The complementary structure of deoxyribonucleic acid

By F. H. C. Crick and J. D. Watson*†
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, University of Cambridge

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 24 *August* 1953)

[Plate 2]

This paper describes a possible structure for the paracrystalline form of the sodium salt of deoxyribonucleic acid. The structure consists of two DNA chains wound helically round a common axis, and held together by hydrogen bonds between specific pairs of bases. The assumptions made in deriving the structure are described, and co-ordinates are given for the principal atoms. The structure of the crystalline form is discussed briefly.

Introduction

The basic chemical formula of DNA is now fairly well established. It is a very long chain molecule formed by the joining together of complex monomeric units called nucleotides. Four main types of nucleotides are found in DNA, and it is probable that their sequence along a given chain is irregular. The relative amounts of the four nucleotides vary from species to species. The linkage between successive nucleotides is regular and involves 3′-5′-phospho-di-ester bonds.

Information about the three-dimensional shape is much less complete than that about its chemical formula. Physical-chemical studies, involving sedimentation, diffusion and light-scattering measurements, have suggested that the DNA chains exist in the form of thin rather rigid fibres approximately 20 Å in diameter and many thousand of ångströms in length (Jordan 1951; Sadron 1953). Very recently these indirect inferences have been directly confirmed by the electron micrographs of Williams (1952) and of Kahler & Lloyd (1953). Both sets of investigators have presented very good evidence for the presence in preparations of DNA of very long thin fibres with a diameter of 15 to 20 Å, and so there now appears little doubt about the general asymmetrical shape of DNA.

The only source of detailed information about the configuration of the atoms within the fibres is X-ray analysis (Astbury 1947; Wilkins, Stokes & Wilson 1953; Franklin & Gosling 1953*a*). DNA's from various sources can be extracted, purified and drawn into fibres which are highly birefringent and give remarkably good X-ray diagrams. The same type of X-ray pattern is obtained from all soures of DNA, and the unit cell found is many times larger than that of the fundamental chemical unit, the nucleotide.

It seems improbable that the structure can be solved solely by modern crystallographic methods such as inequalities or vector superposition. These methods have so far been successfully used with relatively simple compounds. The DNA unit cell, however, is very large, and in fact contains a larger number of atoms than in any

* Aided by a Fellowship from the National Foundation for Infantile Paralysis (U.S.A.).
† Present address, Biology Division, California Institute of Technology, Pasadena 4, California.

structure, crystalline or fibrous, so far determined. Moreover, the number of X-ray reflexions is small, as there are few reflexions at spacings less than 3 Å, and so the classical method of trial and error seems the most promising approach.

It has therefore seemed worth while for us to build models of idealized polynucleotide chains to see if stereochemical considerations might tell us something about their arrangement in space. In doing so we have utilized interatomic distances and bond angles obtained from the simpler constituents of DNA and have only attempted to formulate structures in which configurational parameters assume accepted dimensions. We have only considered such structures as would fit the preliminary X-ray data of Wilkins, Franklin and their co-workers. Our search has so far yielded only one suitable structure. This structure, of which a preliminary account has already appeared (Watson & Crick 1953*a*), consists of two intertwined polynucleotide chains helically arranged about a common axis. The two chains are joined together by hydrogen bonds between a purine base on one chain and a pyrimidine base on the other. This structure appears to us most promising, and in fact we believe that its broad features are correct. In this paper we shall present the assumptions used in formulating this structure and give precise co-ordinates for the principal atoms. We shall make no attempt to test the structure with the experimental X-ray evidence as this is being done by others.

Chemical background

The DNA molecule can be formally divided into two parts, the backbone and the side groups. The backbone, as shown in figure 1, is very regular and is made up of alternate sugar (2-deoxy-D-ribose) and phosphate groups joined together in regular, 3′, 5′-phosphate-di-ester linkages (Brown & Todd 1952; Dekker, Michelson & Todd 1953). The side groups consist of either a purine or a pyrimidine base, only one of which is attached to any given sugar. Two purines, adenine and guanine, and two pyrimidines, cytosine and thymine, are commonly present. In addition, a third pyrimidine 5-methyl-cytosine (Wyatt 1952) occurs in small amounts in certain organisms, while in the *T*-even phages cytosine is absent and is replaced by a fourth pyrimidine, 5-hydroxy-methyl-cytosine (Wyatt & Cohen 1952).

The glycosidic combination of the base and the sugar is known as a nucleoside, while the phosphate ester of a nucleoside is called a nucleotide. The deoxyribose residue in each of the nucleotides is in the furanose form (Brown & Lythgoe 1950) and is glycosidically bound to N_3 in the pyrimidine nucleosides and to N_9 in the purine nucleosides (for a review, see Tipson 1945). The configuration at the glycosidic linkage has been shown to be β in deoxyadenosine and deoxycytidine (Todd *et al.* unpublished) and is considered by analogy to be the same in the other natural deoxyribonucleosides.

A DNA chain may contain thousands of nucleotides and is thought in view of the regular internucleotide linkage to be unbranched. Very little is known about the precise sequence of the different nucleotides, but as far as can be now ascertained the order is irregular and any sequence of nucleotides is possible.

At pH values > 2, the primary phosphoryl groups are ionized, and so most investigations have utilized the sodium salt. The crystallographic analysis has so far

dealt exclusively with this salt, and our structural suggestions are correspondingly limited to this form.

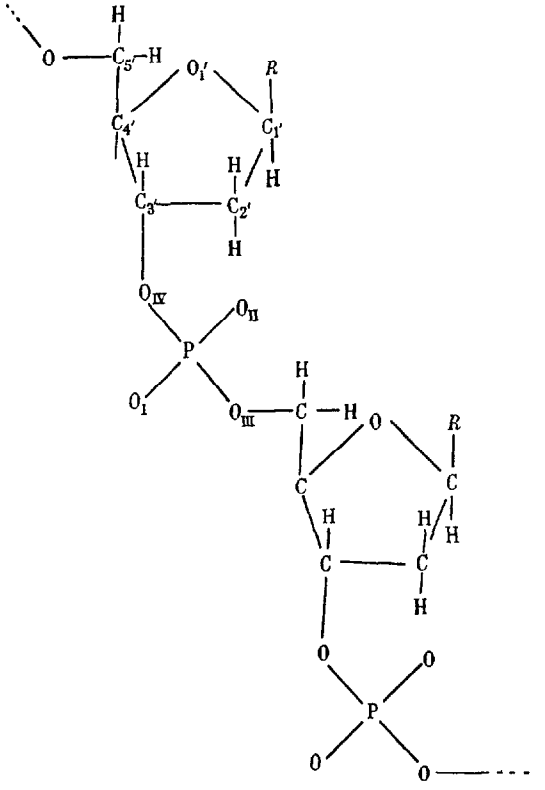

FIGURE 1. The general formula of DNA. R is a purine or pyrimidine base.

CRYSTALLOGRAPHIC CONSIDERATIONS

X-ray photographs of DNA fibres were obtained in 1938 by Astbury & Bell (1938) and more recently by Wilkins & Franklin and their collaborators at King's College, London (Wilkins *et al.* 1953; Franklin & Gosling 1953*a*, *c*). The photographs were taken of purified samples which had been drawn into birefringent fibres in which the DNA molecules are orientated approximately parallel to the fibre axis. The photographs of Wilkins & Franklin and their collaborators are appreciably sharper than those of Astbury & Bell, and we shall restrict our discussion to their work.

It is observed* that DNA can exist in two different forms†, a crystalline form structure A, and a paracrystalline form structure B. The crystalline form occurs at

* The information reported in this section was very kindly reported to us prior to its publication by Drs Wilkins and Franklin. We are most heavily indebted in this respect to the King's College Group, and we wish to point out that without this data the formulation of our structure would have been most unlikely, if not impossible. We should at the same time mention that the *details* of their X-ray photographs were not known to us, and that the formulation of the structure was largely the result of extensive model building in which the main effort was to find any structure which was stereochemically feasible.

† The existence of the two forms was first suggested by powder photographs of DNA gells (Riley & Oster 1951).

75 % relative humidity and contains about 30 % water by weight. Its repeat distance along the fibre axis is 28 Å. At higher humidities this form takes up more water, increases in length by about 30 % and assumes the alternative paracrystalline form. In contrast to the crystalline form which lacks any strong meridional reflexion the paracrystalline form gives a very strong meridional reflexion at 3·4 Å. In conjunction with the increase in fibre length, the repeat along the fibre axis increases to 34 Å. Both forms give equatorial reflexions corresponding to sideways repeats of 22 to 25 Å, and it appears that their diameters are approximately the same. The transition between the two forms is freely reversible, and it seems likely that they are related in a simple manner.

They have further shown (Wilkins *et al.* 1953) that the X-ray pattern of both the crystalline and paracrystalline forms is the same for all sources of DNA ranging from viruses to mammals. At first sight this seems surprising, as the ratios of the various nucleotides vary from one source to another and it might have been expected that the size and shape of the structural unit would vary correspondingly. On the other hand, we should recall that the sequence of nucleotides within a given DNA chain is irregular, and so the fact that DNA forms a repetitive structure (much less a crystalline structure!) is itself unusual.

It seemed to us that the most likely explanation of these observations was that the structure was based upon features common to all nucleotides. This suggested that in the first instance one should consider mainly the configuration of the phosphate-sugar chain, with an 'average' base attached to each sugar. In other words, an idealized polynucleotide with all the monomers the same.

For such a model it is stereochemically plausible to assume that all the sugar and phosphate groups are in equivalent positions and have identical environments irrespective of which nucleotide is being considered. This implies that one nucleotide is related to another by a symmetry operation, and in the case of a single optically active chain, this operation is necessarily a rotation about an axis accompanied by a translation along the axis. This corresponds to a screw axis, and the operation if repeated leads in general to a helix, as pointed out before by Pauling, Corey & Branson (1951) and by Crane (1950).

The idea that the DNA structure is helical* is supported by two general features of the experimental data. First, it provides a simple explanation of the fact that the fibre axis repeat ($\simeq 30$ Å) is many times longer than the probable axial spacing between nucleotides ($\simeq 3$ Å), since a helical structure composed of identical monomers will give a spacing related to the pitch of the helix (Cochran, Crick & Vand 1952). Secondly, the unit-cell dimensions of the crystalline form (Franklin & Gosling 1953*c*) are pseudo-hexagonal in cross-section, as one might expect if the structure was based on helical bundles approximately cylindrical in shape.

We have therefore attempted to build helical structures in which the repeat distance along the fibre axis is that reported by Wilkins, Franklin and co-workers. Before doing so, however, it was necessary to decide whether to build models of the

* We should mention that on several occasions Dr Wilkins in personal conversation indicated that the paracrystalline X-ray pattern had helical features. Our postulation of a helical structure was, however, the consequence of the above reasons, and we feel independent of Dr Wilkins's suggestion.

crystalline form structure *A* or the paracrystalline form structure *B*. We had no hesitation in choosing the latter, mainly because of its extremely strong 3·4 Å meridional reflexion (discussed below), since this gives information which can be of direct help in building models.

Formulation of a structure for the paracrystalline form

The X-ray pattern of structure *B* is dominated by a very strong reflexion on the meridian at a spacing of 3·4 Å (Wilkins *et al.* 1953; Franklin & Gosling 1953*a*). This distance, as first pointed out by Astbury, corresponds to the thickness of a purine or pyrimidine base, and suggests that the nucleotide bases on a given chain are arranged at right angles to the fibre axis and spaced 3·4 Å above each other. The idea that the bases are roughly perpendicular to the fibre axis is supported qualitatively by the ultra-violet dichroism (Wilkins, Gosling & Seeds 1951).

It is difficult to imagine any other arrangement producing such a strong reflexion. This reflexion corresponds to a spacing approximately twice that of the covalent bonds present in DNA, and so most probably arises from a regular arrangement of internucleotide van der Waals contacts. It is worth noting why this reflexion cannot arise from a staggered arrangement of chains containing successive nucleotides spaced 6·8 Å above each other. This distance is approximately the internucleotide length of an extended polynucleotide chain, and if present in DNA should result in reversibly inextensible fibres. Now, Wilkins *et al.* (1951) have reported that DNA fibres can be reversibly stretched by a factor 1·5, and so the fibre axis per nucleotide must be considerably less than the fully extended internucleotide length. We thus have little doubt that the fibre axis translation per nucleotide is 3·4 Å, and (assuming equivalence) that a given polynucleotide chain contains 10 nucleotide residues per 34 Å fibre axis repeat.

It is difficult, nevertheless, to account for the rather high density (Astbury 1947) of DNA on the basis of a helical structure containing but 10 nucleotides within the unit cell. In fact, density consideration suggests the presence of a structure containing two to three times as many residues.

The most plausible way to explain this is to assume that the DNA molecule contains several polynucleotide chains and that they are helically coiled about a common axis. Density considerations immediately rule out the presence of more than three chains, and so we are left to decide between two or three chains. At first sight it appears that three chains is the correct answer, as the density of DNA is generally reported (Astbury 1947) as about 1·65 g cm^{-3}, a value corresponding to approximately 30 nucleotides within a cylinder of radius 10 Å and height 34 Å. We must remember, however, that the density measurements are generally reported from dry specimens (from which only very disordered X-ray patterns can be obtained; Wilkins, personal communication) and that as yet we do not know the effective density of the paracrystalline form.

The density of structure *A*, however, has been measured by Franklin & Gosling (1953*c*), and indicates the presence of approximately 24 nucleotides per lattice point, a value which superficially is incompatible with either two or with three chains. This incompatibility disappears, however, when we consider that the

translation from structure B to structure A is accompanied by a visual shortening of the fibre by roughly 30% (Franklin & Gosling 1953*a*). The longitudinal component is thus no longer 3·4 Å but 3·4 Å × 0·70 = 2·4 Å. The unit cell of structure A, therefore, contains two polynucleotide chains each of which contains about 12 nucleotides per fibre axis repeat, since $2{\cdot}4 \times 12 \simeq 28$. As the transformation from A to B is readily reversible, it seems most improbable that the chains would be grouped in threes in structure B, and we believe that in this form also the fundamental structural unit contains two helically arranged polynucleotide chains.

It is necessary to decide what part of the nucleotide to place in the centre of the helix. Initially, it seemed reasonable to believe that the basic structural arrangement would be dictated by packing consideration at the centre and that the core would contain atomic groups common to all the nucleotides. Our first attempts, therefore, involved possible models with the phosphate groups in the centre, the sugar groups further out and with the bases on the outside (the alternative arrangement of placing the sugar in the centre, is very improbable due to the irregular shape of the deoxyribofuranose group.)

Now the phosphate group carries a negative charge which is neutralized by the presence of a Na^+ ion. We thought it possible that this electrostatic attraction might dominate the structure and that the correct solution to DNA structure might fall out if we found a satisfactory way of packing the charged groups. We decided momentarily to ignore the sugar and base constituents and to build up regular patterns of co-ordination for the Na^+ and phosphate groups. In particular, we tried arrangements in which both of these ions were at the same distance from the fibre axis. No difficulty was found in obtaining repeat distances of 3·4 Å in the fibre direction as long as we considered only the charged groups. When, however, we attempted the next step of joining up the phosphate groups with the sugar groups we ran into difficulty. The phosphate groups tended to be either too far apart for the sugars to reach between them, or to be so close together that the sugars would fit in only by grossly violating van der Waals contacts. At first this seemed surprising, as the sugar-phosphate backbone contains, per residue, five single bonds, about all of which free rotation is possible. It might be thought that such a backbone would be very flexible and compliant. On the contrary, we came to realize that because of the awkward shape of the sugar, there are relatively few configurations which the backbone can assume. It therefore seemed that our initial approach would lead nowhere and that we should give up our attempt to place the phosphate groups in the centre. Instead, we believe it most likely that the bases form the central core and that the regular sugar-phosphate backbone forms the circumference.

Before building models of this type, it is necessary to know the approximate radius at which to place the backbone. As mentioned before, both the crystalline and paracrystalline forms give equatorial reflexions corresponding to sideways spacings of 22 to 24 Å (Wilkins *et al.* 1953; Franklin & Gosling 1953*a*), and so it seems very likely that both have effective radii of approximately 10 Å. This imposes a severe restriction on the types of models, for the polynucleotide chain has a maximum length. The distance between successive phosphorus atoms in a fully

extended chain is only about 7 Å, and so the maximum length of the ten nucleotide repetitive unit is but 70 Å. This is almost exactly the length of one repeat of a helical chain of radius 10 Å and pitch 34 Å, and so we can immediately conclude that the polynucleotide chain can have at most one revolution per fibre axis repeat. If the DNA molecule contained only one chain we could be more definite and conclude that the X-ray evidence demands one turn in 34 Å. As the molecule, however, contains two chains, the possibility remains that they are related by a diad parallel to the fibre axis and that each chain makes only half a revolution in 34 Å.

These possibilities can be differentiated by building models. We find that we can build models of one chain with a rotation of approximately 40° per residue but that it is difficult, if not impossible, with a rotation of only 20°. The van der Waals contacts in this latter case are much too close, and it appears probable that no structure of this type can exist. It, therefore, seems probable that each chain is in a nearly fully extended condition and makes one revolution every 34 Å. It should be noted that this argument rules out the possibility that the two intertwined chains are related by a diad parallel to the fibre axis, for if true, the fibre axis repeat would be halved to 17 Å.

It seems most likely that the two chains will be held together by hydrogen bonds between the bases. Both the purine and pyrimidine bases can form hydrogen bonds at several places on their periphery, and such instability would result from their absence that we may be confident of their presence. These bonds are strongly directional in character and can form only in the plane of the bases. They cannot be formed, however, between bases belonging to the same chain, since successive bases are located approximately on top of each other, and if we would draw a vector joining their centres, it would lie almost perpendicular to the plane in which they can form hydrogen bonds. Instead, we may expect the hydrogen bonds to be formed between bases belonging to the opposing chains and in doing so to unite the bases in pairs. This can be done in a regular manner only if we always join a purine with a pyrimidine. This is accomplished more suitably by forming two hydrogen bonds per pair; one from purine position 1 to pyrimidine position 1, the other from purine position 6 to pyrimidine position 6.

We should note the reason why the two chains cannot be linked together by two purines or by two pyrimidines. It arises from our postulate that each of the sugar-phosphate backbone chains is in the form of a regular helix. This implies that the glycosidic bonds (the link between the sugar and the base) always occur in identical orientation with regard to the helical axis. The two glycosidic bonds of a pair will therefore be fixed in space and have a constant distance between them. This distance, however, is different for each of the three possible types of pairs, purine with purine, pyrimidine with pyrimidine and purine with pyrimidine. The only way, therefore, to keep this distance fixed and to insert both types of bases into the structure is to restrict the pairing to the mixed variety.

We believe that the bases will most likely be present in the tautomeric forms shown in figure 2, and so in general only specific pairs of bases will bond together. These pairs are adenine with thymine (figure 3), and guanine with cytosine (figure 4).

When 5-methyl cytosine is present it should also pair with guanine as the methyl group is located on the side opposite to that involved in the pairing process. For similar reasons, 5-hydroxy-methyl-cytosine should likewise pair with guanine. It is easy to see why the other types of pairs will not occur. If, for instance, adenine is paired with cytosine, there are two hydrogen atoms between the amino nitrogens and none between the two ring nitrogens. For similar reasons guanine cannot be paired with thymine.

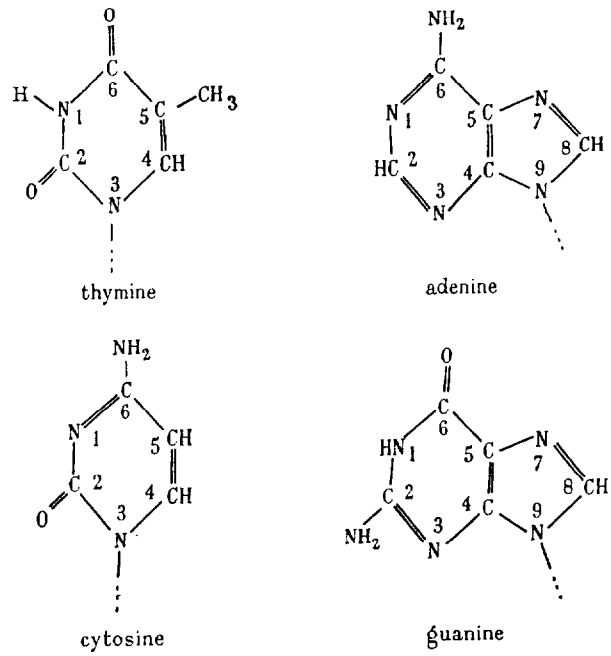

FIGURE 2. The formulae of the four common bases of DNA, showing the tautomeric forms assumed.

When models employing this pairing arrangement are built, several additional structural features become apparent. In the first place, we find by trial that the model can only be built in the right-handed* sense. Left-handed helices can be constructed only by violating the permissible van der Waals contacts. Secondly, in order to maintain the equivalence of the sugar and phosphate groups it is necessary to have the two chains (but not the bases) related by a diad perpendicular to the fibre axis. This is possible because the two glycosidic bonds of a purine-pyrimidine pair are not only the same distance apart in both of our chosen pairs, but are found to be related to each other by a diad, and can thus be fitted into the structure either way round (see figures 3 and 4). It is this feature which allows all four bases to occur on both chains. The insertion of the perpendicular diad requires the chains to run in opposite directions (a chain has a direction determined by the sequence of the atoms in it) and places the sugar-phosphate backbone

* The Fischer convention has recently been shown to be correct (Bijvoet, Peerdeman & van Bommel 1951).

of each chain in identical orientations with regard to the purine and pyrimidine side groups.

The structure can be built with any sequence of bases on a given chain. We should note, however, that the postulate of specific pairs introduces a definite relationship between the sequence of bases on the opposing chains. For instance, if on one chain we find at some point the sequence adenine, cytosine, thymine and adenine, then the corresponding sequence on the other chain must be thymine, guanine, adenine and thymine. The two chains thus bear a complementary relationship to each other.

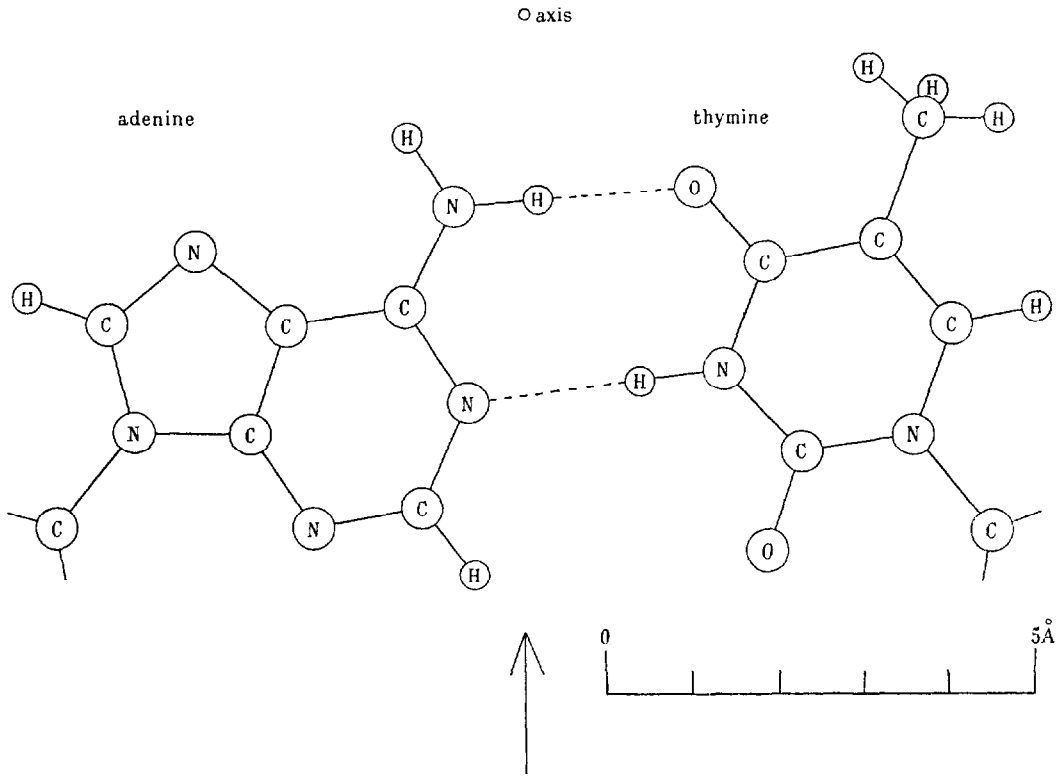

FIGURE 3. The pairing of adenine and guanine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown. The arrow represents the crystallographic diad.

The structure appears to satisfy all of the requirements which we initially postulated for the DNA molecule. The arrangement of the sugar-phosphate backbone which occupies the outer regions of the molecule is extremely regular, and it is possible to imagine it forming a crystalline pattern with neighbouring molecules. On the other hand, it permits an irregular sequence of nucleotides to exist on a given chain and thus allows for a large variety of DNA molecules. This fusion of regular and irregular features is achieved admittedly only at the expense of the additional restrictive postulate of complementary chains. The necessity for this postulate might be considered a severe, if not fatal objection to our structure, but as mentioned later, it is strongly supported by the recent analytical data.

DETAILED CONFIGURATION OF THE DOUBLE HELIX

We shall refer first to the specific pairs of bases. Adenine and thymine are shown paired in figure 3, while guanine and cytosine are shown paired in figure 4. These drawings are to scale and have been constructed as far as possible by utilizing bond angles and bond lengths which have been reported to occur in these compounds.

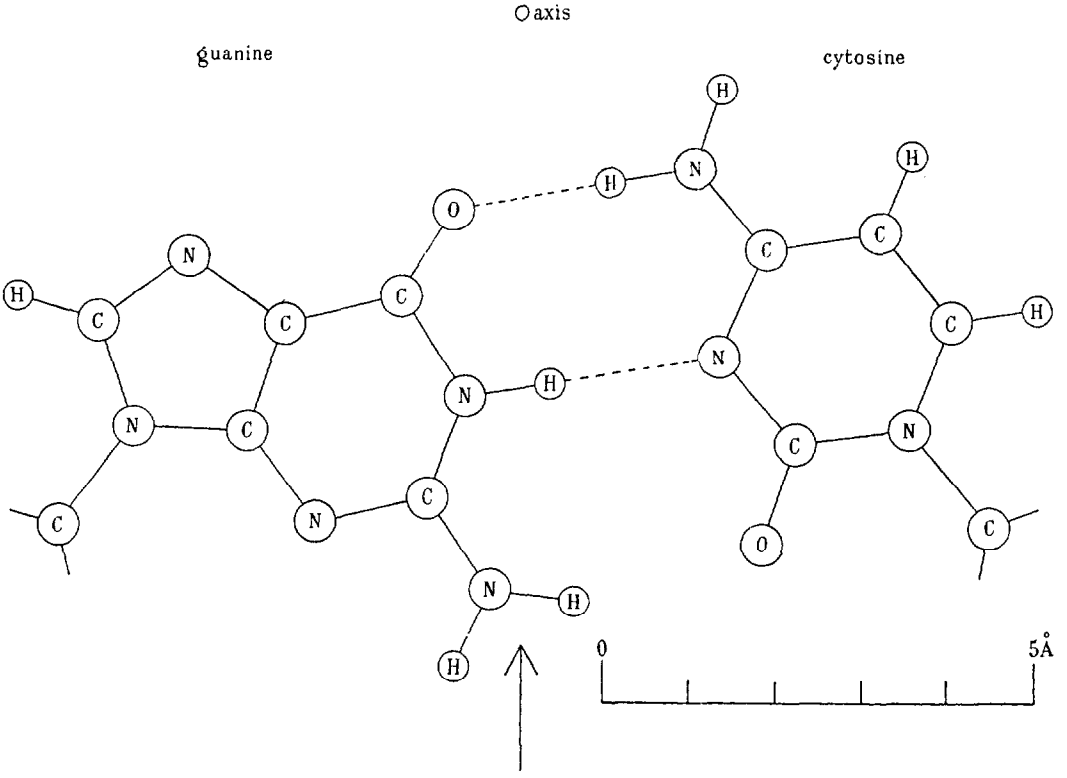

FIGURE 4. The pairing of guanine and cytosine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown. The arrow represents the crystallographic diad.

The crystal structures of both adenine and guanine have been studied by Broomhead (1948, 1951), while the structure of cytosine is known through Furberg's (1950) analysis of the crystal structure of cytidine. More recently Broomhead's data on adenine have been refined by Cochran (1951) and the atomic parameters of this compound are now accurate to within 0·02 Å.

As yet, no determination has been made of the structure of thymine, but it seems unlikely that its ring configuration will differ markedly from cytosine. Any deviations which might occur would have only a negligible effect on the pairing configuration, and we have utilized the idealized thymine configuration of figure 3. We also lack information about the exact angles at the β-glycosidic bond. There is no reason, however, to believe that they should differ significantly from those in cytidine or in the cyclic adenosine nucleoside studied by Zussman (1953), and they likewise have been assigned symmetrically.

The configuration of the adenine-thymine pair is stereochemically most satisfactory. The direction of the vector from the amino nitrogen to the keto oxygen lies exactly in the NH direction, as does the vector from the purine nitrogen atom 1 to the pyrimidine nitrogen atom 1. Both of the hydrogen bonds should therefore be of maximum stability (Donohue 1952). In addition, the two glycosidic bonds of the pair are related by a diad to within 1°, which is less than the accuracy to which the configuration of the bases is known. The distance apart of the $C_{1'}$ carbon atoms of the two sugars is close to 11 Å.

There is more ambiguity about the guanine-cytosine pair. This arises largely from doubt about the exact structure of guanine (Broomhead 1951). In particular, we are doubtful about the exact position of the keto oxygen atom. In figure 4 we have used the published position, and this makes the relative positions of the glycosidic bonds different from the adenine-thymine pair by about 2°. This difference would be negligible if the guanine keto oxygen was symmetrically placed. It is also uncertain as to whether this pair might form a third hydrogen bond between the amino group of guanine and the keto oxygen of cytosine. This point is unlikely to be settled until the configurations of both these bases are known to a greater accuracy. It seems clear, nevertheless, that these uncertainties are only of second-order importance, and that for all practical considerations the two pairs should be considered structurally equivalent.

The phosphate-sugar backbones were constructed utilizing a sugar configuration reported for ribose by Furberg (1950). A similar configuration for a pentose ring has also been reported by Beevers & Cochran (1947) in the fructofuranoside ring of sucrose. It seems probable that the furanose ring is puckered, and we have tentatively placed the $C_{3'}$ atom out of the ring in such a direction that its oxygen atom $O_{3'}$ is brought closer to the common plane. A tetrahedral arrangement has been assumed for the bond angles around the phosphorus atom. The bond lengths about the phosphorus have been assigned unsymmetrically following the suggestion of Pauling & Corey (1953), the two P—O bonds in the backbone have lengths of 1·65 Å while the remaining non ester P—O bonds are thought to have the shorter length of 1·45 Å. As a result of Furberg's analysis of cytidine (1950) there seems little doubt that the glycosidic bond is a single bond. We can thus be sure that the sugar group instead of being coplanar with the nitrogen base, as postulated by Astbury (1947), is more nearly perpendicular to it.

The paired bases are arranged so as to be approximately perpendicular to the fibre axis. This places the glycosidic bonds in a similar arrangement, while the puckered plane of the sugar ring assumes a position nearly parallel to the fibre axis. Each backbone chain completes one revolution after 10 residues in 34 Å, and so the rotation per residue is 36°. The phosphorus atoms are at radii of 10 Å, and the backbone has a configuration roughly similar to that described by Furberg (1952) in his paper dealing with suitable configurations for single helically arranged polynucleotide chains.

General views of the structure are shown in the photographs of figures 5 and 6, plate 2, which illustrate the salient features of a scale model. The drawings in figures 7 and 8 are given to demonstrate more accurately the exact configuration

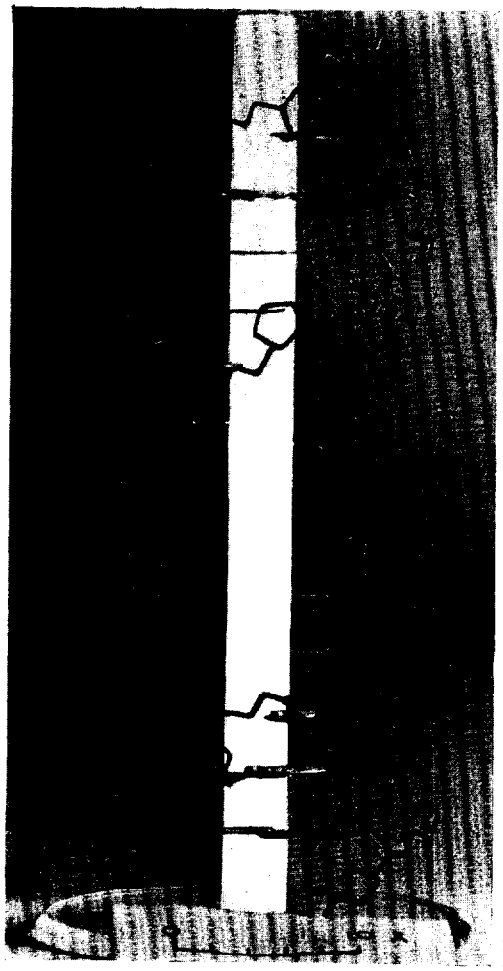

FIGURE 5.

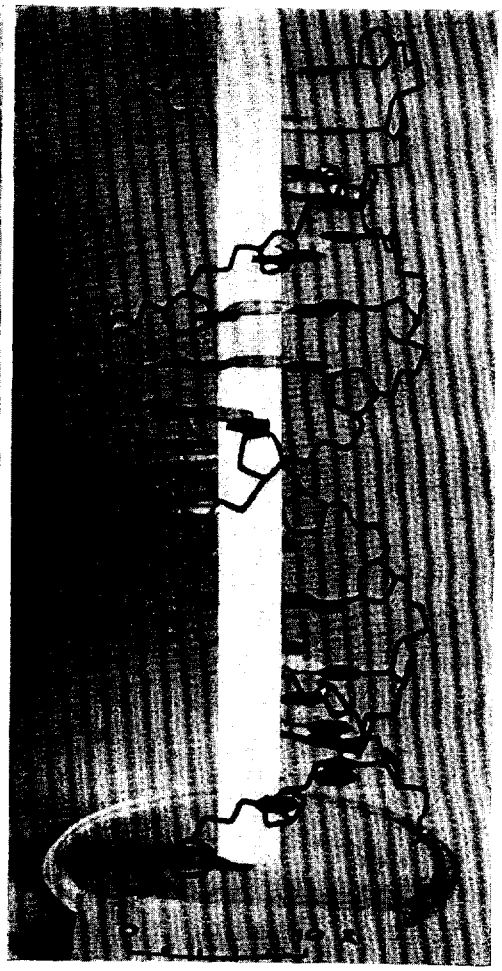

FIGURE 6.

FIGURE 5. Photograph of a rough scale model of the structure. The chemical bonds in the phosphate sugar backbone are represented by wire. (All the hydrogen atoms and the two oxygen atoms of the phosphate group not in ester linkage have been omitted.) The pairs of bases are represented by metal plates. The fibre axis is represented by a Perspex rod.

FIGURE 6. Another view of the model shown in figure 5. The white plates represent the area between the bases in which hydrogen bonding takes place.

(*Facing p.* 90)

of the backbone. Figure 7 shows two successive residues on the same chain projected on to a plane perpendicular to the fibre axis, while in figure 8 is shown a projection of a sugar-phosphate residue on to a plane whose normal is perpendicular to the fibre axis. It can be seen that the atoms forming the sequence

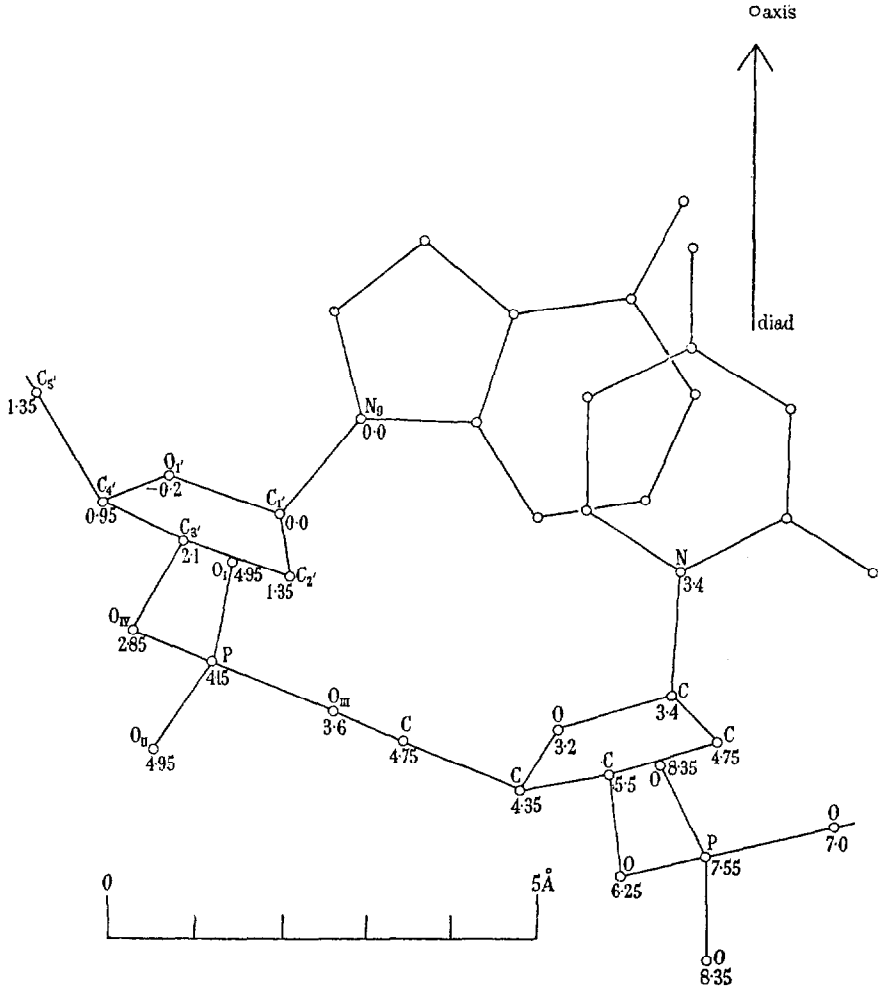

FIGURE 7. A projection of two successive residues of *one* chain of the structure. The direction of projection is parallel to the fibre axis. The figures show the height of each atom (in ångströms) above the level of the lower base.

$C_{4'}$—$C_{5'}$—$O_{5'}$—P—$O_{3'}$ all lie in such a plane; co-ordinates of the principal backbone atoms are given to 0·05 Å in table 1. No attempt has been made to place the sodium ion or the water molecules, though it is possible that some of these groups are located in relatively constant positions.

Because the two backbones are related by a diad, the distance between their effective 'centres of gravity' is much greater than might be imagined from the location of the glycosidic bonds. Instead of being separated by only $\frac{1}{4}$ of the fibre axis repeat (the angle of the pair of glycosidic bonds is close to 90°), they are

separated by approximately $\frac{3}{8}$ of the 34 Å repeat. In contrast to the outside of the molecule, the centre tends to give the impression of a one-stranded helix. This is a consequence of the intimate pairing of the bases.

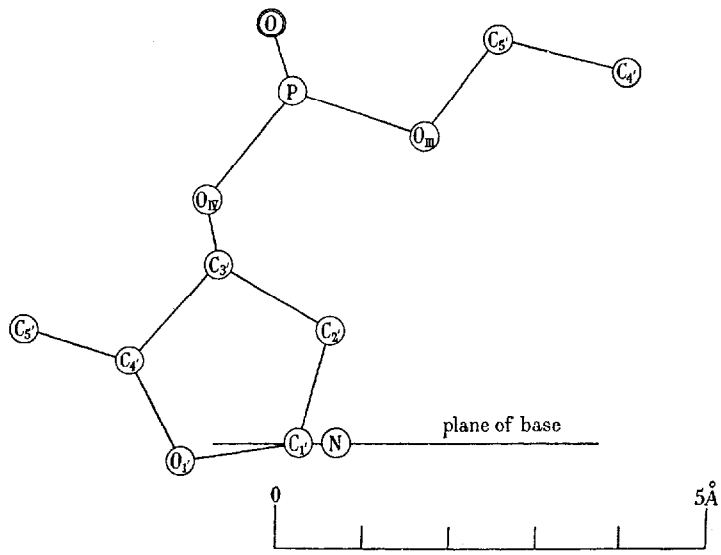

FIGURE 8. A projection of one residue in a direction perpendicular to both the fibre axis and to the plane containing the atoms $C_{4'}$—$C_{5'}$—O_5—P—$O_{3'}$.

TABLE 1. CO-ORDINATES FOR THE ATOMS OF THE BACKBONE, FOR A SINGLE RESIDUE

atom	ρ (Å)	ϕ	Z (Å)
P	10·0	0·0°	0·0
O_I	8·95	− 3·6°	+0·8
O_{II}	11·25	+ 0·7°	+0·8
O_{III}	9·65	+ 8·9°	−0·5
O_{IV}	10·35	− 5·3°	−1·3
$C_{5'}$	9·6	−22·2°	−2·8
$C_{4'}$	9·65	−13·2°	−3·2
$C_{3'}$	9·2	− 7·3°	−2·05
$C_{2'}$	8·65	+ 0·4°	−2·8
$C_{1'}$	8·2	− 3·5°	−4·15
$O_{1'}$	8·8	−11·8°	−4·35
N	6·7	− 4·2°	−4·15
diad	—	+39·0°	−4·15

Each of the van der Waals contacts appears to be acceptable. They are five relatively short contacts between the phosphate oxygen atoms and hydrogen atoms. None, however, is less than 2·5 Å, a quite acceptable length for side-by-side contacts. The position of the plane of the bases with respect to the sugar does not appear to be the optimum, but it is nevertheless within the range stated by Furberg as possible. Another short contact is found between the hydrogen atoms attached to the $C_{3'}$ and $C_{5'}$ atoms of the sugar. This contact, however, is also side by side, and so the postulated length (2·1 Å) appears permissible. The stagger of

hydrogen atoms between the $C_{4'}$—$C_{5'}$ bond is not optimal, but the deviation is only 25° and so allowable.

We can therefore conclude that the model is stereochemically feasible. Nevertheless, it is certainly not ideal, and it is possible that it could be improved by slightly altering the assumptions made about the configuration of the phosphorus atoms, especially its bond lengths, and by altering the configuration of the sugar. We have assumed that the puckering of the sugar ring is achieved by throwing the $C_{3'}$ atom out of the plane of the ring; a better model might result by choosing a different shape. Alternatively, it may be that an attraction between the rings of the bases is pulling the backbone out of its potential minimum.

The crystalline form

The transition to the crystalline form is accompanied by a decrease in water content (Franklin & Gosling 1953*a*), and it seems very probable that this form exists in a more tightly packed condition than the paracrystalline form. It is thus

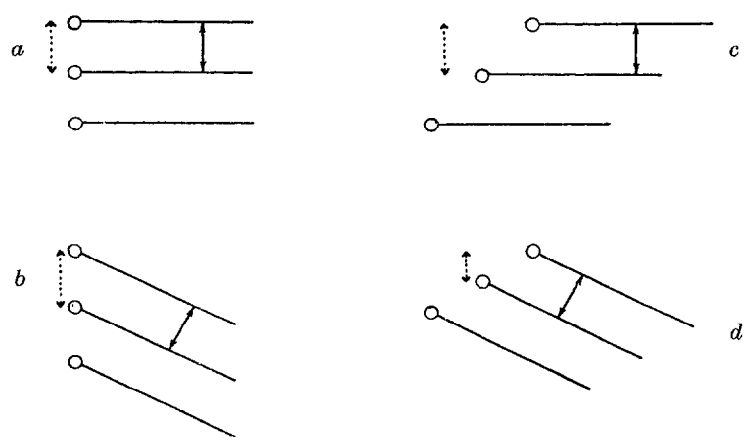

Figure 9. To show that if the bases are staggered, tilting will reduce the translation in the axis direction (represented by a dotted arrow). The solid arrow represents the perpendicular distance between the bases, which remains constant. *a* and *b*, not staggered; *c* and *d*, staggered; *a* and *c*, before tilting; *b* and *d*, after tilting.

not surprising to observe that the change to the crystalline state is characterized by a visual shortening of the fibre length of about 30% (Franklin & Gosling 1953*a*). There is little if any change in the diameter of the fibre, and so it seems likely that the fibre axis translation per nucleotide is reduced from 3·4 to approximately 2·5 Å. This conclusion might appear difficult to believe, as the van der Waals separation of the rings of the bases must remain the same and thus might appear to oppose a fibre shortening, but in fact the vertical translation can be reduced if the paired bases are tilted anti-clockwise (when viewed from the fibre axis).

The manner in which this might occur is shown in figure 9. It can be seen that shortening will only take place if successive pairs of bases are not stacked directly on top of one another, but are displaced to one side. In fact, if the bases are not

displaced, tilting will result in an increase of the fibre-axis translation. Of course, in our structure the successive pairs are displaced helically, not simply sideways as in figure 9, but this in no way destroys the general argument.

We should note that the hydrogen bonding arrangement remains unchanged by the tilting, as both members of a pair are similarly rotated about the perpendicular diad between the bases. This would not be so if the bases were instead related by a diad *parallel* to the fibre axis. In this latter case, the configuration of the backbone could be made equivalent only by tilting the two members of a pair in opposite directions and thus by effectively destroying the hydrogen bonds. Thus, if tilting is shown to occur in the crystalline state, we should have strong reasons for believing that the backbones are related by perpendicular diads.

We have not attempted to construct a detailed model with tilted bases, as we feel that this could be done more suitably in conjunction with the detailed X-ray evidence. Nevertheless, for the reasons outlined above, we believe that such a model can be built and that it will involve the same basic structural features proposed here for the paracrystalline form.

Discussion

Our structure bears only superficial resemblances to the majority of structures previously suggested. Most of these earlier formations (Astbury 1947; Furberg 1952) have involved single stranded structures and must be rejected on the basis of the density considerations outlined in the beginning of this paper. The only multi-stranded structure which previously has been seriously proposed is that of Pauling & Corey, who very kindly sent their manuscript to us prior to its publication. Their structure involved three intertwined helical chains in which the core of the molecule was formed by phosphate groups. Their proposal was submitted without knowledge of the work at King's College, London, by Wilkins and Franklin and their co-workers, and appears in the light of their experimental results to be untenable. The main objection to their proposal involves the number of chains. As indicated earlier the density of the crystalline form (Franklin & Gosling 1953*c*) strongly suggests the presence of two chains, and we find it difficult to imagine that any three-chained proposal can be made which will fit the experimental evidence.

The structure accounts in a nice way for the analytical data on the composition of DNA. By requiring specific pairing of purine and pyrimidine groups, it provides for the first time a suitable explanation for the recent chemical data (Chargaff 1951; Wyatt 1952; Chargaff, Crampton & Lipschitz 1953), which indicated not only a molar equivalence of the purines and pyrimidines, but also the molar equivalence of adenine and thymine, and of guanine and cytosine. The ratio of adenine to guanine varies greatly in DNA's from different sources, and it is difficult to imagine a structural explanation for the equivalence of adenine with thymine and of guanine with cytosine which does not involve specific pairing.

As far as we can tell our structure is compatible with the X-ray evidence of Wilkins and Franklin and their co-workers (Wilkins *et al.* 1953; Franklin & Gosling 1953*a*). In a preliminary report on their work, they have independently

suggested that the basic structure of the paracrystalline form is helical and contains two intertwined chains. They also suggest that the sugar-phosphate backbone forms the outside of the helix and that each chain repeats itself after one revolution in 34 Å.* Nevertheless, these crystallographic conclusions are tentative, and the structure can in no sense be considered proved until a satisfactory solution to the structure of the crystalline form is obtained.

In conclusion, we may mention that the complementary relationship between the two chains is very likely related to the biological role of DNA. It is generally assumed that DNA is a genetic substance and in some way possesses the capacity for self-duplication. It seems to us that the presence of a complementary structure strongly suggests that the self-duplicating process will be found to involve the alternative formation of complementary chains, and that each chain will be found capable of serving as a template for the formation of its complement. A fuller exposition of these latter ideas is given elsewhere (Watson & Crick 1953*b*,*c*).

We are most indebted to Dr M. H. F. Wilkins both for informing us of unpublished experimental observations and for the benefit of numerous discussions. We are also grateful to Dr J. Donohue for constant advice on the problems of tautomerism and van der Waals contacts, and to Professor A. R. Todd, F.R.S., for advice on chemical matters, and for allowing us access to unpublished work.

One of us (J.D.W.) wishes in addition to acknowledge the very kind hospitality provided during his stay at the Cavendish Laboratory by Sir Lawrence Bragg, F.R.S., and by the members of the Medical Research Council Unit located there. He is especially grateful to the encouragement provided by Dr J. C. Kendrew and Dr M. F. Perutz. In conclusion he would like to mention Professor S. E. Luria of the University of Illinois to whom he is indebted for both the opportunity to come to and to remain in Cambridge.

References

Astbury, W. T. 1947 *Symp. Soc. Exp. Biol.* **1**, Nucleic Acid, p. 66. Cambridge University Press.
Astbury, W. T. & Bell, F. O. 1938 *Nature, London*, **141**, 747.
Beevers, C. A. & Cochran, W. 1947 *Proc. Roy. Soc.* A, **190**, 257.
Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. 1951 *Nature, Lond.*, **168**, 271.
Broomhead, J. M. 1948 *Acta Cryst.* **1**, 324.
Broomhead, J. M. 1951 *Acta Cryst.* **4**, 92.
Brown, D. M. & Lythgoe, B. 1950 *J. Chem. Soc.* p. 1990.
Brown, D. M. & Todd, A. R. 1952 *J. Chem. Soc.* p. 52.
Chargaff, E. 1951 *J. Cell. Comp. Physiol.* **38**, 41.
Chargaff, E., Crampton, C. F. & Lipschitz, R. 1953 *Nature, Lond.*, **172**, 289.
Cochran, W. 1951 *Acta Cryst.* **4**, 81.
Cochran, W., Crick, F. H. C. & Vand, V. 1952 *Acta Cryst.* **5**, 581.
Crane, H. R. 1950 *Sci. Mon.* **70**, 376.

* More recently, Franklin & Gosling (1953*b*) have suggested that the X-ray data for the crystalline form also supports a structure of this general type. They also mention that the equatorial reflexions for the paracrystalline form suggest that the diameter of our model is a little too large. *Note added in proof*: Wilkins, Seeds, Stokes & Wilson (1953) have also presented X-ray evidence for the crystalline form being a pair of helices.

Dekker, C. R., Michelson, A. M. & Todd, A. R. 1953 *J. Chem. Soc.* p. 947.
Donohue, J. 1952 *J. Phys. Chem.* **56**, 502.
Franklin, R. E. & Gosling, R. G. 1953*a* *Nature, Lond.*, **171**, 740.
Franklin, R. E. & Gosling, R. G. 1953*b* *Nature, Lond.*, **172**, 156.
Franklin, R. E. & Gosling, R. G. 1953*c* *Acta Cryst.* **6**, 673, 678.
Furberg, S. 1950 *Acta Cryst.* **3**, 325.
Furberg, S. 1952 *Acta chem. scand.* **6**, 634.
Jordan, D. O. 1951 *Progr. biophys.* **2**, 51.
Kahler, H. & Lloyd, B. J. 1953 *Biochim. Biophys. Acta*, **10**, 355.
Michelson, A. M. & Todd, A. R. 1953 *J. Chem. Soc.* p. 951.
Pauling, L., Corey, R. B. & Branson, H. R. 1951 *Proc. Nat. Acad. Sci., Wash.*, **37**, 205.
Pauling, L. & Corey, R. B. 1953 *Proc. Nat. Acad. Sci., Wash.*, **39**, 84.
Riley, D. P. & Oster, G. 1951 *Biochim. Biophys. Acta*, **7**, 526.
Sadron, C. 1953 *Progr. biophys.* **3**, .
Tipson, R. S. 1945 *Advanc. Carbohyd. Chem.* **1**, 238. New York: Academic Press Inc.
Watson, J. D. & Crick, F. H. C. 1953*a* *Nature, Lond.*, **171**, 737.
Watson, J. D. & Crick, F. H. C. 1953*b* *Nature, Lond.*, **171**, 964.
Watson, J. D. & Crick, F. H. C. 1953*c* *Cold Spr. Harb. Symp. Quant. Biol.* (in the Press).
Wilkins, M. H. F., Gosling, R. G. & Seeds, W. E. 1951 *Nature, Lond.*, **167**, 759.
Wilkins, M. H. F., Seeds, W. E., Stokes, A. R. & Wilson, H. R. 1953 *Nature, Lond.*, **172**, 759.
Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. 1953 *Nature, Lond.*, **171**, 738.
Williams, R. C. 1952 *Biochim. Biophys. Acta*, **9**, 237.
Wyatt, G. R. 1952 *The chemistry and physiology of the nucleus.* New York: Academic Press.
Wyatt, G. R. & Cohen, S. S. 1952 *Nature, Lond.*, **170**, 846.
Zussman, J. 1953 *Acta Cryst.* **6**, 504.