

ment, and to Dr. G. E. R. Deacon and the staff and officers of R.R.S. *Discovery II* for their assistance in making the observations.

ing, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, 40, 140 (1920).

quet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, 225 (1949).

Ark, W. S., *Woods Hole Papers in Phys. Oceanog. Meteor.*, 11 (3) (1950).

son, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, 2(11) (1936).

MOLECULAR STRUCTURE OF DEOXYRIBOSE NUCLEIC ACID

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.

The structure for nucleic acid has already been proposed by Pauling and Corey¹. They kindly made our manuscript available to us in advance of publication. Their model consists of three interwoven 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 ray diagrams is the salt, not the free acid. Without acidic hydrogen atoms it is not clear what forces could 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's² 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

figure is purely schematic. The two lines symbolize the phosphate-sugar backbone, and the horizontal rods the pairs of deoxyribose chains.

is a residue on each chain every 3.4 Å. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. 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 chains to tilt so that the structure could be more compact.

The novel feature of the structure is the manner in which the chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined 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 form (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 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, 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 may 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 co-ordinates assumed in building it, together with a list 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.

The double helix

Sydney Brenner

A brief note published in the middle of the last century described a possible structure of DNA: it included a throwaway comment that the structure suggested how DNA could be copied, and so could act as heritable genetic material. The proposal did not meet with instant acceptance, except among a small group of enthusiasts. But its explanatory power was quickly tested and extended, and the structure has assumed iconic status in both biology and the wider world.

J. D. Watson and F. H. C. Crick's paper,¹ "A structure for deoxyribose nucleic acid," published on 25 April 1953, ushered in the modern era of biology. The little figure it included, showing the double-helical structure, has become a symbol of that era and an icon of our time (Fig. 6.1). Today, when DNA, genes, and genetics are on everybody's lips, it is hard to believe that there was a time when many viewed DNA as an inert polymer that was unlikely to have anything to do with heredity.

In the early 1950s, Watson and Crick were at a Medical Research Council unit in the Cavendish Laboratory of Cambridge University. I was working at Oxford University then, and vividly remember the day in early April 1953 when I and others went over to Cambridge to see the Watson-Crick model of DNA. I found it a revelation, and immediately knew what direction I wanted to follow in science. But when the paper appeared later that month, it by no means received universal acclaim. The scientific establishment of the time consisted largely of biochemists who were preoccupied with the transformation of energy and matter in biological systems, and they failed to see how the structure would revolutionize the study of biological information. But that it did. It revealed the basis on which genetic material is passed



Fig. 6.1. An icon of our time: the double-helical structure of DNA as portrayed on the cover of *Nature* of 15 February 2001. This issue⁴ contained milestone papers from the Human Genome Project announcing a draft DNA sequence of the entire human genome. Simultaneously, a draft sequence produced by a company, Celera Genomics, appeared in *Science*.⁵ The “rungs” on the double-helical ladder are pairs of the four molecules (adenine, thymine, guanine, and cytosine) known as bases, while the strands that constitute the outside of the ladder are chains composed of sugar and phosphate. (The image, by Eric Lander, was created by Runaway Technology Inc. using PhotoMosaic by Robert Silvers from artwork by Darryl Leja. Reproduced courtesy of the Whitehead Institute)

from one generation to the next, and it showed the principle behind the genetic code through which proteins are built from a DNA blueprint.

Although most of modern biology stems from this brief paper, there is also a long prehistory of thought in the subject that we now know as genetics. It begins in the 1850s and 1860s with the famous pea-breeding experiments of Gregor Mendel, a monk living in what is now the Czech Republic. His results were published in an obscure journal, and only in 1900 were

they rediscovered and their significance recognized. Put simply, Mendel postulated that there were internal factors that specified external characters; and that to identify a factor specifying a normal or “wild-type” character in an organism, one had to find an example of that organism displaying a variant character. Today the factors are called genes. The characters are the phenotype, an organism’s observable characteristics. And a variant character is a mutant phenotype, which results from the action of a mutant rather than wild-type form of a gene.

Following the rediscovery of Mendel’s work, genetics flourished largely through work on the fruit fly, *Drosophila melanogaster*. This tiny insect is easily bred in large numbers and has large numbers of mutants (in eye color, for instance, among many other characteristics). The pioneer was Thomas Hunt Morgan, who spent most of his career at Columbia University. As a frustrated embryologist, he had entered the new field of genetics around the turn of the century in the hope that knowledge about genes would throw light on the then intractable problems of developmental biology.

Morgan’s research was largely responsible for revealing the relationship between Mendel’s findings and genes and chromosomes (of which the fruit fly has four pairs; we have twenty-three pairs). Most phenotypic characters are inherited independently of each other, but some are “linked”: that is, they tend to be inherited together, implying that they occur on the same chromosome. Linkage can be broken, because pairs of chromosomes can swap segments in a process known as recombination. The closer genes are to each other on a chromosome, the stronger the linkage, and the more likely they or their mutant forms are to be inherited together. In consequence, from the breeding and “gene-mapping” work in *Drosophila*, analyzing patterns of inheritance of mutant phenotypes, Morgan could assign a physical reality to the gene as part of a chromosome and with a defined location on that chromosome.

Meanwhile, research on enzymes was developing in the new field of biochemistry. Enzymes are proteins that catalyze chemical transformations in a cell. Their chemical and physical constitution were to remain enigmas for some time. But in 1909 a British physician, Archibald Garrod, proposed a connection between genes and enzymes: he observed that certain inherited diseases reflected a patient’s production of malfunctioning enzymes (he called these “inborn errors of metabolism”). In the 1940s this line of thinking culminated in the research of G. W. Beadle and E. L. Tatum in the United States, who investigated nutritional deficiencies due to mutant enzymes in the bread mold *Neurospora*. From the results Beadle formulated the “one

gene, one enzyme” principle, which holds that a single gene specifies a single enzyme.

As early as the last years of the nineteenth century, E. B. Wilson had stated in a book that DNA (which he called “chromonucleic acid”) was the physical carrier of heredity. And in 1944, in what at least in retrospect were highly convincing experiments, Oswald Avery and colleagues showed that a heritable alteration to one strain of the bacterium *Pneumococcus* could be caused by adding an extract of a second strain to the growth medium of the first. They identified the active material in the extracts as DNA. Nonetheless, as the 1950s dawned, most people still believed that proteins would have to have the leading role in copying genes: only proteins seemed to have the required chemical specificity.

Thus was the stage set. Genetics had arrived at the conclusion that the gene was the fundamental unit of heredity, indivisible like the atom of early physics. Biochemistry had shown that the work of cells was performed by enzymes, specialized proteins of elaborate but unknown structures, performing all the various steps of biosynthesis and energy metabolism. Each enzyme was specified by a gene, but no enzymes seemed to be concerned with specifying genes. DNA was a candidate as the genetic material, but by no means was this widely accepted.

When Watson and Crick’s paper¹ appeared, it was not at all immediately evident that it provided an answer to any of the problems of genetics confronting us at that time. The structure was described and the notion of the specificity of the base pairs stated: the authors proposed that the four types of molecule (bases) of which DNA was known to be composed—adenine, thymine, guanine, and cytosine—could pair off only as A–T and G–C. However, the only comment about the structure’s potential biological significance was contained in the famous sentence: “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.”

In an article² marking the twenty-first anniversary of the paper’s appearance, we learn from Francis Crick that the sentence was inserted as a compromise between the two authors. Crick was keen to discuss the genetic implications.

Watson [writes Crick]² was against it. He suffered from periodic fears that the structure might be wrong and that he had made an ass of himself. I yielded to his point of view but insisted that something be put in the paper, otherwise someone else would certainly write to make the suggestion, assuming we had been too blind to see it.

I was an early convert: my own memory of having seen the structure earlier in Cambridge is that it was certain to be right, explaining as it did instantly to me the central puzzle of genetics. But it didn't take the authors themselves long to present their argument in more detail, for a second paper³ appeared on 30 May 1953 which spelled out the biological implications. In it Watson and Crick give more detail about "complementary replication"—how the invariable pairing of A with T and C with G (fig. 6.2) could lead to reconstruction of a double helix from the information contained in one strand only. They explain how mutations could occur through one base being substituted for another. And they mention that "it . . . seems likely that the precise sequence of the bases is the code which carries the genetical information." This generated the whole field of study of the genetic code.

Before these possibilities registered with most scientists, the proposed structure of DNA had to be combined with two other pieces of research to result in the birth of modern biology. The first, which preceded the double helix, was Frederick Sanger's work in Cambridge on the structure of the protein insulin. Sanger showed that proteins have a precise chemical structure as defined by the sequence of amino acids linked in a chain. As became evident in later work on the genetic code in which I was involved, the sequence of bases in the DNA specifies the sequence of amino acids in a protein via an intermediate molecule, messenger RNA.

The second piece of research was Seymour Benzer's subsequent investigations on the fine structure of the gene. Benzer was at Purdue University, working with bacteriophages—viruses that infect bacteria. He isolated hundreds of mutants of a gene; by using techniques to detect viruses that were the result of genetic recombination, he was able to resolve and map mutant phenotypes onto many different sites on a DNA sequence and show that they corresponded to distances between the base pairs in the DNA structure. This destroyed the classical, bead-on-a-string notion of genes as indivisible units of function, mutation, and recombination. It was replaced by the concept of the gene as a stretch of DNA, with hundreds of possible mutational states caused by alterations of single base pairs.

Today we no longer have to study genes indirectly through their effects on the phenotypes of organisms, but can use the technologies of DNA cloning and sequencing to characterize genes directly. Today the genetic code is no longer a daring speculation, but a working tool that allows us to derive the sequences of amino acids in a protein from the DNA sequence of the gene encoding it. Today we know in great detail the molecular machinery that copies DNA, repairs it and expresses it.

It all started with this little paper of Watson and Crick.¹ The fusion of

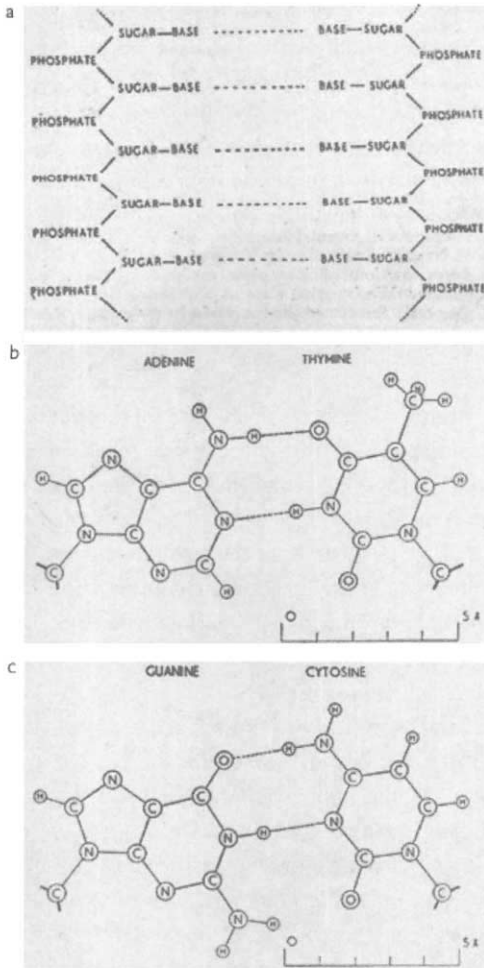


Fig. 6.2. The greater detail about DNA structure and base-pairing that followed in Watson and Crick's second paper,³ of 30 May 1953. These diagrams are facsimiles of the originals. a, The chemical formula of a pair of DNA strands, with sugar-phosphate chains and bases attached to the chains. Forces known as hydrogen bonds (dotted lines) connect each pair of bases and so the two strands of the double helix. b, The pairing of adenine with thymine, and c, that of guanine with cytosine. Watson and Crick's most telling insight¹ was how, because A always pairs with T and G with C, the double helix can be re-created from a single strand—that is, how the genetic material can be copied.

genetics and biochemistry was then already under way. But the structure of DNA greatly intensified the process, which resulted in the creation of molecular biology and the tools to study the chemistry of information in biological systems. The route to the double helix was strewn with Nobel Prizes (see Further reading). And in 1962 Watson and Crick, along with Maurice Wilkins for his analysis of DNA structure using the X-ray technique, received the prize in physiology or medicine. The fiftieth anniversary of the structure's publication was widely celebrated in 2003 with various events and publications.⁶

Watson and Crick, of course, are among those rare scientists whose

names are known to the general public. Watson has been director, then president, of Cold Spring Harbor Laboratory, New York, and was a prime mover behind the Human Genome Project. In recent years Crick has been at the Salk Institute in California, his interests centering on neurobiology and the problem of consciousness.

Let Crick have the last word on the double helix. Here he is again, in his 1974 article,² responding to the argument that a scientific discovery is akin to a work of art, and style is as important as content:

Rather than believe that Watson and Crick made the DNA structure, I would rather stress that the structure made Watson and Crick. After all, I was almost totally unknown at the time and Watson was regarded, in most circles, as too bright to be really sound. But what I think is overlooked in such arguments is the intrinsic beauty of the DNA double helix. It is the molecule which has style, quite as much as the scientists.

References

1. Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Crick, F. The double helix: a personal view. *Nature* **248**, 766–769 (1974).
3. Watson, J. D. & Crick, F. H. C. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967 (1953).
4. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
5. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
6. Dennis, C. (ed.) The double helix: 50 years. *Nature* **421**, 395–453 (2003).

Further reading

- Judson, H. F. *The Eighth Day of Creation: Makers of the Revolution in Biology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996; expanded edition).
- Nobel e-Museum. *The Nobel Prize in Physiology or Medicine 1962* (<http://www.nobel.se/medicine/laureates/1962/>). (Also physiology or medicine prizes in 1933 and 1958, and chemistry prize in 1958.)
- Olby, R. C. *The Path to the Double Helix: The Discovery of DNA* (Dover, Mineola, NY, 1994; reprint of the edition of 1974).
- Watson, J. D. *The Double Helix* (Penguin, London, 1999; new edition with introduction by S. Jones).
- Watson, J. D. with Berry, A. *DNA: The Secret of Life* (Knopf, New York, 2003).