THE STRUCTURE OF DNA

F. H. C. Crick

The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, The Cavendish Laboratory, Cambridge, England

THE STRUCTURE proposed for DNA (deoxyribonucleic acid) has been described before (for a review, see Jordan, 21) and will only be outlined here. It consists of two polynucleotide chains running in opposite directions and twined round one another. The two chains are held together by hydrogen bonds between the bases, each base being joined to a companion base on the other chain. This pairing of bases is specific, adenine going with thymine, and guanine with cytosine. The structure is not only found in extracted DNA from a wide variety of sources, but is also present in intact biological material such as sperm heads and bacteriophage (36).

The x-ray work up to 1954 has already been briefly summarized, with references (6). Since then, the group at King's College, London, under Dr. M. H. F. Wilkins have published an interim report (17) on their studies of the B form of DNA and on their work on nucleoprotamine. Two reviews (9, 37) have also appeared; the one by Wilkins (37) touching on the very recent work of his group on nucleohistone. Some studies of the B form of DNA have also been carried out by Wykoff (38), who has shown that the B structure obtained by stretching the A form is slightly different from that produced by swelling it. The King's College group have obtained the B form in a crystalline state, and in particular, have shown that the lithium salt gives a good lattice (37).

According to Wilkins (37), the structure is now firmly established. The detailed structure described by Crick and Watson (8) has been shown to have too large a diameter, and a drawing has been given of an improved model which is in fairly good agreement with the x-ray data. No coordinates have so far been published.* (It should be noted

* Dr. Wilkins has offered to make his provisional coordinates available to anyone who would like to have them.

that there is some minor disagreement between Wykoff and the King's College group.) The published work on the A form is even more preliminary. This is unfortunate, as the ultimate test of a model must be its agreement with the very beautiful and detailed pictures which Wilkins and his coworkers have obtained from the A form. It is encouraging to learn that isomorphous replacements of Na, K, and Rb ions have been possible (37). The full publication of the experimental data, together with the coordinates of the proposed models and their calculated Fourier transforms is awaited with interest.

There have been two independent suggestions (19, 24) that the polynucleotide chains of the DNA structure are paranemically coiled; that is to say, that the two chains are not truly intertwined, but merely lie side by side in an intimate but distorted embrace. Gamow (19) does not use the term paranemic coiling to describe his proposal, but his suggestion that "the long helical molecule is wound into a coil possessing the same repetition period as the original helix" is a description of paranemic coiling. This has already been discussed and rejected (35), as being very difficult to reconcile with the data. The two authors putting forth this idea are apparently unaware of the convention that for a structure to be given serious consideration it must be possible to build a scale model of it having acceptable bond distances and angles—inspiration by itself is not enough. Until a satisfactory model has been presented this idea must be regarded as incorrect.

The experimental evidence, other than the x-ray data, which supports the double helical model, falls into two classes: the chemical evidence, and the physico-chemical evidence. The chemical evidence shows that the molar ratio of adenine/thymine and of guanine/cytosine + 5-methyl cytosine are very close to unity for all sources of DNA. This striking experimental fact was originally pointed out by Chargaff. The latest evidence is reviewed by Chargaff elsewhere in this volume (3).

Unusual Bases

One feature of the recent analytical data which appears to be causing a certain amount of confusion is the occurrence in DNA of various bases other than the usual four. The replacement of cytosine by 5-hydroxymethyl cytosine (probably with glucose attached) in the T-even phages presents no problem, since it will fit into the structure without difficulty. The smaller amount of 5-methyl cytosine which occurs in the DNA of various organisms is again not a structural

problem, since it appears to replace cytosine, in the sense that the molar sum of cytosine and 5-methyl cytosine is closely equal to the number of moles of guanine. The same applies to the very small amounts of 6-methyl aminopurine (15) found in certain bacteria. However, the larger amounts of this base which apparently replace thymine in the thymineless mutant T15— of *Escherichia coli,* when growing without a supply of thymine (14), do constitute a problem; but the effect is lethal to the cells and the DNA may well be abnormal.

A possible explanation of this peculiar behavior is suggested by the data on 5-methyl cytosine. This base does not replace cytosine at random, since it has been shown that different fractions of DNA have large variations in the cytosine/5-methyl cytosine ratio (3, 4, 25). However, it has been found that 5-methyl cytosine occurs almost entirely next to guanine (30-32).

This relationship suggests that the precursors of DNA are not mononucleotides (or related molecules) but resemble dinucleotides or higher nucleotides. Applying this idea to the case of 6-methyl aminopurine, one might surmise that 6-methyl aminopurine (denoted by X) is incorporated not by itself, but attached to some other nucleotide, probably thymidylic acid (denoted by T) either as XT or as TX, and that, in particular, it goes into places which would normally accommodate TT. The amounts incorporated are crudely compatible with this idea. This hypothesis has the unusual advantage of being testable—for example, by analyzing the various amounts of various dinucleotides, from a partial digest, which contains X. In general, whenever an unusual base is incorporated into DNA it would be well worth-while to see if it occurs with certain preferred neighbors (including the ends of the polynucleotide chains).

Physical Chemical Evidence

There is now much physical evidence to support a two-chain structure. Briefly, this includes the titration curve (20-23, 27), which suggests that the bases form hydrogen bonds, and that these are bonds within the structure, since the titration hysteresis persists to very low dilutions (22); the shape and size of the molecule in solution, obtained from a combination of light scattering, viscosity, and sedimentation measurements (12, 28), which show that DNA in solution is highly extended, but not completely straight, and that its diameter is compatible with the double helix model (12); and also studies of the rates

at which the structure is broken down by gamma rays (5), acid (34), or enzymatic attack (33), which are consistent with there being two strands in the DNA, so that the molecule does not come apart until there are breaks in both backbones almost opposite one another (opposite to within two nucleotides, according to Thomas, 33).

The claim of Alexander and Stacey (1) to have separated the two chains by the use of various treatments, such as exposure to 4 *M* urea, has not been accepted by other workers (12). Similarly the suggestion of Dekker and Schachman (10) that there were occasional breaks (say 1 in 50) in the phosphate-sugar backbone has not been supported by the recent evidence (27, 33, 34), which seems to show that breaks, if they exist at all, must occur only very infrequently (less than 1 in 500, say).

General Remarks

It is important to notice the combination of symmetry and pseudosymmetry, and of repetition and non-repetition, in the DNA structure. The phosphate-sugar backbone repeats regularly, both chemically and structurally. This repetition necessarily implies that the phosphate-sugar groups are related by symmetry, in this case by a screw axis, and it is this which makes the backbone a simple helix. Again, the two separate phosphate-sugar backbones are related to each other by symmetry, in this case by two-fold rotation axes perpendicular to the fiber axis.

The arrangement of the bases, however, does not repeat, and only shows pseudosymmetry; that is, the region occupied by a pair of bases is fixed, and successive regions are related to each other by symmetry, but there is no restriction on which pair of bases occurs at any point, as long as one of the allowed pairs is used.

There are many different ways of pairing the four common bases, using two hydrogen bonds, and these have recently been systematically described by Donohue (11). As far as we know, these are all equally likely in solution. The base pairing described is the only satisfactory way which allows all four bases to occur on one chain, and which will fit the x-ray data. It remains to be seen whether there are other structural reasons (e.g., that the glycosidic bond must point roughly toward the axis) which favor the particular pairing suggested.

It should be noted that while x-ray diffraction shows that a substantial portion of the DNA must be in the double helix form it is an extremely poor method for deciding how much of the DNA is in this

configuration. The titration curve and the analytical data suggest that the great majority of bases are paired. However, as has been stressed before (35), it seems certain that the molecule is folded in its biological condition, and there may conceivably be occasional regions where the configuration is somewhat modified. Before this idea becomes too popular, however, it would be nice to have some positive (rather than indirect) evidence for it.

Replication and Gene Action

The function of DNA is outside the scope of this paper, but there have recently been a number of suggestions for mechanisms which are based on features of the molecular structure. For example, the mechanism of Lockingen and DeBusk (26) depends on the presence of breaks in the backbones, which are now thought to be absent, or at least very rare. The suggestion of Block (2) depends upon being able to turn over the bases—as does the scheme of Dounce, Morrison, and Monty (13) for RNA. This may be possible for adenine, but molecular models suggest that it is difficult if not impossible to do it for guanine, thymine, and cytosine because the resulting van der Waals contact would be too close (between the pyrimidine O, in position 2, and the ribose ring oxygen, or between the NH_2 of guanine and the adjacent sugar). The suggestion of Schwartz (29), that aromatic amino acids are sandwiched between pairs of bases, can be criticized (apart from the rather doubtful idea that 5-membered rings attract only 5-membered rings, and 6-membered rings only 6-membered rings) on the ground that it seems unlikely that such a model could be constructed, since the argument that the bases can be packed 7 Å apart is based on a misconception of the structure.

In both these cases, it would have been better if the authors had attempted to build scale models* to show that their ideas were structurally feasible.

Incidentally, it is possible to produce schemes similar to that of Block by separating the base-pairs by *rotating* one of the DNA chains about the fiber axis, relative to the other. This is structurally possible if the bases are not too close to the axis of the molecule, but this may not normally be the case (17, 36).

* A recent comment that "Real Research is done at the Bench and not playing about with Metal Models," should be ignored here.

Although not directly supported by the experimental evidence, it seems very probable that the two phosphate-sugar backbones run in opposite directions and are related by two-fold axes. This means that if we regard the sequence of base pairs as a code, there is nothing in the structure to tell us in which direction to read it, *except* the sequence of the bases themselves. Thus, a code of the type described by Gamow (18) does not specify the *direction* of the polypeptide chain for which the DNA is supposed to be coding. For example, the sequence gly-ala-leu would correspond in his code to the same piece of DNA as would leu-ala-gly, using the usual convention. There are several ways out of the dilemma—for example, that the base-sequence makes "sense" if read one way and "nonsense" if read the other—but the point is a fundamental one and should not be overlooked.

It has been pointed out elsewhere (7) that the most obvious way in which the sequence of bases could express itself in terms of physical chemistry is in the patterns of sites for hydrogen bonding presented by the structure. The significance of 6-methylaminopurine in this respect should be noted, since the introduction of the methyl group changes this pattern radically, at least on the inner surfaces of the structure.

Nucleoprotein

Essentially all the x-ray work on nucleoprotein has been done by Wilkins and his coworkers (17, 37). Much of it is still in a very preliminary stage.

Nucleoprotamine.

The evidence suggests that the protamine chain is wound helically round the DNA structure in the *smaller* of the two grooves between the backbones (17). Models show (17) that an extended polyarginine chain can be fitted in there without difficulty, with the positively charged basic groups of the side-chains going alternately up and down to the negatively charged phosphate groups of the DNA backbones. In nucleoprotamine there appears to be one arginine for every phosphate, yet only two-thirds of the protamine side-chains are basic. This suggests that the polypeptide chain is folded whenever the non-polar amino acids occur. Model building shows (17) that it is difficult to construct a fold with one non-polar residue, but relatively easy with two in succession. The data on the amino acid sequence (16) show that the

non-polar residues do indeed occur in pairs. It is not yet clear whether the interaction between the non-polar residues and the bases of the DNA structure is specific or non-specific, nor whether the non-polar folds go inwards or outwards (17).

Nucleohistone.

The main features of the preliminary x-ray work (37) are that (1) the DNA—or at least part of it—maintains its characteristic structure; and that (2) some larger repeating structure is also present. These results are obtained with nuclei, swollen in water and drawn into fibers, and also with artificial combinations of DNA with (lysine-rich) histone. One equatorial spacing (of about 60 Å) changes little on drying; another, around 40 Å, alters with humidity. The significance of these results is not yet clear. Their main importance is to show that nucleohistone has a structure of some sort.

ACKNOWLEDGMENT

I should like to thank Dr. M. H. F. Wilkins for allowing me to read the manuscript of his latest paper (37) before publication.

REFERENCES

1. Alexander, P., and Stacey, K. A., *Biochem. J.*, **60**, 194 (1955).
2. Block, D. P., *Proc. Natl. Acad. Sci. U. S.*, **41**, 1058 (1955).
3. Chargaff, E., this volume.
4. Chargaff, E., Crampton, C. F., and Lipshitz, R., *Nature*, **172**, 289 (1953).
5. Cox, R. A., Overend, W. G., Peacocke, A. R., and Wilson, S., *Nature*, **176**, 919 (1955).
6. Crick, F. H. C., *Proc. Natl. Acad. Sci. U. S.*, **40**, 756 (1954).
7. Crick, F. H. C., *Biochem. Soc. Symposia* (in press).
8. Crick, F. H. C., and Watson, J. D., *Proc. Roy. Soc. (Lond.)*, A, **223**, 80 (1954).
9. Crick, F. H. C., and Watson, J. D., *Rend. ist. Lombardo Sci., Pt. I.*, **89**, 52 (1955).
10. Dekker, C. A., and Schachman, H. K., *Proc. Natl. Acad. Sci. U. S.*, **40**, 894 (1954).
11. Donohue, J., *Proc. Natl. Acad. Sci. U. S.*, **42**, 60 (1956).
12. Doty, P., *Intern. Congr. Biochem.*, **3** (Brussels, 1955), 135 (1956).
13. Dounce, A. L., Morrison, M., and Monty, K. J., *Nature*, **176**, 597 (1955).
14. Dunn, D. B., and Smith, J. D., *Nature*, **175**, 336 (1955).
15. Dunn, D. B., and Smith, J. D., *Biochem. J.*, **60**, XVII (1955).
16. Felix, K., Fischer, H., and Krekels, A., *Progr. Biophys. and Biophys. Chem.*, **6**, 1 (1956).
17. Feughelman, M., Langridge, R., Seeds, W. E., Stokes, A. R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F., Barclay, R. K., and Hamilton, L. D., *Nature*, **175**, 834 (1955).
18. Gamow, G., *Kgl. Danske Videnskab. Selskab, Biol. Medd.*, **22**, 1 (1954).

19. Gamow, G., *Proc. Natl. Acad. Sci. U. S.*, **41**, 7 (1955).
20. Gulland, J. M., Jordan, D. O., and Taylor, H. F. W., *J. Chem. Soc.*, 1131 (1947).
21. Jordan, D. O., in *The Nucleic Acids*, Vol. I (E. Chargaff and J. N. Davidson, eds.), Academic Press, New York (1955).
22. Jordan, D. O., Mathieson, A. R., and Matty, Sheila, *J. Chem. Soc.*, 154 and 158 (1956).
23. Lea, W. A., and Peacocke, A. R., *J. Chem. Soc.*, 3361 (1951).
24. Linser, H., *Biochim. et Biophys. Acta*, **16**, 295 (1955).
25. Lipshitz, R., and Chargaff, E., *Biochim. et Biophys. Acta*, **19**, 256 (1956).
26. Lockinger, L. S., and DeBusk, A. G., *Proc. Natl. Acad. Sci. U. S.*, **41**, 925 (1955).
27. Peacocke, A. R., *J. Chem. Soc.* (in press).
28. Sadron, C., *Intern. Congr. Biochem.*, **3** (Brussels, 1955), 125 (1956).
29. Schwartz, D., *Proc. Natl. Acad. Sci. U. S.*, **41**, 300 (1955).
30. Sinsheimer, R. L., *J. Biol. Chem.*, **208**, 445 (1954).
31. Sinsheimer, R. L., *J. Biol. Chem.*, **215**, 579 (1955).
32. Smith, J. D., and Markham, R., *Nature*, **170**, 120 (1952).
33. Thomas, C. A., *J. Am. Chem. Soc.*, **78**, 1861 (1956).
34. Thomas, C. A., and Doty, P., *J. Am. Chem. Soc.*, **78**, 1854 (1956).
35. Watson, J. D., and Crick, F. H. C., *Cold Spring Harbor Symposia Quant. Biol.*, **18**, 123 (1953).
36. Wilkins, M. H. F., Seeds, W. E., Stokes, A. R., and Wilson, H. R., *Nature*, **172**, 759 (1953).
37. Wilkins, M. H. F., *Biochem. Soc. Symposia* (in press).
38. Wykoff, H. W., Thesis, Mass. Inst. Technology (1955).