The Discovery of the Double Helix

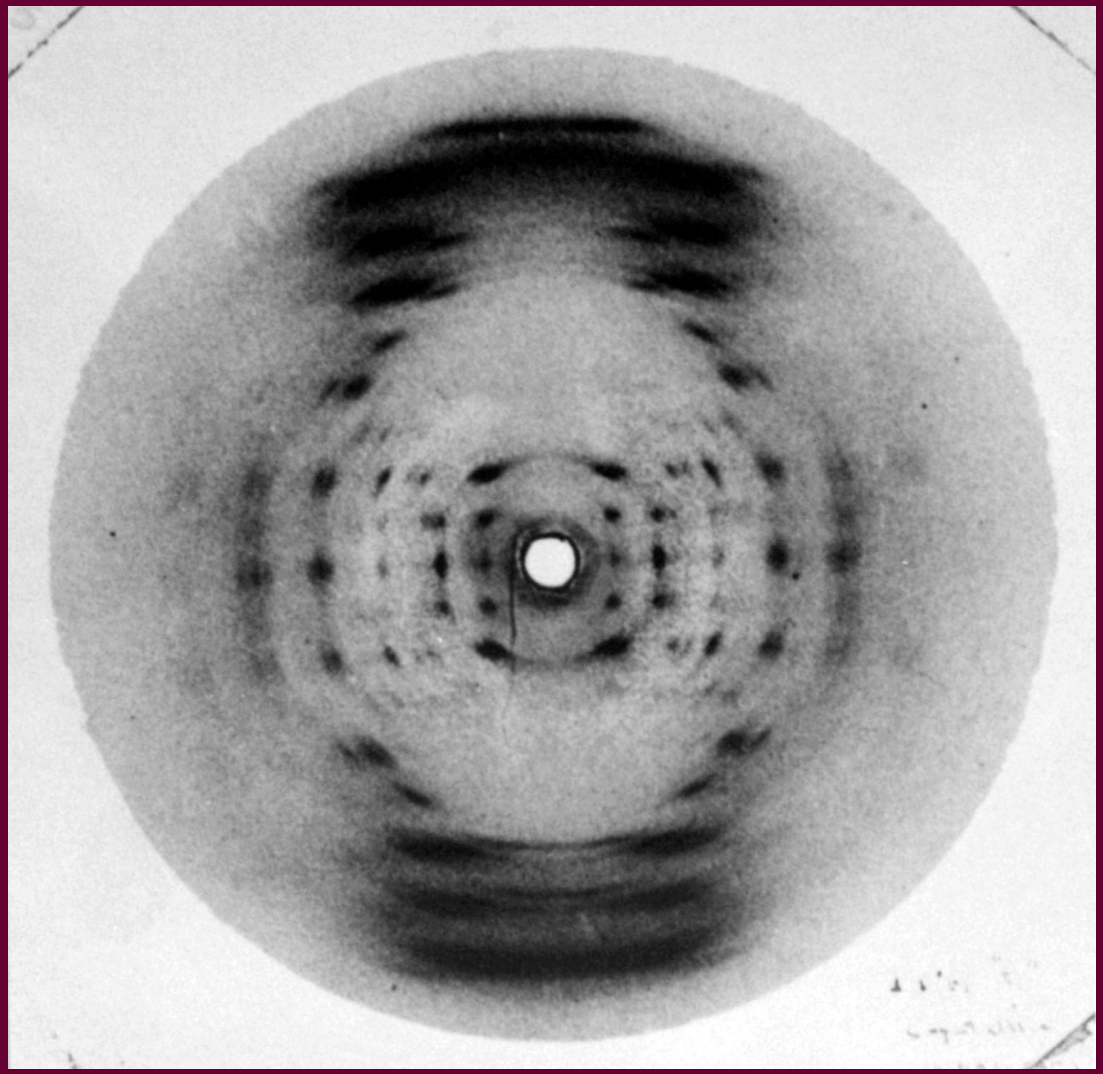
James D. Watson

Chancellor Emeritus
Cold Spring Harbor Laboratory

CERN

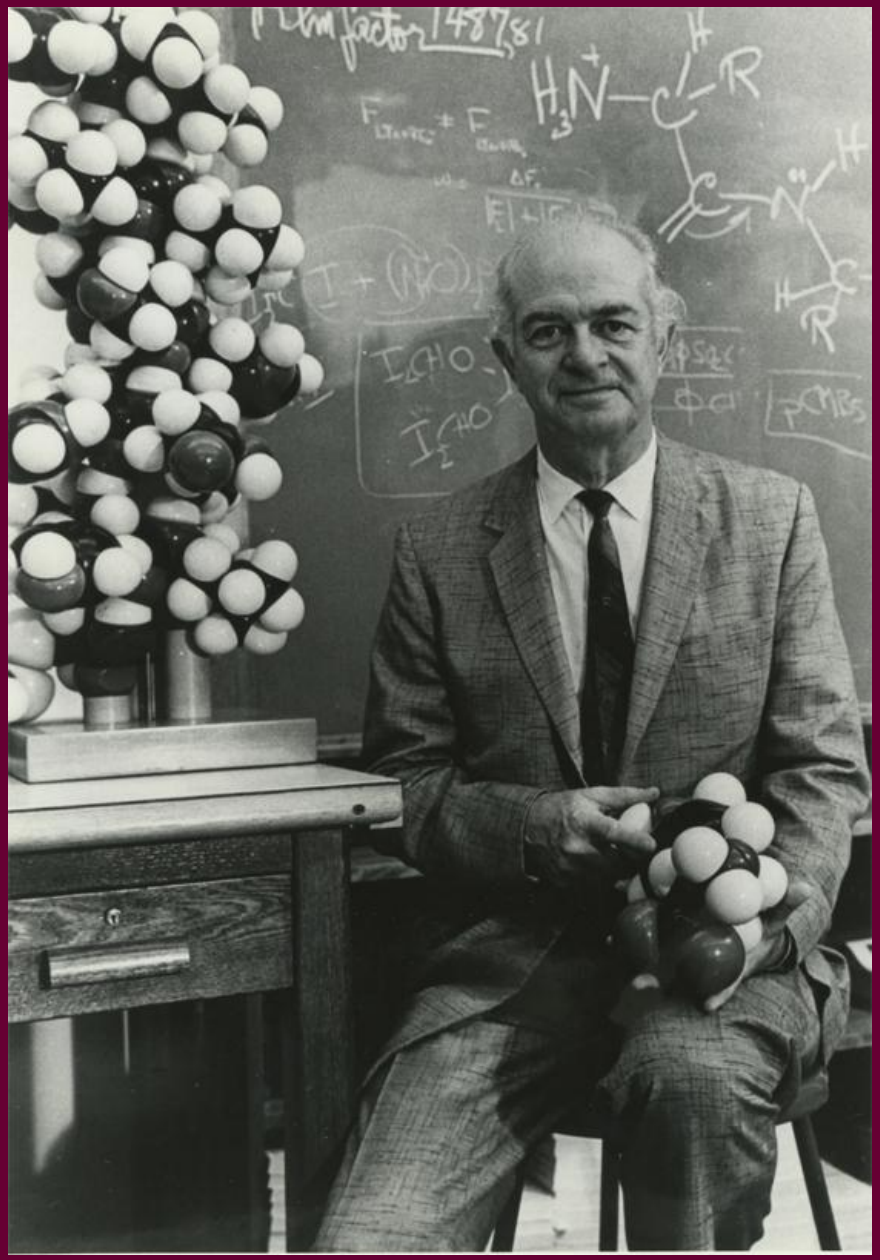
Geneva, Switzerland

6 September 2011

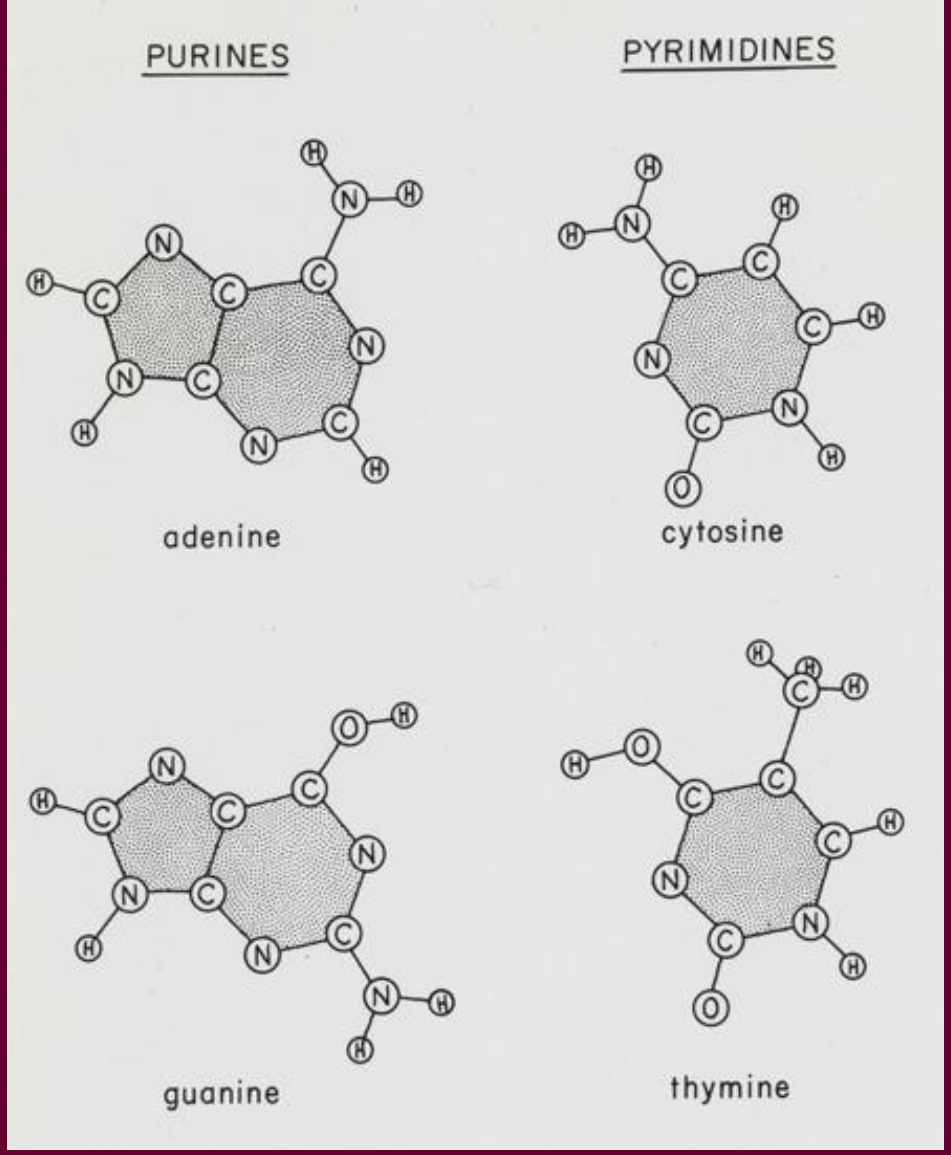


An X-ray photograph of crystalline DNA in the A form.

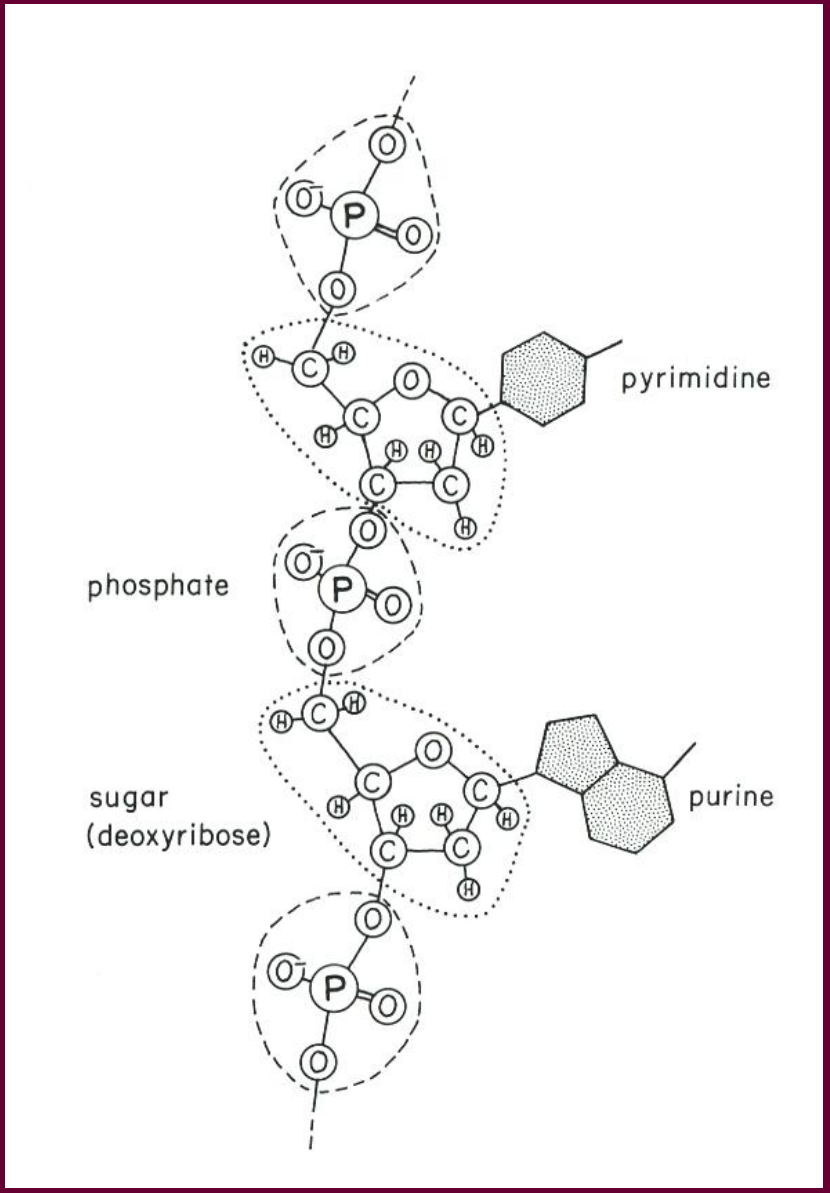
Francis next to a Cavendish X-ray tube



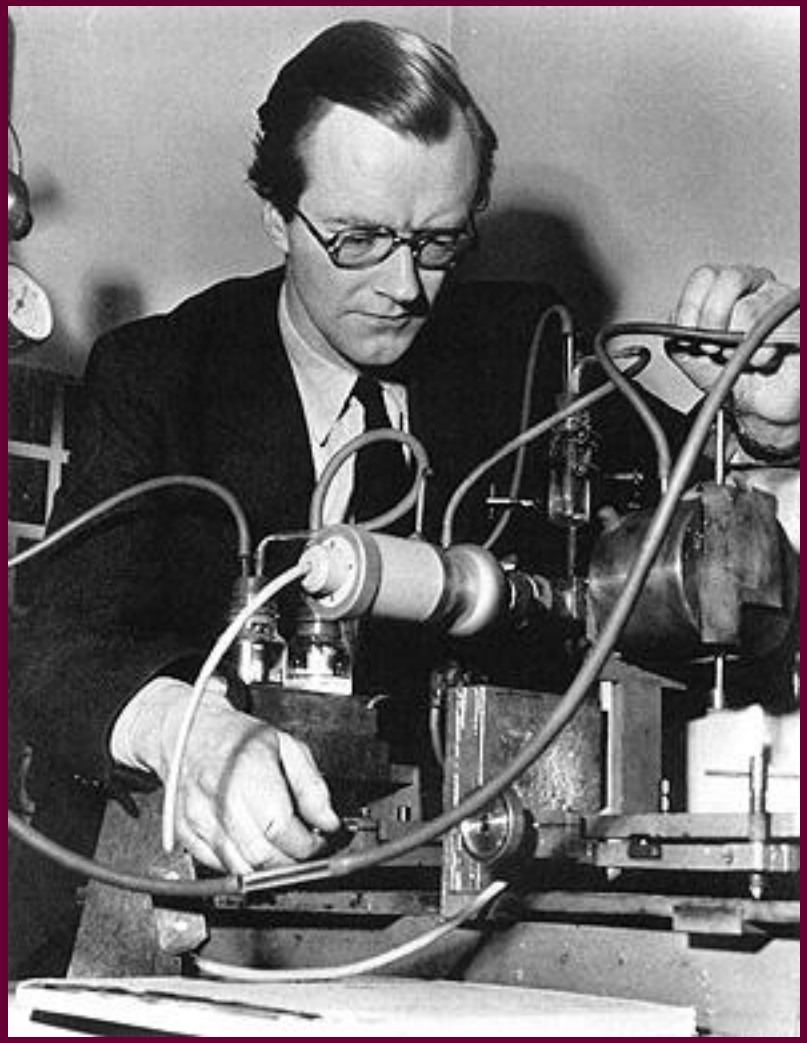
Linus Pauling with his atomic models.



The chemical structures of the four DNA bases as they were often drawn about 1951. Because the electrons in the five- and six-membered rings are not localized, each base pair has a planar shape with a thickness of 3.4 A.



A more detailed view of the covalent bonds of the sugar-phosphate backbone.



Maurice Wilkins with X-ray crystallographic equipment about 1954. Courtesy King's College, London, and Horace Freeland Judson.

Rosalind Franklin, while in Paris, serving afternoon coffee in evaporating dishes.

Acta Cryst. (1952). **5**, 581

The Structure of Synthetic Polypeptides. I. The Transform of Atoms on a Helix

By W. Cochran

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

F. H. C. Crick

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

and V. Vand*

Chemistry Department, The University, Glasgow W. 2, Scotland

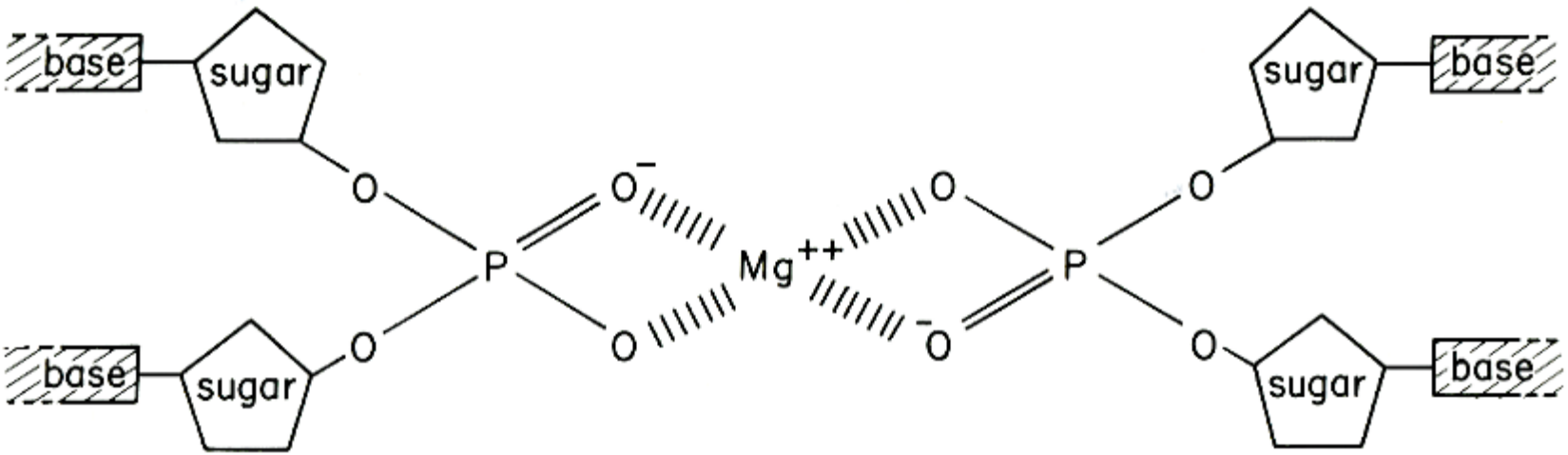
(*Received* 16 *February* 1952)

The formulæ are given for the Fourier transforms of a number of helical structures; namely, a thin helical wire, a set of identical atoms spaced at regular intervals on a helix, and the general case of a group of atoms repeated by the operation of a non-integer screw. General predictions are made concerning the intensities of the X-ray diffraction pattern of the synthetic polypeptide poly-γ-methyl-L-glutamate, assuming that its structure is based on the α-helix suggested by Pauling & Corey.

1. Introduction

The following calculations were undertaken because of current interest in the structures of certain synthetic polypeptides. The preliminary X-ray data for these polypeptides have been described by Bamford, Hanby & Happey (1951) and their infra-red behaviour by Ambrose & Elliott (1951). Pauling & Corey (1951) have interpreted the structures of the two polypeptides which have so far given the best X-ray diffraction pictures, namely poly-γ-methyl-L-glutamate and poly-γ-benzyl-L-glutamate, in terms of the α-helix described by Pauling, Corey & Branson (1951). In this structure the residues repeat along the helix with a spacing of about 1·5 Å in the chain direction, and Perutz (1951) has found that a strong meridional reflexion of spacing 1·5 Å is given by poly-γ-benzyl-

* Imperial Chemical Industries Fellow.



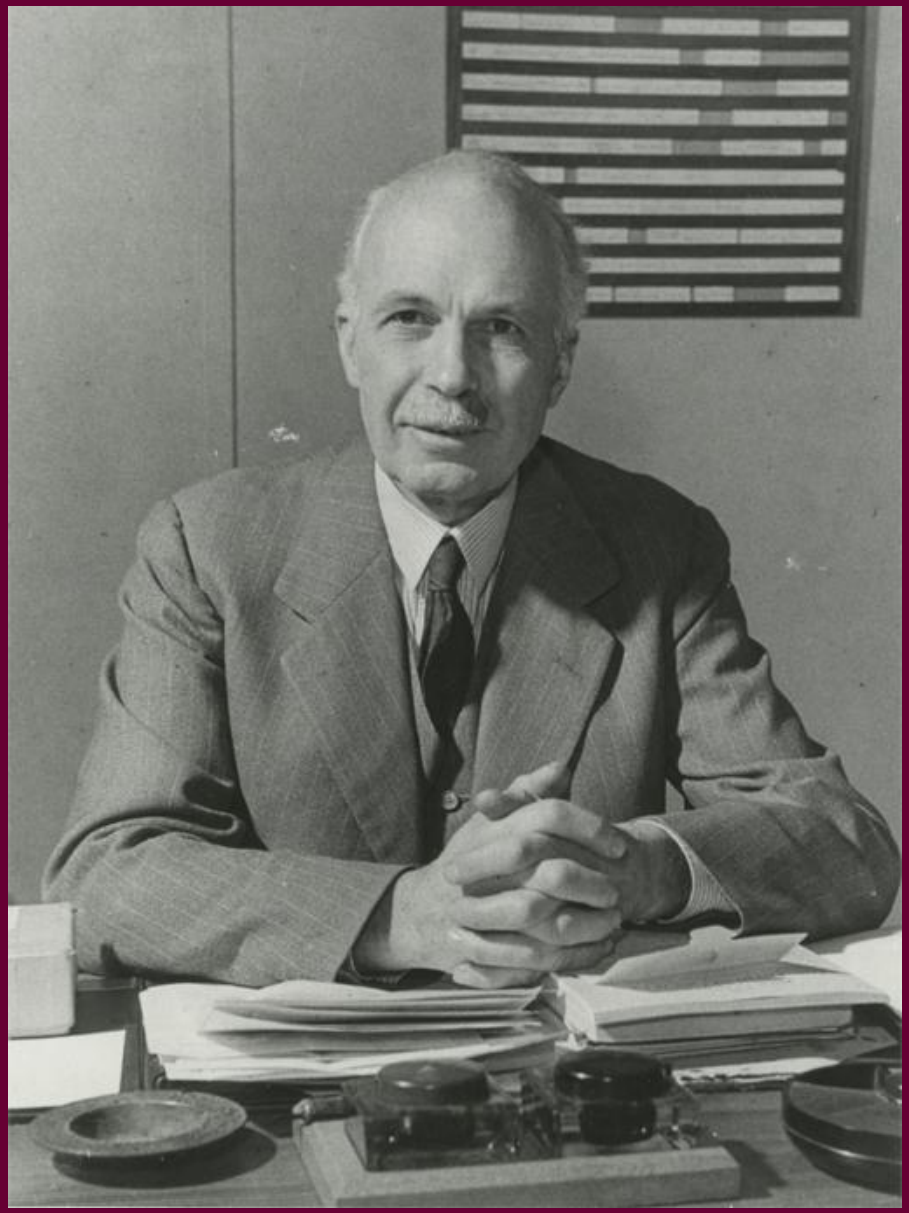
How Mg++ ions might be used to bind negatively charged phosphate groups in the center of a compound helix.

Lost Crick Correspondence

Dear Francis, This is just to say how bloody browned off I am entirely & how rotten I feel about it all & how entirely friendly I am (though it may possibly appear differently). We are really between forces which may grind all of us into little pieces … I had to restrain Randall from writing to Bragg complaining about your behaviour. Needless to say I did restrain him, but so far as your security with Bragg is concerned it is probably much more important to pipe down & build up the idea of a quiet steady worker who never creates 'situations' than to collect all the credit for your excellent ideas at the expense of good will.

And you see it does make me a bit confused about our discussions if you get too interested in everything which is important; where I say confused I mean confused, I am now largely incapable of any logical thinking in relation to polynucleotide chains or anything. And poor Jim — may I shed a crocodile & very confused tear?

- Maurice Wilkins
11 December 1951



Sir Lawrence Bragg sitting at his Cavendish desk.

Lost Crick Correspondence

Dear Maurice,
Just a brief note to thank you for the letters and to try to cheer you up. We think the best thing to get things straight is for us to send you a letter setting out in a mild manner our point of view. This will take a day or so to do, so we hope you'll excuse the delay. Please don't worry about it, because we've all agreed that we must come to an amicable arrangement… so cheer up and take it from us that even if we kicked you in the pants it was between friends. We hope our burglary will at least produce a united front in your group!

Yours ever
Francis
Jim

13 December 1951

In Paris on the way to the Riveria, spring 1952

THE STRUCTURE OF TOBACCO MOSAIC VIRUS

I. X-RAY EVIDENCE OF A HELICAL ARRANGEMENT OF SUB-UNITS AROUND THE LONGITUDINAL AXIS

by

J. D. WATSON*

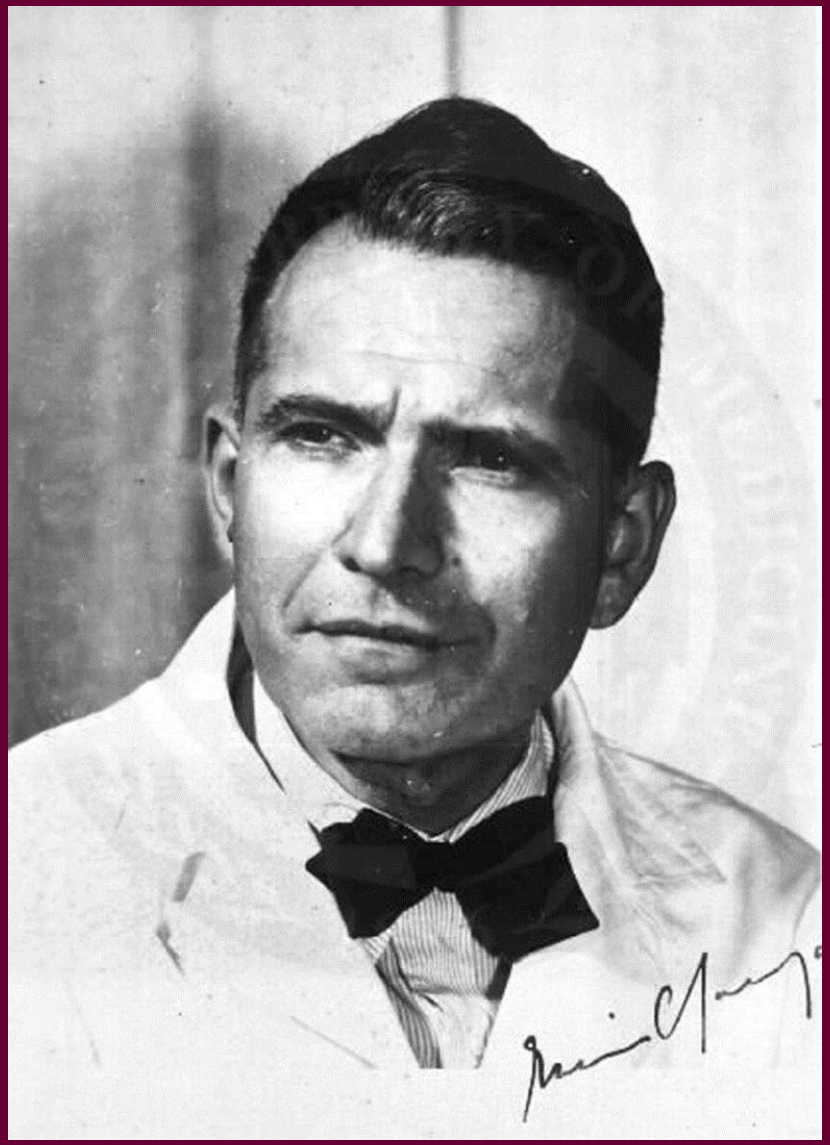
The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge (England)

INTRODUCTION

Tobacco Mosaic Virus (TMV) is a well known structure. It has a molecular weight of about $4 \cdot 10^7$ and is rod shaped with a diameter of 150 A and a length of 2800 A. Like other plant viruses, it contains both ribonucleic acid (6%) and protein (94%). In liquid solution, the virus may aggregate lengthwise and form regions of parallel orientation in which the particles are arranged in hexagonal close packing at equal distances from one other. No regularity of arrangement exists in the direction of the particle length; the oriented regions appear to be liquid crystalline. Upon drying, the oriented arrangement of liquid solutions is retained and it is possible to obtain oriented dry specimens.

X-rays were first used to investigate the structure of TMV by BERNAL AND FANKUCHEN[1]. They obtained highly oriented X-ray photographs containing a very large number of distinct reflexions. By varying the inter-particle distance, they were able to show that the X-ray pattern arose not only from the regular arrangement of the virus particles in a hexagonal lattice but in addition from the presence of a complex structure within each virus particle. The repeat distance along the fibre axis was found to be 68 A, *i.e.* much shorter than the length of the particle; they concluded that each virus particle contains a large number of equivalent sub-units. In this paper we shall deal with the internal structure of TMV and shall present newly-obtained X-ray evidence indicating that the internal regularity is based on a division of the virus into crystallographically equivalent sub-units helically arranged around the longitudinal axis of the particle.

The meeting at Royaumont, July 1952



Erwin Chargaff

Photo courtesy of the National Library of Medicine.

A PROPOSED STRUCTURE FOR THE NUCLEIC ACIDS

By Linus Pauling and Robert B. Corey

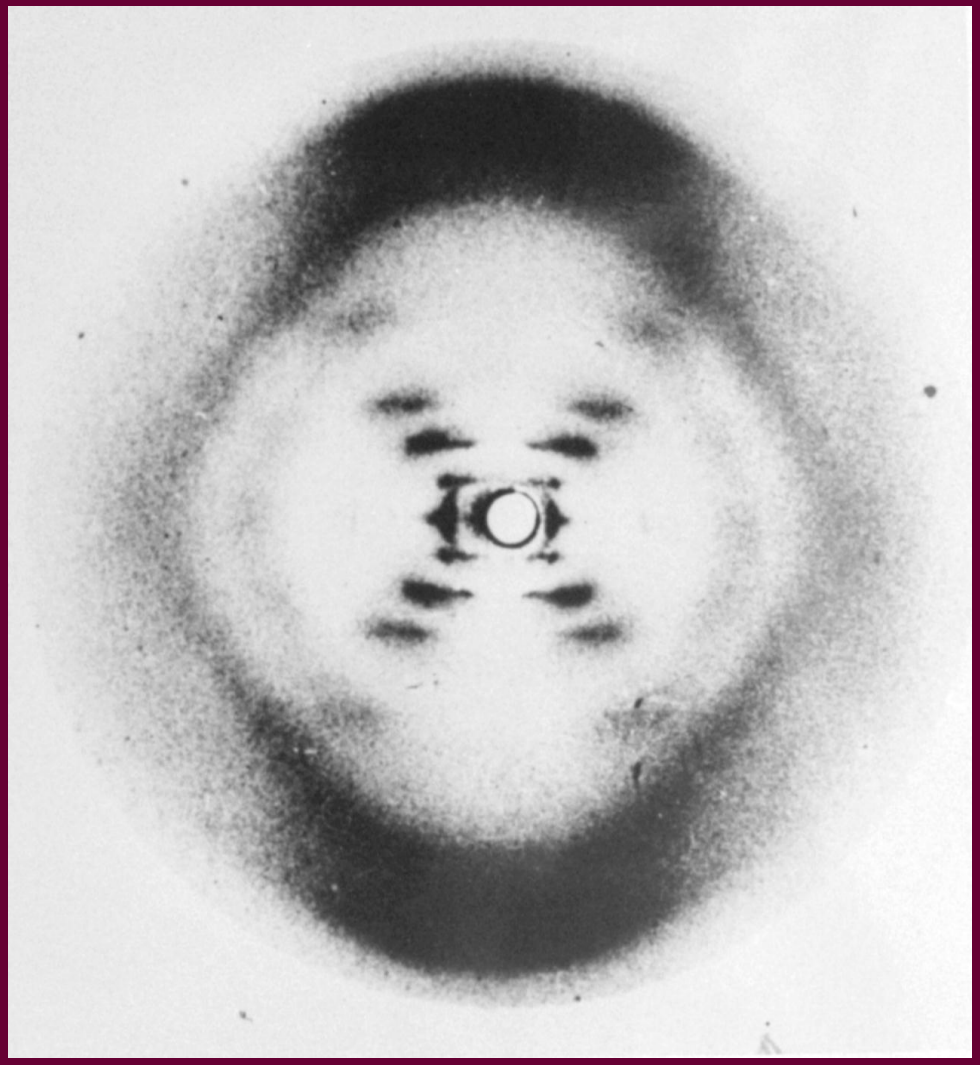
Gates and Crellin Laboratories of Chemistry,* California Institute of Technology

Communicated December 31, 1952

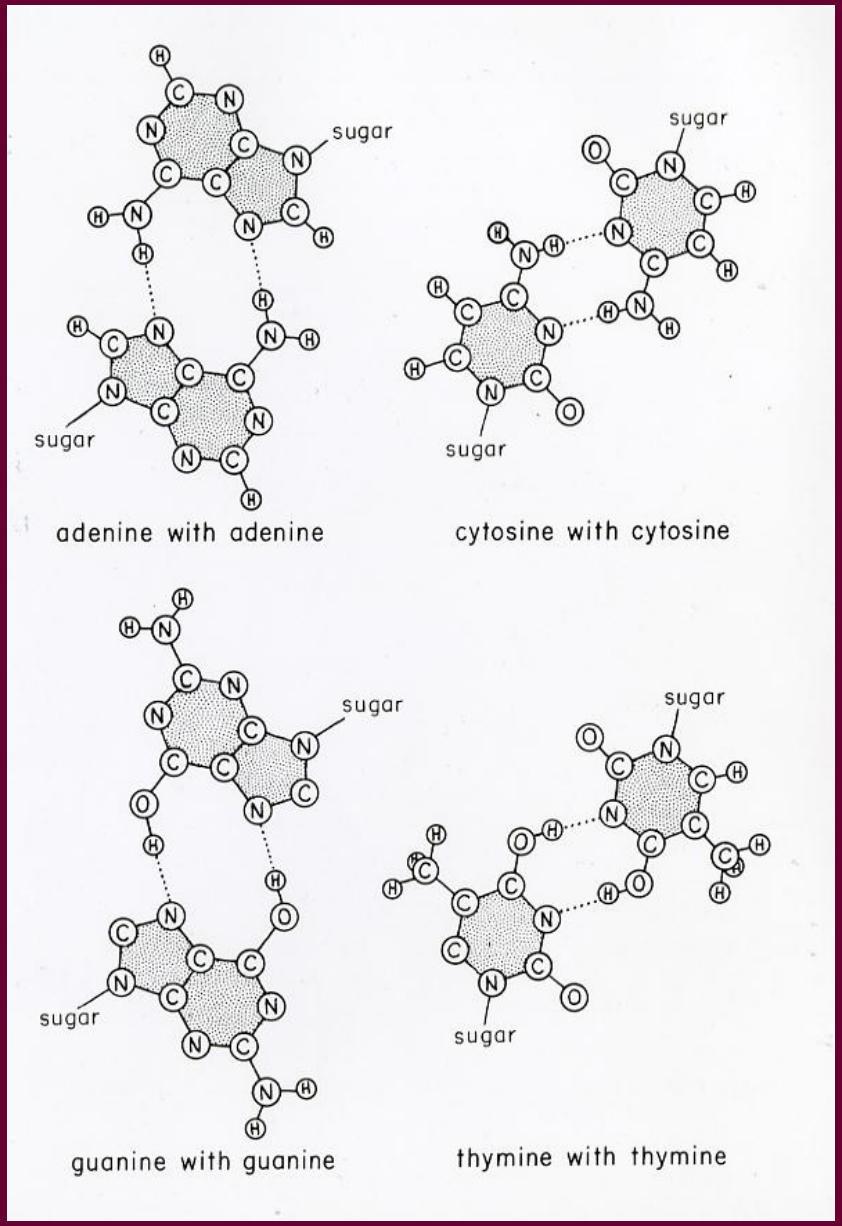
The nucleic acids, as constituents of living organisms, are comparable in importance to the proteins. There is evidence that they are involved in the processes of cell division and growth, that they participate in the transmission of hereditary characters, and that they are important constituents of viruses. An understanding of the molecular structure of the nucleic acids should be of value in the effort to understand the fundamental phenomena of life.

We have now formulated a promising structure for the nucleic acids, by making use of the general principles of molecular structure and the available information about the nucleic acids themselves. The structure is not a vague one, but is precisely predicted; atomic coordinates for the principal atoms are given in table 1. This is the first precisely described structure for the nucleic acids that has been suggested by any investigator. The structure accounts for some of the features of the x-ray photographs; but detailed intensity calculations have not yet been made, and the structure cannot be considered to have been proved to be correct.

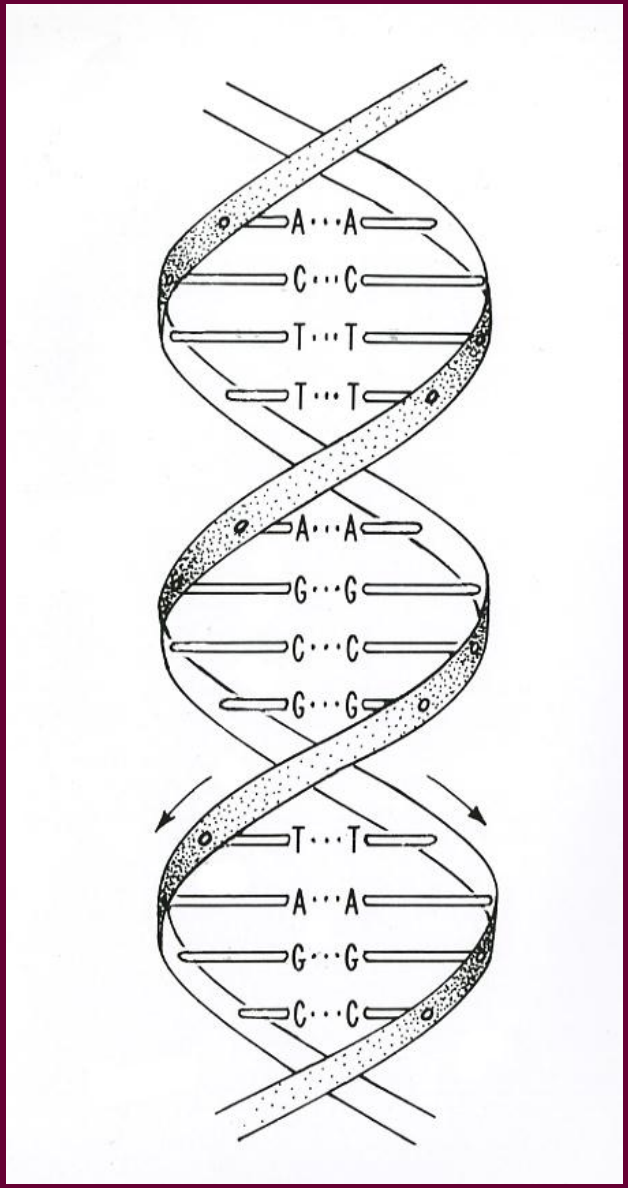
The Formulation of the Structure.—Only recently has reasonably complete information been gathered about the chemical nature of the nucleic acids. The nucleic acids are giant molecules, composed of complex units. Each unit consists of a phosphate ion, HPO_4^{--}, a sugar (ribose in the ribonucleic



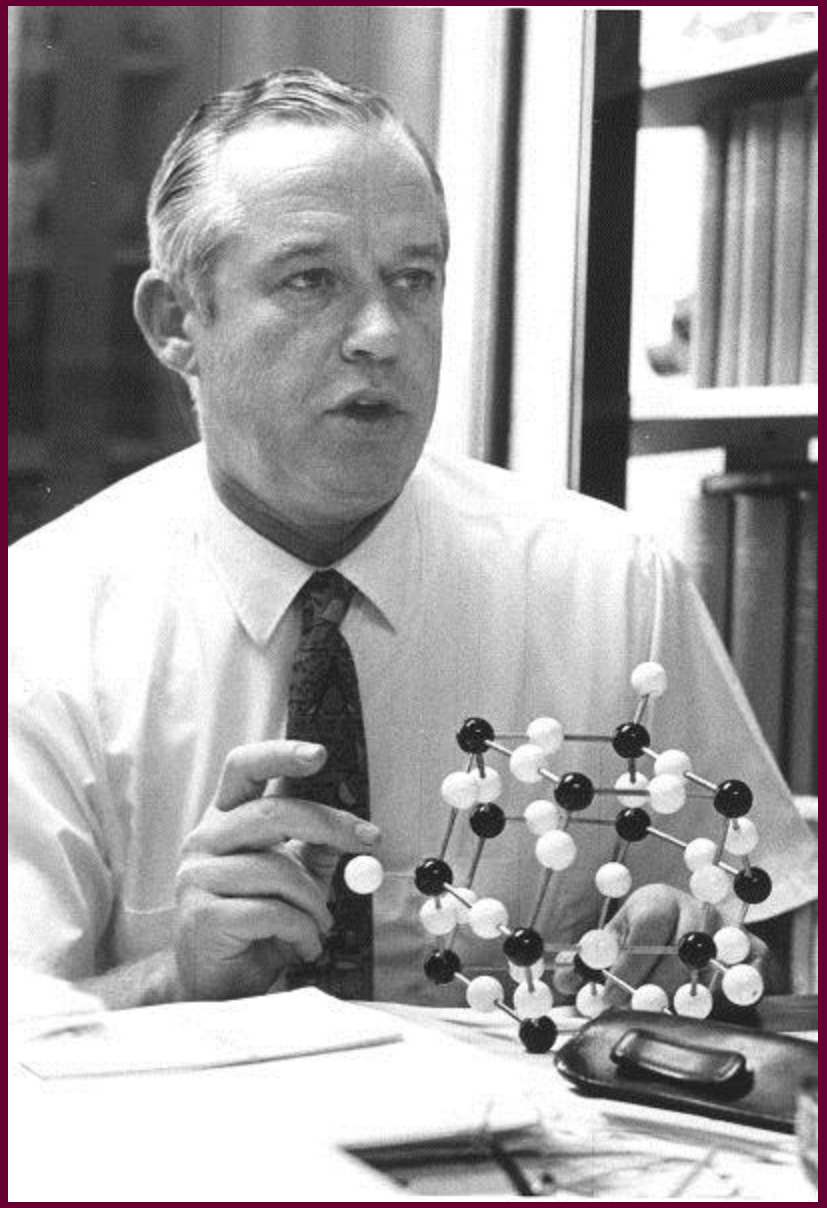
An X-ray photograph of DNA in the B form, taken by Rosalind Franklin late in 1952.



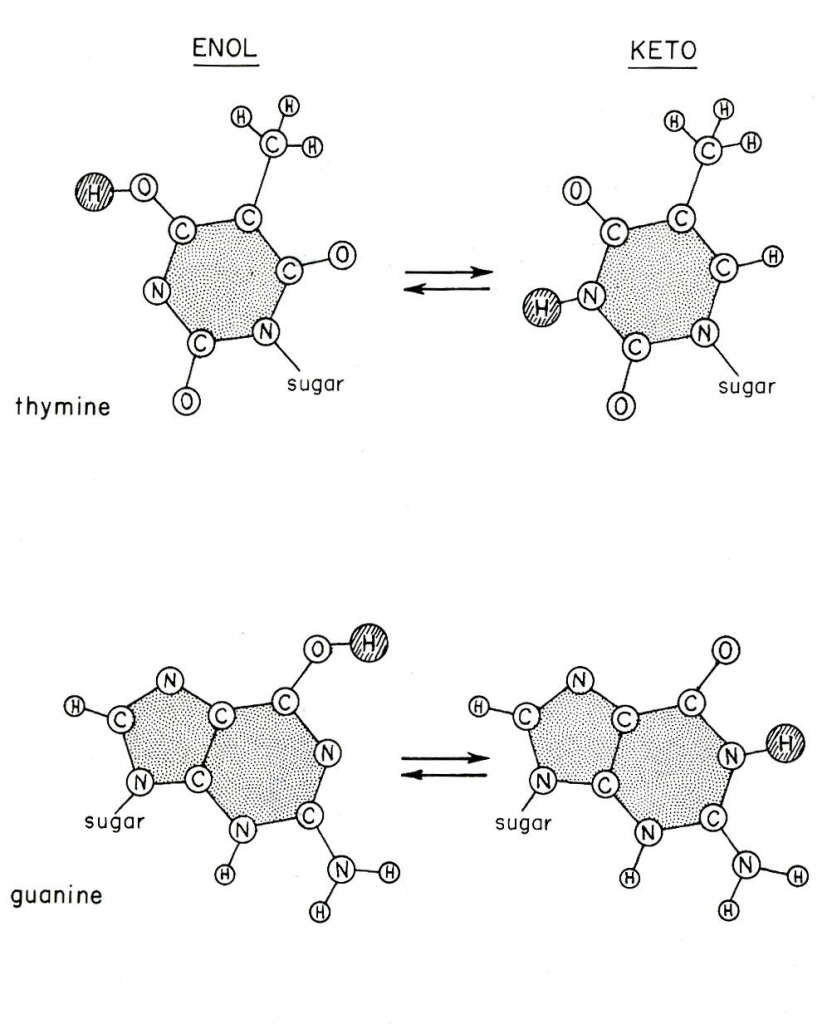
The four base pairs used to construct the like-with-like structure (hydrogen bonds are dotted)



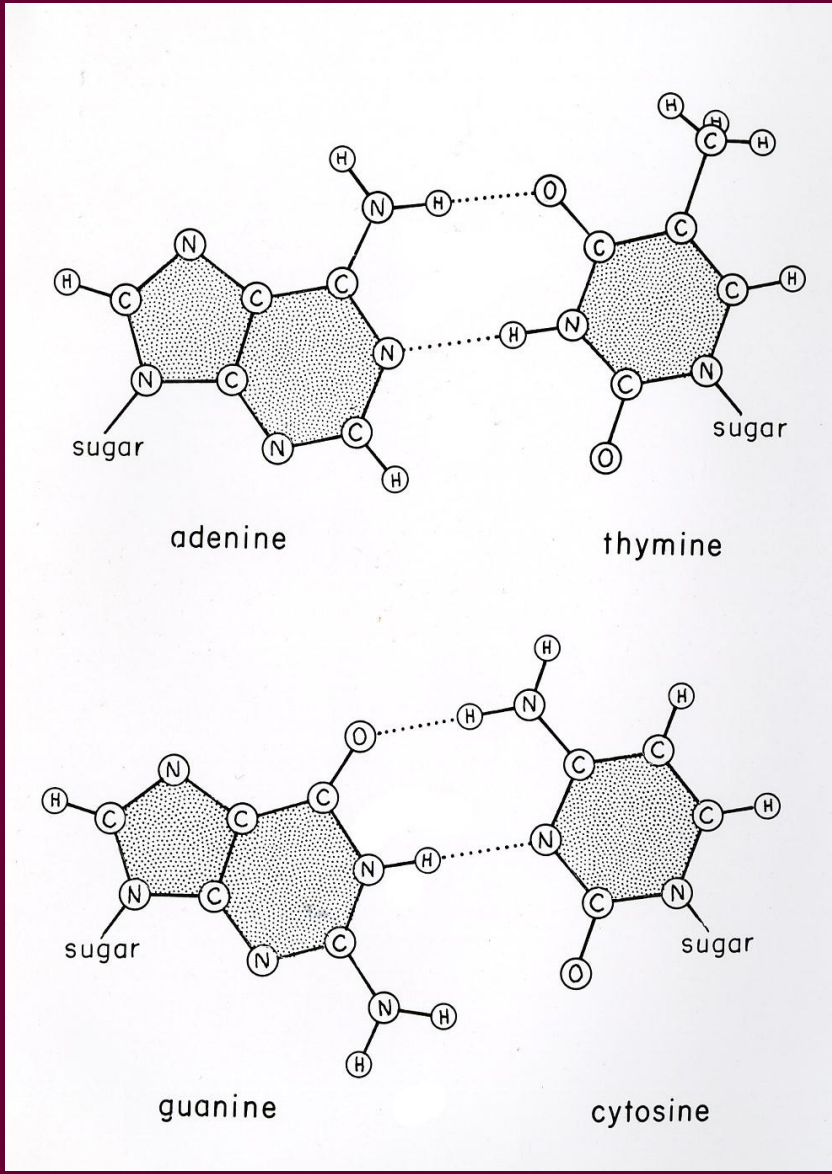
A schematic view of a DNA molecule built up from like-with-like base pairs.



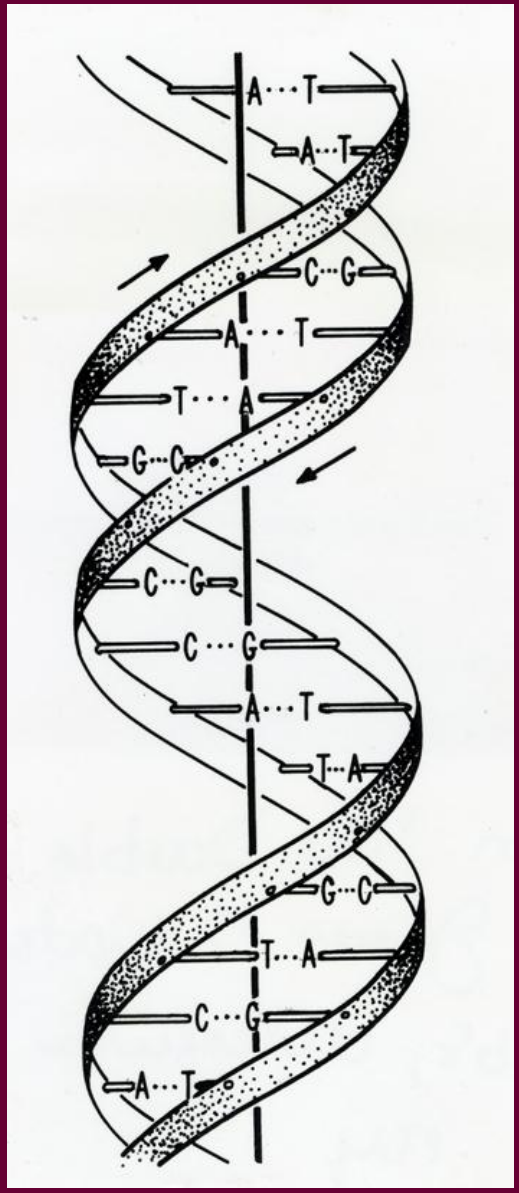
Jerry Donohue



The contrasting tautometric forms of guanine and thymine which might occur in DNA. The hydrogen atoms that can undergo the changes in position (a tautometric shift) are shaded.

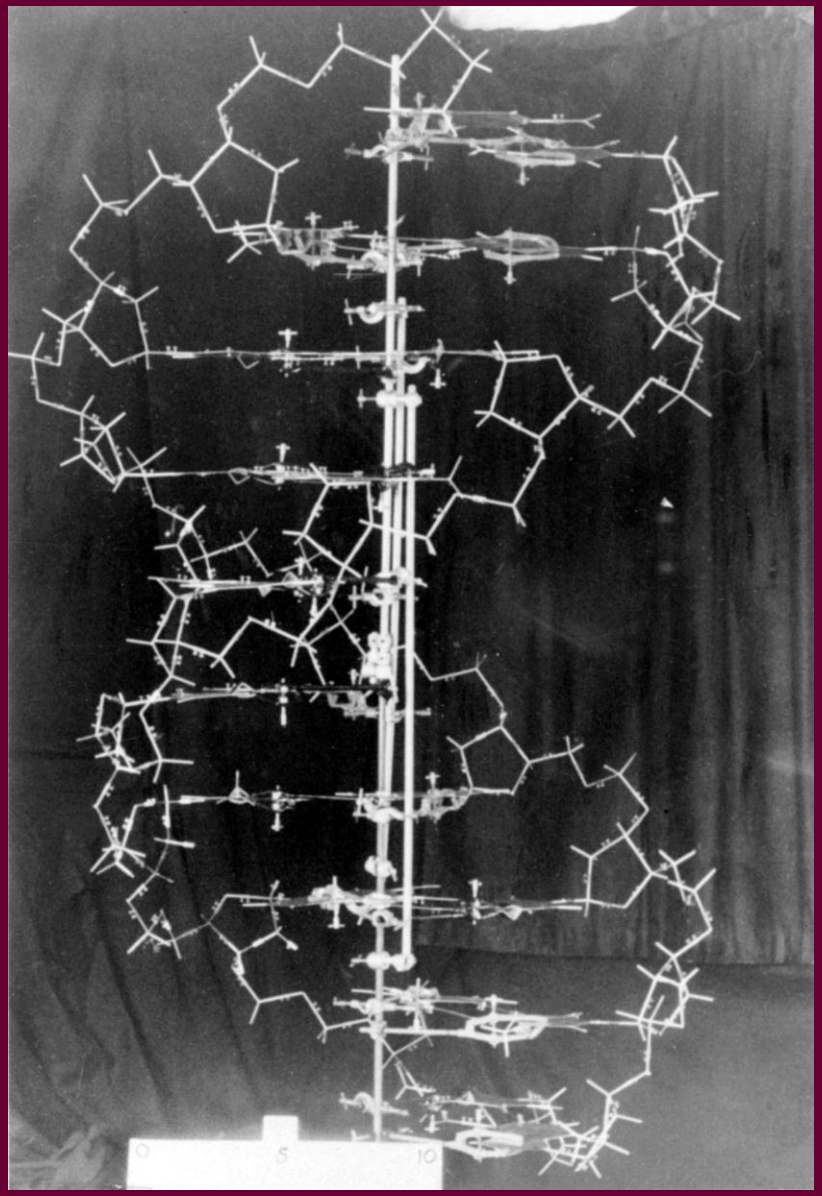


The adenine-thymine and guanine-cytosine base pairs used to construct the double helix (hydrogen bonds are dotted). The formation of a third hydrogen bond between guanine and cytosine was considered, but rejected because a crystallographic study of guanine hinted that it would be very weak. Now this conjecture is known to be wrong. Three strong hydrogen bonds can be drawn between guanine an cytosine.



A schematic illustration of the double helix.

The Eagle, Cambridge



The original demonstration model of the double helix

Lost Crick Correspondence

Dear Francis,
Herewith almost uncorrected draft. How should we refer to your note? Welcome suggestions & [illegible] acknowledgements.
It looks very much as though I will be too late tonight so maybe we had better cancel the supper idea which is a pity but anyway I got the bloody thing finished.
Have you a structure for collagen yet?

- Maurice Wilkins
March 1953

Lost Crick Correspondence

Dear Francis,
It looks as though the only thing is to send Rosy's & my letters as they are & hope the Editor doesn't spot the duplication. I am so browned off with the whole madhouse I don't really care much what happens.
If Rosy wants to see Pauling, what the hell can we do about it? If we suggested it would be nicer if she didn't that would only encourage her to do so. Why is every body so terribly interested in seeing Pauling ... Now Raymond wants to see Pauling too! To hell with it all.

- Maurice Wilkins
March 1953

Elizabeth “Betty” Watson, with Clare Bridge in the background.

(*Reprinted from Nature, Vol.* 171, *p.* 737, *April* 25, 1953)

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

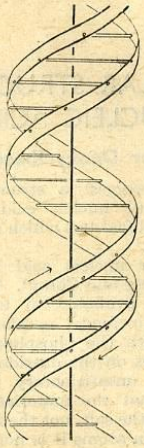
A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows : purine position 1 to pyrimidine position 1 ; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms

(that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are : adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine ; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick

Medical Research Council Unit for the
Study of the Molecular Structure of
Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

WHILE the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1-3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing

as distinct from that of crystalline nucleoprotein. This confirms current ideas of phage structure.

Transforming principle (in collaboration with H. Ephrussi-Taylor). Active deoxypentose nucleate allowed to dry at ~ 60 per cent humidity has the same crystalline structure as certain samples[3] of sodium thymonucleate.

We wish to thank Prof. J. T. Randall for encouragement; Profs. E. Chargaff, R. Signer, J. A. V. Butler and Drs. J. D. Watson, J. D. Smith, L. Hamilton, J. C. White and G. R. Wyatt for supplying material without which this work would have been impossible; also Drs. J. D. Watson and Mr. F. H. C. Crick for stimulation, and our colleagues R. E. Franklin, R. G. Gosling, G. L. Brown and W. E. Seeds for discussion. One of us (H. R. W.) wishes to acknowledge the award of a University of Wales Fellowship.

M. H. F. Wilkins
Medical Research Council Biophysics Research Unit,
A. R. Stokes
H. R. Wilson
Wheatstone Physics Laboratory,
King's College, London.
April 2.

[1] Astbury, W. T., Symp. Soc. Exp. Biol., 1, Nucleic Acid (Cambridge Univ. Press, 1947).
[2] Riley, D. P., and Oster, G., *Biochim. et Biophys. Acta*, **7**, 526 (1951).
[3] Wilkins, M. H. F., Gosling, R. G., and Seeds, W. E., *Nature*, **167**, 759 (1951).
[4] Astbury, W. T., and Bell, F. O., Cold Spring Harb. Symp. Quant. Biol., **6**, 109 (1938).
[5] Cochran, W., Crick, F. H. C., and Vand, V., *Acta Cryst.*, **5**, 581 (1952).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

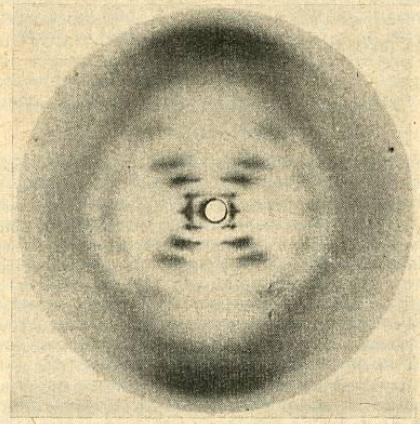
Molecular Configuration in Sodium Thymonucleate

Sodium thymonucleate fibres give two distinct types of X-ray diagram. The first corresponds to a crystalline form, structure *A*, obtained at about 75 per cent relative humidity; a study of this is described in detail elsewhere[1]. At higher humidities a different structure, structure *B*, showing a lower degree of order, appears and persists over a wide range of ambient humidity. The change from *A* to *B* is reversible. The water content of structure *B* fibres which undergo this reversible change may vary from 40–50 per cent to several hundred per cent of the dry weight. Moreover, some fibres never show structure *A*, and in these structure *B* can be obtained with an even lower water content.

The X-ray diagram of structure *B* (see photograph) shows in striking manner the features characteristic of helical structures, first worked out in this laboratory by Stokes (unpublished) and by Crick, Cochran and Vand[2]. Stokes and Wilkins were the first to propose such structures for nucleic acid as a result of direct studies of nucleic acid fibres, although a helical structure had been previously suggested by Furberg (thesis, London, 1949) on the basis of X-ray studies of nucleosides and nucleotides.

While the X-ray evidence cannot, at present, be taken as direct proof that the structure is helical, other considerations discussed below make the existence of a helical structure highly probable.

Structure *B* is derived from the crystalline structure *A* when the sodium thymonucleate fibres take up quantities of water in excess of about 40 per cent of their weight. The change is accompanied by an increase of about 30 per cent in the length of the fibre, and by a substantial re-arrangement of the molecule. It therefore seems reasonable to suppose that in structure *B* the structural units of sodium thymonucleate (molecules on groups of molecules) are relatively free from the influence of neighbouring molecules, each unit being shielded by a sheath of water. Each unit is then free to take up its least-energy configuration independently of its neighbours and, in view of the nature of the long-chain molecules involved, it is highly likely that the general form will be helical[3]. If we adopt the hypothesis of a helical structure, it is immediately possible, from the X-ray



Sodium deoxyribose nucleate from calf thymus. Structure *B*

diagram of structure B, to make certain deductions as to the nature and dimensions of the helix.

The innermost maxima on the first, second, third and fifth layer lines lie approximately on straight lines radiating from the origin. For a smooth single-strand helix the structure factor on the nth layer line is given by :

$$F_n = J_n(2\pi rR) \exp i\, n(\psi + \tfrac{1}{2}\pi),$$

where $J_n(u)$ is the nth-order Bessel function of u, r is the radius of the helix, and R and ψ are the radial and azimuthal co-ordinates in reciprocal space[2]; this expression leads to an approximately linear array of intensity maxima of the type observed, corresponding to the first maxima in the functions J_1, J_2, J_3, etc.

If, instead of a smooth helix, we consider a series of residues equally spaced along the helix, the transform in the general case treated by Crick, Cochran and Vand is more complicated. But if there is a whole number, m, of residues per turn, the form of the transform is as for a smooth helix with the addition, only, of the same pattern repeated with its origin at heights mc^*, $2mc^*$ . . . etc. (c is the fibre-axis period).

In the present case the fibre-axis period is 34 A. and the very strong reflexion at 3·4 A. lies on the tenth layer line. Moreover, lines of maxima radiating from the 3·4-A. reflexion as from the origin are visible on the fifth and lower layer lines, having a J_5 maximum coincident with that of the origin series on the fifth layer line. (The strong outer streaks which apparently radiate from the 3·4-A. maximum are not, however, so easily explained.) This suggests strongly that there are exactly 10 residues per turn of the helix. If this is so, then from a measurement of R_n the position of the first maximum on the nth layer line (for n $5\leqslant$), the radius of the helix, can be obtained. In the present instance, measurements of R_1, R_2, R_3 and R_5 all lead to values of r of about 10 A.

Since this linear array of maxima is one of the strongest features of the X-ray diagram, we must conclude that a crystallographically important part of the molecule lies on a helix of this diameter. This can only be the phosphate groups or phosphorus atoms.

If ten phosphorus atoms lie on one turn of a helix of radius 10 A., the distance between neighbouring phosphorus atoms in a molecule is 7·1 A. This corresponds to the P . . . P distance in a fully extended molecule, and therefore provides a further indication that the phosphates lie on the outside of the structural unit.

Thus, our conclusions differ from those of Pauling and Corey[4], who proposed for the nucleic acids a helical structure in which the phosphate groups form a dense core.

We must now consider briefly the equatorial reflexions. For a single helix the series of equatorial maxima should correspond to the maxima in $J_0(2\pi rR)$. The maxima on our photograph do not, however, fit this function for the value of r deduced above. There is a very strong reflexion at about 24 A. and then only a faint sharp reflexion at 9·0 A. and two diffuse bands around 5·5 A. and 4·0 A. This lack of agreement is, however, to be expected, for we know that the helix so far considered can only be the most important member of a series of coaxial helices of different radii ; the non-phosphate parts of the molecule will lie on inner co-axial helices, and it can be shown that, whereas these will not appreciably influence the innermost maxima on the layer lines, they may have the effect of destroying or shifting both the equatorial maxima and the outer maxima on other layer lines.

Thus, if the structure is helical, we find that the phosphate groups or phosphorus atoms lie on a helix of diameter about 20 A., and the sugar and base groups must accordingly be turned inwards towards the helical axis.

Considerations of density show, however, that a cylindrical repeat unit of height 34 A. and diameter 20 A. must contain many more than ten nucleotides.

Since structure *B* often exists in fibres with low water content, it seems that the density of the helical unit cannot differ greatly from that of dry sodium thymonucleate, 1·63 gm./cm.^{3} [1,5], the water in fibres of high water-content being situated outside the structural unit. On this basis we find that a cylinder of radius 10 A. and height 34 A. would contain thirty-two nucleotides. However, there might possibly be some slight inter-penetration of the cylindrical units in the dry state making their effective radius rather less. It is therefore difficult to decide, on the basis of density measurements alone, whether one repeating unit contains ten nucleotides on each of two or on each of three co-axial molecules. (If the effective radius were 8 A. the cylinder would contain twenty nucleotides.) Two other arguments, however, make it highly probable that there are only two co-axial molecules.

First, a study of the Patterson function of structure *A*, using superposition methods, has indicated[6] that there are only two chains passing through a primitive unit cell in this structure. Since the $A \rightleftharpoons B$ transformation is readily reversible, it seems very unlikely

that the molecules would be grouped in threes in structure *B*. Secondly, from measurements on the X-ray diagram of structure *B* it can readily be shown that, whether the number of chains per unit is two or three, the chains are not equally spaced along the fibre axis. For example, three equally spaced chains would mean that the *n*th layer line depended on J_{3n}, and would lead to a helix of diameter about 60 A. This is many times larger than the primitive unit cell in structure *A*, and absurdly large in relation to the dimensions of nucleotides. Three unequally spaced chains, on the other hand, would be crystallographically non-equivalent, and this, again, seems unlikely. It therefore seems probable that there are only two co-axial molecules and that these are unequally spaced along the fibre axis.

Thus, while we do not attempt to offer a complete interpretation of the fibre-diagram of structure *B*, we may state the following conclusions. The structure is probably helical. The phosphate groups lie on the outside of the structural unit, on a helix of diameter about 20 A. The structural unit probably consists of two co-axial molecules which are not equally spaced along the fibre axis, their mutual displacement being such as to account for the variation of observed intensities of the innermost maxima on the layer lines ; if one molecule is displaced from the other by about three-eighths of the fibre-axis period, this would account for the absence of the fourth layer line maxima and the weakness of the sixth. Thus our general ideas are not inconsistent with the model proposed by Watson and Crick in the preceding communication.

The conclusion that the phosphate groups lie on the outside of the structural unit has been reached previously by quite other reasoning[1]. Two principal lines of argument were invoked. The first derives from the work of Gulland and his collaborators[7], who showed that even in aqueous solution the —CO and $—NH_2$ groups of the bases are inaccessible and cannot be titrated, whereas the phosphate groups are fully accessible. The second is based on our own observations[1] on the way in which the structural units in structures *A* and *B* are progressively separated by an excess of water, the process being a continuous one which leads to the formation first of a gel and ultimately to a solution. The hygroscopic part of the molecule may be presumed to lie in the phosphate groups ($(C_2H_5O)_2PO_2Na$ and $(C_3H_7O)_2PO_2Na$ are highly hygroscopic[8]), and the simplest explanation of the above process is that these groups lie on the outside of the structural units. Moreover, the ready availability of the phosphate groups for interaction

with proteins can most easily be explained in this way.

We are grateful to Prof. J. T. Randall for his interest and to Drs. F. H. C. Crick, A. R. Stokes and M. H. F. Wilkins for discussion. One of us (R. E. F.) acknowledges the award of a Turner and Newall Fellowship.

ROSALIND E. FRANKLIN*
R. G. GOSLING

Wheatstone Physics Laboratory,
King's College, London.
April 2.

* Now at Birkbeck College Research Laboratories, 21 Torrington Square, London, W.C.1.

[1] Franklin, R. E., and Gosling, R. G. (in the press).
[2] Cochran, W., Crick, F. H. C., and Vand, V., *Acta Cryst.*, **5**, 501 (1952).
[3] Pauling, L., Corey, R. B., and Branson, H. R., *Proc. U.S. Nat. Acad. Sci.*, **37**, 205 (1951).
[4] Pauling, L., and Corey, R. B., *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[5] Astbury, W. T., Cold Spring Harbor Symp. on Quant. Biol., **12**, 56 (1947).
[6] Franklin, R. E., and Gosling, R. G. (to be published).
[7] Gulland, J. M., and Jordan, D. O., Cold Spring Harbor Symp. on Quant. Biol., **12**, 5 (1947).
[8] Drushel, W. A., and Felty, A. R., *Chem. Zent.*, **89**, 1016 (1918).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.

(*Reprinted from Nature, Vol.* 171, *p.* 964, *May* 30, 1953)

GENETICAL IMPLICATIONS OF THE STRUCTURE OF DEOXYRIBONUCLEIC ACID

By J. D. WATSON and F. H. C. CRICK

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

THE importance of deoxyribonucleic acid (DNA) within living cells is undisputed. It is found in all dividing cells, largely if not entirely in the nucleus, where it is an essential constituent of the chromosomes. Many lines of evidence indicate that it is the carrier of a part of (if not all) the genetic specificity of the chromosomes and thus of the gene itself.

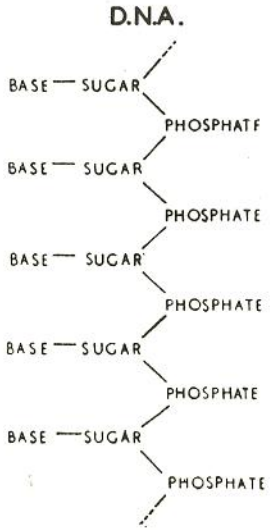


Fig. 1. Chemical formula of a single chain of deoxyribonucleic acid

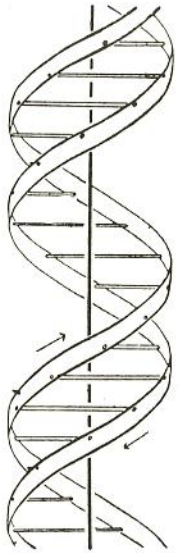


Fig. 2. This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

Until now, however, no evidence has been presented to show how it might carry out the essential operation required of a genetic material, that of exact self-duplication.

We have recently proposed a structure[1] for the salt of deoxyribonucleic acid which, if correct, immediately suggests a mechanism for its self-duplication. X-ray evidence obtained by the workers at King's College, London[2], and presented at the same time, gives qualitative support to our structure and is incompatible with all previously proposed structures[3]. Though the structure will not be completely proved until a more extensive comparison has been made with the X-ray data, we now feel sufficient confidence in its general correctness to discuss its genetical implications. In doing so we are assuming that fibres of the salt of deoxyribonucleic acid are not artefacts arising in the method of preparation, since it has been shown by Wilkins and his co-workers that similar X-ray patterns are obtained from both the isolated fibres and certain intact biological materials such as sperm head and bacteriophage particles[2,4].

The chemical formula of deoxyribonucleic acid is now well established. The molecule is a very long chain, the backbone of which consists of a regular alternation of sugar and phosphate groups, as shown in Fig. 1. To each sugar is attached a nitrogenous base, which can be of four different types. (We have considered 5-methyl cytosine to be equivalent to cytosine, since either can fit equally well into our structure.) Two of the possible bases—adenine and guanine—are purines, and the other two—thymine and cytosine—are pyrimidines. So far as is known, the sequence of bases along the chain is irregular. The monomer unit, consisting of phosphate, sugar and base, is known as a nucleotide.

The first feature of our structure which is of biological interest is that it consists not of one chain, but of two. These two chains are both coiled around a common fibre axis, as is shown diagrammatically in Fig. 2. It has often been assumed that since there was only one chain in the chemical formula there would only be one in the structural unit. However, the density, taken with the X-ray evidence[2], suggests very strongly that there are two.

The other biologically important feature is the manner in which the two chains are held together. This is done by hydrogen bonds between the bases, as shown schematically in Fig. 3. The bases are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the

other. The important point is that only certain pairs of bases will fit into the structure. One member of a pair must be a purine and the other a pyrimidine in order to bridge between the two chains. If a pair consisted of two purines, for example, there would not be room for it.

We believe that the bases will be present almost entirely in their most probable tautomeric forms. If this is true, the conditions for forming hydrogen bonds are more restrictive, and the only pairs of bases possible are :

adenine with thymine ;
guanine with cytosine.

The way in which these are joined together is shown in Figs. 4 and 5. A given pair can be either way round. Adenine, for example, can occur on either chain ; but when it does, its partner on the other chain must always be thymine.

This pairing is strongly supported by the recent analytical results[5], which show that for all sources of deoxyribonucleic acid examined the amount of adenine is close to the amount of thymine, and the amount of guanine close to the amount of cytosine, although the cross-ratio (the ratio of adenine to guanine) can vary from one source to another. Indeed, if the sequence of bases on one chain is irregular, it is difficult to explain these analytical results except by the sort of pairing we have suggested.

The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of

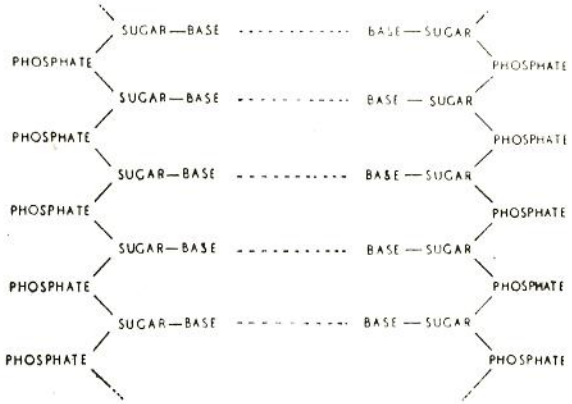


Fig. 3. Chemical formula of a pair of deoxyribonucleic acid chains. The hydrogen bonding is symbolized by dotted lines

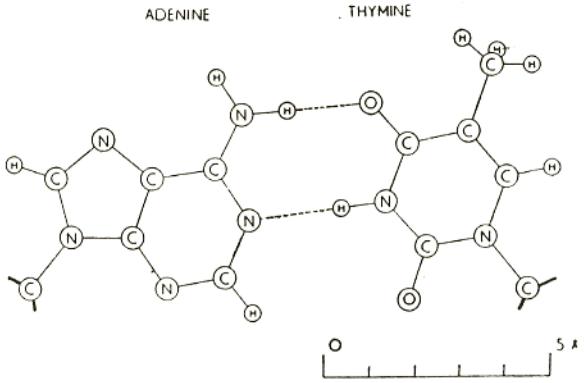


Fig. 4. Pairing of adenine and thymine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown

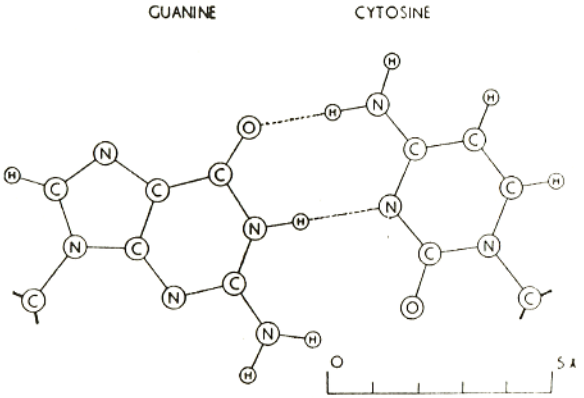


Fig. 5. Pairing of guanine and cytosine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown

bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information. If the actual order of the bases on one of the pair of chains were given, one could write down the exact order of the bases on the other one, because of the specific pairing. Thus one chain is, as it were, the complement of the other, and it is this feature which suggests how the deoxyribonucleic acid molecule might duplicate itself.

Previous discussions of self-duplication have usually involved the concept of a template, or mould. Either

the template was supposed to copy itself directly or it was to produce a 'negative', which in its turn was to act as a template and produce the original 'positive' once again. In no case has it been explained in detail how it would do this in terms of atoms and molecules.

Now our model for deoxyribonucleic acid is, in effect, a *pair* of templates, each of which is complementary to the other. We imagine that prior to duplication the hydrogen bonds are broken, and the two chains unwind and separate. Each chain then acts as a template for the formation on to itself of a new companion chain, so that eventually we shall have *two* pairs of chains, where we only had one before. Moreover, the sequence of the pairs of bases will have been duplicated exactly.

A study of our model suggests that this duplication could be done most simply if the single chain (or the relevant portion of it) takes up the helical configuration. We imagine that at this stage in the life of the cell, free nucleotides, strictly polynucleotide precursors, are available in quantity. From time to time the base of a free nucleotide will join up by hydrogen bonds to one of the bases on the chain already formed. We now postulate that the polymerization of these monomers to form a new chain is only possible if the resulting chain can form the proposed structure. This is plausible, because steric reasons would not allow nucleotides 'crystallized' on to the first chain to approach one another in such a way that they could be joined together into a new chain, unless they were those nucleotides which were necessary to form our structure. Whether a special enzyme is required to carry out the polymerization, or whether the single helical chain already formed acts effectively as an enzyme, remains to be seen.

Since the two chains in our model are intertwined, it is essential for them to untwist if they are to separate. As they make one complete turn around each other in 34 A., there will be about 150 turns per million molecular weight, so that whatever the precise structure of the chromosome a considerable amount of uncoiling would be necessary. It is well known from microscopic observation that much coiling and uncoiling occurs during mitosis, and though this is on a much larger scale it probably reflects similar processes on a molecular level. Although it is difficult at the moment to see how these processes occur without everything getting tangled, we do not feel that this objection will be insuperable.

Our structure, as described[1], is an open one. There is room between the pair of polynucleotide chains

(see Fig. 2) for a polypeptide chain to wind around the same helical axis. It may be significant that the distance between adjacent phosphorus atoms, 7·1 A., is close to the repeat of a fully extended polypeptide chain. We think it probable that in the sperm head, and in artificial nucleoproteins, the polypeptide chain occupies this position. The relative weakness of the second layer-line in the published X-ray pictures[3a,4] is crudely compatible with such an idea. The function of the protein might well be to control the coiling and uncoiling, to assist in holding a single polynucleotide chain in a helical configuration, or some other non-specific function.

Our model suggests possible explanations for a number of other phenomena. For example, spontaneous mutation may be due to a base occasionally occurring in one of its less likely tautomeric forms. Again, the pairing between homologous chromosomes at meiosis may depend on pairing between specific bases. We shall discuss these ideas in detail elsewhere.

For the moment, the general scheme we have proposed for the reproduction of deoxyribonucleic acid must be regarded as speculative. Even if it is correct, it is clear from what we have said that much remains to be discovered before the picture of genetic duplication can be described in detail. What are the polynucleotide precursors ? What makes the pair of chains unwind and separate ? What is the precise role of the protein ? Is the chromosome one long pair of deoxyribonucleic acid chains, or does it consist of patches of the acid joined together by protein ?

Despite these uncertainties we feel that our proposed structure for deoxyribonucleic acid may help to solve one of the fundamental biological problems—the molecular basis of the template needed for genetic replication. The hypothesis we are suggesting is that the template is the pattern of bases formed by one chain of the deoxyribonucleic acid and that the gene contains a complementary pair of such templates.

One of us (J.D.W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis (U.S.A.).

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

[2] Wilkins, M. H. F., Stokes, A. R., and Wilson, H. R., *Nature*, **171**, 738 (1953). Franklin, R. E., and Gosling, R. G., *Nature*, **171**, 740 (1953).

[3] (*a*) Astbury, W. T., Symp. No. 1 Soc. Exp. Biol., 66 (1947). (*b*) Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952). (*c*) Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953). (*d*) Fraser, R. D. B. (in preparation).

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

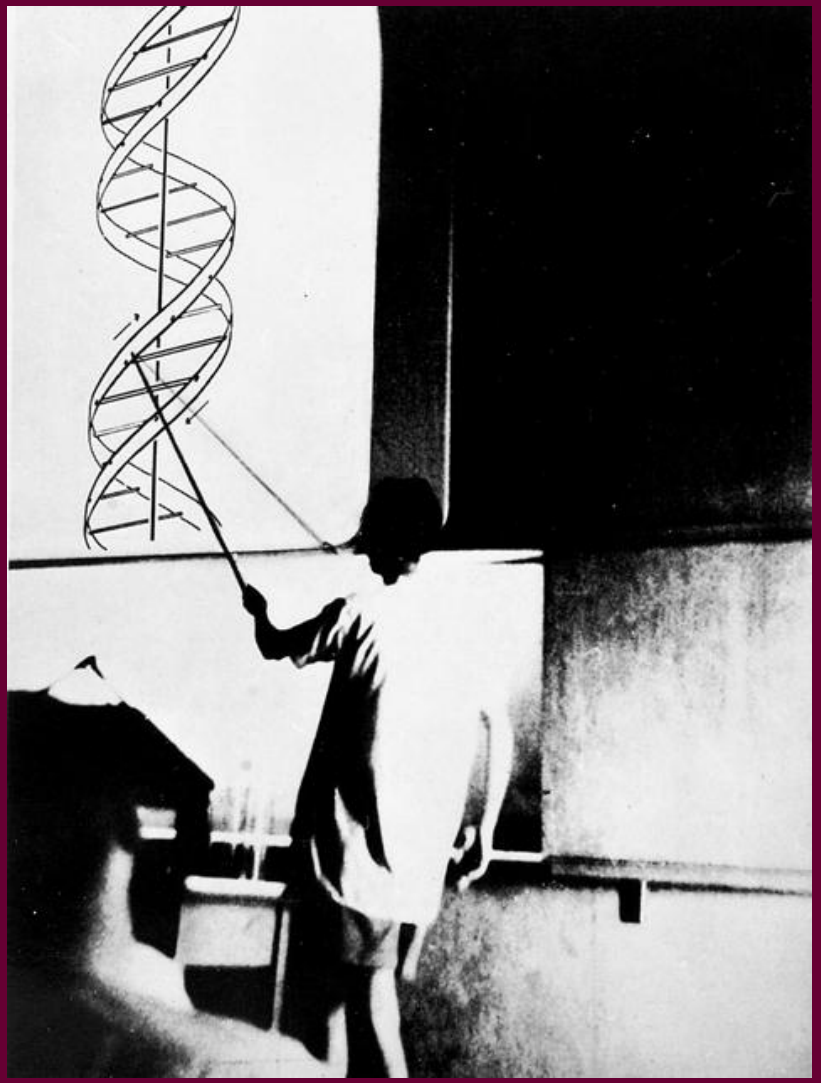
[5] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952). Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.

Lost Crick Correspondence

My dear Francis,
I gather you have got the coordinates of your model or some worked out. Do you think we could have a copy of what you have? The crystalline data is clearing up nicely. To think that Rosie had all the 3D data for 9 months & wouldn't fit a helix to it and there was I taking her word for it that the data was anti-helical. Christ. We have redone a lot of the 3D more accurately on mouse & will need all the extra accuracy for dealing with some of the finer points. Regards & to Odile too.
Yours
M
P.S. I think I have a flat.

3 June 1953

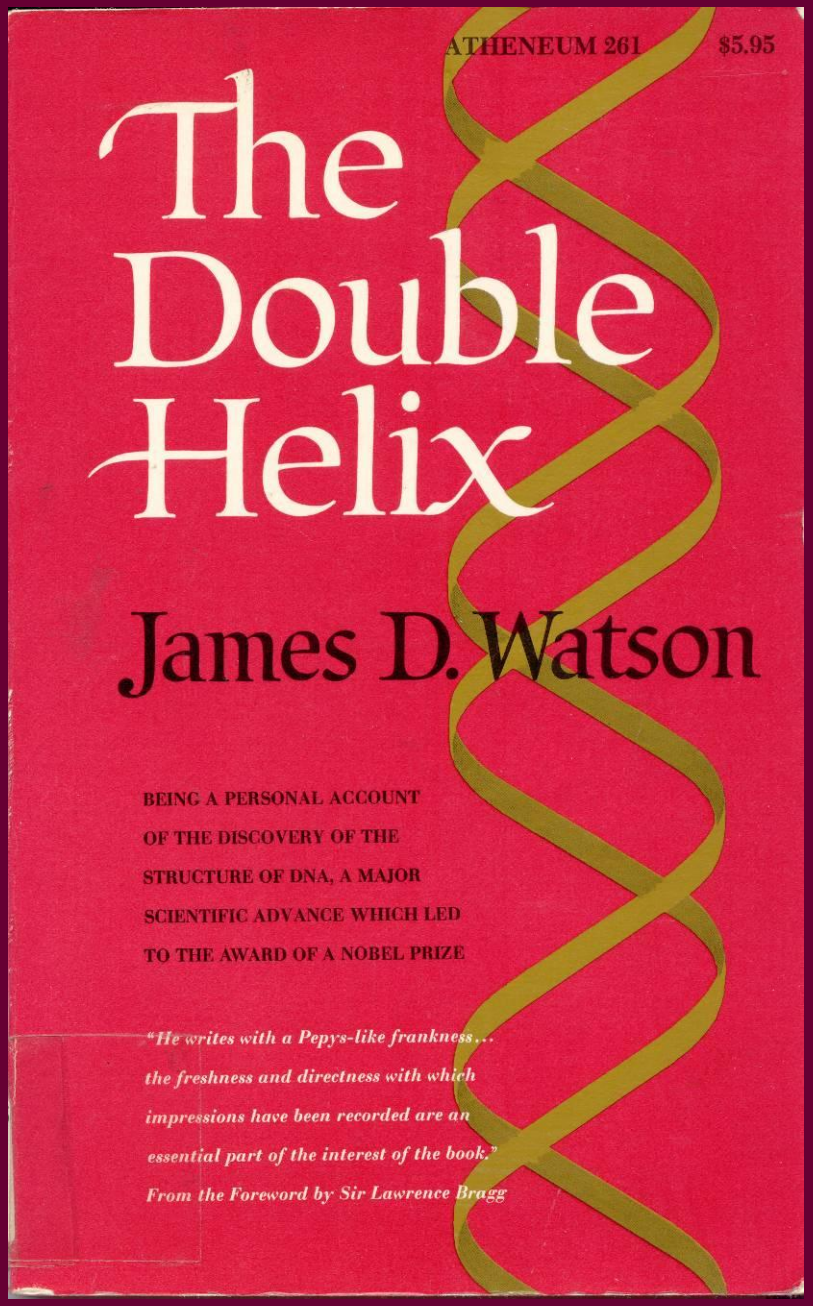


Watson explains the double helix with an overhead projector at the 1953 CSH Symposium.

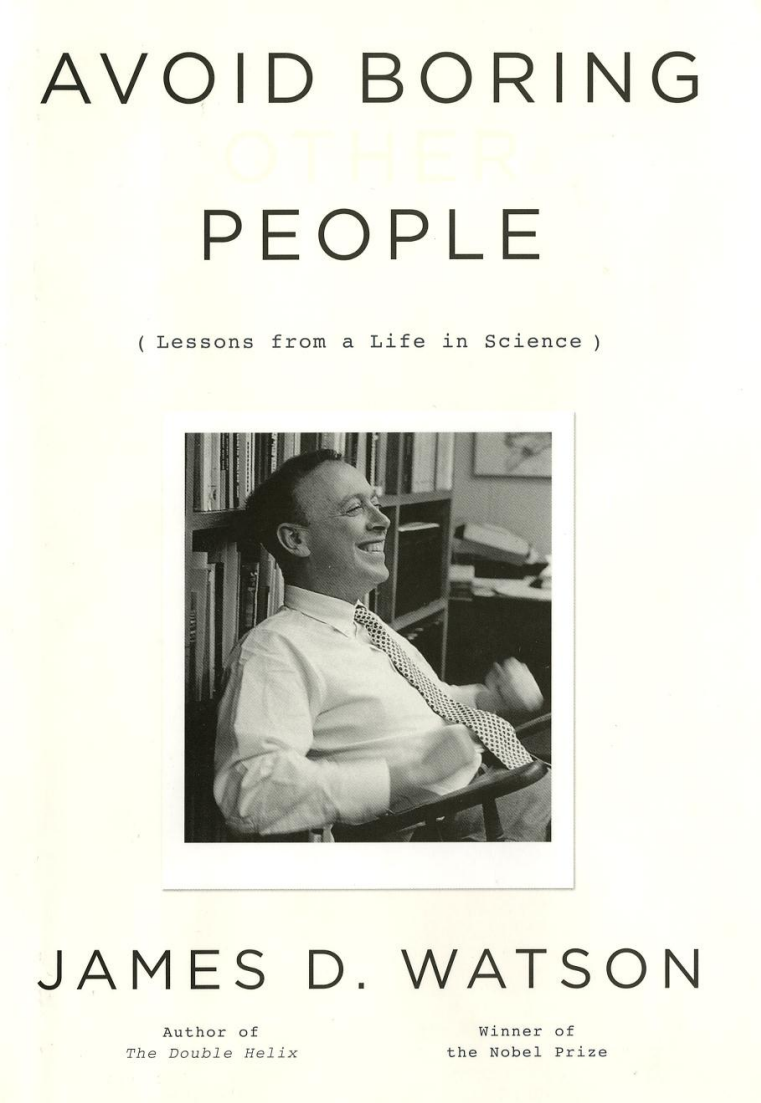
Pasadena Conference on the Structure of Proteins

September 1953

1962 Nobel Prize Ceremony


ATHENEUM 261 $5.95
The Double Helix
James D. Watson
BEING A PERSONAL ACCOUNT
OF THE DISCOVERY OF THE
STRUCTURE OF DNA, A MAJOR
SCIENTIFIC ADVANCE WHICH LED
TO THE AWARD OF A NOBEL PRIZE
"He writes with a Pepys-like frankness...
the freshness and directness with which
impressions have been recorded are an
essential part of the interest of the book."
From the Foreword by Sir Lawrence Bragg

Aiming for Gold



1. Choose an objective apparently ahead of its time

2. Work on problems only when you feel tangible success may come in several years

3. Work with a teammate who is your intellectual equal

4. Stay in close contact with your intellectual competitors

5. Never be the brightest person in the room

6. Always have someone to save you