THE MOLECULAR CONFIGURATION OF NUCLEIC ACIDS

by

M. H. F. WILKINS

Medical Research Council Biophysics Research Unit, Biophysics Department, King's College, University of London

Nobel Lecture, December 11, 1962

Nucleic acids are basically simple. They are at the root of very fundamental biological processes, growth and inheritance. The simplicity of nucleic acid molecular structure and of its relation to function expresses the underlying simplicity of the biological phenomena, clarifies their nature, and has given rise to the first extensive interpretation of living processes in terms of macromolecular structure. These matters have only become clear by an unprecedented combination of biological, chemical and physical studies, ranging from genetics to hydrogenbond stereochemistry. I shall not discuss all this here but concentrate on the field in which I have worked, and show how X-ray diffraction analysis has made its contribution. I shall describe some of the background of my own researches, for I suspect I am not alone in finding such accounts often more interesting than general reviews.

Early Background

I took a physics degree at Cambridge in 1938, with some training in X-ray crystallography. This X-ray background was influenced by J. D. Bernal, then at the Cavendish. I began research at Birmingham, under J. T. Randall, studying luminescence and how electrons move in crystals. My contemporaries at Cambridge had mainly been interested in elementary particles, but the organisation of the solid state and the special properties which depended on this organisation interested me more. This may have been a forerunner of my interest in biological macromolecules and how their structure related to their highly specific properties which so largely determine the processes of life.

During the war I took part in making the atomic bomb. When the war was ending, I, like many others, cast around for a new field of research. Partly on account of the bomb, I had lost some interest in physics. I was therefore very interested when I read Schrödinger's book "What is Life" and was struck by the concept of a highly complex molecular structure which controlled living processes. Research on such matters seemed more ambitious than solid-state physics. At that time many leading physicists such as Massey, Oliphant, and Randall (and later I learned that Bohr shared their view) believed that physics would contribute significantly to biology; their advice encouraged me to move into biology.

I went to work in the Physics Department at St. Andrews, Scotland, where Randall had invited me to join a biophysics project he had begun. Stimulated by Muller's experimental modification, by means of X-radiation, of genetic substance, I thought it might be interesting to investigate the effects of ultrasonics; but the results were not very encouraging.

The biophysics work then moved to King's College, London, where Randall took the Wheatstone Chair of Physics and built up, with the help of the Medical Research Council, an unusual laboratory for a Physics Department, where biologists, biochemists and others worked with the physicists. He suggested I might take over some ultra-violet microscope studies of the quantities of nucleic acids in cells. This work followed that of Caspersson, but made use of the achromatism of reflecting microscopes. By this time, the work of Caspersson (1941) and Brachet (1941) had made the scientific world generally aware that nucleic acids had important biological roles which were connected with protein synthesis. The idea that DNA might itself be the genetic substance was, however, barely hinted at. Its function in chromosomes was supposed to be associated with replication of the protein chromosome thread. The work of Avery, MacLeod and McCarty, showing that bacteria could be genetically transformed by DNA, was published in 1944, but even in 1946 seemed almost unknown, or if known its significance was often belittled.

It was fascinating to look through microscopes at chromosomes in cells, but I began to feel that as a physicist I might contribute more to biology by studying macromolecules isolated from cells. I was encouraged in this by Gerald Oster who came from Stanley's virus laboratory and interested me in particles of tobacco mosaic virus. As Caspersson had shown, ultra-violet microscopes could be used to find the orientation of ultra-violet absorbing groups in molecules as well as to measure quantities of nucleic

acids in cells. Bill Seeds and I studied DNA, proteins, tobacco mosaic virus, vitamin B12, etc. While examining oriented films of DNA prepared for ultraviolet dichroism studies, I saw in the polarising microscope extremely uniform fibres giving clear extinction between crossed nicols. I found the fibres had been produced unwittingly while I was manipulating DNA gel. Each time that I touched the gel with a glass rod and removed the rod, a thin and almost invisible fibre of DNA was drawn out like a filament of spider's web. The perfection and uniformity of the fibres suggested that the molecules in them were regularly arranged. I immediately thought the fibres might be excellent objects to study by X-ray diffraction analysis. I took them to Raymond Gosling, who had our only X-ray equipment (made from war-surplus radiography parts) and who was using it to obtain diffraction photographs from heads of ram spermatozoa. This research was directed by Randall, who had been trained under W. L. Bragg and had worked with X-ray diffraction. Almost immediately, Gosling obtained very encouraging diffraction patterns (see fig. 1). One reason for this success was that we kept the fibres moist. We remembered that, to obtain detailed X-ray patterns from proteins, Bernal had kept protein crystals in their mother liquor. It seemed likely that the configuration of all kinds of water-soluble biological macromolecules would depend on their aqueous environment. We obtained good diffraction patterns with DNA made by Signer and Schwander (1949), which Signer brought to London to a Faraday Society meeting on nucleic acids and which he generously distributed so that all workers, using their various techniques, could study it.

Realisation that the genetic material was a pure chemical substance, and signs that its molecular structure was singularly simple

Between 1946 and 1950 many lines of evidence were uncovered indicating that the genetic substance was DNA, not protein or nucleoprotein. For instance, it was found that the DNA content of a set of chromosomes was constant, and that DNA from a given species had a constant composition although the nucleotide sequence in DNA molecules was complex. It was suggested that genetic information was carried in the polynucleotide chain in a complicated sequence of the four nucleotides. The great significance of bacterial transformation now became generally recognised, and the
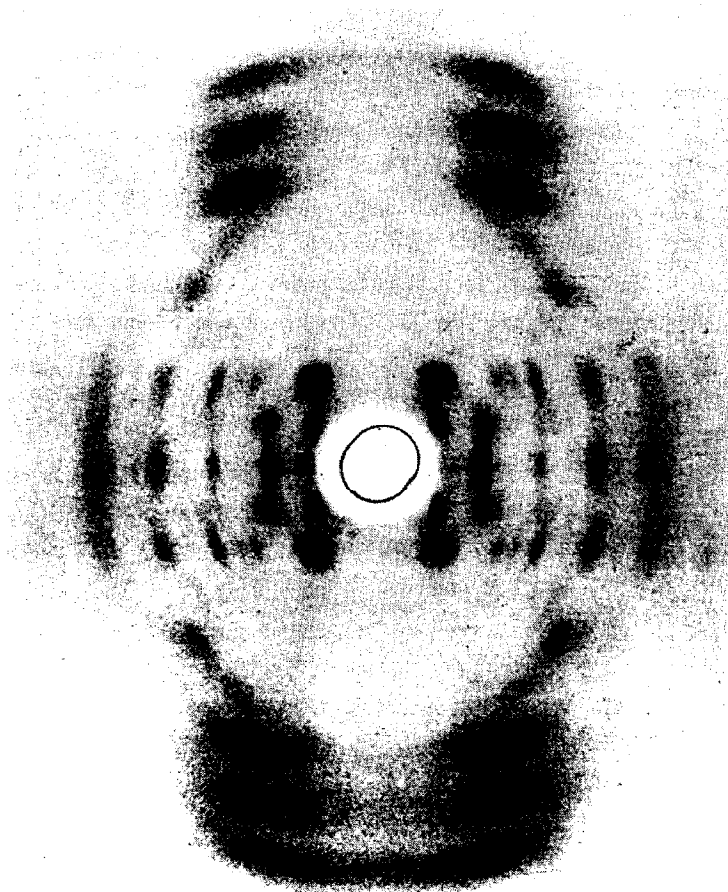
Fig. 1. One of the first X-ray diffraction photographs of DNA taken in our laboratory. This may be compared with the later photograph in Fig. 9. (photograph with R. Gosling; DNA by R. Signer).

demonstration by Hershey and Chase (1952) that bacteriophage DNA carried the viral genetic information from parent to progeny helped to complete what was a fairly considerable revolution in thought.

The prospects of elucidating genetic function in terms of molecular structure were greatly improved when it was known that the genetic substance was DNA, which had a well defined chemical structure, rather than an ill-defined nucleoprotein. There were many indications of sim-

plicity and regularity in DNA structure. The chemists had shown that DNA was a polymer in which the phosphate and deoxyribose parts of the molecule were regularly repeated in a polynucleotide chain with 3'— 5' linkages. Chargaff (1950) discovered an important regularity: although the sequence of bases along the polynucleotide chains was complex and the base composition of different DNA's varied considerably, the numbers of adenine and thymine groups were always equal, and so were the numbers of guanine and cytosine. In the electron microscope, DNA was seen as a uniform unbranched thread of diameter about 20 Å. Signer, Caspersson and Hammarsten (1938) showed by flow-birefringence measurements that the bases in DNA lay with their planes roughly perpendicular to the length of the thread-like molecule. Their ultra-violet dichroism measurements gave the same results and showed marked parallelism of the bases in the DNA in heads of spermatozoa. Earlier Schmidt (1937) and Pattri (1932) had studied optically the remarkable ordering of the genetic material in sperm heads. Astbury (1947) made pioneer X-ray diffraction studies of DNA fibres and found evidence of considerable regularity in DNA; he correctly interpreted the strong 3.4 Å reflection as being due to planar bases stacked on each other. The electro-titrometric study by Gulland and Jordan (1947) showed that the bases were hydrogen-bonded together, and indeed Gulland (1947) suggested that the polynucleotide chains might be linked by these hydrogen bonds to form multi-chain micelles.

Thus the remarkable conclusion that a pure chemical substance was invested with a deeply significant biological activity coincided with a considerable growth of many-sided knowledge of the nature of the substance. Meanwhile we began to obtain detailed X-ray diffraction data from DNA. This was the only type of data that could provide an adequate description of the 3-dimensional configuration of the molecule.

The need for combining X-ray diffraction studies of DNA with molecular model-building

As soon as good diffraction patterns were obtained from fibres of DNA, great interest was aroused. In our laboratory, Alex Stokes provided a theory of diffraction from helical DNA. Rosalind Franklin (who died some years later at the peak of her career) made very valuable contributions to the X-ray analysis. In Cambridge, at the Medical Research Council laboratory

where structures of biological macromolecules were studied, my friends Francis Crick and Jim Watson were deeply interested in DNA structure. Watson was a biologist who had gone to Cambridge to study molecular structure. He had worked on bacteriophage reproduction and was keenly aware of the great possibilities that might be opened up by finding the molecular structure of DNA. Crick was working on helical protein structure and was interested in what controlled protein synthesis. Pauling and Corey, by their discovery of the protein α-helix, had shown that precise molecular model-building was a powerful analytical tool in its own right. The X-ray data from DNA were not so complete that a detailed picture of DNA structure could be derived without considerable aid from stereochemistry. It was clear that the X-ray studies of DNA needed to be complemented by precise molecular model-building. In our laboratory we concentrated on amplifying the X-ray data. In Cambridge, Watson and Crick built molecular models.

The paradox of the regularity of the DNA molecule

The sharpness of the X-ray diffraction patterns of DNA showed that DNA molecules were highly regular — so regular that DNA could crystallise. The form of the patterns gave clear indications that the molecule was helical, the polynucleotide chains in the molecular thread being regularly twisted. It was known, however, that the purines and pyrimidines of various dimensions were arranged in irregular sequence along the polynucleotide chains. How could such an irregular arrangement give a highly regular structure? This paradox pointed to the solution of the DNA structure problem and was resolved by the structural hypothesis of Watson and Crick.

The helical structure of the DNA molecule

The key to DNA molecular structure was the discovery by Watson and Crick (1953a) that, if the bases in DNA were joined in pairs by hydrogen-bonding, the overall dimensions of the pairs of adenine and thymine and of guanine and cytosine were identical. This meant that a DNA molecule containing these pairs could be highly regular in spite of the sequence of the bases being irregular. Watson and Crick proposed that the DNA molecule consisted of two polynucleotide chains joined together by base-
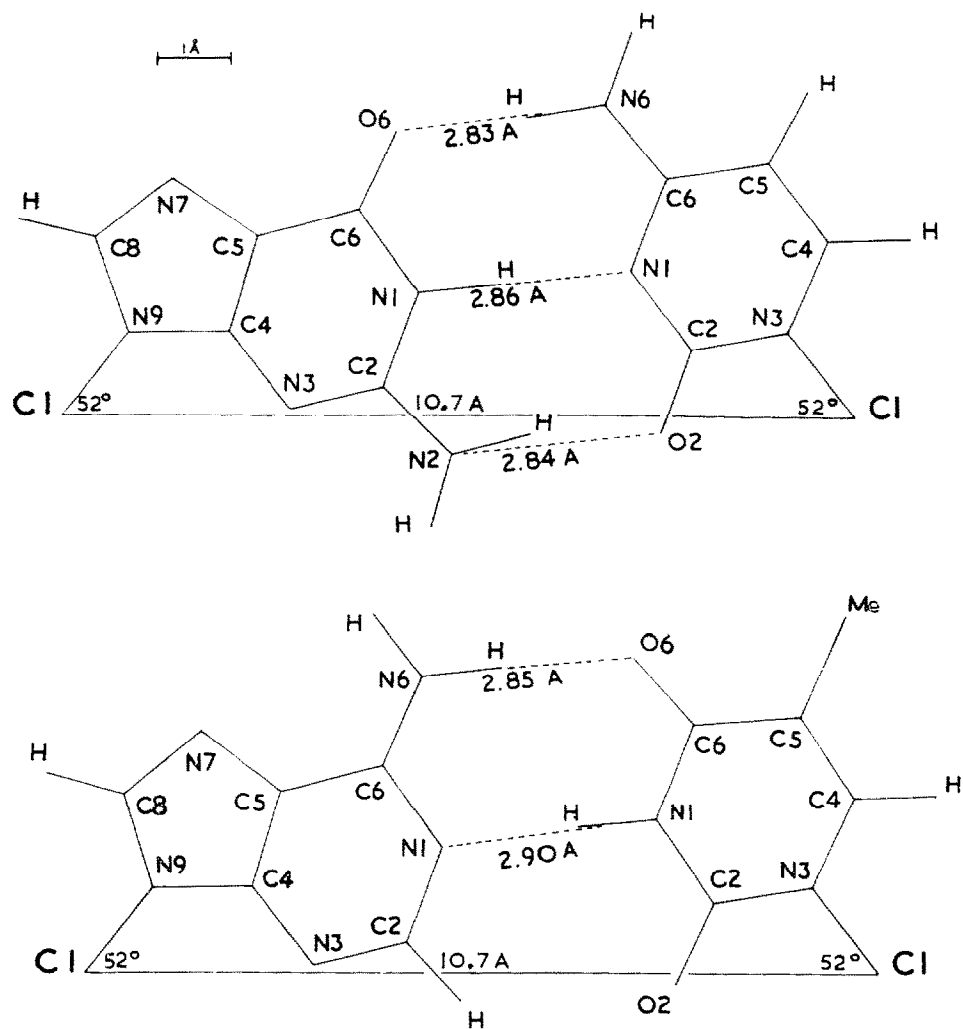
Fig. 2. Watson-Crick base-pairs (revised by S. Arnott). Top, guanine hydrogen-bonded to cytosine. Bottom, adenine hydrogen-bonded to thymine. The distances between the ends of the C_1N_9 and C_1N_1 bonds are 10.7 Å in both pairs, and all these bonds make an angle of 52° with the C_1C_1 line.

pairs. These pairs are shown in Fig. 2. The distance between the bonds joining the bases to the deoxyribose groups is exactly (within the uncertainty of 0.1 Å or so) the same for both base-pairs, and all those bonds make exactly (within the uncertainty of 1° or so) the same angle with the line joining the C_1 atoms of the deoxyribose (see Fig. 2). As a result, if two polynucleotide chains are joined by the base-pairs, the distance be-

tween the two chains is the same for both base-pairs and, because the angle between the bonds and the C_1C_1 line is the same for all bases, the geometry of the deoxyribose and phosphate parts of the molecule can be exactly regular.

Watson and Crick built a two-chain molecular model of this kind, the chains being helical and the main dimensions being as indicated by the X-ray data. In the model one polynucleotide chain is twisted round the other and the sequence of atoms in one chain runs in opposite direction to that in the other. As a result, one chain is identical with the other if turned upside down, and every nucleotide in the molecule has identical structure and environment. The only irregularities are in the base sequences. The sequence along one chain can vary without restriction, but base-pairing requires that adenine in one chain be linked to thymine in the other, and similarly guanine to cytosine. The sequence in one chain is, therefore, determined by the sequence in the other, and is said to be complementary to it.

The structure of the DNA molecule in the B configuration is shown in Fig. 3. The bases are stacked on each other 3.4 Å apart and their planes are almost perpendicular to the helix axis. The flat sides of the bases cannot bind water molecules; as a result there is attraction between the bases when DNA is in an aqueous medium. This hydrophobic bonding, together with the base-pair hydrogen-bonding, stabilises the structure.

The Watson-Crick hypothesis of DNA replication, and transfer of information from one polynucleotide chain to another

It is essential for genetic material to be able to make exact copies of itself; otherwise growth would produce disorder, life could not originate, and favourable forms would not be perpetuated by natural selection. Base-pairing provides the means of self-replication (Watson and Crick 1953 b). It also appears to be the basis of information transfer during various stages in protein synthesis.

Genetic information is written in a four-letter code in the sequence of the four bases along a polynucleotide chain. This information may be transferred from one polynucleotide chain to another. A polynucleotide chain acts as a template on which nucleotides are arranged to build a new chain. Provided that the two-chain molecule so formed is exactly
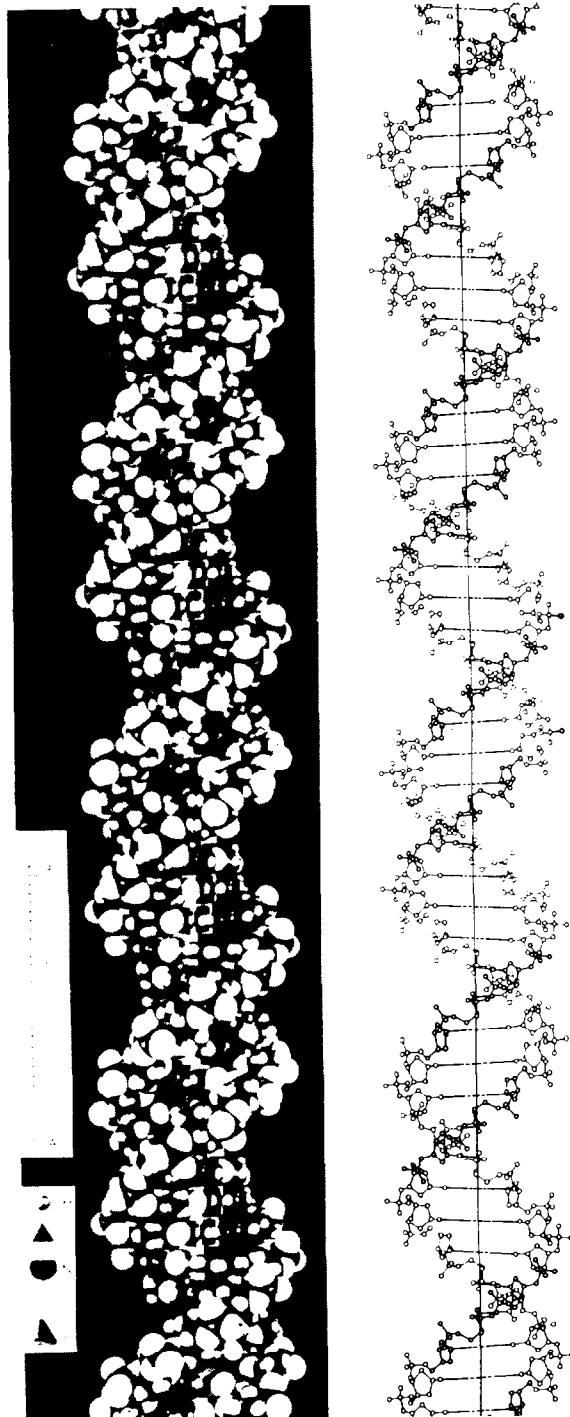
134

Fig. 3. Left. Molecular model of the *B* configuration of DNA. The sizes of the atoms correspond to van der Waals diameters. Right. Diagram corresponding to the model. The two polynucleotide chains, joined by hydrogen-bonded bases, may be seen clearly.

regular, base-pairing ensures that the sequence in the new chain is exactly complementary to that in the parent chain. If the two chains then separate, the new chain can act as a template, and a further chain is formed; this is identical with the original chain. Most DNA molecules consist of two chains; clearly, the copying process can be used to replicate such a molecule. It can also be used to transfer information from a DNA chain to an RNA chain (as is believed to be the case in the formation of messenger RNA).

Base-pairing also enables specific attachments to be made between part of one polynucleotide chain and a complementary sequence in another. Such specific interaction may be the means by which amino acids are attached to the requisite portions of a polynucleotide chain that has encoded in it the sequence of amino acids that specifies a protein. In this case the amino acid is attached to a transfer RNA molecule and part of the polynucleotide chain in this RNA pairs with the coding chain.

Since the base-pairs were first described by Watson and Crick in 1953, many new data on purine and pyrimidine dimensions and hydrogen bond lengths have become available. The most recent refinement of the pairs (due to S. Arnott) is shown in Fig. 2. We now take the distance between C_1 atoms as 10.7 Å instead of the value used recently of 11.0 Å, mainly because new data on N-H . . . N bonds show that this distance is 0.2 Å shorter between ring nitrogen atoms than between atoms that are not in rings. The linearity of the hydrogen bonds in the base-pairs is excellent and the lengths of the bonds are the same as those found in crystals (these lengths vary by about 0.04 Å).

The remarkable precision of the base pairs reflects the exactness of DNA replication. One wonders, however, why the precision is so great, for the energy required to distort the base-pairs so that their perfection is appreciably less, is probably no greater than one quantum of thermal energy. The explanation may be that replication is a co-operative phenomenon involving many base-pairs. In any case, it must be emphasised that the specificity of the base-pairing depends on the bonds joining the bases to the deoxyribose groups being correctly placed in relation to each other. This placing is probably determined by the DNA polymerising enzyme. Whatever the mechanics of the process are, the exact equivalence of geometry and environment of every nucleotide in the double-helix should be conducive to precise replication. Mistakes in the copying process will be produced if there are tautomeric shifts of protons involved

in the hydrogenbonding, or chemical alterations of the bases. These mistakes can correspond to mutations.

The universal nature and constancy of the helical structure of DNA

After our preliminary X-ray studies had been made, my friend Leonard Hamilton sent me human DNA he and Ralph Barclay had isolated from human leukocytes of a patient with chronic myeloid leukemia. He was studying nucleic acid metabolism in man in relation to cancer and had prepared the DNA in order to compare the DNA of normal and leukaemic leukocytes. The DNA gave a very well-defined X-ray pattern. Thus began a collaboration that has lasted over many years and in which we have used Hamilton's DNA, in the form of many salts, to establish the correctness of the double helix structure. Hamilton prepared DNA from a very wide range of species and diverse tissues. Thus it has been shown that the DNA double helix is present in inert genetic material in sperm and bacteriophage, and in cells slowly or rapidly dividing or secreting protein (Hamilton *et al.* 1959). No difference of structure has been found between DNA from normal and from cancerous tissues, or in calf thymus DNA separated into fractions of different base composition by my colleague Geoffrey Brown.

We also made a study, in collaboration with Harriet Ephrussi-Taylor, of active transforming principle from pneumococci, and observed the same DNA structure. The only exception to double helical DNA so far found is in some very small bacteriophages where the DNA is single-stranded. We have found, however, that DNA, with an unusually high content of adenine, or with glucose attached to hydroxymethylcytosine, crystallised differently.

DNA structure is not an artefact

It did not seem enough to study X-ray diffraction from DNA alone. Obviously one should try to look at genetic material in intact cells. It was possible that the structure of the isolated DNA might be different from that *in vivo*, where DNA was in most cases combined with protein. The optical studies indicated that there was marked molecular order in sperm heads and that they might therefore be good objects for X-ray study, whereas chromosomes in most types of cells were complicated objects with little sign of ordered structure. Randall had been interested in this
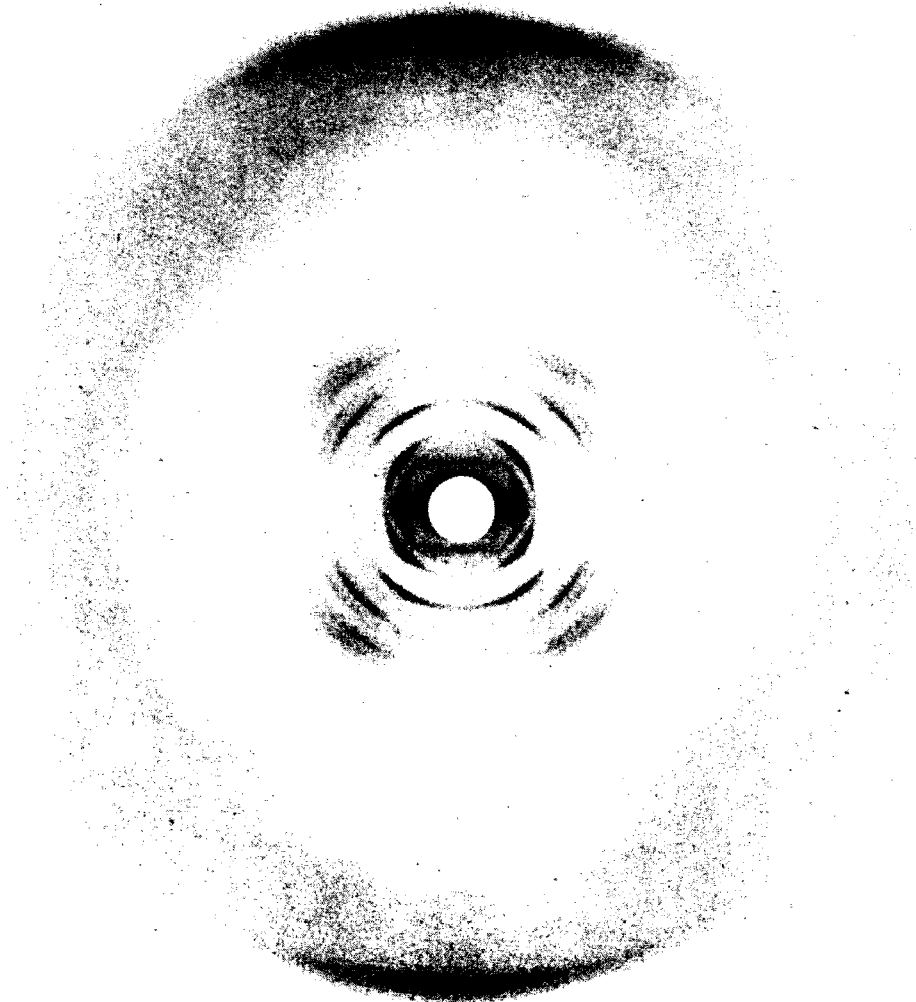
Fig. 4. X-ray diffraction pattern of cephalopod sperm. The DNA molecules in the sperm heads have their axes vertical. The 3.4 Å internucleotide spacing corresponds to the strong diffraction at the top and bottom of the pattern. The sharp reflections in the central part of the pattern show that the molecules are in crystalline array.

matter for some years and had started Gosling studying ram sperm. It seemed that the rod-shaped cephalopod sperm, found by Schmidt to be highly anisotropic optically, would be excellent for X-ray investigation. Rinne (1933), while making a study of liquid crystals from many branches of Nature, had already taken diffraction photographs of such sperm; but presumably his technique was inadequate, for he came to the mistaken con-
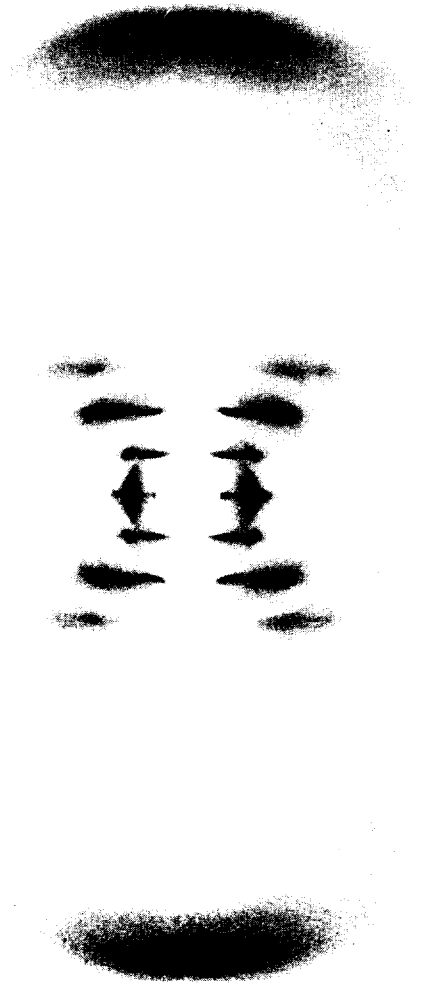
Fig. 5. X-ray diffraction photograph of DNA fibres (*B* configuration) at high humidity. The fibres are vertical. The 3.4 Å reflection is at the top and bottom. The angle in the pronounced X shape, made by the reflections in the central region, corresponds to the constant angle of ascent of the polynucleotide chains in the helical molecule. (Photograph with H. R. Wilson; DNA by L. D. Hamilton.)

clusion that the nucleoprotein was liquid-crystalline. Our X-ray photographs (Wilkins and Randall 1953) showed clearly that the material in the sperm heads had 3-dimensional order, i. e. it was crystalline and not liquid-crystalline. The diffraction pattern (Fig. 4) bore a close resem-

blance to that of DNA (Fig. 5), thus showing that the structure in fibres of purified DNA was basically not an artefact. Working at the Stazione Zoologica in Naples, I found it possible to orient the sperm heads in fibres. Intact wet spermatophore, being bundles of naturally-oriented sperm, gave good diffraction patterns. DNA-like patterns were also obtained from T2 bacteriophage given me by Watson.

The X-ray diffraction patterns of DNA and the various configurations of the molecule

X-ray diffraction analysis is the only technique that can give very detailed information about the configuration of the DNA molecule. Optical techniques, though valuable as being complementary to X-ray analysis, provide much more limited information — mainly about orientation of bonds and groups. X-ray data contributed to the deriving of the structure of DNA at two stages. First, in providing information that helped in building the Watson-Crick model; and second, in showing that the Watson-Crick proposal was correct in its essentials, which involved readjusting and refining the model.

The X-ray studies (e. g. Langridge *et al.* 1960, Wilkins 1961) show that DNA molecules are remarkable in that they adopt a large number of different conformations, most of which can exist in several crystal forms. The main factors determining the molecular conformation and crystal form are the water and salt contents of the fibres and the cation used to neutralise the phosphate groups (see Table 1).

I shall describe briefly the three main configurations of DNA. In all cases the diffraction data are satisfactorily accounted for in terms of the same basic Watson-Crick structure. This is a much more convincing demonstration of the correctness of the structure than if one configuration alone were studied. The basic procedure is to adjust the molecular model until the calculated intensities of diffraction from the model correspond to those observed (Langridge *et al.* 1960).

As with most X-ray data, only the intensities, and not the phases, of the diffracted beams from DNA are available. Therefore the structure cannot be derived directly. If the resolution of X-ray data is sufficient to separate most of the atoms in a structure, the structure may be derived with no stereochemical assumption except that the structure is assumed to consist of atoms of known average size. With DNA, however, most of

TABLE 1. Summary of various forms of DNA in fibres

| Configuration of molecule	Number of nucleotide pairs per turn of helix	Inclination of base pairs in molecule	Salt	Relative humidity and condition necessary	Crystal class	Crystallinity	Molecular positions	Unit-cell dimensions			
								a (Å)	b (Å)	c (Å)	β
A	11·0	20	Na K Rb	75 %	mono-clinic	crystalline	$0, 0, 0$ $\frac{1}{2}, \frac{1}{2}, 0$	22·24	40·62	28·15	97·0
B	10	~0	Li	66 % 3 % LiCl in fibre	ortho-rhombic	crystalline	$0, 0, \frac{1}{6}$ $\frac{1}{2}, \frac{1}{2}, -\frac{1}{6}$	22·5	30·9	33·7	—
			Li	75 – 90 %	ortho-rhombic	semi-crystalline	$0, 0, \frac{1}{8}$ $\frac{1}{2}, \frac{1}{2}, -\frac{1}{8}$	21·4	38·5	33·6	–
			Li Na K Rb	92 %	hexagonal	semi-crystalline	$0, 0, 0$ $\frac{1}{3}, \frac{2}{3}, \frac{1}{6}$ $\frac{2}{3}, \frac{1}{3}, -\frac{1}{6}$	46	—	34·6	–
B$_2$	9·9	0?	Na	75 % under tension	tetragonal	semi-crystalline	$0, 0, 0$ $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$	27·4	-	33·8	—
C	9·3	–5	Li	44 % no LiCl	ortho-rhombic	semi-crystalline	$0, 0, \frac{1}{5}$ $\frac{1}{2}, \frac{1}{2}, -\frac{1}{5}$	20·1	31·9	30·9	–
			Li	44 % in some specimens only. No LiCl in fibre	hexagonal	semi-crystalline	$0, 0, 0$ or $\frac{1}{2}$ $\frac{1}{3}, \frac{2}{3}, \frac{1}{6}$ or $-\frac{1}{3}$ $\frac{2}{3}, \frac{1}{3}, -\frac{1}{6}$ or $\frac{1}{3}$	35·0		30·9	
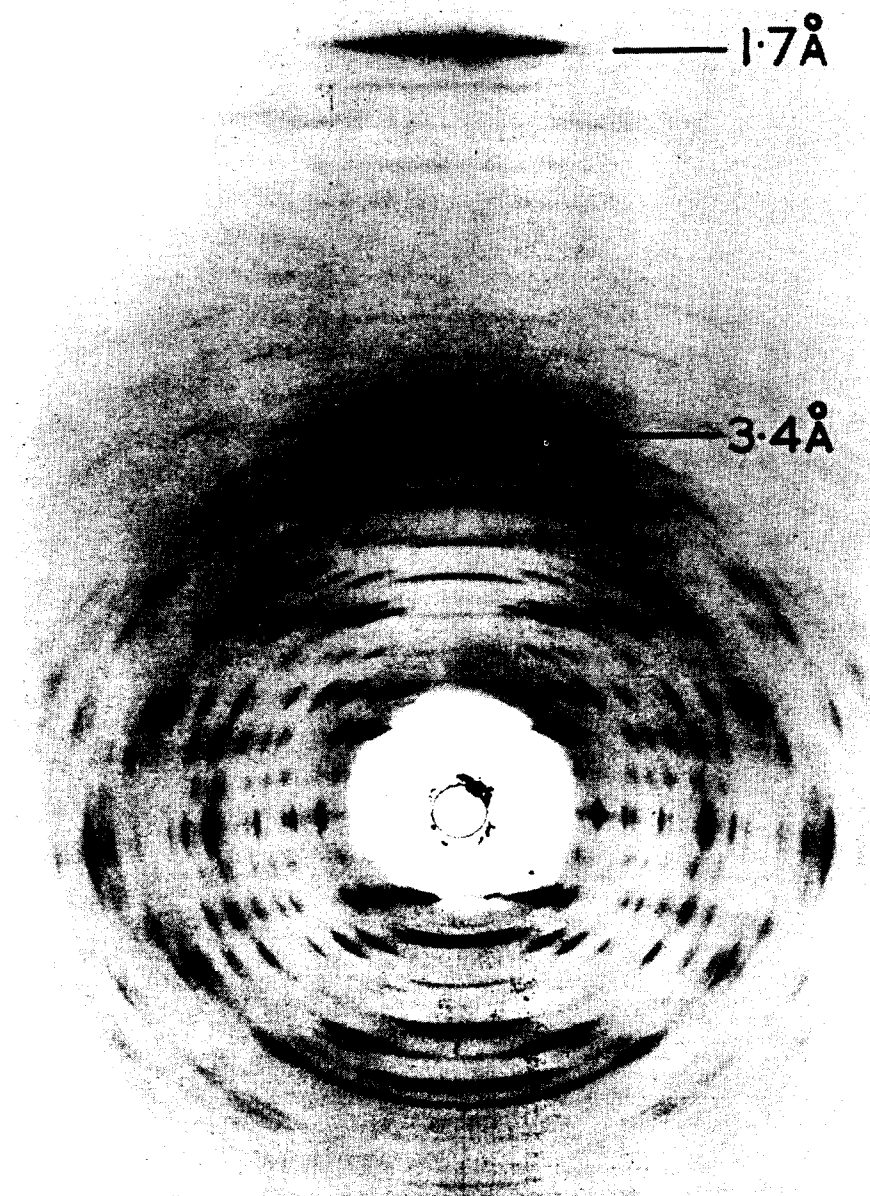
Fig. 6. X-ray pattern of microcrystalline fibres of DNA. The general intensity distribution is similar to that in Plate 4 but the diffraction is split into sharp reflections, owing to the regular arrangement of the molecules in the crystals. Sharp reflections extend to spacings as small as 1.7 Å. (Photograph with N. Chard; DNA by L. D. Hamilton.)
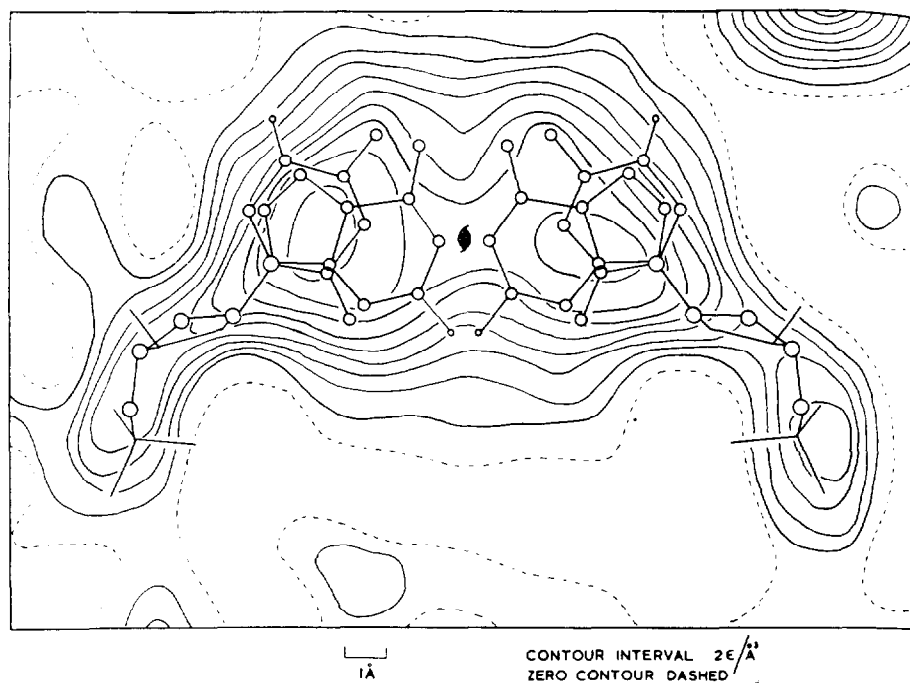
142

CONTOUR INTERVAL 2 ϵ/$\mathring{A}^3$
ZERO CONTOUR DASHED

Fig. 7. Fourier synthesis map (by S. Arnott) showing the distribution of electron density in the plane of a base-pair in the *B* configuration of DNA. The distribution corresponds to an average base-pair. The shape of the base-pair appears in the map, but individual atoms in a base-pair are not resolved. (The Fourier synthesis is being revised and the map is subject to improvement.)

the atoms cannot be separately located by the X-rays alone (see Fig. 7). Therefore, more extensive stereochemical assumptions are made: these take the form of molecular model-building. There are no alternatives to most of these assumptions; but where there might be an alternative, e. g. in the arrangement of hydrogen bonds in a base-pair, the X-ray data should be used to establish the correctness of the assumption. In other words, it is necessary to establish that the structure proposed is unique. Most of our work in recent years has been of this nature. To be reasonably certain that the DNA structure was correct, X-ray data, as extensive as possible, had to be collected.

The B Configuration

Fig. 5 shows a diffraction pattern of a fibre of DNA at high humidity when the molecules are separated by water and, to a large extent, behave

Fig. 8. Molecular model of DNA in the *A* configuration. The base-pairs may be seen inclined 20° to the horizontal.

independently of each other. We have not made intensive study of DNA under these conditions. The patterns could be improved, but they are reasonably well defined, and the sharpness of many of their features shows that the molecules have a regular structure. The configuration is known as B (see also Fig. 3); it is observed *in vivo*, and there is evidence that it exists when DNA is in solution in water. There are 10 nucleotide pairs per helix turn. There is no obvious structural reason why this number should be integral; if it is exactly so, the significance of this is not yet apparent.

When DNA crystallises, the process of crystallisation imposes restraints on the molecule and can give it extra regularity. Also, the periodic arrangement of the molecules in the microcrystals in the fibre causes the diffrac-
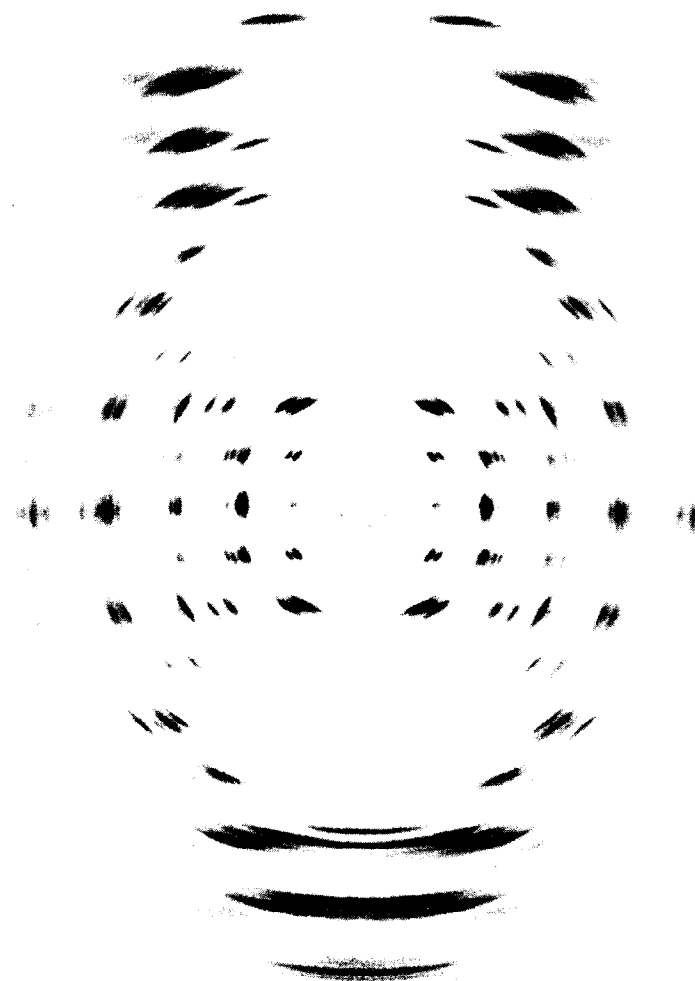
Fig. 9. X-ray diffraction pattern of microcrystalline fibres of DNA in the A configuration. (Photograph with H. R. Wilson; DNA by L. D. Hamilton.)

tion pattern to be split into sharp reflections corresponding to the various crystal planes (Fig. 6). Careful measurement of the positions of the reflections and deduction of the crystal lattice enables the directions of the reflections to be identified in three dimensions. Diffraction patterns from most fibrous substances resemble Fig. 5 in that the diffraction data are

2-dimensional. In contrast, the crystalline fibres of DNA give fairly complete 3-dimensional data. These data give information about the appearance of the molecule when viewed from all angles, and are comparable with those from single crystals. Techniques such as 3-dimensional Fourier synthesis (see Fig. 7) can be used and the structure determination made reasonably reliable.

The A configuration

In this conformation, the molecule has 11 nucleotide pairs per helix turn; the helix pitch is 28 Å. The relative positions and orientations of the base, and of the deoxyribose and phosphate parts of the nucleotides differ considerably from those in the B form; in particular the base-pairs are tilted 20° from perpendicular to the helix axis (Fig. 8).

The A form of DNA was the first crystalline form to be observed (Fig. 1). Although it has not been observed *in vivo*, it is of special interest because helical RNA adopts a very similar configuration. A full account of A DNA will shortly be available. A good photograph of the A pattern is shown in Fig. 9.

The C configuration

This form may be regarded as an artefact formed by partial drying. The helix is non-integral, with about 9 1/3 nucleotide pairs per turn. The helices pack together to form a semicrystalline structure: there is no special relation between the position of one nucleotide in a molecule and that in another. The conformation of an individual nucleotide is very similar to that in the B form. The differences between the B and C diffraction patterns are accounted for by the different position of the nucleotides in the helix. Comparison of the forms provides further confirmation of the correctness of the structures. In a way, the problem is like trying to deduce the structure of a folding chair by observing its shadow: if the conformation of the chair is altered slightly, its structure becomes more evident.

The helical structure of RNA molecules

In contrast to DNA, RNA gave poor diffraction patterns, in spite of much effort by various workers including ourselves. There were many indications that RNA contained helical regions, e. g. optical properties of RNA

solutions strongly suggested (e. g. Doty 1961) that parts of RNA molecules resembled DNA in that the bases were stacked on each other and the structure was helical; and X-ray studies of synthetic polyribonucleotides suggested that RNA resembled DNA (Rich 1959). The diffraction patterns of RNA (Rich and Watson 1954) bore a general resemblance to those of DNA, but the nature of the pattern could not be clearly distinguished because of disorientation and diffuseness. An important difficulty was that there appeared to be strong meridional reflections at 3.3 Å and 4 Å. It was not possible to interpret these in terms of one helical structure.

In early work, many RNA preparations were very heterogeneous. We thought that the much more homogeneous plant virus RNA might give better patterns, but this was not so. However, when preparations of ribosomal RNA and 'soluble' RNA became available, we felt the prospects of structure analysis were improved. We decided to concentrate on 'soluble' RNA largely because Geoffrey Brown in our laboratory was preparing large quantities of a highly-purified transfer RNA component of soluble RNA for his physical and chemical studies, and because he was fractionating it into various transfer RNA's specific for incorporation of particular amino acids into proteins. This RNA was attractive for other reasons: the molecule was unusually small for a nucleic acid, there were indications that it might have a regular structure, its biochemical role was important, and in many ways its functioning was understood.

We found it very difficult to orient transfer RNA in fibres. However, by carefully stretching RNA gels in a dry atmosphere under a dissecting microscope, I found that fibres with birefringence as high as that of DNA could be made. But these fibres gave patterns no better than those obtained with other types of RNA, and the molecules disoriented when the water content of the fibres was raised. Watson Fuller, Michael Spencer, and myself worked for many months trying to make better specimens for X-ray study. We made little progress until Spencer found a specimen that gave some faint but sharp diffraction rings in addition to the usual diffuse RNA pattern. This specimen consisted of RNA gel that had been sealed for X-ray study in a small cell, and he found that it had dried slowly owing to a leak. The diffraction rings were so sharp that we were almost certain that they were spurious diffraction due to crystalline impurity — this being common in X-ray studies of biochemical preparations. A specimen of RNA had given very similar rings due to DNA impurity.

We were therefore not very hopeful about the rings. However, after several weeks Spencer eliminated all other possibilities: it seemed clear that the rings were due to RNA itself. By controlled slow drying, he produced stronger rings; and, with the refined devices we had developed for stretching RNA and with gels slowly concentrated by Brown, Fuller oriented the RNA without destroying its crystallinity. These fibres gave clearly defined diffraction patterns, and the orientation did not disappear when the fibres were hydrated. It appeared that the methods I had been using earlier, of stretching the fibres as much as possible, destroyed the crystallinity. If instead, the material was first allowed to crystallise slowly, stretching oriented the microcrystals and the RNA molecules in them. Single molecules were too small to be oriented well unless aggregated by crystallisation. It was rather unexpected that, of all the different types of RNA we had tried, transfer RNA which had the lowest molecular weight, oriented best.

The diffraction patterns of transfer RNA were clearly defined and well oriented (Spencer, Fuller, Wilkins and Brown, 1962). These improvements revealed a striking resemblance between the patterns of RNA and A DNA (Fig. 10). The difficulty of the two reflections at 3.3 Å and 4 Å was resolved (Fig. 11): in the RNA pattern the positions of reflections on three layer-lines differed from those in DNA; as a result, when the patterns were poorly-oriented, the three reflections overlapped and gave the impression of two. There was no doubt that the RNA had a regular helical structure almost identical with that of A DNA. The differences between the RNA and DNA patterns could be accounted for in terms of small differences between the two structures.

An important consequence of the close resemblance of the RNA structure to that of DNA is that the RNA must contain base sequences that are largely or entirely complementary. The number of nucleotides in the molecule is about 80. The simplest structure compatible with the X-ray results consists of a single polynucleotide chain folded back on itself, one half of the chain being joined to the other by base-pairing. This structure is shown in Fig. 12. While we are certain the helical structure is correct, it must be emphasised that we do not know whether the two ends of the chain are at the end of the molecule. The chain might be folded at both ends of the molecule with the ends of the chain somewhere along the helix. It is known that the amino acid attaches to the end of the chain terminated by the base sequence cytosine-cytosine-adenine.
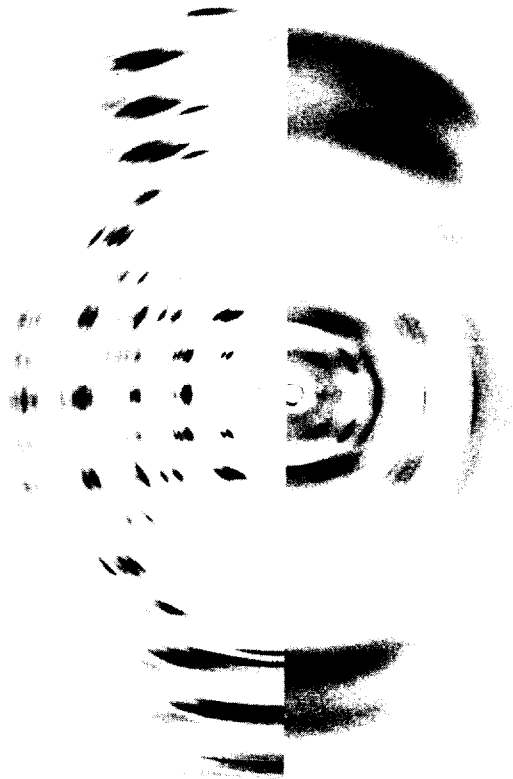
148

Fig. 10. Comparison of the X-ray diffraction patterns of fibres of DNA in the *A* configuration (left) and transfer RNA (right). The general distribution of intensity is very similar in both patterns, but the positions of the sharp crystalline reflections differ because the molecular packing in the crystals is different in the two cases. (Photograph with W. Fuller and M. Spencer: RNA by G. L. Brown.)

Relation of the molecular structure of RNA to function

Molecular model-building shows that the number of nucleotides forming the fold at the end of a transfer RNA molecule must be three or more. In our model, the fold consists of three nucleotides, each with an unpaired base. It might be that this base-triplet is the part of the molecule that attaches to the requisite part of the coding RNA polynucleotide chain that determines the sequence of amino acids in the polypeptide chain of a protein. It is believed that a base-triplet in the coding RNA corresponds to each amino acid. The triplet in the transfer RNA could attach itself specifically to the coding triplet by hydrogen-bonding and formation of base-pairs. It must be emphasised, however, that these ideas are speculative.

We suppose that part of the transfer RNA molecule interacts specifically with the enzyme that is involved in attaching the amino acid to the RNA; but we do not know how this takes place. Similarly, we know little of
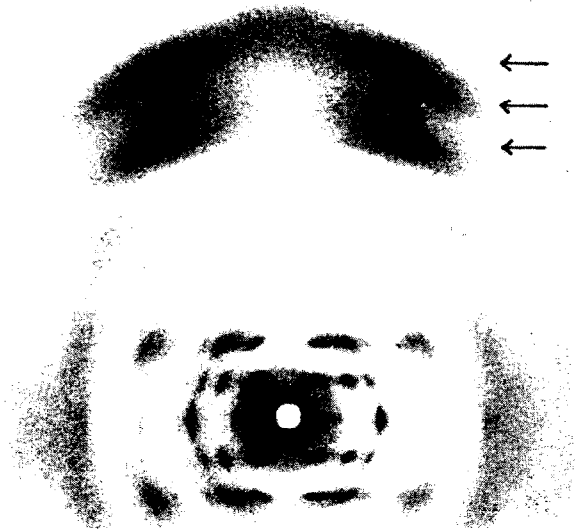
Fig. 11. Diffraction pattern of transfer RNA showing resolution of diffraction, in the regions of 3.3 Å and 4 Å. into three layer-lines indicated by the arrows and corresponding to the *A* DNA pattern. (Photograph with W. Fuller and M. Spencer; RNA by G. L. Brown.)

the way in which the enzyme involved inDNAreplication interacts withDNA, or of other aspects of the mechanics of DNA replication. The presence of complementary base sequences in the transfer RNA molecule, suggests that it might be self-replicating like DNA; but there is at present little evidence to support this idea. The diffraction patterns of virus and ribosome RNA show that these molecules also contain helical regions; the functions of these are uncertain too.

In the case of DNA, the discovery of its molecular structure led immediately to the replication hypothesis. This was due to the simplicity of the structure of DNA. It seems that molecular structure and function are in most cases less directly related. Derivation of the helical configuration of RNA molecules is a step towards interpreting RNA function; but more complete structural information, e. g. determination of base sequences, and more knowledge about how the various kinds of RNA interact in the ribosome, will probably be required before an adequate picture of RNA function emerges.

The possibility of determining the base sequence of transfer RNA by X-ray diffraction analysis

Since the biological specificity of nucleic acids appears to be entirely determined by their base sequences in them, determination of these sequences
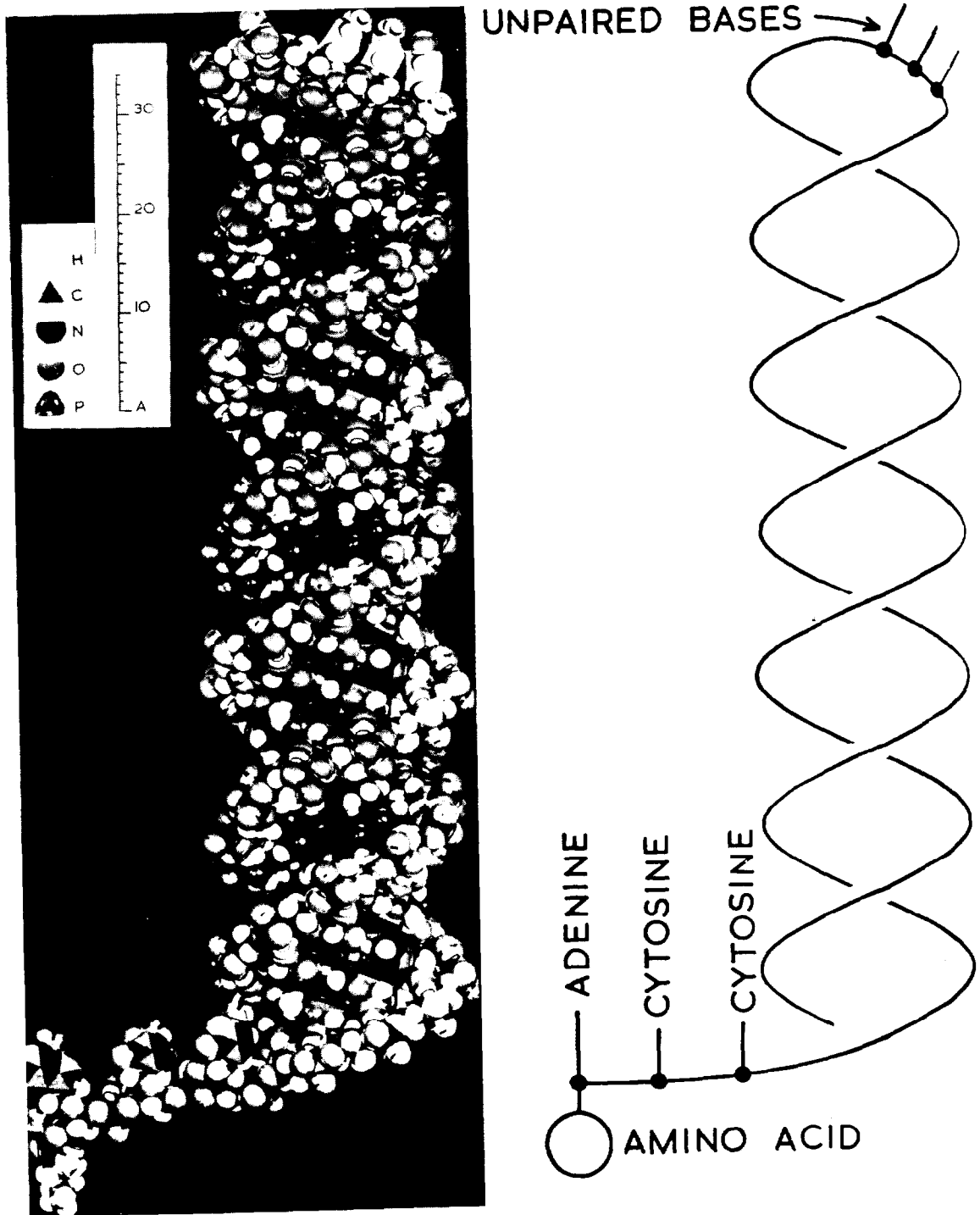
Fig. 12. Molecular model and diagram of a transfer RNA molecule.

is probably the most fundamental problem in nucleic acid research today. The number of bases in a DNA molecule is too large for determination of base sequence by X-ray diffraction to be feasible. However, in transfer RNA the number of bases is not too large. The possibility of complete structure analysis of transfer RNA by means of X-rays is indicated by two observations. First, we have observed (Fig. 13), in X-ray patterns of transfer RNA, separate spots, each corresponding to a single crystal of RNA. We estimated their size to be about 10 μ and have confirmed this estimate by observing, in the polarising microscope, birefringent regions that probably are the crystals. It should not be too difficult to grow crystals several times larger, which is large enough for single-crystal X-ray analysis.

The second encouraging observation is that the X-ray data from DNA have restricted resolution almost entirely on account of disorientation of the microcrystals in DNA fibres. The DNA intensity data indicate that the temperature factor (B $=$ 4 Å) is the same for DNA as for simple compounds. It thus appears that DNA crystals have fairly perfect crystallinity and that, if single crystals of DNA could be obtained, the intensity data would be adequate for precise determination of all atomic positions in DNA (apart from the non-periodic base sequence).

We are investigating the possibility of obtaining single crystals of DNA, but the more exciting problem is to obtain single crystals of transfer RNA with crystalline perfection equal to that of DNA, and thereby analyse base sequence. At present, the RNA crystals are much less perfect than those of DNA. However, most of our experiments have been made with RNA that is a mixture of RNA's specific for different amino acids. We have seldom used RNA that is very largely specific for one amino acid only. We hope that good preparations of such RNA may be obtained consisting of one type of molecule only. We might expect such RNA to form crystals as perfect as those of DNA. If so, there should be no obstacle to the direct analysis of the whole structure of the molecule, including the sequence of the bases and the fold at the end of the helix. We may be over-optimistic, but the recent and somewhat unexpected successes of X-ray diffraction analysis in the nucleic acid and protein fields, are cause for optimism.
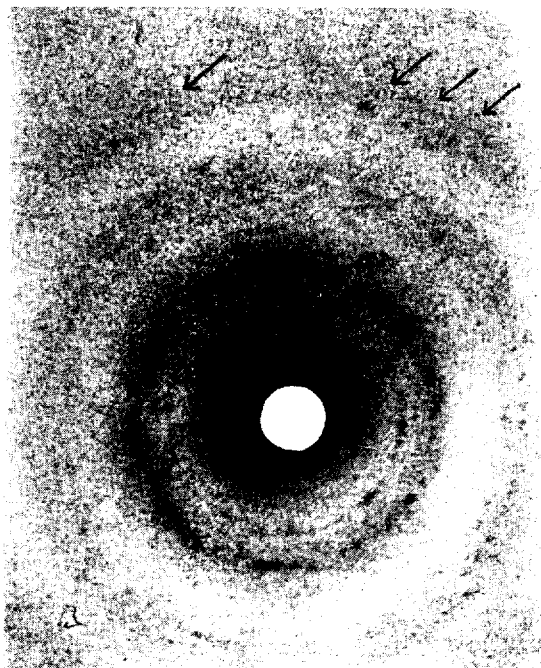
152

Fig. 13. Diffraction pattern of unoriented transfer RNA, showing diffraction rings with spots corresponding to reflections from single crystals of RNA. The arrows point to reflections from planes ∼ 6 Å apart.

Acknowledgements

During the past twelve years, while studying molecular structure of nucleic acids, I have had so much help from so many people that all could not be acknowledged properly here. I must, however, thank the following:

Sir John Randall, for his long-standing help and encouragement, and for his vision and energy in creating and directing a unique laboratory;

all my co-workers at various times over the past twelve years; first, Raymond Gosling, Alex Stokes, Bill Seeds and Herbert Wilson, then Bob Langridge, Clive Hooper, Max Feughelman, Don Marvin and Geoffrey Zubay, and at present, Michael Spencer, Watson Fuller and Struther Arnott, who with much ability, skill and persistence (often through the night) carried out the X-ray, molecular model-building, and computing studies; my late colleague Rosalind Franklin who, with great ability and experience of X-ray diffraction, so much helped the initial investigations on DNA;

Leonard Hamilton for his constant encouragement and friendly cooperation, and for supplying us with high-quality DNA isolated in many forms and from many sources; Geoffrey Brown for giving me moral and

153

REFERENCES

Astbury, W. T. (1947). Symp. Soc. Exp. Biol. I Nucleic Acid p. 66. Cambridge: University Press.

Avery, O. T., MacLeod, C. M., and McCarty, M. (1944). J. Exp. Med. *79*, 137.

Brachet, J. (1941). Arch. Biol. Liège *53*, 207.

Caspersson, T. (1941). Naturwissenschaften *29*, 33.

Chargaff, E. (1950). Experientia *6*, 201.

Doty, P. (1961). Biochem. Soc. Symp. No. 21, 8.

Gulland, J. M. (1947). Cold. Spr. Harb. Symp. Quant. Biol. XII, p. 95.

Gulland, J. M. & Jordan, D. O. (1947). Symp. Soc. Exp. Biol. I Nucleic Acid, Cambridge: University Press.

Hamilton, L. D., Barclay, R. K., Wilkins, M. H. F., Brown, G. L., Wilson, H. R., Marvin, D. A., Ephrussi-Taylor, H. and Simmons, N. S. (1959). J. Biophys. Biochem. Cytol. *5*, 397.

Hershey, A. D. & Chase, M. (1952). J. Gen. Physiol. *36*, 39.

Hoogsteen. K. (1959). Act. Cryst. *12*, 822.

Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. and Hamilton, L. D. (1960). J. Mol. Biol. *2*, 19.

Pattri, H. O. E. (1932). Z. f. Zellforschung u. mikrosk. Anat. *16*, 723.

Rich, A. (1959) in A Symposium on Molecular Biology, edit. by Zirkle. 47. (University Chicago Press).

Rich, A. and Watson, J. D. (1954). Nature *173*, 995.

Rinne, F. (1933). Trans. Farad. Soc. *29*, 1016.

Schmidt, W. J. (1937). Die Doppelbrechung von Karyoplasma Zytoplasma und Metaplasma (Berlin).

Signer, R., Caspersson, T. and Hammarsten, E. (1938). Nature *141*, 122.

Signer, R. and Schwander, H. (1949). Helv. chim. Acta. *32*, 853.

Spencer, M., Fuller, W., Wilkins, M. H. F. and Brown, G. L. (1962). Nature *194*, 1014.

Watson, J. D. and Crick, F. H. C. (1953 a). Nature *171*, 737.

Watson, J. D. and Crick, F. H. C. (1953 b). Nature *171*, 964.

Wilkins, M. H. F. (1961). J. Chim. Phys. 891.

Wilkins, M. H. F. and Randall, J. T. (1953). Biochim. Biophys. Acta *10*, 192.