

STRUCTURE AND FUNCTION OF DNA

F. H. C. CRICK B.Sc.

For a number of years a new scientific unity has been developing from a synthesis in which chemistry and physics have become closely linked with biology. On the biological side cytologists are concerned with what happens to the nucleo-proteins and the nucleic acids present in plant and animal cells. These compounds are of great interest to biochemists and biophysicists, and during 1953 important advances have been made on this front as a result of discoveries in British, U.S. and Canadian laboratories.

One of the fundamental problems of biology is the manner in which the hereditary factors are copied and passed on from one generation to another. In particular we should like to know *in terms of atoms and molecules* just how these factors carry the genetical information, how the cell produces an exact copy of them, and how they exert their influence on the cell.

The hereditary factors are believed to be carried by the chromosomes, the rather fibrous bodies found inside the nucleus of a living cell. Now chromosomes mainly consist of two kinds of substance—*protein* and *nucleic acid*—and much experimental work has been carried out to discover their chemical nature. As far as nucleic acid is concerned this work has been conspicuously successful, and its general chemical formula is now known. Until recently, however, hardly anything has been found out about its 'structure'—the arrangement of the molecule in space—but within the last year there have been exciting developments in this direction, and it is with these that this article is mainly concerned.

Two sorts of nucleic acid are found in living cells. The one in which we are interested in this article is called *deoxyribonucleic acid*, or DNA for short. Its general formula is very simple to grasp. It consists of a very long chain made up of alternate sugar and phosphate groups. The sugar is always the same sugar (known as deoxyribose, or ribose with one oxygen missing) and it is always joined on to the phosphates in the same way—by ester linkages—so that this long chain is perfectly regular, repeating the same phosphate-sugar sequence over and over again.

This chain is only part of the molecule, however, for every sugar has a 'base' attached to it, as shown in Fig. 1, but the base is not always the same base. Commonly four different types are found. They are all flat heterocyclic rings—two purines, known as adenine and guanine, and two pyrimidines, known as thymine and cytosine: (their formulae appear later, in Figs. 5 and 6). As far as is known the order in which they follow one another along the chain is irregular, and a typical bit of DNA might have the formula shown in Fig. 2, in which the names of the bases have been written in at random. It is because the exact sequence of the bases is not known that one can only say that the *general* formula of DNA is established.

It should not be thought that this rather simple formula was found in a day. It has taken more than twenty-five years' work by organic chemists to prove it, and it should be reckoned as one of the major achievements of organic chemistry applied to biology. It is the foundation for all the ideas described in the rest of this article.

During the last few years biochemists, using improved modern methods (in particular, chromatography and

ultra-violet absorption), have tackled the problem of the relative amounts of the four bases in DNA from different species. The leaders in this field have been Dr. E. Chargaff and his colleagues at Columbia University, New York, and Dr. G. R. Wyatt in Canada. They have shown that the relative amounts of the various bases can vary from species to species, but appear to be fixed (within the limits of experimental error) for a given species, irrespective of which individual or which organ the DNA is taken from.

The chemical formula does not by itself tell one the shape of the molecule. This is because there are many single bonds in the phosphate-sugar chain, and as rotation is in theory possible about all of them, one might expect

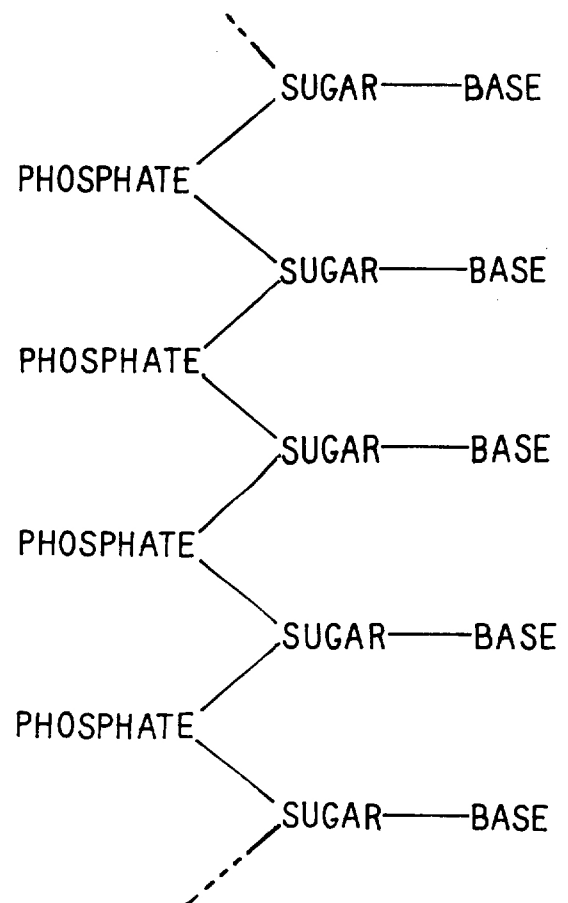


FIG. 1. The general chemical formula of a single chain of DNA.

the chain to coil about in a rather random manner. Curiously enough, both measurements of the viscosity and light-scattering of DNA in solution, and pictures of dry DNA in the electron microscope all suggest that the molecule is long, thin, and fairly straight, rather like a stiff bit of cord. The width of the molecule, as measured for example by the length of the 'shadow' in the electron microscope, is about 20 Å. The length of the DNA inside the cell is perhaps very great indeed, and even after it has been extracted, a process which may break it up somewhat, it is still fairly long. A typical figure would be, say, 30,000 Å, or 3μ.*

None of these methods tells us anything about the detailed arrangement in space of the atoms inside the molecule (the typical distance between atoms bonded together in organic molecules is 1½ Å). For this it is necessary to use X-ray diffraction. The DNA from a tissue—the favourite is the thymus gland of the calf—can be extracted by mild methods, and then drawn into long fibres. The fibres can be mounted in an X-ray diffraction camera in the usual manner, and the diffraction pictures recorded on a photographic plate. The pioneer work was done by Professor W. L. Astbury and Dr. Florence Bell before the war, but almost all the recent work has been done by Dr. M. H. F. Wilkins, Dr. Rosalind Franklin and their co-workers at King's College, London. The diffraction pictures they have obtained are of an extremely high quality. If the structure could be deduced unambiguously from the X-ray photographs the solution would have been easy, but as is well known, this is not possible. From a postulated structure one can work out mathematically the expected diffraction pattern, but there is no direct way of going from the X-ray picture to the structure. In mathematical language this is because the diffraction pattern gives the amplitudes of the Fourier components of the electron density, but not their relative phases.

Nevertheless, certain facts emerged straight away from the X-ray work. Firstly, it was found that there were two distinct X-ray patterns, depending upon the humidity. One of these, which occurred when the water content was about 40%, was crystalline; that is, there was three-dimensional order present. When the humidity was raised the fibres took up more water, increased in length by about 30%, and gave a different pattern which tended to be paracrystalline; that is, the molecules were all parallel to each other, but packed side by side in a less regular manner.

Secondly, it was found that DNA from different sources, containing different amounts of the four bases, gave apparently identical X-ray patterns. This was rather surprising. The X-ray reflections do not extend to very small spacings, so that the picture they would give of the molecule might be rather fuzzy, and one base might look rather like another. What the identity of the different X-ray pictures suggested was that the broad arrangement of the molecule was independent of the exact sequence of the bases. But then the fact that DNA, with its irregular sequence of bases, gave a truly crystalline picture at all was rather surprising.

The third thing the X-ray pictures showed was that the crystallographic repeat distance in the fibre direction was rather long—28 Å in the crystalline form, 34 Å in the paracrystalline form—compared with the maximum possible

* 1 Å = 10⁻⁴μ = 10⁻⁸ cm.

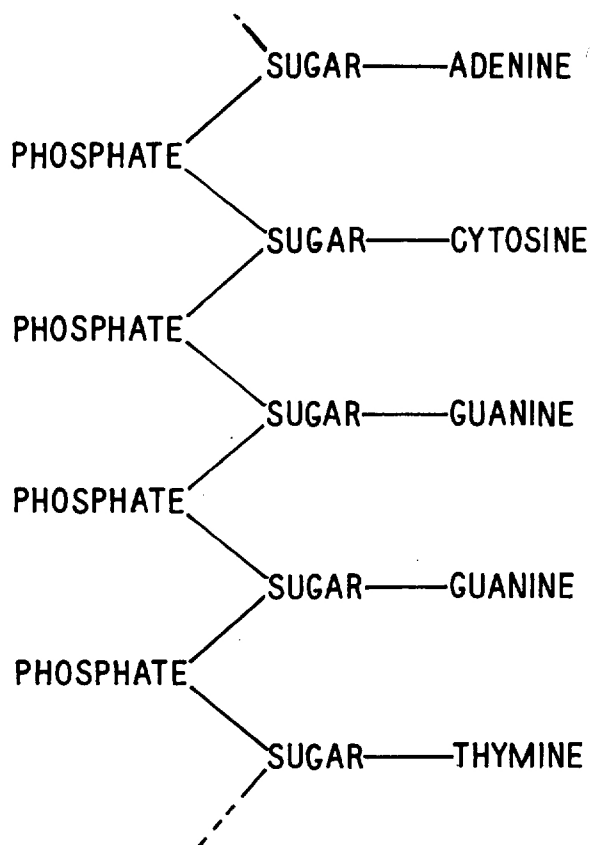


FIG. 2. A typical stretch of a DNA chain. The names of the bases have been written in at random.

repeat distance of the chain, which when fully extended is only about 7 Å from phosphate to phosphate. This showed that there were several chemical repeats of the phosphate-sugar chain in one structural repeat.

Dr. J. D. Watson and I, working in the Medical Research Council Unit in the Cavendish Laboratory, were convinced that one could get somewhere near the structure by building models. There is a great deal of information available about the bond distances between atoms, about the angles between bonds, and also about the size of atoms—the so-called van der Waals distances. All this can be embodied in scale models. Because rotation is possible about single bonds, the models when first built are not stiff, but somewhat flexible. The problem is rather like a three-dimensional jig-saw puzzle with curious pieces joined together by rotatable joints.

Stimulated by the preliminary X-ray results of the King's College workers, we attempted to build models which would be consistent with their data for the paracrystalline form. We assumed that since the phosphate-sugar chain was chemically regular it would probably take up a regular arrangement in space. In other words we assumed that the configuration of any one phosphate-sugar group would look exactly like all the others. It can easily be shown that the only possible form for a chain, the links of which fulfil

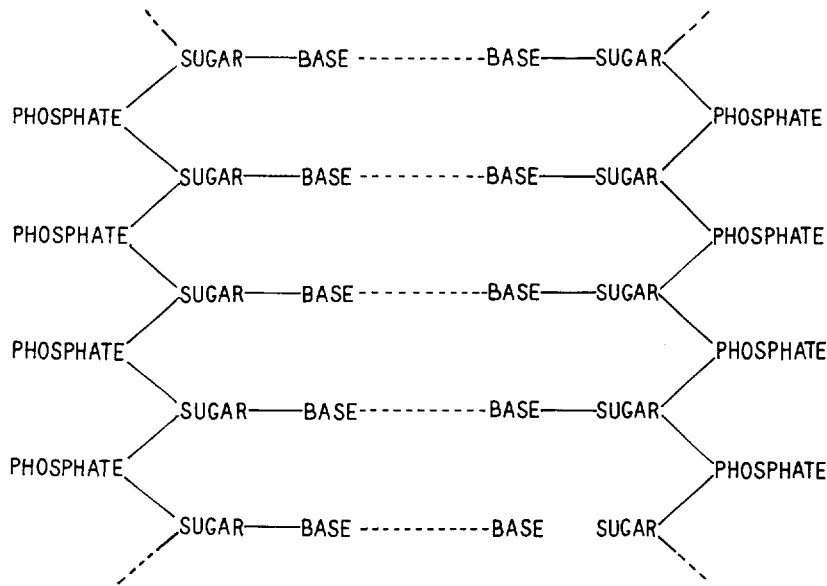


FIG. 3. The general formula of a pair of DNA chains. The dotted lines symbolise the hydrogen bonds holding the two chains together.

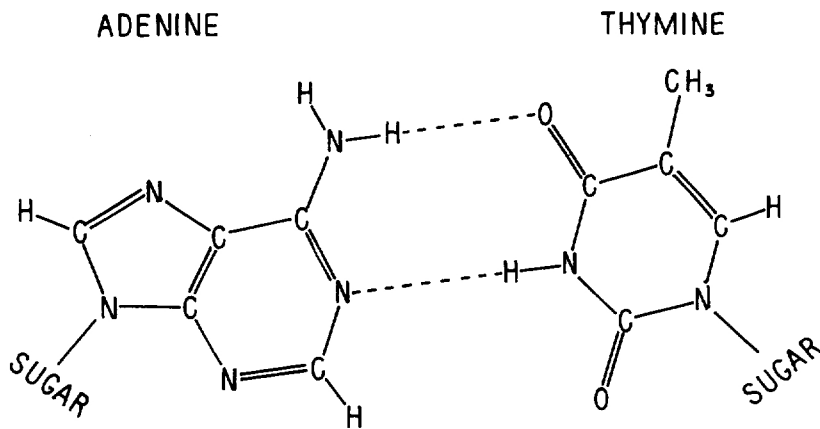


FIG. 5. The pairing of adenine and thymine. The hydrogen bonds are shown dotted. The two sugars belong to different chains.

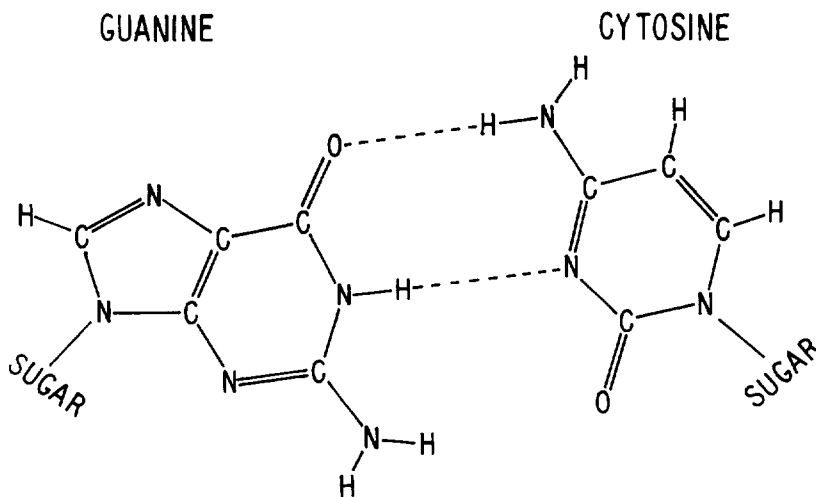


FIG. 6. The pairing of guanine and cytosine. The hydrogen bonds are shown dotted. The two sugars belong to different chains.

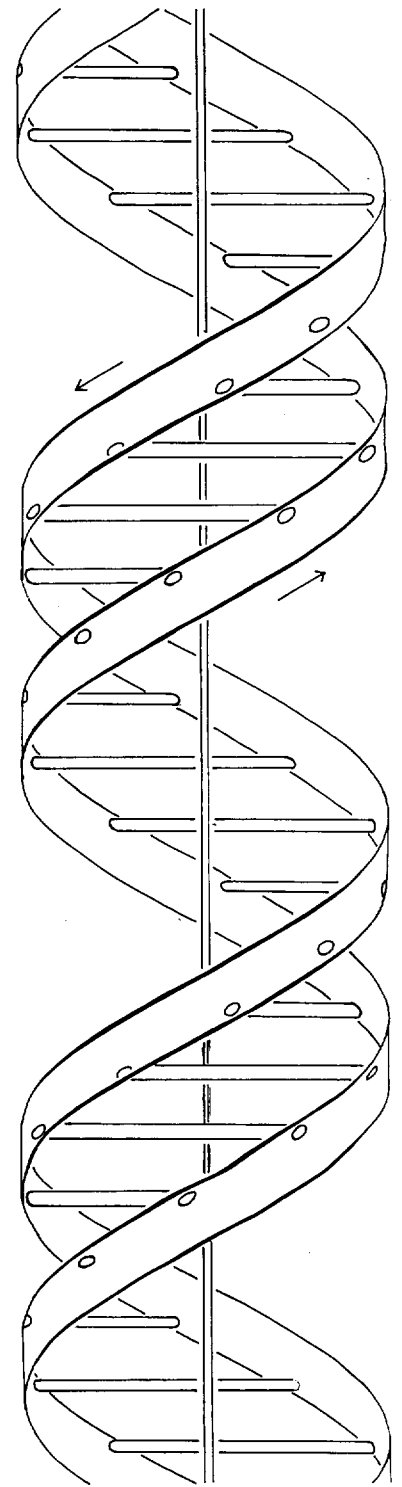


FIG. 4. The proposed structure for DNA shown diagrammatically. The two phosphate-sugar chains are symbolised by ribbons, and the pairs of bases holding the chains together are represented as horizontal rods. The vertical line marks the imaginary fibre axis.

this condition, is a helix,* or a degenerate helix, such as a straight line or a circle. Notice that a helix accounts rather naturally for the long repeat distance of the structure, as this would correspond to one turn of the helix. This restriction—that the phosphate-sugar groups are spatially uniform—is a great help in model building as it reduces very considerably the number of possibilities that have to be explored. Indeed, at first we were unable to build any satisfactory model consistent with our assumptions, but eventually we arrived at a structure which we now believe to be correct in its broad outlines.

This particular model does not contain just one DNA chain, but a pair of them, wound round a common axis. These two chains are linked together by their bases. A base on one chain is joined by weak *physical* bonds to a base at the same level on the other chain, and all the bases are paired off in this way right along the structure. This is shown diagrammatically in Fig. 3. The general appearance of the structure is shown in a symbolic manner in Fig. 4, in which the two ribbons represent the phosphate-sugar chains, and the pairs of bases holding them together are symbolised as horizontal rods. It will be found that this figure looks exactly the same upside down, and to preserve this feature we have built our model so that the actual sequence of atoms in one phosphate-sugar chain is in the opposite direction to the corresponding sequence in the other. This is shown symbolically by the two small arrows.

Now it is found that one cannot build this model with any bases one pleases; only certain pairs of the four bases will fit into the structure. In any pair there must always be one big one (purine) and one little one (pyrimidine). If one tries to put in two purines—two big ones, that is—there is not sufficient room for them. Conversely, a pair made of two pyrimidines is too small to bridge the gap between the two chains. Moreover when one examines in detail how the hydrogen bonds are formed between the bases it is found that (making certain plausible assumptions) the pairing is even more restricted. The only possible pairs that will fit in are:

Adenine with Thymine
and
Guanine with Cytosine

The way these pairs are formed is shown in Figs. 5 and 6. The dotted lines show the weak physical bonds, known as hydrogen bonds, which hold the two bases of a pair together. (Hydrogen bonds are, for example, the main forces holding different water molecules together, and it is because of them that water is a liquid at room temperatures and not a gas.)

These specific pairs can be built into the structure either way round. We can have adenine on the first chain paired with thymine on the second, or vice versa. But if we do have adenine at some point on one of the chains, it is essential to have thymine paired with it on the other. It is impossible to fit in guanine or cytosine or a second adenine. In the same way guanine must always be paired with cyto-

On the other hand the model places no restriction on the

* A helix is often loosely called a spiral; thus a spiral staircase is usually a helix, not a spiral.

sequence of pairs of bases as one proceeds along the structure. Any specific pair can follow any other specific pair. This is because a pair of bases is flat, and as in this model they are stacked one above another like a pile of coins, it does not matter which pair goes above which.

This specific pairing of the bases is the direct result of the assumption that both phosphate-sugar chains are helical. This implies that the distance apart of two sugar groups at the same level (one belonging to each chain) is always the same, no matter where one is along the chain. It follows that the bases, which are of course linked to the sugars, have always the same amount of space in which to fit, as can be seen from studying Fig. 4. If it were not for this restriction the bases could hydrogen-bond together in many different ways. It is the regularity of the phosphate-sugar chains, therefore, which is at the root of the specific pairing.

EVIDENCE FOR THE MODEL

The experimental evidence in support of a model of this general type is now considerable. Measurements of the density and water content of the DNA fibres, taken with the evidence showing how the fibres can be extended in length, strongly suggest that there are two DNA chains in the structural unit. The X-ray patterns have a large number of places where the diffraction intensity is zero and these occur exactly where one expects them from helical structures of this type. Moreover the X-ray diffraction data approximates quite closely to cylindrical symmetry, as it should. Recently Wilkins and his co-workers have given a brilliant analysis of the details of the X-ray pattern of the crystalline form, and have shown that they are consistent with a structure of this type, though in this form the bases are not perpendicular to the fibre axis, but tilted away from it.

As the structure is a relatively stiff one it easily explains the extended shape of the DNA in solution. It is also consistent with the titration curve. This has irreversible features which suggest that the bases are hydrogen-bonded together. However, the most striking support for the specific pairing of the bases comes from the recent analytical data. These show that for every species so far examined—and there are over forty of them—the number of adenines in some given DNA is closely equal to the number of thymines, and the number of guanines equal to the number of cytosines, although the cross-ratio (between say adenine and guanine) can vary considerably from species to species. This remarkable fact, which is exactly what one would expect from a model containing only the specific pairs, was first pointed out by Dr. Chargaff. Indeed, since the sequence of bases along a single chain is believed to be irregular this result is very difficult to explain except by specific pairing.

It might be thought that while this model might be correct for the DNA extracted from the cell and made into fibres the DNA inside the cell was in a radically different form. This seems unlikely since it is difficult to see how the very characteristic features of the model could be produced merely by the extraction procedure. However, Dr. Wilkins has shown that it is possible to get very similar X-ray pictures from intact biological material, such as

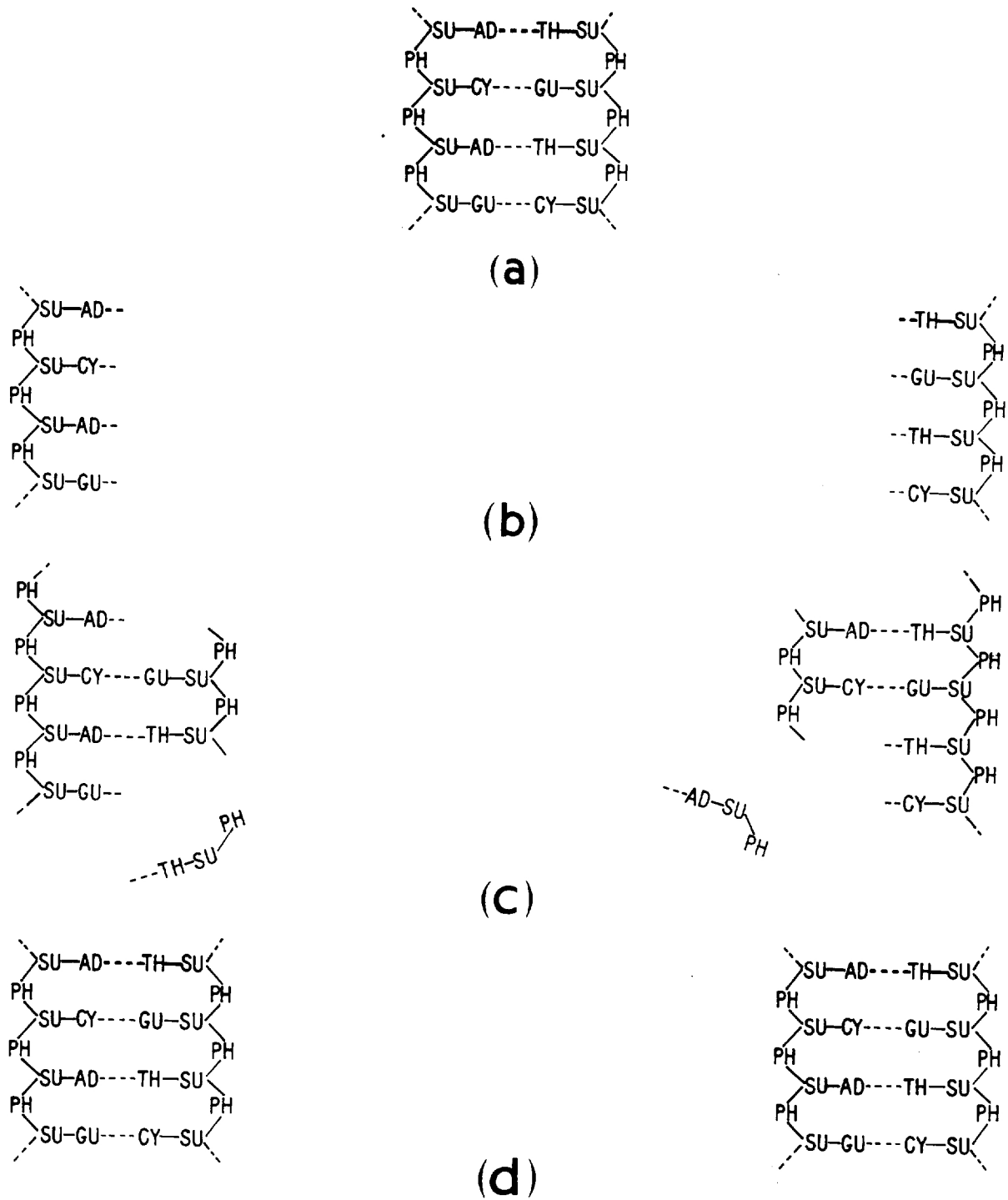


FIG. 7 (A) A typical stretch of the DNA structure.
 (B) The two chains separate.
 (C) The formation of two new chains from loose nucleotides.
 (D) The process complete. Note that the sequence of the bases has been copied exactly.

The letters represent the first two letters of the words PHosphate, SUGar, ADenine, GUanine, THymine and CYtosine.

sperm heads and bacteriophage, so that there seems little doubt that the structure is biologically significant.

The present position is therefore that while the details of the structure remain to be worked out—and until this is done the model cannot be considered as proved—it seems very probable that the following statements will stand the test of time:

1. The structure consists of two chains.
2. The chains are helical and wound round a common axis.
3. They are held together by hydrogen bonds between specific pairs of bases.
4. The structure occurs in biologically intact material.

A POSSIBLE REPLICATION MECHANISM

Now the exciting thing about a model of this type is that it immediately suggests how the DNA might produce an exact copy of itself. This is because the model consists of two parts, each of which is the *complement* of the other. The basic idea is that the two chains in the structure unwind and separate. Each chain then acts as a sort of mould on to which a new complementary chain can be synthesised. When this process is complete there will be *two* pairs of chains where we only had one before. Moreover, because of the specific pairing of the bases the sequence of the pairs of bases will have been duplicated exactly.

As an analogy consider two photographic films, one a positive and the other a negative of the same scene. Now if one gives the positive to one person, and asks him to print a negative from it, and also gives the original negative to another person, and asks him to print a positive from it, they will end up with two pairs of photographs, each pair like the original pair. We shall, in effect, have made an exact copy of our original pair in one step.

To see how this works out in the case of DNA let us consider the process in rather more detail. Since we have to synthesise two new chains we require some new material. The exact precursors of DNA are not known, but let us assume for simplicity that it is built up from nucleotides, which is the name given to the small molecules which contain one phosphate, one sugar and one base.

Imagine, then, that we have a *single* helical chain of DNA, and that floating around it, inside the cell, there is a supply of the four sorts of nucleotides. Every now and then a loose nucleotide will attach itself by its base to one of the bases of the DNA chain. Now if this happens to two adjoining

bases, and if the loose nucleotides are the type which can form specific pairs with those already there, they will be in just the right position to be joined together, and, eventually, to form part of the new chain. If one or both of them is *not* the correct type to go in at that point, it will be impossible to join them together and before long they will diffuse elsewhere. Thus only the nucleotides with the proper bases will get joined together to form the new chain.

While this process is going on, the other single chain of the original pair will also be forming, in a similar manner, a new chain complementary to itself. The whole process is illustrated in Fig. 7. In Fig. 7A there is shown a small stretch of the original pair of chains. In Fig. 7B these have separated. In the next figure new chains are being formed from loose nucleotides, and in Fig. 7D the process is complete; it can be seen that the original pair has now been duplicated.

At the moment this idea must be regarded simply as a working hypothesis. Straight away it raises a number of questions. How do the two chains unwind? What holds a *single* chain in a helical configuration? (Watson and I suspect that the replication starts almost as soon as the unwinding, so that only a very short stretch is ever in the 'single' state at one time.) Most important of all, how does the DNA influence the rest of the cell? We believe that the sequence of the bases along the DNA is the code that carries the genetical information, but how does it produce its effect? We can see how the code may be copied, but as yet we cannot read it.

In favour of the idea one can only say that it seems rather an odd coincidence to find in the one material which is most closely associated with replication a structure of exactly the type one would need to carry out a specific replication process, namely, one showing both variety and complementarity. The process is also attractive in its simplicity. While it is obvious that whole chromosomes have a fairly complicated structure it is not unreasonable to hope that the molecular basis underlying them may be rather simple. If this is so it may not prove too difficult to devise experiments to unravel it. It is, after all, remarkable that X-rays, which only show clearly the regular parts of a structure should tell us anything at all about a material whose main purpose, we suspect, is to embody variety. In any event we now have for the first time a well-defined model for DNA and for a possible replication process, and this in itself should make it easier to devise crucial experiments.

REFERENCES

For general background, consult *The Biochemistry of the Nucleic Acids* by J. N. Davidson (Methuen monograph).

Papers on recent research

1953. The helical structure of crystalline deoxyribose nucleic acid, M. H. F. Wilkins, W. E. Seeds, A. R. Stoker and H. R. Wilson,

Nature (1953), **172**, 759. (This also gives all the other X-ray references.)

Electron microscope. R. C. Williams, *Biochim. et Biophys. Acta* (1953), **9**, 237. H. Kahler and B. J. Lloyd, *Biochim. et Biophys. Acta* (1953), **10**, 355.