Historical article: DNA polymorphism and the early history of the double helix

Struther Arnott

Biological Structure and Function Section, Biomedical Sciences Division, Imperial College London, London SW7 2AZ, UK

Early X-ray diffraction patterns from oriented fibres indicated that DNA must have a simple, repetitious structure and encouraged some researchers, who were already convinced that DNA was the genetic material, to undertake more detailed diffraction analyses and speculative modelling. The pioneering experimental work by Wilkins in the Wheatstone Laboratory at King’s College London in the late 1940s first inspired, and then was overtaken by, the conjectural modelling of Watson and Crick in the Cavendish Laboratory at Cambridge. Why this was allowed to happen is still something of a puzzle. Here, I explore the puzzle and expose a peculiar flaw in the details of the original Watson–Crick model that was left for Wilkins to resolve.

Why me and why now?
The Royal Society of London subjects its Fellows to a postmortem examination in the form of a biographical memoir prepared by surviving colleagues. For Maurice Wilkins (1916–2004), some of this duty fell to me. I worked with Wilkins for a decade that saw the award of a Nobel prize in 1962 to Crick, Watson and myself, and with Wilkins for a decade that saw the award of a Nobel prize in 1968 to Crick, Watson and himself, and publication in 1968 of The Double Helix [1] – Watson’s historical fable that provoked so much controversy about the roles of other scientists at King’s College London (KCL). With members of Wilkin’s group, I helped to consolidate the structure of the DNA duplex and to define the structures of RNA duplexes. Along the way, I had to develop more systematic methods for analysing X-ray diffraction patterns from fibrous biopolymers. These methods made it easier for me to explore DNA and RNA polymorphism and to solve the structures of important polypeptides and gel-forming polysaccharides from many plant and animal connective tissues.

Before I joined Wilkins, I had worked with the famous chemists J.M. Robertson and D.H.R. Barton and therefore could be regarded as the first professional chemical crystallographer to look at DNA. My mindset was that solving structures had to be the first priority. To do this, we had to adopt or to adapt methods that had worked before; only when these methods failed did we try to invent new tools.

Distortion
I am only too well aware that publications about personalities can be more exciting and profitable than chronicles confined to science, but on re-visiting Wilkin’s early DNA work I was disconcerted by the extent to which the drama in The Double Helix had overpowered more subtle historical issues and by the persistence of many erroneous details from early reports on DNA that were refuted in subsequent publications. Even in his much later autobiography [2], Wilkins was, to a large extent, in thrall to the agenda set by Watson, and he left unanswered many intriguing questions about his scientific strategy and tactics, especially those relating to the early days of his DNA programme.

In neither source, for example, did it ever become clear why Wilkins did so little to exploit his very good diffraction pattern of A-form DNA in 1950 (sic). This pattern was the one that prompted Alex Stokes to decide – together with Wilkins – that DNA was helical, and to go on to describe so tersely and elegantly the general features of diffraction from helices. It was this same pattern, displayed by Wilkins in Naples in the Spring of 1951, that first enthused Watson with the prospect that the genetic material might have an incredibly simple crystal structure that would be easy to model.

As it turned out, Watson’s enthusiasm for DNA modelling had to be re-ignited in 1953 by a different X-ray pattern from another DNA allomorph, B-DNA. So it came about that it was not the dimensions and symmetry of A-DNA but those of B-DNA that were incorporated into the first base-paired model of the double helix [3]. Ironically, the conformational details of this model were, unwittingly, more appropriate for an A-form DNA [4]; thus, the Watson–Crick structure that so satisfied biologists failed the tests of diffraction physicists. By taking up and resolving this dilemma, Wilkins kept alive and augmented his stake in the discovery of the structure of DNA.

Polymorphism
The diverse structures of globular proteins illustrate just how varied polymorphic biopolymers can be, even when they are composed of chemically similar building blocks. By contrast, the similarity of these blocks is made apparent in, for example, extensive α or β helices, where successive peptides have the same backbone conformation despite their different side chains. In special cases, a repeated sequence gives rise to a unique symmetrical structure; for example, a repeated Pro-Pro-Gly sequence leads to the collagen triple helix.
Figure 1. Fibre diffraction fingerprints and helical structures of some DNAs and RNAs. (a) A-DNA; (b) B-DNA; (c) Z-DNA; (d) A-RNA. In each form shown, the double helical molecules are uniaxially orientated and further ordered into microcrystals, the axes of which are also more or less parallel. The fibres have been tipped in order to record on the meridian (the central vertical axis in each pattern) the reflections from which the screw symmetries of the polynucleotide chains can be diagnosed. For A-DNA (a), the first meridional reflection is on the eleventh layer line above the equator (the central horizontal axis in each pattern) and indicates 11-fold screw symmetry (i). The patterns for B-DNA (b), Z-DNA (c) and A-RNA (d) indicate, respectively, 10-fold, 6-fold and 12-fold symmetry (i). In series (ii) of the images, one pitch length of each double helix is shown alongside its diffraction pattern; the views are perpendicular to the helix axes and are designed to show mainly the groove structures in the different helices. In series (iii), the views perpendicular to the helix axes show that the phosphate groups are invariantly on the molecular periphery but the base pairs are in various positions. A-DNA (a) is the double helix that is most compact vertically with a pitch of only 28 Å associated with base pairs that are tipped away from being perpendicular to the helix axis; the base pairs are near the surface of the duplex and therefore the minor groove of A-DNA is shallow (ii,iii). B-DNA (b) has a very different morphology with a pitch of 34 Å; bases are
Nucleotides are more complex than peptides; thus, polynucleotides, when single-stranded, can be expected to be even more polymorphic than polypeptides. But when Watson–Crick base-pairing is possible, the structural isomorphism of the $\text{A} \cdot \text{T}, \text{G} \cdot \text{C}, \text{T} \cdot \text{A}$ and $\text{C} \cdot \text{G}$ pairs results in just two common types of polynucleotide double helix, A-DNA and B-DNA, which are both right-handed (Figure 1a,b). The analogy with polypeptides is continued with special nucleotide sequences; for example, poly(d(GC))poly(d(GC)) can exist in an exotic left-handed Z helix (Figure 1c), and in A- and B-form DNA.

In each category (A, B and Z), there is enough residual conformational variability to produce a wide range of distinct helical morphologies in DNA and RNA double helices [5]. I have provided a sampler gallery of only four of these structures: namely, the ones that I explicitly discuss in this review. Their X-ray diffraction fingerprints are shown in Figure 1a(i),b(i),c(i),d(i), accompanied by mutually perpendicular views of their helical shapes in Figure 1a(ii,iii),b(ii,iii),c(ii,iii),d(ii,iii).

**A-DNA**

Polymorphism confused the interpretation of earliest X-ray patterns from DNA fibres obtained by Astbury's group at Leeds. We can now see that their specimens were often a mixture of the two phases A and B, which led to the confused conclusion that the two important vertical spacings in DNA were 27 Å (the layer line spacing from A-DNA) and 3.4 Å (the strong meridional reflection from B-DNA). Noticing that $3.4 \approx 27/8$, Astbury's rational deduction was that the ordered form of DNA was a stack of nucleotides with an eight-fold periodicity perhaps deriving from the sequence of bases. But it was Furberg, a PhD student in Bernal's laboratory at Birkbeck College, London (1947–1949), who made an imaginative synthesis and constructive leap forward in his PhD thesis (S. Furberg, PhD thesis, University of London, 1949) by putting together a DNA model that was indeed a (eightfold) helix comprising a single strand of nucleotides with stacked bases and sugar rings that displayed the conformation that he had obtained from his own X-ray crystallographic analysis of cytidine [6] – itself a virtuoso achievement in the late 1940s.

The pattern of crystalline A-DNA that Wilkins and Gosling [7] obtained in the summer of 1950 not only cleanly removed the confusion of mixed phases but also provided a sure foundation for a quasi-orthodox X-ray analysis. Its 100 diffraction spots provided unit-cell dimensions ($a = 22\text{ Å}, b = 40\text{ Å}, c = 28\text{ Å}, \beta = 97^\circ$) and a space group assignment (C2), and the peculiar overall distribution of intensity from the spots suggested a symmetry for the molecular structure that lay behind the diffraction pattern. The odd feature of the pattern (Figure 1a(ii)) is the lozenge-shaped empty space on the meridian that comes to a point on the eleventh layer line.

Crystallographic axes with empty diffraction spaces imply screw (i.e. helix) symmetries. Crystals themselves are permitted only two-, three-, four- or six-fold screw axes; therefore, the 11-fold screw symmetry indicated by the first meridional reflection appearing on the eleventh layer line of A-DNA has to be a property of its molecules and not their packing in the crystal. A C-face-centred unit cell with the ratio $b/a = \sqrt{3}$ further implies that, looking down the fibre axis, the scattering units are equally spaced ~22 Å apart on a quasi-hexagonal net. Moreover, the additional two-fold rotation axes of the provisional space group would be perpendicular to the molecular 11-fold screw axis and would partition each scattering unit into two subunits of opposite polarity.

In retrospect, it is difficult to imagine that a well-trained crystallographer, committed to getting a structure by hook or by crook, would not conclude that the assignment to A-DNA of space group C2, no. 5, in International Tables [8] immediately meant that its structure contained $2 \times 2$ almost identical polynucleotide chains, diadically paired and packed like a bundle of cylinders 22 Å apart. The cylinders, in fact, would be spirals with 11-fold screw symmetry.

Of course, this is not what happened.

**What if…?**

Structural crystallographers in their analyses prefer to escape as soon as possible from diffraction or reciprocal space into real space. The best way of doing this is to make a Fourier transformation of the X-ray amplitudes – that is, the square roots of the intensities, suitably corrected – and their associated phases. From the resulting electron density map, a model can be made. When the X-ray data have high resolution, making a model might be the very straightforward business of either identifying atoms and their positions or draping a polypeptide chain of known sequence along a ridge of high electron density. At the lower resolutions characteristic of DNA fibre data, more structural information must be put into the model because the individual atomic positions are blurred. This information has to come in the form of prefabricated chemical components with dimensions that are well established from high-resolution analyses of model compounds.

Modelling of some kind is inescapable in structural crystallography, even if it is no more than a list of provisional conclusions. Tangible models are better. They challenge vagueness and, when appropriately constructed, can be used to calculate not only diffraction intensities (to compare with those observed) but also diffraction phases that can be combined with the intensity-derived amplitudes in Fourier syntheses to start or to continue a process of imaging and improvement. It is surprising how adequate, from a phasing point of view, incomplete or inaccurate structures can be, and how few data are needed to make perpendicular to the helix axis and near the centre of the duplex, such that both grooves are deep (ii,iii). Z-DNA (c) has a pitch of 44 Å and is rather prism-like (ii,iii). The A-RNA duplex (d) is like B-DNA in having a large pitch (36 Å) and bases that are nearly perpendicular to the helix axis, but conformationally and morphologically it is an A-type structure with C3'-endo puckered sugar rings and with one deep and one shallow groove (ii,iii); in these respects, it resembles the original Watson and Crick model that was intended to be the solution to the B-DNA diffraction pattern.
crucial decisions about improving preliminary models. Bunn et al. [9], working on penicillin in the 1940s, pioneered this kind of adventurous use of Fourier syntheses and thereby were able to convert a provisional, erroneous model for penicillin into what turned out to be the chemically accurate one. In later work with fibre data, my colleagues and I successfully used this approach to determine the presence and precise site of a second chain in hyaluronate [10], to decide that A-RNA was an 11-fold rather than a 10-fold helix [11], and that the Hoogsteen alternative base-pairing was an unacceptable option for B-DNA [12]. Furberg’s model, adjusted to comply with the new and more accurate experimental observations, could have been used as the provisional structure of the polynucleotide chains in A-DNA, and a process of refinement could have been attempted as early as 1950.

**Simplifying complexity**

A crystal model with spirals of nucleotides would have been a bold start, but the next challenge that could not be evaded, and should not have been postponed, was to explain how a relatively simple, highly symmetric structure could accommodate four different kinds of nucleotide in any sequence. This kind of challenge was as old as chemical crystallography itself. It first presented itself in the field of mineralogy, where it was found that many materials of different, complex chemical compositions had the same, simple crystal form. This phenomenon, termed ‘isomorphism’, arises because many minerals have simple lattice structures, which comprise silicate most often but phosphate sometimes! The lattices form cages that are not very discriminating in terms of content – provided that it is the right size and shape.

The related phenomenon of clathrate formation was a source of excitement to organic chemical crystallographers in the late 1940s, when DNA structural studies were being resurrected. Clathrates, so named by Palin and Powell [13] at Oxford in 1945, can occur when substances such as hydroquinone or urea are crystalized in the presence of suitable guest molecules. When these molecules are an appropriate size and shape to fill holes in the simple, but highly hydrogen-bonded, lattices of the host materials, they get taken up to form complexes that are crystallographically isomorphous with the host substance.

Exploiting the first clean diffraction pattern of oriented and polycrystalline DNA by using an updated Furberg model and the methods of chemical crystallographers was an option that could have been pursued as early as 1950. But perhaps it would have been possible only in the right environment? Did the research unit of the KCL Medical Research Council have too many physicists and not enough chemists? Or was this option not pursued because the primitive state of computing at the time required chemical crystallographers to do a great deal of hard pounding on adding machines, and at KCL there were too many generals and not enough gunners to do such hard pounding?

Eventually, in 1952, a graduate student (Gosling) and a new postdoctoral fellow (Franklin) did embark on intensive computation, but their efforts were concentrated on Patterson transforms that involved only diffraction intensities and not phases [14]. This mathematically more sophisticated way of looking at the intensities could do little more than confirm the heuristic interpretation of the A-DNA diffraction that had already been made. In the end, the results would still have needed to be given a concrete form in a model; and the final model would need to explain how a diverse sequence of bases with different sizes and shapes could be incorporated into a simple repetitious structure.

**B-DNA overtakes A-DNA**

Meanwhile, it had become possible to get clean diffraction patterns of the B-DNA allomorph [14,15]. To begin with, the fibres of B-DNA were only uniaxially oriented; that is, the DNA molecules were lined up parallel to one another but not otherwise ordered. Microcrystals combined with uniaxial orientation (Figure 1b[i]) came only later when the subtle interplay of retained salts and water content on the crystallinity of DNA fibres was worked out. In addition, it was initially thought that B-DNA was the wet and more ‘natural’ allomorph, whereas A-DNA was merely an ‘artefact’ of drying and thus less worthy of further investment. Even then, this idea was obviously an oversimplification.

Gosling, who worked with both Wilkins and Franklin, scrupulously describes in his PhD thesis (R.G. Gosling, PhD thesis, University of London, 1953) three kinds of DNA sample: in one, the B-form was the only form seen in damped fibres; in another, only the A-form was found; and in a third, the A-form, B-form or both forms could be obtained, depending on the ambient relative humidity. We now know that A-form crystals are as highly hydrated as B-form crystals, that it is the excess retained salt that shuts down the possibility of a B-to-A transition, and that the absence of retained salt helps to preserve the A-form even at the highest relative humidity. With intermediate levels of retained salt, it is possible to switch between A-DNA and B-DNA by changing the ambient relative humidity.

There was a good reason to concentrate attention on B-DNA, but it was not the lack of hydration in A-DNA. A more attractive property of the B-DNA molecules was their diffraction signature. The diffraction pattern of the B allomorph is strikingly similar to the elementary textbook pattern provided by a simple helix of atoms; thus, however mistakenly it might have been argued that A-DNA molecules could not be helices, it would have been perverse not to accept that the most likely conformation of B-DNA molecules was indeed a helix. In private, even Franklin [16] had come to regard B-DNA as helical, despite having earlier ostentatiously buried this hypothesis for ‘crystalline’ (i.e. A-form) DNA. Certainly, Watson and Crick were inspired to resume DNA modelling within the constraints set by the B-form pattern: a ten-fold screw symmetry, a 34-Å pitch, and nucleotides stacked 3.4-Å apart with their bases approximately perpendicular to the helix axis.

In a tubular cage comprising antiparallel sugar–phosphate bars, hydrogen-bonded A•T, T•A, G•C and

www.sciencedirect.com
C•G pairs could vicariously replace one another because they were isomorphous – in other words, the same size, a similar shape, and containing glycosidic links in the same symmetrical positions.

**Furberg’s ghost**

The Watson and Crick [3] paper announcing their essential conclusions about the general structure of DNA did not fail to acknowledge, albeit dismissively, Furberg’s earlier work. But implicit in the detailed coordinates of every nucleotide of their B-DNA model [4] was the ghost of Furberg’s structure of cytidine with its puckered deoxyribose rings. Five-membered sugar rings are not planar, and ribose rings in particular have two equally common, puckered forms that have come to be known as C3′-endo and C2′-endo. The first is found in A-DNA and the second in B-DNA. Ironically, ribose had the C3′-endo conformation in Furberg’s cytidine crystals; thus, what Watson and Crick had unwittingly done was to produce a model of a credible DNA double helix, but it was an A-form double helix and therefore not an accurate model of their intended target, B-DNA.

For most biologists this did not matter: the Watson–Crick conjecture was accurate enough and its base-pairing component was immediately stimulating and productive. The fact that there was some connection between the model and X-ray diffraction experiments was reassuring, but the appropriateness of the connection was almost irrelevant. Yet the full ratification by diffraction scientists that other researchers thought was needed had to be withheld for the very good reason that the misfit of the diffraction amplitudes calculated from the model was of the same order as would be expected from a completely wrong structure.

This discrepancy had come about because the switch in sugar pucker, which might seem a modest local flaw in a model, gets amplified into much larger shifts in the relative positions and orientation of phosphates and bases. The consequent morphological changes in a double helix are large in terms of atomic positions, and the diffraction signals that derive from them are symmetrical positions.

---

**Concluding remarks**

The relentless pursuit of the best (X-ray) structure for B-DNA was helped by technical improvements, but changes in analytical tactics were long delayed. The process of refining this structure could be described as a slow transition from one kind of sugar ring conformation to another, but throughout the 1950s, and well into the 1960s, the refinement was never overtly recognized as such. Attention was concentrated on the bases and phosphate groups – that is, the main X-ray scatterers – the positions of which were being slowly adjusted to give a better fit with the X-ray intensities observed. The correlation between base and phosphate positions and the precise puckering of the sugar rings was noted only when linked-atom refinement was introduced and the conformation angles at every nucleotide bond had to be stated explicitly [17].

In addition, the repeated rounds of manual model building and calculation of the consequent diffraction were never short-circuited by the real-space approach of examining the provisional images that could be calculated by combining observed diffraction amplitudes with provisional model phases. Interestingly, the original, physically flawed model of Watson and Crick would have provided phases good enough to show that the base pairs in their model needed to be moved ~5-A closer to the helix axis (Figure 2).

Although the meticulous process of refinement maintained interest in DNA diffraction and its most competent practitioner, Wilkins, its downside was that it provided plenty of time for others to speculate on alternative structures. If DNA could be polymorphic, then so could its models. Thus, there arose speculative models incorporating other patterns of base-pairing, different chirality, and even ladder-like structures with no chirality at all! Indeed, there are real examples of many of these conjectured structures: non-Watson–Crick pairs are found in four-stranded poly(G) [19]; Hoogsteen pairs [20] and Watson–Crick pairs [21,22] are present in triple-stranded DNAs; left-handed but Watson–Crick-paired helices can be obtained with poly(d(C))poly(d(G)) [23]; and an achiral DNA duplex has been even observed in a drug complex [24]. In fact, most of the architects of the ‘hopeful monsters’ were not interested in polynucleotide polymorphism, but in toppling the prevailing model of B-DNA and its steady improvements. Nor were they interested in investing time and energy themselves in detailed diffraction tests of their speculations; this task was left to the
No fibrous polynucleotide structure produced in a laboratory not of Wilkins’ school has survived critical re-examination: this fact might not say much for the quality of other laboratories, but it says a great deal for the meticulous standards of fibre diffraction analysis that he initiated and promoted with such important consequences for molecular biology.

Acknowledgements
Currently, I am indebted for the provision of facilities to John Squire (Head of the Biological Structure and Function Section, Imperial College, London, UK, and Editor of Fibre Diffraction Review) and to Stephen Neidle (Director of the CRUK Biomolecular Structure Group, University of London School of Pharmacy, London, UK).

References
1 Watson, J.D. (1968) The Double Helix: A Personal Account of the Discovery of the Structure of DNA, Athenaeum
11 Arnott, S. et al. (1966) Molecular and crystal structures of double-helical RNA. III. An 11-fold molecular model and comparison of the agreement between the observed and calculated three-dimensional diffraction data for 10 and 11-fold models. J. Mol. Biol. 27, 535–548
17 Arnott, S. et al. (1968) Molecular conformations and structural transitions of RNA complementary helices and their possible biological significance. Nature 220, 561–564

Conscientious professionalism of Wilkins and his helpers. Responding to the continuous sniping might have been tiresome, but it surely contributed to the evolution of more robust methods of fibre diffraction analysis.

Figure 2. Fourier syntheses with low-resolution X-ray data obtained from crystalline B-DNA. Data are from Figure 1b(i). Blue contours indicate positive electron density on a plane through a base pair; red contours indicate negative density. Spheres indicate atoms that would contribute electron density in this plane. (a) The result obtained if the calculation is done with the observed X-ray amplitudes and phases calculated from the original, physically flawed, Watson-Crick model of B-DNA. A broad plateau of electron density includes the base pairs of the phasing model but spreads out beyond it. (b) Subtracting the density of the model at the same resolution reveals the bases of the phasing model in a deep negative trough and indicates that a better model should place them on the ridge of positive density at a distance of 5 Å. (c) When the model is revised as indicated in (b), the new calculated phases combined with the observed amplitudes give a more acceptable fit between model and image.