings are best interpreted on the basis of hydrogen bonding of the adenine-thymine and guanine-cytosine type.

Along these lines it is relevant to mention the existence of a naturally occurring "analogue" of cytosine, hydroxymethyl cytosine (HMC), which is found in place of cytosine in the DNA of the coli bacteriophages of the T-even series. In this case the DNA contains equivalent amounts of HMC and guanine and, as usual, equivalent amounts of adenine and thymine. Of additional interest is the fact that the DNA’s of T₅, T₆, and T₇ contain glucose linked to the hydroxymethyl groups of the HMC in characteristic ratios although it is clear that in T₅ and T₆ some of the HMC groups contain no glucose. These characteristics have posed two problems regarding the synthesis of these DNA’s which might appear to be incompatible with the simple base-pairing hypothesis. First, what mechanism is there for preventing the inclusion of cytosine in a cell which under normal conditions has deoxycytidine triphosphate and incorporates it into its DNA? Secondly, how does one conceive of the origin of the constant ratios of glucose to HMC in DNA if the incorporation were to occur via glucosylated and non-glucosylated HMC nucleotides? Our recent experiments have shown that the polymerase reaction in the virus-infected cell is governed by the usual hydrogen-bonding restrictions but with the auxiliary action of several new enzymes developed specifically in response to infection with a given virus. Among the new enzymes is one which splits deoxycytidine triphosphate and thus removes it from the sites of polymerase action. Another is a type of glucosylating enzyme which transfers glucose from uridine diphosphate glucose directly and specifically to certain HMC residues in the DNA.
The third line of evidence is supplied by an analysis of the purine and pyrimidine base composition of the enzymatically synthesized DNA. We may ask two questions. First, will the product have the equivalence of adenine to thymine and of guanine to cytosine that characterize natural DNA? Secondly, will the composition of the natural DNA used as primer influence and determine the composition of the product? In Fig. 8 are the results which answer these two questions. The experiments are identical except that in each case a different DNA primer was used: Mycobacterium phlei, Escherichia coli, calf thymus and phage T₂ DNA. In answer to the first question it is clear that in the enzymatically synthesized DNA, adenine equals thymine and guanine equals cytosine so that the purine content is in every case identical to the pyrimidine. In answer to the second question it is again apparent that the characteristic ratio of adenine-thymine pairs to guanine-cytosine pairs of a given DNA primer is imposed rather faithfully on the product that is synthesized. Whether these measurements are made with isotopic tracers when the net DNA increase is only 1% or if it is 1,000% the results are the same. It can be said further that it has not been possible to distort these base ratios by using widely differing molar concentrations of substrates or by any other means. In the last line of Fig. 8 is a rather novel "DNA" which is synthesized under conditions that I will not describe here. Suffice it to say that after very long lag periods a copolymer of

<table>
<thead>
<tr>
<th>DNA</th>
<th>A</th>
<th>T</th>
<th>G</th>
<th>C</th>
<th>A+G</th>
<th>A+T</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mycobacterium phlei</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primer</td>
<td>0.65</td>
<td>0.66</td>
<td>1.35</td>
<td>1.34</td>
<td>1.01</td>
<td>0.49</td>
</tr>
<tr>
<td>product</td>
<td>0.66</td>
<td>0.65</td>
<td>1.34</td>
<td>1.37</td>
<td>0.99</td>
<td>0.48</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primer</td>
<td>1.00</td>
<td>0.97</td>
<td>0.98</td>
<td>1.05</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>product</td>
<td>1.04</td>
<td>1.00</td>
<td>0.97</td>
<td>0.98</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td><em>Calf thymus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primer</td>
<td>1.12</td>
<td>1.08</td>
<td>0.85</td>
<td>0.85</td>
<td>1.02</td>
<td>1.29</td>
</tr>
<tr>
<td>product</td>
<td>1.12</td>
<td>1.08</td>
<td>0.85</td>
<td>0.85</td>
<td>1.02</td>
<td>1.29</td>
</tr>
<tr>
<td><em>Bacteriophage T₂</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primer</td>
<td>1.33</td>
<td>1.33</td>
<td>0.67</td>
<td>0.70</td>
<td>0.98</td>
<td>1.02</td>
</tr>
<tr>
<td>product</td>
<td>1.33</td>
<td>1.33</td>
<td>0.69</td>
<td>0.70</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td><em>A-T Copolymer</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.99</td>
<td>1.93</td>
<td>&lt; 0.05</td>
<td>&lt; 0.05</td>
<td>1.03</td>
<td>&gt; 40</td>
</tr>
</tbody>
</table>

Fig. 8. Chemical composition of enzymatically synthesized DNA with different primers.
deoxyadenylate and thymidylate (A-T) develops which has the physical size and properties of natural DNA and in which the adenine and thymine are in a perfectly alternating sequence. When this rare form of DNA-like polymer is used as a primer, new A-T polymer synthesis starts immediately and even though all four triphosphates be present, no trace of guanine or cytosine can be detected in the product. The conclusion from these several experiments thus seems inescapable that the base composition is replicated in the enzymatic synthesis and that hydrogen-bonding of adenine to thymine and guanine to cytosine is the guiding mechanism.

Enzymatic replication of nucleotide sequences

The fourth line of evidence which I would like to cite is drawn from current studies of base sequences in DNA and their replication. As I have suggested already, we believe that DNA is the genetic code; the four kinds of nucleotides make up a four-letter alphabet and their sequence spells out the message. At present we do not know the sequence; what Sanger has done for peptide sequence in protein remains to be done for nucleic acids. The problem is more difficult, but not insoluble.

Our present attempts at determining the nucleotide sequences will be described in detail elsewhere and I will only summarize them here. DNA is enzymatically synthesized using $^{32}P$ as label in one of the deoxynucleoside triphosphates; the other three substrates are unlabeled. This radioactive phosphate, attached to the 5-carbon of the deoxyribose, now becomes the bridge between that substrate molecule and the nucleotide at the growing end of the chain with which it has reacted (Fig. 9). At the end of the synthetic reaction (after some $10^{11}$ diester bonds have been formed), the DNA is isolated and digested enzymatically to yield the 3' deoxynucleotides quantitatively. It is apparent (Fig. 9) that the P atom formerly attached to the S-carbon of the deoxynucleoside triphosphate substrate is now attached to the 3-carbon of the nucleotide with which it reacted during the course of synthesis of the DNA chains. The $^{32}P$ content of each of the 3' deoxynucleotides, isolated by paper electrophoresis, is a measure of the relative frequency with which a particular substrate reacted with each of the four available nucleotides in the course of synthesis of the DNA chains. This procedure carried out four times, using in turn a different labeled substrate, yields the relative frequencies of all the sixteen possible kinds of dinucleotide (nearest neighbor) sequences.
Such studies have to date been carried out using DNA primer samples from six different natural sources. The conclusions are:
1. All 16 possible dinucleotide sequences are found in each case.
2. The pattern of relative frequencies of the sequences is unique and reproducible in each case and is not readily predicted from the base composition of the DNA.
3. Enzymatic replication involves base-pairing of adenine to thymine and guanine to cytosine and, most significantly:
4. The frequencies also indicate clearly that the enzymatic replication produces two strands of opposite direction, as predicted by the Watson and Crick model.

These studies and anticipated extensions of them should yield the dinucleotide frequencies of any DNA sample which can serve as an effective primer for enzymatic replication and thus provide some clues for deciphering the DNA code. Unfortunately this method does not provide information about trinucleotide frequencies but we are hopeful that with the improvement of enzymatic tools for analysis and chromatographic techniques for isolation some start can be made in this direction.
Returning to the earlier-stated requirement for all four deoxynucleoside triphosphates and DNA in order to obtain DNA synthesis, we can now regard and understand these requirements as another and final line of evidence for hydrogen bonding. Without added DNA there is no template for hydrogen bonding and without all four triphosphates synthesis stops early and abruptly for lack of a hydrogen bonding mate for one of the bases in the template.

**Summary**

The enzymatic approaches to the problem of DNA replication and the properties of the DNA-synthesizing enzyme purified from Escherichia coli have been sketched. The unifying and basic generalization about the action of this enzyme is that it catalyzes the synthesis of a new DNA chain in response to directions from a DNA template; these directions are dictated by the hydrogen-bonding relationship of adenine to thymine and guanine to cytosine. The experimental basis for this conclusion is derived from the observations of: (1) The double-stranded character of the enzymatically synthesized DNA and its origin from a single-stranded molecule, (2) the pattern of substitution of analogues for the naturally occurring bases, (3) the replication of the chemical composition, (4) the replication of the nucleotide (nearest neighbor) sequences and the antiparallel direction of the strands, and (5) the requirement for all four deoxynucleoside triphosphates (adenine, thymine, guanine, and cytosine) and DNA for DNA synthesis.

In closing may I repeat what was said at the banquet last night: Any credit for the work cited here is shared by my colleagues in New York, Bethesda, Saint Louis and Stanford, and by the whole international community of chemists, geneticists and physiologists, which is truly responsible for the progress in nucleic acid biochemistry.

3. G. W. Beadle, in Chemical Basis of Heredity, W. D. McElroy and B. Glass (Eds.),
Johns Hopkins Press, Baltimore, 1957, p. 3.
Quad. Biol., 18 (1953) 123.
834.
8. A. Kornberg, in The Chemical Basis of Heredity, W. D. McElroy and B. Glass (Eds.),
(1958) 163.
(1958) 171.
17. H. K. Schachman, I. R. Lehman, M. J. Bessman, J. Adler, E. S. Simms, and A.
22. F. Weygand, A. Wacker, and Z. Dellweg, Z. Naturforsch., 7 b (1952) 19; D. B.
Dunn and J. D. Smith, Nature, 174 (1954) 305; S. Zamenhof and G. Grisboff,
197 (1952) 199.
5892.