A-DNA and B-DNA: Comparing their Historical X-Ray Fibre Diffraction Images.

Amand A. Lucas

Facultés Universitaires Notre-Dame de la Paix, 61 rue de Bruxelles, B5000 Namur, Belgium

e-mail : amand.lucas@fundp.ac.be

Abstract.

A-DNA and B-DNA are two quite distinct molecular conformations which double stranded DNA drawn into a fibre can assume, depending on the water content of the fibre. The corresponding historic X-ray fibre diffraction diagrams, measured respectively by Wilkins and Gosling and by Franklin and Gosling in the early 1950's, played an equally crucial role in the discovery of the double helical structure of the molecule by Watson and Crick in 1953. The purpose of the present paper is to provide a didactic and comparative explanation of the structural content of the two diagrams treated on the same footing. This will be accomplished by two methods. The first, intended for science students and professionals, relies on a basic introduction to helical diffraction theory. The second, accessible to a much wider readership, will make use of the optical transform method by which both A and B X-ray images can be simulated optically. The simulations are conducted with a simple laser pointer and a dozen optical diffraction gratings, all held on a single diffraction slide. The gratings have been specially designed to pinpoint just which of the structural element of the molecule is responsible for each of the revealing features of the fibre diffraction images.

Keywords

audience :general public, continuing education

Domain : chemical education research, history, public understanding

Pedagogy *:* hands-on learning, manipulative

Topic *:* biophysical chemistry, conformational analysis, X-ray crystallography

A-DNA and B-DNA: Comparing their Historical X-Ray Fibre Diffraction Images.

Amand A. Lucas

Facultés Universitaires Notre-Dame de la Paix, 61 rue de Bruxelles, B5000 Namur, Belgium

e-mail : amand.lucas@fundp.ac.be

I. Introduction.

The birth of molecular biology in the early 1950's was hastened by a few great contributions in *structural biology*, such as the conception of the protein α-helix by Linus Pauling and of course the discovery of the double helix structure of DNA by James Watson and Francis Crick. The latter, which revealed at once the molecular structure of the genes and the mechanism of their replication [1], caused a complete revolution in all life sciences.

As is well known, the physical method of *X-ray diffraction* has played a key role in these defining moments of science history. In the present paper I propose a *didactic description and interpretation* of the original *X-ray diffraction patterns of DNA*. These historic photos deserve to be explained anew in a language more technically informative than the mere lip service paid to the images in history books or in the abundant popular literature. Yet I will largely avoid the specific jargon of X-ray crystallography which has often kept the understanding of the famous diffraction patterns out of reach of non specialists.

After a brief historical introduction (section II), the paper explains some of the fundamentals of *X-ray scattering* (section III), in particular the all-important *concept of layer line* (section IV) used to analyse the *fibre diffraction patterns* of filamentous biopolymers.

In section V, I review the first major contribution to the validation of the *helical molecular model* conceived by Linus Pauling for proteins [2]. This contribution came in the form of a theoretical paper by William Cochran, Francis Crick and Vladimir Vand (CCV) [3] predicting the way a regular *monoatomic helix* should scatter X-rays. The mathematical and physical elements of this important paper can be explained in simple terms and will greatly help us to

understand the X-ray images of DNA. However the use of the CCV theory for its originally intended purpose [3], namely for the interpretation of X-ray diffraction by protein α-helices, is a bit too advanced and will not be pursued in the present didactic paper (a recent review [4] covers this and further applications to other helical molecules).

In section VI, the most important historical *X-ray pictures of DNA fibres* and their meaning for the structure of the molecule are described in detail on the basis of the understanding provided by the CCV theory. For the first time, I discuss *comparatively the diffraction patterns of both the A-DNA and B-DNA conformations* which are two distinct DNA structures reversibly adopted by the molecule depending on the water content of the fibre. Discussing simultaneously the A and B patterns is important since, as will be shown, the two images revealed complementary information about the DNA structure. In fact they proved equally important for marking out the intellectual path followed by Crick and Watson towards the discovery of the double helix. To my knowledge the A and B patterns have never been explained *on the same footing*, neither in the popular nor in the didactic literatures.

Finally, in section VII, the structural content of the historic DNA X-ray pictures will be analysed once more, this time *without any reference to the CCV mathematical theory*, but with the help of straightforward *optical diffraction experiments* with which one can simulate X-ray diffraction entirely visually. The optical diffraction gratings of the simulations have been designed to build up progressively a complete understanding of all the key features of the historic X-ray fibre patterns. Here again we discuss in parallel the real and simulated diffraction patterns of *both A-DNA and B-DNA*.

The presentation should be suitable for a course in diffraction physics or/and in structural biology and the teaching level and pace can be adjusted from introductory to advanced. The present paper completes an earlier publication [5] devoted exclusively to B-DNA, the conformation believed to be prevalent in the living cell.

II. A Brief Historical Background.

Biological organisms involve giant carbonaceous compounds or *macromolecules*. The concept of large polymers based on carbon is a 19th century idea originating from Kekulé's theory of the tetra valence of this element. However it is only in the 1920's that organic substances of high molecular weight were conceived as being made of *long, covalently linked, flexible hydrocarbon chains*, a conception especially defended by Herman Staudinger [6] who coined the very word "macromolecule" for such long linear chains [6].

At about the time it was proposed, the new macromolecular concept became testable with the powerful physical technique of *X-ray diffraction*. Introduced by von Laue and the Braggs in the mid 1910's, the method was quickly developed to explore the various possible states of aggregation of organic as well as inorganic matter at the atomic scale. Broadly speaking the diffracting sample could be presented to the X-ray beam in essentially one of three forms depending on its degree of internal order: i) as a large *single crystal* which produces a pattern of sharp diffraction spots, ii) as a *powder of randomly oriented micro crystals,* giving discrete diffraction rings or iii) as a *fibre* either of axially oriented mini crystals (crystallites) or of *parallel macromolecular chains,* giving a pattern somewhat intermediate between the previous two. Only the latter case will be considered in the present paper since *fibre diffraction* was the technique historically used for the early structural studies of fibrous proteins, DNA and other long chain polymers.

When filamentous macromolecules are packed in a fibre along a fixed direction, the X-ray intensities diffracted by the fibre fall on the observation screen along approximately straight and equidistant lines, the so-called *layer lines*, perpendicular to that direction. This important concept, introduced by Michael Polanyi already in 1921 [7] for the X-ray study of cellulose, will be explained in section IV.

In the 1930's, while Staudinger's macromolecular concept was gaining wider acceptance, the question arose as to the possible *state of aggregation* or *spatial mode of folding* of the flexible polymer chains in their natural, functional forms. For example, which folding mode could best explain the ease with which some of the natural fibres, such as wool or silk, could be *reversibly stretched* to a very large extent? Was it a planar zigzag or spiral, a random coil, a helical fold, ...? William Astbury and co-workers [8] made pioneering attempts at investigating this question by X-ray diffraction, in particular for the keratins of animal hair. But although specific models were proposed, the rather poorly resolved X-ray fibre data could not make the actual folding mode absolutely certain [6].

On another front of the study of macromolecular chains, in the mid-1930's, Pauling and Mirsky [9] proposed i) that the flexible amino acid backbone of *globular proteins or enzymes* in their native states does not fold randomly but meanders in space along an invariable and reproducible path specific to each protein; and ii) that this unique, so-called tertiary structure is stabilized against denaturation, that is disordering, by *multiple intra molecular hydrogen bonds* between CO and NH groups of the folded protein backbone. However, this brilliant insight could not be confirmed until much later, when quantitative data were accumulated, mainly by X-ray crystallography, on the stereochemistry of individual amino acids, small peptides and other simple molecules related to proteins [10].

DNA, also a flexible filamentous polymer, was first studied by Astbury and Bell [11] in the 1930's by X-ray fibre diffraction. But again, although specific models were proposed, the true structure could not be deduced from the original, poorly resolved X-ray patterns and the discovery of the actual folding mode had to wait for information on the detailed stereo chemical arrangement of the elementary nucleic acid monomers.

After the lull of the Second World War, great advances in the knowledge of macromolecular structures were accomplished in quick succession, in particular by Pauling and his school of

structural biology in the USA and by several British schools of X-ray crystallography [12]. By the early 1950's, X-ray diffraction data of greatly improved resolution were obtained for polymers of both amino acids and nucleic acids. Spurred by Pauling and Corey's theoretical conception of the α-helix conformation of proteins [2], Cochran, Crick and Vand [3], published an extraordinary fruitful mathematical paper which predicted the diffraction pattern of a regular monoatomic helix and used this model to interpret the X-ray diffraction produced by crystalline fibres of artificial polypeptide chains. It is difficult to overestimate the importance of this calculation for, in addition to its usefulness for understanding helical proteins, this particular piece of mathematical physics, in the hands of Crick and Watson, was soon to prove decisive for their momentous discovery of the DNA double helix structure in 1953 [1].

Since the heroic beginnings summarized above in the application of X-ray crystallography to the basic helical molecules of biology, an immense progress has been achieved in the last five decades in the determination of the structure of biological macromolecules of ever increasing complexity. While X-ray diffraction alone used to dominate the field, beginning in the late 1950's and in the 1960's, electron diffraction combined with high-resolution electron microscopy [13] progressively grew in importance (in parallel with the increasing involvement of computers). Spectacular examples in the application of these tools have been the determination, at atomic-resolution, of the structures of large helical assemblies such as microtubules, bacterial flagellar filaments, helical viruses, muscle fibres, receptor pores of the cell membrane, and many other complex systems of the living world.

III. Atomic Scattering Factors.

In this section and in the next one, I present a few elementary facts about X-ray diffraction. A more detailed exposition is given in ref. [4]. There are numerous books devoted to the

foundations of X-ray crystallography but here I will cover only the bare, self contained minimum necessary to understand the X-ray fibre diagrams of DNA.

In the kinematical (first Born) approximation which is adequate for describing the weak process of X-ray diffraction, the differential cross section for scattering from the initial k_i to the final k_f wave vectors by a collection of atoms at nuclei positions r_j is the square modulus of a complex amplitude A(q):

$$|A(q)|^2 = \left| \sum_j f_j(q) e^{iq.r_j} \right|^2 \qquad (1)$$

where $q = k_f - k_i$ is the scattering wave vector and where $f_j(q)$ is the Fourier transform of the electron density of atom j, a scalar quantity called the atomic form factor. The behaviour of f(q) vs q for a few light elements is shown in Fig.1.

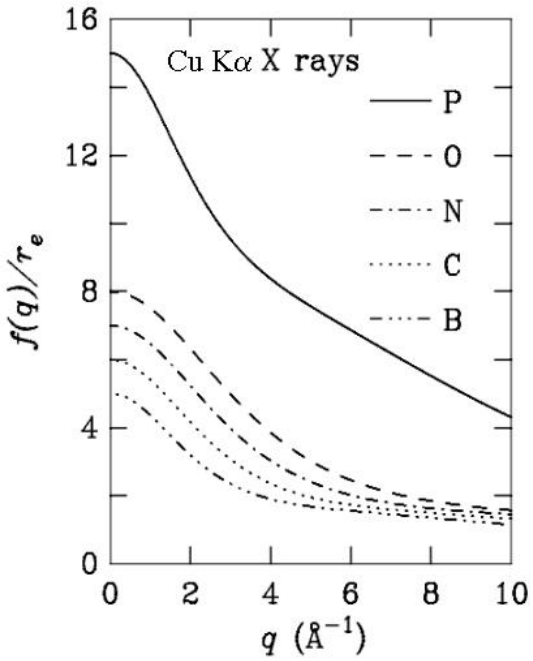


Fig.1

Fig.1 *The atomic scattering factors of a few elements for Cu K_α X-rays (0.15 nm wavelength) in units of the classical electron radius $r_e = 2.8\ 10^{-15}$ m.*

The form factor peaks in the forward direction (q = 0) and, in units of the classical electron radius $r_e = e^2/4\pi\varepsilon_0 mc^2 = 2.8\ 10^{-15}$ m., f(0) coincides with the atomic number Z of the element.

From Fig.1, we note that while Hydrogen (Z = 1) scatters hardly at all, the *"life elements"* Carbon (6), Nitrogen (7) and Oxygen (8), making up the amino acid backbones of proteins as well as much of DNA, scatter X-ray with comparable strengths. However, Phosphorus (15) which is present in the phosphate group of the DNA backbone, scatters about four times more intensely than the other atoms, a feature which, as we shall see, turned out to be crucial for the interpretation of the X-ray images of a DNA fibre.

IV. Layer Lines.

For the interpretation of X-ray diagrams of fibres, an important unifying structural concept was first introduced by Michael Polanyi [7], namely that of *layer-planes* and *layer-lines* in reciprocal space. The concept is most easily explained with a simple model frequently used in the standard teaching of elementary diffraction theory.

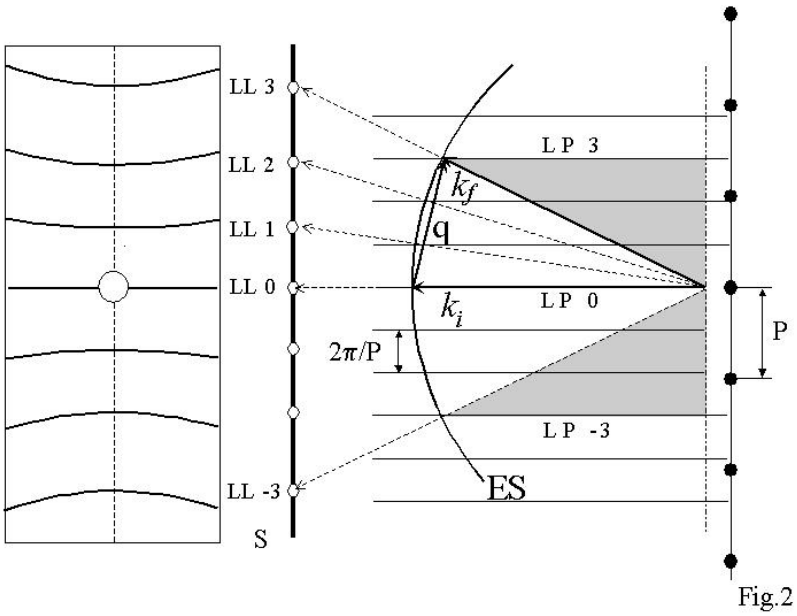


Fig.2

Fig.2 *Principle of the geometrical construction of the hyperbolic layer lines in the X-ray diffraction pattern of a linear array of point scatterers (dark circles) with period P. ES: Ewald Sphere; LP: Layer Planes; S: observation Screen; LL: Layer Lines. The shaded regions are cones whose intersections by the screen generate the layer lines.*

Let us first consider (Fig.2) a *single linear, regular "polymer"* consisting of identical "atoms" or point scatterers occupying equidistant positions z_j=jP on the z-axis, where j is an integer and P is the polymer repeat period. For normal incidence, the scattering amplitude in Eq. (1) reduces to

$$A(q) = f(q)\sum_j e^{iq_z jP} = f(q)\frac{2\pi}{P}\sum_{l=-\infty}^{+\infty}\delta(q_z - l\frac{2\pi}{P}) \qquad (2)$$

where we have made use of the identity (the fundamental relation for diffraction by ordered structures):

$$\sum_{n=-\infty}^{+\infty} e^{2\pi inx} = \sum_{m=-\infty}^{+\infty}\delta(x - m). \qquad (3)$$

If the reader has forgotten how this relation arises, it will suffice to remember that the exponential in the left-hand side is represented by a point on the unit circle in Gauss plane of complex numbers. *If x is not an integer*, the summation over n produces an infinite number of such points distributed evenly over the circle; so by judiciously grouping the points (in diametrically opposed pairs or as apexes of regular polygons), *the sum is seen to vanish*. On the other hand, *if x is any integer m*, all terms on the left are equal to 1 and *the sum is then infinite*. These italicised statements are just what is meant by the delta function on the right-hand side.

The vertical component q_z of the transfer wave vector is therefore quantized to integer multiples of $2\pi/P$. Eq. (2) expresses the familiar fact that the Fourier transform of a vertical set of points separated by P is a set of horizontal planes separated by $2\pi/P$ in reciprocal space. These planes of equation $q_z = 2\pi l/P$ are called *layer planes* and are labelled by the integer l.

Now *for elastic scattering*, relevant for the diffraction of monochromatic X-rays, the allowed wave vector k_f of the scattered X-ray, in addition to lying on one of the layer planes, must also lie on the so-called *Ewald sphere*, that is a sphere in reciprocal space of radius $|k_f| = |k_i|$. The intersections of the latter by the former are latitudes on the sphere and constitute the bases of cones with a common apex at the sphere centre (Fig.2). These conical surfaces

intersect the observation screen along the loci of nonzero diffracted intensity. Such conical sections are *hyperbolae* called *layer lines* (see the front view of the screen S in Fig.2). Due to the finite radius of the Ewald sphere, the distance between the layer lines is chirped, that is it increases towards higher-l (Fig.2). Along each layer-line the diffracted intensity is governed by the square atomic form factor $|f(q)|^2$ which, for a point scatterer, is a monotonously decreasing function of q (see Fig.1).

More generally, the diffraction pattern (at normal incidence; see ref. [4] for oblique incidence) of any linear periodic polymer whose repeated monomer may have any 3-D shape whatsoever and may contain any number of atoms, will also be organized in hyperbolic layer-lines similar to those of the monoatomic example discussed above. Suppose indeed that the j^{th} monomer has its atoms at positions $r_{j\mu} = r_\mu + jP$. One immediately sees from Eq. (1) that the scattering amplitude is obtained by simply replacing f(q) by the *monomer form factor* $F(q)=\Sigma_\mu f_\mu(q)\exp(iq.r_\mu)$ in Eq. (2). The sharp layer-line organization subsists. $|F(q)|^2$ governs the intensity along the layer-line which is now distributed into a pattern of maxima and minima reflecting the interferences of the waves scattered by the monomer atomic content.

Consider now a *fibre* containing a large number N of such identical parallel polymers at "positions" ρ in the fibre (this symbolic ρ might include a rotation around the polymer axis and an axial translation within the repeat period, i.e. a screw operation). The diffraction amplitude of the fibre will be obtained by multiplying the scattering amplitude A(q) of the ρ'th polymer by a phase factor exp(iq.ρ) and summing over all the polymers

$$\Sigma = \sum_\rho A(q) e^{iq.\rho} \qquad (4)$$

If the fibre is a *gel, sometimes called a "paracrystal"* (parallel molecules but *disordered* distribution of ρ's), due to the randomness of the phases q.ρ, the intensity $|\Sigma|^2$ will reduce approximately to $N<|A(q)|^2>$ where the pointed brackets indicate averaging over the angular

orientation of the molecules around their axis. *So for a disordered, gel-like fibre the diffraction pattern is representative of a single, angularly averaged molecule*.

At the other extreme, that is if the fibre is *fully crystalline* (parallel molecules with positions ρ on a 2-D lattice) the sum over ρ in Eq. (4), on account of a two-dimensional version of Eq. (2), amounts to a sum of delta functions at the nodes of the 2-D reciprocal lattice. In this case the intensity along the layer-lines is broken up into such sharp, discrete spots as fall onto the Ewald sphere. *So for a single-crystal fibre the overall pattern is just that of an ordinary 3-D crystal*. The positions of the spots are determined by Bragg's law and depend on the fibre angular orientation around its axis.

In practice however, given that even a very thin, 10 μm diameter fibre may contains as many as 10^8 parallel polymers, it is highly unlikely that it would crystallize into a single crystal. Instead the fibre will usually comprise *very many crystallites* all aligned with a single axis, here the polymer axis, parallel to the fibre direction but with otherwise *random angular orientation around that fixed direction*. The diffraction pattern is then similar to the so-called rotation diagram of a single crystal, that is the pattern obtained by rotating a crystal, while being under the X-ray beam, a full 360° around one axis. So for a fibre of singly oriented crystallites, all possible Bragg spots consistent with the Ewald sphere construction are observed along the layer lines, independent of the fibre angular orientation.

V. Diffraction by a Monoatomic Helix.

In 1952*,* Cochran, Crick and Vand (CCV) [3] developed an analytical theory for X-ray diffraction by a monoatomic helix. This work was intended to help with the interpretation of X-ray images measured by Bamford *et al* [15] in 1951 on crystalline fibres of artificial polypeptides made of identical amino acids. If the polypeptide has a helical conformation, as was conjectured by Pauling *et al* [2] just about that time, it can be considered as a collection of

monoatomic helices sharing a common axis, one helix for every distinct atom in the amino acid monomer . Then, from Eq. (1), the total diffraction amplitude of the complete protein is the sum of the amplitudes of its individual atomic helices. The immediate interest of the CCV theory was to give a transparent, analytical expression of these amplitudes at a time, in 1952, when computers, if at all available, were barely capable of a brute force calculation of the total diffraction intensity from its definition in Eq. (1). In addition, the concepts developed in the CCV paper proved to be of the utmost importance for the then imminent discovery of the DNA double helix in early 1953: commenting on the eminent role of the theory in the search for the structure of DNA, Crick himself exclaimed *"it did mean that I had the expertise at my fingertips"* [12].

The authors of the CCV theory [3] arrived at their result by a highly original, albeit rather lengthy derivation. Apparently they were not aware of a mathematical formula of the 19th century, the so called *Jacobi-Anger expansion*, which allows to obtain the CCV result in just two lines. The formula gives the Fourier development of a plane wave in cylindrical waves as follows

$$e^{iq.r} = \sum_n e^{in(\psi_q + \pi/2)} J_n(q_\perp r) e^{-in\varphi} e^{iq_z z} \qquad (5)$$

where (r,φ,z) and $(q_\perp,\psi_q,q_z)$ are, respectively, the cylindrical coordinates of r and q around a vertical axis and where the $J_n(x)$ are the regular cylindrical Bessel functions of integer order n. Deriving this formula is a standard exercise in Fourier series (see note [14]). Now, as shown on Fig.3, the cylindrical coordinates $\mathbf{r}_j = (\rho_j, \varphi_j, z_j)$ of atoms on a regular circular helix of radius r, period P and axial atomic repeat p_a are $\rho_j = r$, $\varphi_j = \varphi_0 + 2\pi(z_j - z_0)/P$ and $z_j = z_0 + jp_a$, where (r, φ_0, z_0) are the coordinates of an atom taken as origin. Introducing these coordinates into Eq. (1) in which the plane waves are replaced by Eq. (5) and using Eq. (3) once more, one obtains the CCV result :

$$A(q) = \frac{2\pi}{P} f(q) e^{iq_z z_0} \sum_{n,m} J_n(q_\perp r)\, e^{in(\psi_q - \varphi_0 + \pi/2)}\, \delta(q_z - n\frac{2\pi}{P} - m\frac{2\pi}{p_a}) \qquad (6)$$

The atomic helix need not be an integer helix, that is P is not necessarily an integer multiple of p_a. For instance Pauling α–helix has 3.6 amino acid residues per period, meaning 18 residues over 5 helix turns. Let P_a designates the true axial translation repeat of the polymer. Then, according to the previous section, layer lines must appear at $q_z = 2\pi l/P_a$ where l is integer. The diffraction amplitude along a layer line of given order l is governed by the double summation of Eq.(6) where the couples of allowed integers (n,m) must be selected to satisfy the *l^{th} layer line "selection rule"* of the delta function:

$$\frac{l}{P_a} = \frac{n}{P} + \frac{m}{p_a} \qquad (7)$$

Note that the intensity, the square modulus of Eq. (6), is independent of z_0 but does depend on φ_0, that is the diffraction pattern changes as one rotates the helix around its axis.

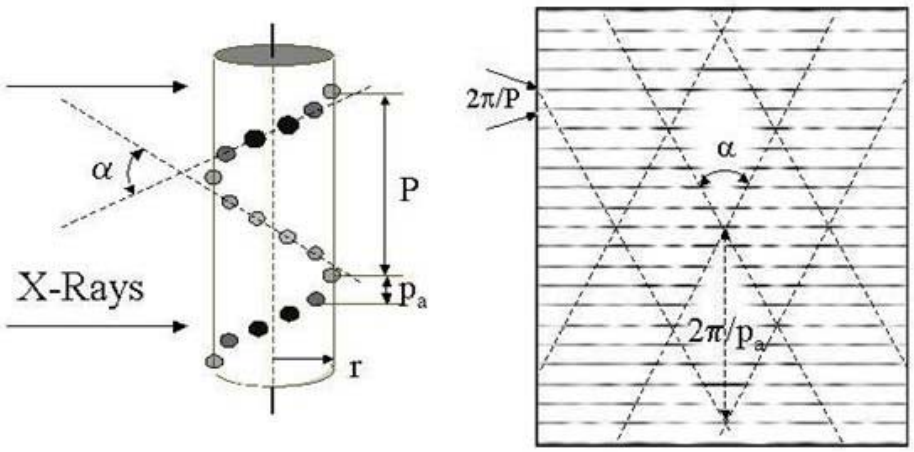


Fig. 3

Fig.3 *Computer simulated pattern of X-ray diffraction by a phosphorus helix of period P, radius r and atomic repeat p_a (see Table I for the leading Bessel contributions). Notice the central St Andrew's cross, the Bessel oscillations along the layer lines and the diamond pattern (highlighted by broken lines) with nearly empty North and South diamonds. The structural parameters of the helix (P, r, p_a) are readable from the geometrical parameters of the pattern ($2\pi/P$, α, $2\pi/p_a$).*

Fig.3 shows an important example. It gives the calculated CCV intensity, averaged over φ_0 (as in a gel fibre), of diffracted Cu Kα X-rays (1.5 nm wavelength) by an integer helix whose parameters are those of the Phosphorus helix in a single strand of B-DNA, namely $P_a = P = 10p_a = 3.4$ nm. The selection rule is then simply $l = n + 10m$ which allows one to construct Table I.

l	0	1	2	3	4	5	6	7	8	9	10
n	0 -10	1 -9	2 -8	3 -7	4 -6	5, -5	-4 6	-3 7	-2 8	-1 9	0 10
m	0 1	0 1	0 1	0 1	0 1	0, 1	1 0	1 0	1 0	1 0	1 0
$\mathbf{J_n}$	$\mathbf{J_0}$ J_{-10}	$\mathbf{J_1}$ J_{-9}	$\mathbf{J_2}$ J_{-8}	$\mathbf{J_3}$ J_{-7}	$\mathbf{J_4}$ J_{-6}	$\mathbf{J_5}$, $\mathbf{J_{-5}}$	$\mathbf{J_{-4}}$ J_6	$\mathbf{J_{-3}}$ J_7	$\mathbf{J_{-2}}$ J_8	$\mathbf{J_{-1}}$ J_9	$\mathbf{J_0}$ J_{10}

Table I. *Line 1 of the Table lists the l = 0 to l = 10 layer lines of the computed diffraction pattern of the Phosphorus helix of a single strand DNA (see Fig.3). Lines 2 and 3 lists the corresponding values of the integers n and m selected by the "selection rule" l = n + 10m and entering the CCV amplitude formula of Eq.(6) in the text. Line 4 lists the two leading Bessel terms (dominant in bold) contributing to the CCV amplitudes of the corresponding layer lines.*

To understand from Table I the distribution of intensity in Fig.3, the crucial properties of the Bessel functions (see pictures of Bessel functions in any calculus textbook or on Internet) are i) the oscillating nature of $J_n(x)$ for increasing x, where $x = q_\perp r$ is here a measure of the distance from the meridian along the layer line; ii) the fact that *$J_n(x)$ remains small from x = 0 up to x ≈ n*; and iii) $J_{-n} = (-1)^n J_n$. Therefore, as one progresses from the equatorial layer line l = 0 upward to the fifth layer line l = 5 (or downward to the l = -5 layer line), the first and strongest maximum of the oscillating intensity $|J_n(x)|^2$ moves outward from x = 0 towards x ≈ 5. Then from the l = ±5 to the l = ±10 layer lines, the first maximum of $|J_n(x)|^2$ marches back from x ≈ 5 towards the meridian x = 0 again (Table I). From l = ±10 to l = ±15, the first maximum move outward again, etc…. Hence the set of first (and strongest) maxima of the Bessel intensities *define a kind of Saint Andrew cross folded back every 5 layer lines*, as indicated in Fig.3 by the broken lines.

From reading the value of the meridian angle α between the arms of the central cross ($\alpha \approx 60°$ in Fig.3) it is possible to obtain an estimate of the helix radius r. Indeed, due to the reciprocal relationship between $q_\perp$ and r expressed by the dimensionless argument $x = q_\perp r$ of the Bessel functions, it is easy to understand that the arms of the cross in reciprocal space are about perpendicular to the slope of the helix in real space (Fig.3). From this, the desired relationship is found to be $r = (P/2\pi)\text{cotan}(\alpha/2)$ [5].

The folded and repeated Saint Andrew crosses create a diamond pattern throughout the diffraction image (Fig.3). *The meridian diamonds are mostly empty of intensity* due to the above mentioned property ii) of the Bessel functions (one exception in Fig.3 is the strong meridian intensity on the 8^{th} layer line, a further effect of the finite radius of the Ewald sphere which has been explained in [4]). This spectacular feature, along with the central cross itself, were clearly visible in the experimental X-ray pattern of B-DNA (see section VI, Fig.4b; a B-DNA molecular model is shown in Fig.7). This lead Crick and Watson, who had seen that unpublished pattern and who were familiar with the CCV theory, to become strongly convinced *that DNA was indeed helical*, and absolutely indispensable piece of information to build a mechanical model of the molecule. Furthermore, the *Phosphate groups of atoms* (i.e. the strongest X-ray scatterers in the molecule, see Fig.1) to which the central cross of the experimental image was assigned *were likely to be at the periphery of the molecule rather than inside it*. Indeed in a polyatomic helical molecule such as DNA, the helices of each type of atom or group of atoms produce approximately additive Saint Andrew crosses and *the helices of smaller radii produce more open crosses, i.e. of larger meridian angles*. Therefore all inner atomic helices of the molecule contribute intensities only within the east and west angles of the central cross arising from the most exterior helix, namely the Phosphate helix in DNA. This important point will be illustrated again in the optical simulation experiments of section VII.

VI. Comparing the A-DNA and B-DNA Patterns.

In late 1951, Rosalind Franklin and her student-collaborator Raymond Gosling at King's College, London made a crucial observation, namely that the X-ray diffraction pattern produced by a DNA fibre could be made to change reversibly from one type to another, very different type by changing the ambient humidity in the X-ray chamber. Changing the humidity resulted in removing or adding water molecules to the fibre [12]. Fig.4a,b shows the two historic patterns for a drier or a wetter fibre containing about 7 to 9 water molecules per nucleotide, respectively [16].

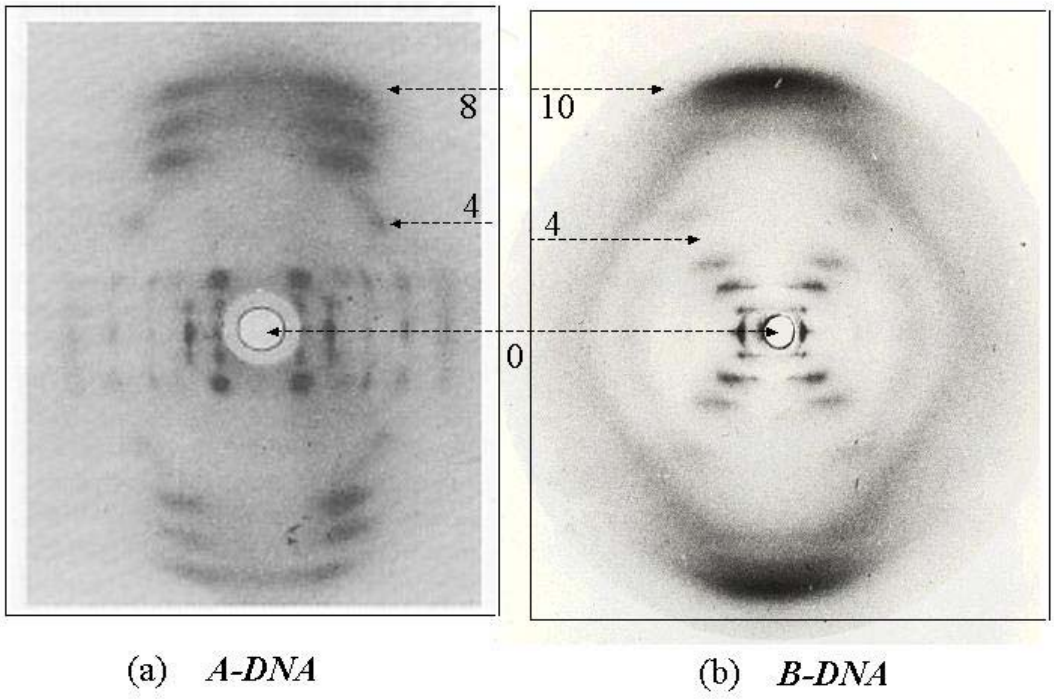


Fig.4

Fig.4 *The historic X-ray fibre diffraction patterns (a) of A-DNA and (b) of B-DNA plotted on the same scale. The 8th layer line of pattern (a) occurs at about the 10th layer line of pattern (b) reflecting a 20% increase in the DNA helix period. The A-DNA pattern shows crystalline spots on the first few layer lines.*

The two pictures are drawn on the same scale for comparison. Franklin and Gosling realised that these patterns reflected two distinct conformations of the molecule which they christened A-DNA and B-DNA, respectively. Today B-DNA is thought to be the normal resting state of

chromosomal DNA in the high-humidity condition of living cells, while the occurrence in vivo of the A-DNA conformation is less clear, although a form similar to it is known to be adopted by the double helical portions of t-RNA molecules and by the double-stranded RNA genome of reoviruses.

First consider the A-type pattern of Fig.4a. Images similar to this one had already been obtained previously, but not published in print, by Maurice Wilkins and Gosling at King [17]. On account of the 2-D lattice of relatively sharp, discrete spots observed near the centre of the picture along the first few layer lines (Fig.4a), these researchers had correctly surmised that *the DNA molecules were arranged in crystalline order in the fibre*. Crystallisation implied that the DNA molecules must have a definite, regular structure susceptible of determination by classical crystallographic methods. The possibility of such a crystalline state of DNA greatly excited Watson when he saw the A-pattern publicly shown by Wilkins at a conference in Naples in 1951 [12]. This serendipitous historical circumstance caused Watson to change his postdoctoral program and to move to Cambridge, England where he could learn elements of X-ray crystallography from Bragg's group. That group included Crick who, with Watson, ended up discovering the double helix thanks to information provided not only by A-type patterns re-measured by Franklin and Gosling but also by the entirely novel B-type [16] newly observed by the King group.

Second consider the B-pattern of Fig.4b. In the high-humidity B-type state, the extra water molecules must have invaded the space between the DNA molecules, freeing them from being locked into crystallites. This separated the molecules and *destroyed their regular positional as well as angular order,* causing the crystalline spots of A to disappear. In B only broad streaks remain along the layer lines, characteristic of X-ray scattering by an angularly averaged molecule (see the discussion in section IV).

The organization in regularly spaced layer lines of the patterns in Fig.4, means that the filamentous molecule has a periodically repeating structural unit in the fibre direction. *The layer line separation reveals the value of the period*. The spectacular reversible change in the type and distribution of spots on going from A to B (Fig.4) was taken as a manifestation not only of the perturbation of the long range spatial order just discussed but also of *a reversible change of internal structure of the DNA molecules* themselves. In particular the large decrease of over 20% in the layer line spacing on going from A to B (Fig.4) implied an increase of about as much in the polymer repeat period: P = 2.8 nm for A-DNA and P = 3.4 nm for B-DNA.

In earlier, pioneering X-ray works on DNA fibres by Astbury [11], the big North and South meridian arcs seen in the B pattern had already been observed (the arcing is due partly to the deviation from parallelism of the molecules and partly to the deviation from horizontality of the bases; see below). Astbury's patterns were much more confused than either one of Fig.4 because, as it turned out, his fibres were in a mixed A and B state. But they did show the big meridian smears which he correctly ascribed to *a nucleotide repeat of 0.34 nm along a filamentous molecule*. The newly observed, pure B pattern in Fig.4b confirms this number since it has ten layer line intervals separating the centre from the North smear, which implies that the molecule has ten repeating units within one period of 3.4 nm. It turned out that the thick arcs were produced by the coherent scattering of X-rays by the equidistant, nearly horizontal flat bases separated by 0.34 nm (a molecular model of B-DNA showing the stack of bases is depicted in Fig.7 which will be discussed further later). Although the base sequence is random and the bases themselves are quite variable in both their geometry and chemical composition, *their flat electron density is "seen" edge on by the X-ray beam,* and so they act as a set of narrow equidistant "slits" (the word "slit" will be used here and later in this paper by virtue of the fact that the diffraction by a narrow slit in a plain material is equivalent to that of a narrow material slab in vacuum). On the basis of the 0.34 nm meridian arc and of the

(erroneous) assumption that the pentose sugar is coplanar with the base to which it is attached in a nucleotide (in reality the two chemical groups turned out to be perpendicular to each other), Astbury proposed a single strand, straight model of DNA in which the flat horizontal base-sugar groups are merely piled up vertically on top of each other, like "a pile of pennies", separated and covalently linked by the phosphate groups. This early model was quite interesting but lacked any compelling functional value.

Note that *the A pattern lacks the big meridian smears* just discussed in the B image, implying that the bases of A-DNA are not at all horizontal. A further immediately apparent, big difference between the two patterns *is the presence near the centre of a Saint Andrew cross in B and its absence in A*. We now endeavour to explain these differences.

As already stated, pattern B reflects the scattering by individual molecules. So the cross must be produced by the regularly repeated part of the molecule, namely by its sugar-phosphate backbone. In his popular autobiographic novel [18], Watson reported having been dumb struck upon seeing pattern B for the first time, because for him that cross screamed "helix". Indeed, having learned from Crick the predictions of helical diffraction theory (e.g. for his earlier X-ray studies of a helical virus), Watson identified the central cross as being the Saint Andrew cross expected from a helical molecule. Furthermore the B image showed a diamond pattern which not only meant helix but could have revealed at once to Watson and Crick (if only in a confused way since, after all, they did not know the structure yet) the three fundamental dimensions of the phosphate helix: its pitch P (3.4 nm, from the layer line interval), its radius r (1 nm, from the meridian angle of the cross) and its nucleotide repeat p_a (0.34 nm, from the vertical diagonal of the meridian diamonds). The absence of intensity in the meridian diamonds and the large radius r indicated that the backbone helix was likely to be at the periphery of the molecule (see the discussion in section V and the optical simulations

below) and the bases inside, the inverse of unsuccessful models constructed earlier by Watson and Crick [12] and also by Pauling and Corey [19] shortly before the double helix discovery.

The A spot pattern (Fig.4a) did not show a recognizable central Saint Andrew cross. Although, like the B-pattern, it lacked intensity in the meridian areas, it exhibited no other so obvious clue as to the helical form of the molecule. Franklin took this and other evidences that the A-form molecule was not helical after all. And since the crystalline spot pattern was much more detailed and hence promising for working out the molecular structure by traditional X-ray crystallographic methods, she temporarily put pattern B aside, ignoring or forgetting its strong helical message, and decided to concentrate exclusively on a quantitative interpretation of the A picture. It is this decision which, with the benefit of hindsight, is said to have cost her the great discovery [12],[17].

Amongst the most significant and decisive information that she and Gosling produced by analysing A-type patterns was the determination of the space group of the crystallographic unit cell in which A-DNA crystallizes. Thanks to highly resolved spot patterns [16] similar to the sharp image of Fig.5 (notably the observed splitting of all spots) [22], it was found that the A-DNA crystal lattice belongs to the *C-face centred monoclinic space group.*

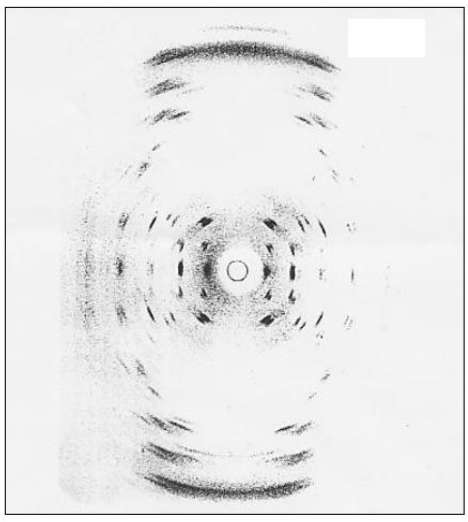


Fig.5

Fig.5 *High resolution X-ray fibre diffraction patterns of A-DNA showing the splitting of crystalline spots associated with the monoclinic symmetry of the crystallites [16].*

This means that the unit cell has monoclinic symmetry ($a \neq b \neq c$; $\alpha = \gamma = 90°$, $\beta = 97°$) and contains two molecules, one running along the slightly off-vertical c-axis, and the other, parallel one threading through the centre of the ab face, as depicted in Fig.6.

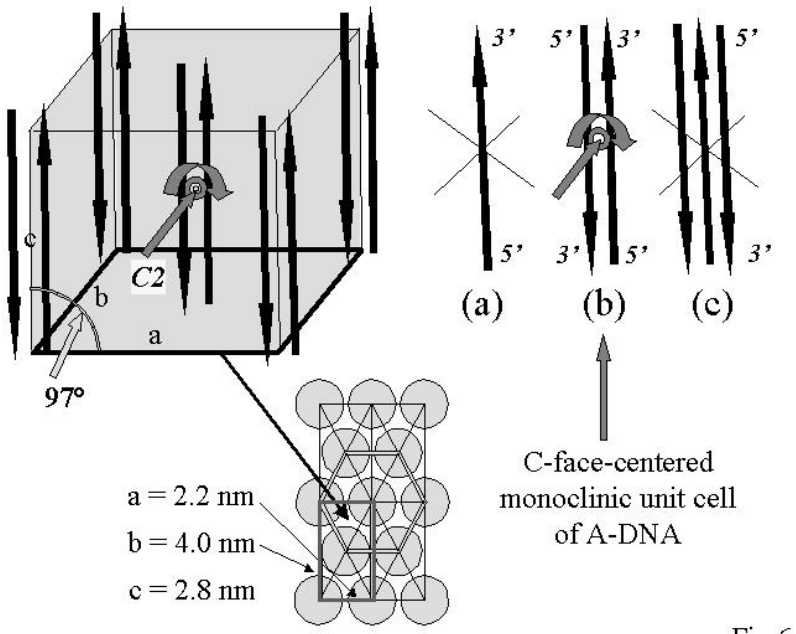


Fig.6

Fig.6 *The C-face-centred monoclinic unit cell of crystallites in a A-DNA crystal fibre [16]. The monoclinic cell, which is close to a hexagonal structure (one hexagon in the a-b face is highlighted), has the dimensions indicated [22]. The C2 symmetry axis demands that the DNA molecule must itself be dyadic and therefore have two anti parallel polar strands (b). The dyad excludes the possibility of a single strand (a) as well as a triple strand (c).*

In effect the structure was quite close to a hexagonal arrangement of long rod-like molecules (Fig.6), packed like a bunch of pencils. The deviations from exact hexagonal symmetry (in particular the c-axis not being perpendicular to the ab face) arose from interactions between the groves of the twisted backbones of neighbouring molecules brought closer to each other by the partial removal of water from the fibre.

That *C2 monoclinic crystallographic fact* also came to Crick's knowledge prior to publication although he hadn't seen the picture itself. It implied an absolutely crucial

consequence that, apparently, only Crick realized at the time [12], [20]. He reasoned that because the unit cell possesses a *C2 dyad axis* (that is a 180° rotation symmetry axis) perpendicular to the ac face (Fig.6), *its content*, namely *the DNA molecules themselves, must also have a twofold symmetry axis perpendicular to the backbone axis*. Otherwise the unit cell as a whole would not be invariant under the C2 symmetry of the monoclinic space group. But the sugar-phosphate backbone of nucleic acids was known to be an *oriented polymer*, due notably to the covalent 3'-5' asymmetric attachments of each pentose to its two phosphate neighbours. Fig.7 shows a model of the familiar double strand helix where the polar nature of each backbone is clearly indicated.

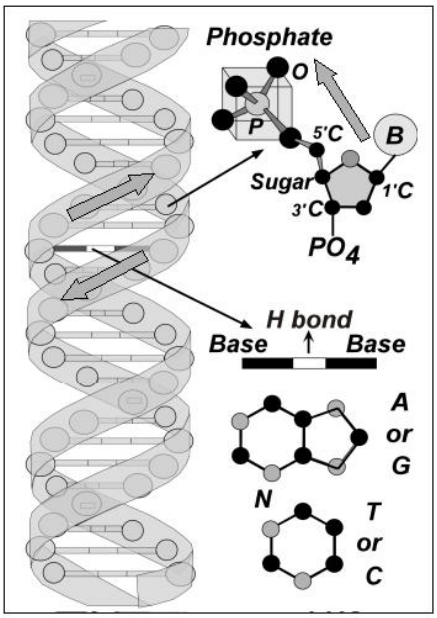


Fig.7

Fig.7 *A schematic representation of the DNA double helix in its B conformation. Only the heterogeneous carbon-nitrogen rings of the flat bases (lower right) are shown, not their fringing chemical groups. A detailed representation of the 3'-5' oriented sugar-phosphate backbone is shown (upper right). The arrows indicate the counter orientation of the strands.*

The only possible way, Crick concluded, for having a dyadic DNA molecule with such polar strands was to assume that *it possessed two counter oriented strands* (Fig.6). The dyad operation would merely convert one strand of the duplex into the other. *One or three strands could be excluded as not being C2 invariant*. Four strands per molecule could also be C2

invariant but were *incompatible with the results of density measurements*, while the two-strand model was fully consistent with such data. Of course, for such reasoning to hold, based as it was on the observed X-ray diffraction pattern, it was necessary to assume that the disordered sequence of inner bases (as well as the solvated metal ion background of the salt) could be safely ignored as producing only diffuse scattering (except for the big off-meridian smears at high-l layer lines, to be discussed below).

So a spacious cylindrical cage made of two counter oriented, intertwined helical sugar-phosphate strands was built by Watson and Crick in a renewed attempt at model building [12]. *A helical structure with two anti parallel backbones*: that was the crucial intermediate step towards arriving at the final double helix. Inside that cage, Watson tried to install pairs of the four planar bases A, T, G, C. The base pairs were set about perpendicular to the cage axis, as demanded by the big meridian smears of the B-pattern, and were made to interact via hydrogen bonds across the duplex axis (Fig.7). Such hydrogen linkages were known to form between pure bases or related chemicals in their condensed phase. Exploring the limited set of six possible couples, Watson stumbled on the A-T, G-C pairings as the only combinations with the correct hydrogen bonding arrangement and with nearly identical overall sizes which could be made to fit exactly between the rigid outer helical backbones. That clinching discovery, the keystone of the model building, rapidly lead to the completion of the beautiful and definitive DNA double helix.

The so-called Watson-Crick A-T, G-C base pairs instantly elucidated the earlier biochemical analysis of Chargaff [21] according to which, irrespective of the source of DNA, the concentrations of the four bases satisfy the rules [A]/[T] = [G]/[C] = 1. In addition to revealing *the molecular structure of the genes and the static mode of their conservation in double copies of complementary base sequences*, the Watson-Crick model suggested at once a possible

dynamical molecular mechanism for gene replication, by strand separation and complementary copying of each strand [1].

I now complete the qualitative description of the A pattern in comparison to the B. Upon drying the fibre, as noted before, the double helix period shrinks by about 20%, changing from 3.4 nm in B-DNA to 2.8 nm in A-DNA (Fig.4). Franklin and Gosling [16] and later workers [22] further determined that in A-DNA the number of nucleotides per helical period is close to 11 per strand, instead of 10 in B-DNA. Removing water thus compresses the molecules along their axis. But because of repulsion between nearest neighbour bases, this forces all of them to become inclined over the horizontal plane. Later X-ray studies [22] revealed a tilt angle as large as 20° from the horizontal (as will be shown in Fig.10 in the next section). *So the bases no longer act as horizontal electron slabs seen edge-on by the X-ray beam; this explains why the big meridian smears, so strong in the B pattern, have disappeared altogether from the A pattern*. The effect of the 20° tilt is to produce instead the two sets of strong arcs on the 6^{th}, 7^{th} and 8^{th} layer lines on each side of the meridian axis (Fig.4a and Fig.5). This can be understood by noting that only *one or two base pairs* in every half turn of the helix continue to be seen by the beam approximately as edge-on slabs, not horizontal but inclined at 20° from the horizontal. These two base pairs of each half helix turn form a kind of see-through slot (see Fig.10 in the next section) which scatter X-ray strongly and coherently while the *remaining seven base pairs per turn are seen partly face on* at varying angles and cause diffuse scattering. In effect what the smears represent is *the vestige of a Saint Andrew cross* created by the diffraction by these edge-on base pairs making up a kind of zigzagging, double slit diffraction grating (see Fig.11 in the next section). The reason why in Fig.4a and Fig.5 the Saint Andrew cross appears only along the high-l layer lines is because along low-l layer lines the waves scattered by the two adjacent base pairs of each couple interfere destructively. This will be demonstrated below with optical simulations.

VII. Optical Simulations.

In this section I follow the methodology developed in a previous publication for B-DNA [5] and extend it to A-DNA. It consists in performing optical diffraction experiments to simulate the real X-ray DNA images in order to explain their structural content. *Here no recourse to helical diffraction theory will be necessary*: a diffraction slide and a simple laser pointer will suffice to do the experiments which are accessible to any teaching level, from high school to graduate studies. The simulations reveal how the powerful X-ray diffraction method contributed to the great discovery of the double helix (the slide is available on demand from the author; see acknowledgements).

I have designed simplified DNA planar models to be used as motifs for a new set of 12 optical gratings, all held on a single slide. The motifs are conceived to pinpoint one by one the diffraction effects created by each of the structural elements of the molecule [5],[24]. A photo of the slide content is shown in Fig.8 [24] where the gratings are numbered from P1 to P12.

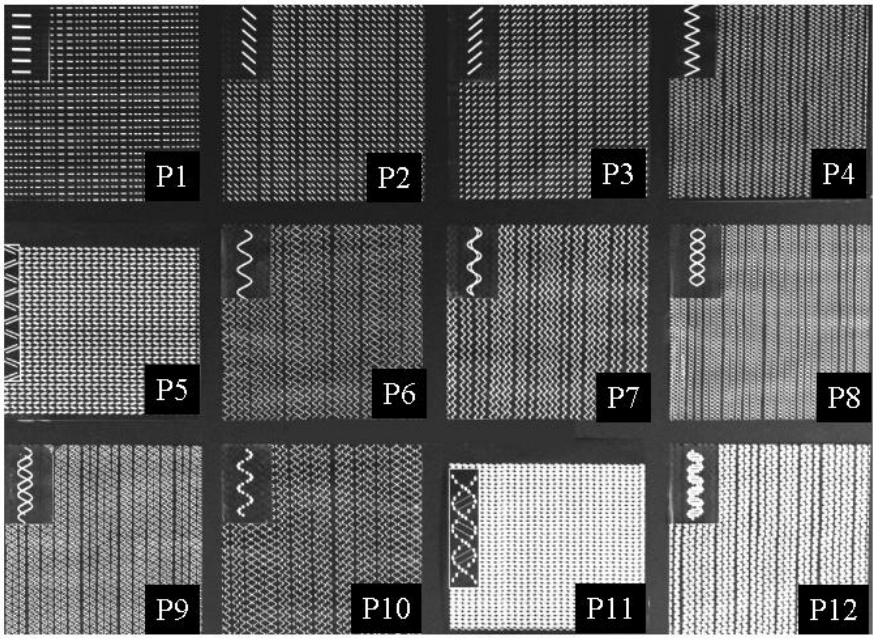


Fig.8

Fig.8 *A photograph of the diffraction slide used in this paper for the optical simulations of X-ray diffraction by A- and B-DNA. The slide comprises 12 diffraction panels in 3 lines of 4 columns numbered P1 to P12. The motifs of the panels are shown inset and in expanded form in Figures P1 to P12.*

The motif of each grating is shown as an inset and can be easily examined on the slide itself with a magnifier. The narrow beam of a good laser pointer held in one hand is passed through each successive grating of the slide held in the other hand, and the scattered light is projected on a distant white screen or wall. The simulations are best performed in a dimmed class room although, with a good laser, they work even in broad day light when the pattern is made visible by projecting it in a darker corner of the room. The simulation can be conducted with a red laser pointer, as in ref. [5], or a green pointer which generally produces diffraction patterns of better visibility to the human eye. Eventually a laser, if convenient and instructive, may not be permitted in the class room for safety reasons. If the case arises I remind the reader that a laser is not indispensable at all to observe diffraction effects: the diffraction patterns can be observed vividly by placing the slide close to the eye and looking through the successive panels at a distant *point source of classical light* ***(never use the direct beam of a laser for that method of viewing)*** [5].

For B-DNA, the final, most elaborate motif of the slide (last panel, P12) is shown in Fig.9a in comparison to a standard molecular model in Fig.9b [23]

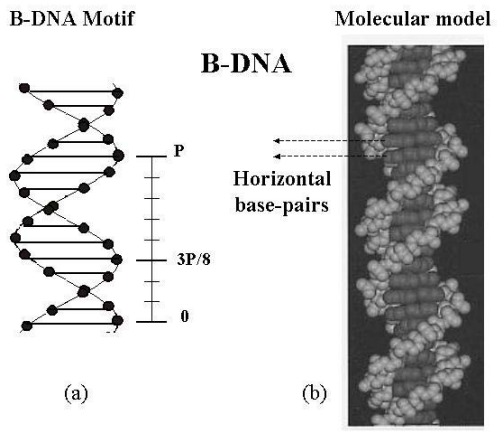
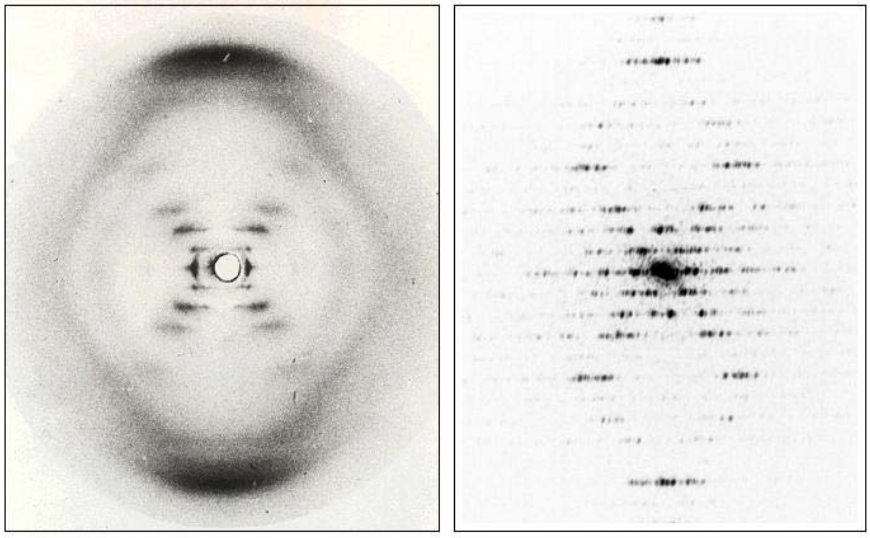


Fig.9

Fig.9 *(b): a molecular, space filling model of B-DNA. The backbones are in light grey and the bases in dark. The base pairs are perpendicular to the axis and are all seen edge on. (a): the diffraction motif used to simulate the B-DNA diffraction pattern at normal incidence in panel P12 of the slide. The backbone is represented by 10 black dots per period. The base pairs are represented as horizontal bars. The two sine backbones are separated by 3P/8.*

The motif consists of two coaxial sine waves (planar projection of a double helix perpendicular to its axis) each made of 10 discrete sugar-phosphate "atoms" per period P, axially separated by 3P/8, and connected by 10 equidistant, horizontal "base pairs" simulated by simple bars (electron slabs seen edge-on). For *gel* B-DNA the motifs in panel P12 are *disordered*, leading to broad diffraction streaks. In Fig.10, the optical transform of that motif (Fig.10b) is compared to the real B-DNA X-ray picture (Fig.10a).



(a) (b)

Fig.10

Fig.10 *(a) X-ray diffraction pattern of B-DNA (same as in Fig.4b). (b) optical diffraction simulation produced by panel P12 of the slide with the motif shown on Fig.9a.*

Similarly for A-DNA, the final motif of the slide (panel P11) is shown in Fig.11a along with a standard molecular model in Fig.11b [23]

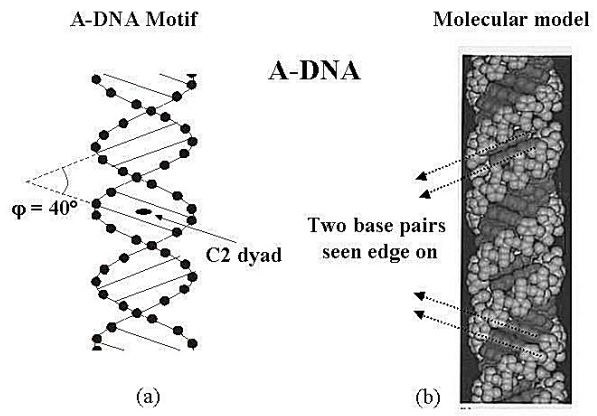
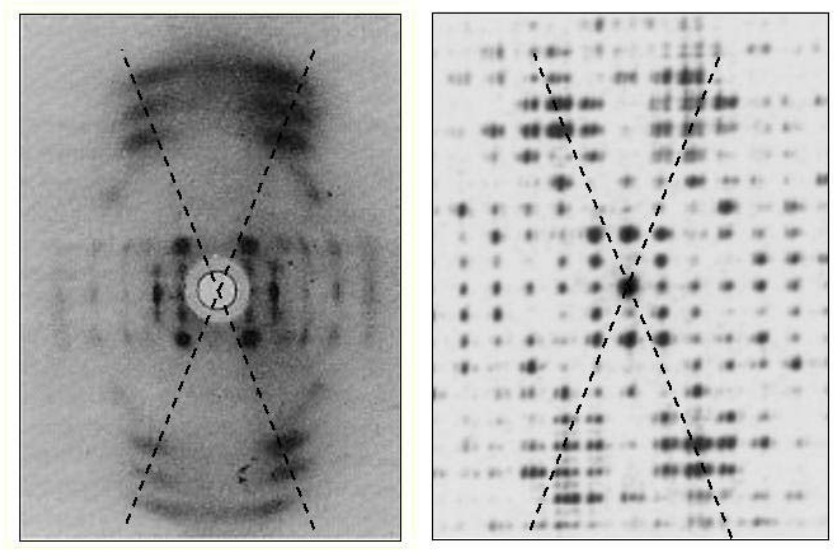


Fig.11

Fig.11 *(b) a molecular, space filling model of A-DNA. The backbones are in light grey and the bases in dark grey. Notice that a couple of base pairs define a see-through slot at each half period (dotted arrows). (a): the diffraction motif used to simulate the A-DNA diffraction pattern in panel P11 of the slide where the motifs are arranged on a 2-D lattice. The backbone is represented by 11 black dots per period. The base pairs are represented as bars inclined at 20° from the horizontal. Only the two base pairs seen edge on in each half-period are included in the model. The motif, as the molecule backbone, is C2 invariant.*

The motif now has 11 sugar-phosphate "atoms" per period of each strand but only 2 parallel base pairs for each half period of the double helix. The base pairs are inclined at 20° from the horizontal, alternatively clockwise or counter clockwise, and are meant to represent *the two and only two adjacent base pairs which are seen edge on in the molecular model (Fig.11b)*. This choice of motif ought to be clear after our discussion at the end of the previous section. Note that for *crystal* A-DNA the motifs in panel P11 have been repeated on a regular *2-D lattice* (a simple rectangular lattice) which leads to a spotty diffraction pattern. Fig.12 compares the optical transform of that motif (Fig.12b) to the real A-DNA X-ray picture (Fig.12a).



(a) (b)

Fig.12

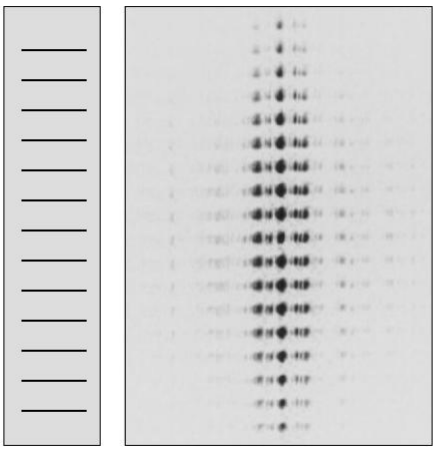
Fig.12 *(a) X-ray diffraction pattern of A-DNA (same as in Fig.4a). (b) optical diffraction simulation produced by panel P11 of the slide with the motif shown on Fig.11a. The broken lines in both patterns highlight the Saint Andrew cross produced by the zigzagging, edge-on base pairs of Fig.11. Note that the optical pattern in (b) is C2 invariant like the motif itself in Fig.11a.*

Although the spots and streaks of the X-ray patterns are not rendered in their actual positions and intensities, it is obvious that the optical transforms of the planar models do reproduce faithfully *the general appearance* of the X-ray images. This means that the models incorporate and successfully mimic the principal structural elements of the real double helix which are responsible for the observed features of the X-ray photos. In fact the models are so elaborate that their optical transforms are as complex as the real X-ray images. So much so that, if the simulations were limited at just presenting the spectacular similarity of the diffraction patterns, they would not bring any additional understanding beyond that already achieved by reference to the CCV helical diffraction theory (section V). However, even simplified as we have tried to present it, the CCV theory might still be beyond the grasp of mathematically unprepared audiences, e.g. high-school classes to which complex numbers, Fourier series, reciprocal space,

etc… have not yet been taught. For such cases, the simulations of panels P11 and P12 alone would not be of much help.

This is why I have devised the progressive optical simulation method, largely free of diffraction theory, based on the single diffraction slide described here. The 12 diffraction motifs and their optical transforms are presented and discussed below in the figures numbered P1 to P12.

P1. A set of equidistant horizontal bars, which we shall call "slits" for short, diffracts along a reciprocal set of equidistant horizontal lines.



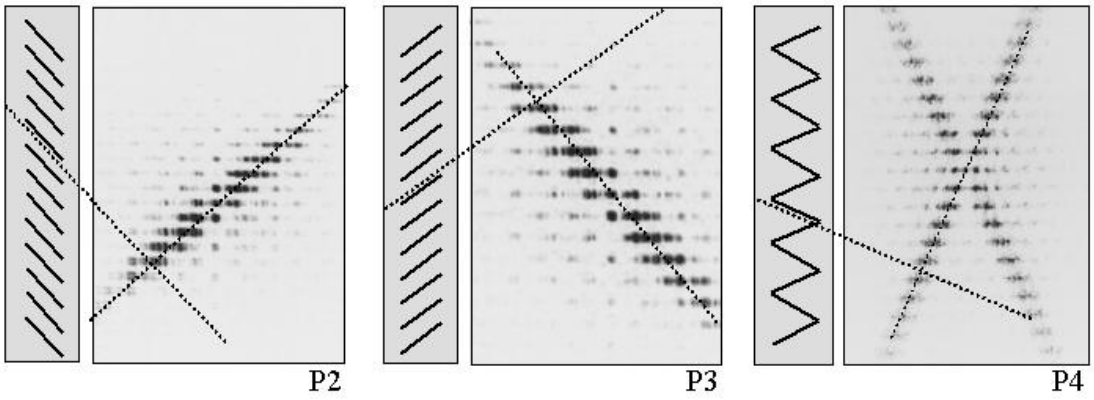
P1

P1. *A set of equidistant, horizontal slits diffracts light along regularly spaced layer lines. The band of maximal intensities is perpendicular to the slits.*

This introduces the concept of *layer lines in diffraction space*. All 12 panels of the slide diffract along horizontal layer lines since this is what one expects from any vertical *periodic "polymer"*. The layer lines are loci where light rays scattered from the slits interfere constructively, that is in angular directions α_n given by Thomas Young's diffraction law in transmission, $P\sin\alpha_n \approx P\alpha_n = n\lambda$ where n is integer and where λ is the laser wavelength [5]. The vertical repeat P of the motifs can be directly read from the layer line spacing S and the distance D from the slide to the observation screen, according to the formula $P = \lambda/\alpha_1 = \lambda D/S$.

It is important to *emphasize the reciprocity* of the relationship between P and S brought about by diffraction and to warn the student that the layer lines never result from straight, shadow-like projections of the slits.

P2, P3, P4. Here the slits are inclined one way or the other over the horizontal. *Introducing this slanting is the crucial pedagogical step* to understand the basic reason for the formation of Saint Andrew crosses common to all patterns from P4 onward, as well as to the X-ray images.



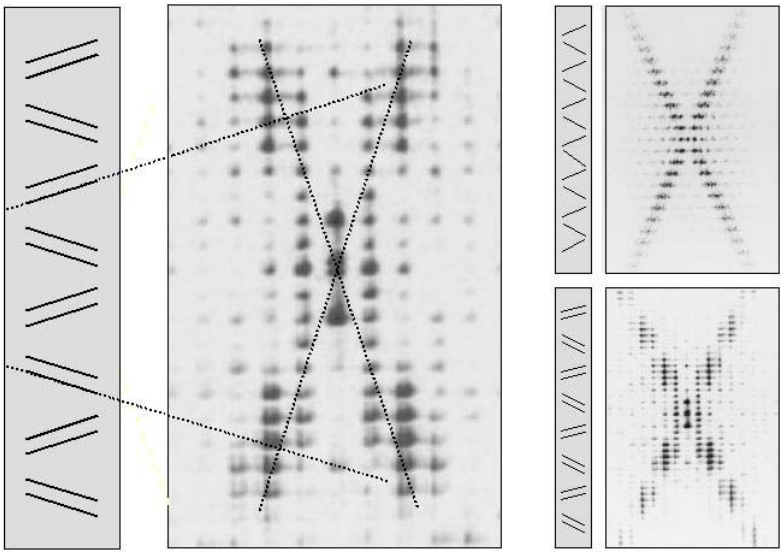
P2. *A set of equidistant slits rotated clockwise diffracts light along regularly spaced layer lines. The band of maximal intensities is perpendicular to the slits (thin dotted lines).*

P3. *A set of equidistant slits rotated anti clockwise diffracts light along regularly spaced layer lines. The band of maximal intensities is perpendicular to the slits (thin dotted lines).*

P4. *A zigzag motif diffracts light to form a Saint Andrew cross. The arms of the cross are perpendicular to the zigzagging slits (thin dotted lines).*

On symmetry ground alone, each individual slit in P2, P3, if it was isolated, would diffract a maximum of light in a direction perpendicular to it. This explains qualitatively why *the bands of maxima in P2, P3 and the arms of the cross in P4 are perpendicular to the zigzag slits*. *So the cross angle in P4 equals the zigzag angle*. The measurement of that angle combined with the knowledge of the period from the layer line spacing therefore reveals the zigzag amplitude of the motif.

P5. This panel shows the simulation with a *double zigzag motif* which P4 has prepared us to understand. The double slit stands for two base pairs seen edge on in A-DNA.



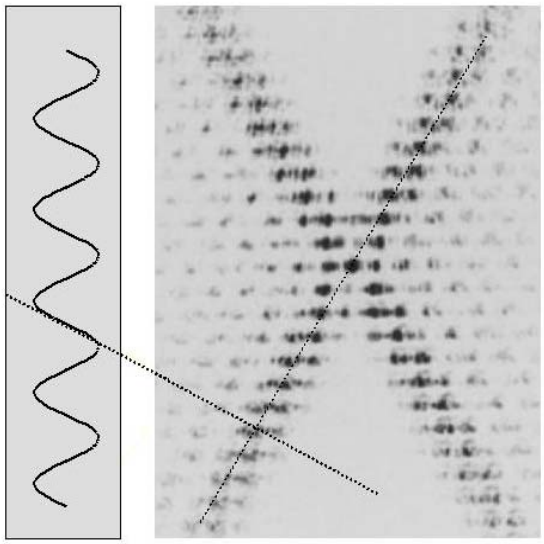
P5

P5. *A zigzag double slit motif diffracts light along regularly spaced layer lines. The motifs are arranged on a regular 2-D lattice on panel P5 of the slide, resulting in a spotty diffraction pattern. The maximal diffracted intensities define a Saint Andrew cross whose arms are perpendicular to the slits (thin dotted lines). The cross arms are clearly visible only at high-l layer lines (see text). The two reduced patterns on the right reproduce P4 (upper right) and P5 (lower right) for comparison. Notice the modulation of the arms of the Saint Andrew cross of the double slit motif (lower right).*

First note that in grating P5 the parallel motifs have been arranged on a regular 2-D lattice rather than in random horizontal and vertical positions. The spectacular result of the ordering is a *diffraction pattern of sharp spots* instead of broad streaks. However the intensity distribution over the spots delineates again a clear Saint Andrew cross. The angle of the cross arms (about 40° in P5) equals the angle between the zigzagging, edge on base pairs in A-DNA. In addition, the cross intensity is weak along the lower-l layer lines (3 to 5) and strong along the higher-l layer lines (6 to 8). The reason for this is that the intensity of the Saint Andrew cross is *modulated by the envelope of the Thomas Young diffraction of each double slit*. To illustrate this, the diffraction patterns for the single and the double slit motifs are reproduced again, in

reduced size, and compared in P5 (right-hand side of the figure). While the intensity along the arms of the Saint Andrew cross of the single slit pattern is seen to fade away uniformly towards high diffraction orders (due to the fading of the single slit form factor), one observes vividly that the fading intensity is modulated in the double slit pattern. That proves the basic origin of the large, off meridian intensities observed along the high-l layer lines of the A-DNA patterns (Fig.12), as announced in section VI.

P6. The optical transform of a single, continuous sine wave, which here represents the planar projection of a *continuous helix* [5], also shows a prominent Saint Andrew cross.

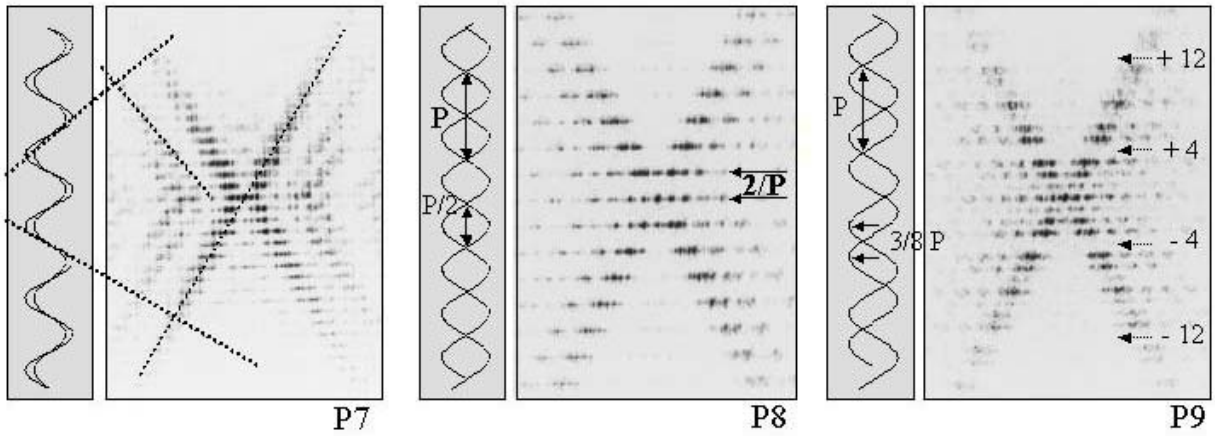


P6

P6. *Diffraction by a sinusoidal motif. The band of maximal intensities is perpendicular to the zigzagging slope of the sine wave (thin dotted lines). There is no intensity in the meridian angle of the cross.*

The arms of the cross are perpendicular to the zigzagging slopes of the sine function and whose opening angle thereby reveals the helix amplitude in relation to its period (see section V) [5]. For the B-DNA pattern the meridian angle of the central cross was 60° as in P6 and the resulting backbone radius was r = 1 nm.

P7, P8, P9. This set of panels simulate the way *two sine waves*, i.e. two continuous coaxial helices, behave in diffraction.



P7. *Two coaxial sinusoidal motifs with different amplitudes. The intensities diffracted by the small sine wave fall within the east and west angles of the Saint Andrew cross of the large sine wave. There is no intensity diffracted in the meridian angles of the large sine wave cross.*

P8. *A motif with two out-of-phase sine waves. Note that the true period of the motif is half the sine wave period. The layer line spacing is doubled as compared to the spacing of a single sine wave.*

P9. *A motif with two sine waves separated by 3P/8. Note the extinction of the 4th layer lines.*

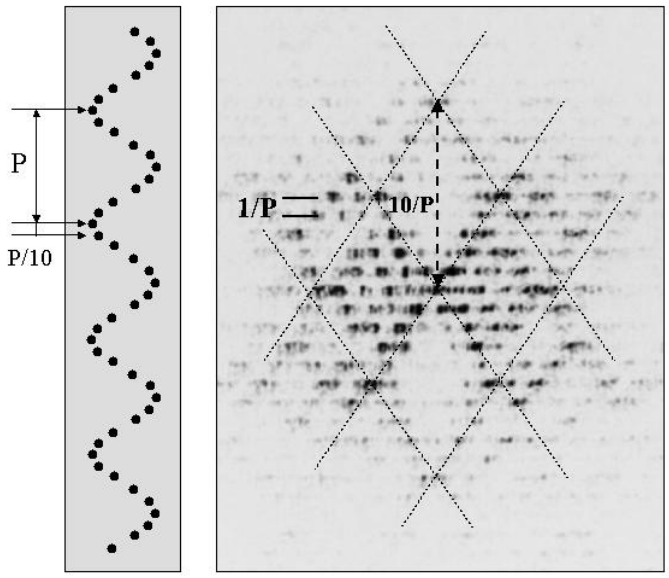
P7 represents *two coaxial helices of different radii*, such as the phosphate helix and the sugar helix in one DNA strand. One observes that the intensity of the Saint Andrew cross created by the small-radius helix falls exclusively within the east and west angles of the cross associated with the large-radius helix. This is consistent with the arms of the crosses being perpendicular to the slopes of their respective helices. The message here is that for a polyatomic helical molecule, there will be no intensity to fall in the meridian angles of the largest, i.e. most exterior atomic helix arising from any of the smaller, interior atomic helices. The scattering power of the phosphate helix in DNA being substantially stronger than that of the rest of the molecule, the spectacular cross of Franklin's X-ray B-DNA photograph (Fig.10a) was naturally attributed to the phosphate groups. Then the absence of intensity in the meridian angles of that cross established that *the phosphate helix must sit at the periphery of the molecule* and not inside it.

In P8 the *two identical sine waves have opposite phases* and represent two identical, coaxial helices *axially separated by half a period*. Such a model would be adequate, on symmetry ground again, to represent *two identical, co-oriented DNA strands*. This parallel conformation was indeed tried in model building by Watson and Crick a few weeks before their final, anti parallel solution, on the speculative but reasonable assumption that in replication, each strand could be copied identical to itself. But that configuration was invalidated for the following technical reason. As shown by the motif in P8, the *true axial period* of the planar model is in fact *half the period of each isolated sine wave*. This also holds true for the model double helix itself: the true axial period of the co-oriented double helix is one half that of each single strand, one helix taking the place of the other upon a translation by half a period (or a rotation by 180° around the vertical axis). Therefore the layer line spacing of the simulated pattern *is doubled that of a single strand* (compare P8 to P7) and likewise for the expected X-ray pattern of such a co-oriented double helix. In effect, all odd layer lines of a single strand are eliminated by destructive interferences between the radiation scattered by the two out-of-phase strands (see [25] for an algebraic demonstration), while all even layer line receive 4 times as much intensity. But because the observed B-DNA X-ray pattern (Fig.10) indicated that there must be 10 base pairs per true period of the double strand molecule, the model builders tried to construct a double co-oriented helix with 10 adjacent nucleotides within half the period or *half a turn* of each single strand helix. The angular separation between one nucleotide and the next would then have to be 18° only instead of 36° when the 10 nucleotides are distributed over a complete single helix turn. This turned out to be quite impossible to realize without severely violating the known stereo chemical constrains of the awkward sugar-phosphate units. The model was quickly abandoned for this and for other, unrelated reasons [12].

In P9, the motif has *two identical continuous sine wave axially shifted by 3/8 (or 5/8) of a period*. This model is meant to represent approximately the true situation in B-DNA in which

the two counter oriented helical strands must be separated in this uneven way in order to establish the hydrogen bonds between the opposite base pairs attached to the two backbones. The spectacular effect of this shift is to *extinguish the 4th layer line*. Ref.[5] gives a straightforward explanation of this extinction caused by destructive interferences between the waves scattered by the two backbones which arrive out of phase at all point of the 4th layer line (and the 12th, 20th, …) (see [25] for an algebraic demonstration). It should be mentioned that Crick and Watson were not aware of this highly technical and particularly revealing feature of the B-DNA pattern while building their model, whereas Franklin and Gosling did mention it when they finally publish their paper [26] and duly drew from it the correct conclusion on the 3P/8 shift.

P10. This shows the optical transform of a *monoatomic sine wave* simulating the Phosphorus helix of a single strand helical DNA.

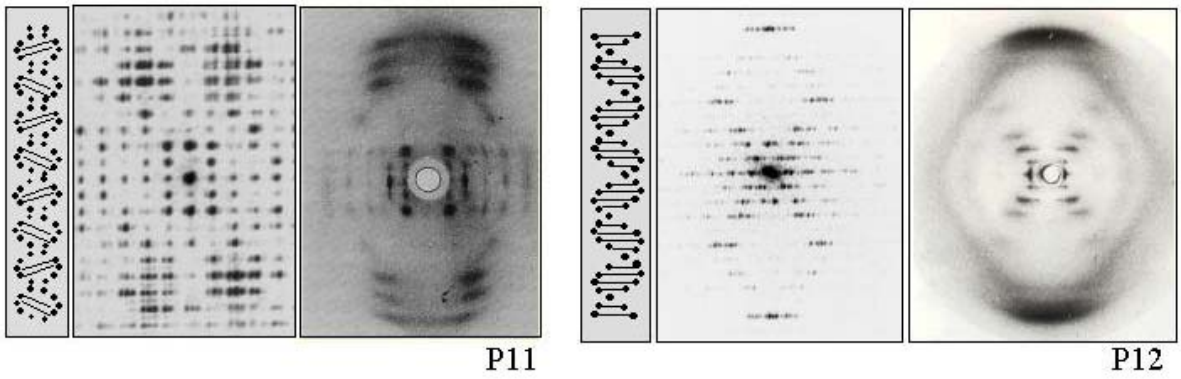


P10

P10. *Diffraction by an atomic sinusoidal motif. Note the diamond pattern (dotted lines) and the absence of intensity in the meridian diamonds.*

One observes vividly the *diamond pattern* caused by the atomic periodicity and the *emptiness of the meridian diamonds* in perfect agreement with the observation in P7 (and the CCV theory, see section V).

P11, P12. These optical transforms have already been discussed in Fig.12 and Fig.10 for A-DNA and B-DNA respectively.



P11. *Optical simulation of the X-ray diffraction by an A-DNA fibre. The motif on the left is arranged on a 2-D lattice in panel P11 of the slide. The strong features on the 6 to 8th layer lines of both the X-ray picture and the simulated pattern arise from the inclined base pairs seen edge on and forming a zigzagging double slit grating. (compare with P5).*

P12. *Optical simulation of the X-ray diffraction by a B-DNA fibre. The strong streaks on the 10th layer lines of both the X-ray picture and the simulated pattern arise from the horizontal base pairs seen edge on.*

The diffraction motifs of these panels are concatenations of the structural elements explored separately in the previous panels of the diffraction slide. Although these elements create intensities which are not strictly additive (only their amplitudes are), their individual influence on the general pattern is still clearly recognizable.

Conclusions.

I have explained the historic X-ray fibre diffraction patterns of A-DNA and B-DNA on an equal footing and have pointed out the origin of their prominent similarities and differences. The most essential *structural information* revealed jointly by the two A and B images can be

summarized as follows. The molecule backbone has a regular helical structure (pattern B); there are about 10 repeating units per period (B); the phosphate groups are at the periphery of the molecule (A and B); the backbone has two counter oriented polar strands (A); at high relative humidity of the fibre the inner bases are near a plane perpendicular to the helical axis (B), whereas at low relative humidity, the bases are inclined 20° away from this plane (A).

The functionally potent scheme discovered by Watson-Crick, namely the complementary base pairing A-T, G-C of the genetic letters, was not a structural information flowing directly from the X-ray fibre patterns which, fortunately, were rather insensitive to the randomness of the base sequence. But once the regular and rather rigid backbone structure had been revealed by such patterns, the pairing mechanism became a highly probable consequence of that structure. In turn, the enormous biological significance of base pairing strongly reinforced the likely correctness of the double helix model itself.

All these conclusions are of course far easier to state, now that the DNA double helix is known in exquisite details. What the present paper has achieved is a didactic exercise in reverse crystallography, that is explaining the observed diffraction images from the known molecular structure. Needless to say, the meaning of those images was far from being obvious for the researchers groping in the dark in their search for the molecular nature of the genes. In addition, it is important for the students to realize that the solution to the DNA structural problem did not rely solely on the X-ray diffraction data examined here but involved a great deal of information coming from many a scientific horizon [12].

In the abundant popular literature on the double helix, it is almost always the B-DNA pattern, (the famous photograph 51 [5],[27],[28]) which is emphasized as the key unpublished information which helped Crick and Watson along the path to their great discovery. The present paper tries to re-establish a more balanced view according to which the A-DNA pattern did play an equally crucial, if more subtle role for the discovery.

This paper offers two levels of explanations, one based on the beautiful helical diffraction theory [3] perhaps accessible only to more advanced students or teachers of the physical sciences, and another one based on simple optical diffraction simulations accessible to any lay readership [5]. Hopefully, the structural meaning of the real X-ray pictures can now be fully understood by a much wider public.

Acknowledgments

I thank Prof. Philippe Lambin for his considerable help throughout the conception of the present and previous related works [4],[5]. I acknowledge the unfailing and expert technical assistance of Michel Mathot for the realisation of the diffraction slides and of the photos in figures P1-P12. I am most grateful to Prof. Sir Harry Kroto for giving to my earlier work on B-DNA [5] the vast audience of the website of his excellent VEGA Science Trust (http://www.vega.org.uk/). The 9-panels slide used there and in ref. [5] has served as the basis for a DNA optical transform kit produced and distributed by the Institute for Chemical Education (http://ice.chem.wisc.edu/) whose Director, Prof. John Moore, I thank most sincerely. The 9-panels slide (B-DNA) and/or the present 12-panels slide (A-DNA and B-DNA) can be ordered by e-mail from amand.lucas@fundp.ac.be. Also obtainable from my e-mail address is a video film, now available in DVD format, entitled *"From Light to Life"* (also available in French) where I relate the discovery of the double helix and the optical simulations of B-DNA [5].

References

[1] Watson F.W. and Crick F.H.C., "A Structure for deoxyribose nucleic acid", Nature **171**, 737-738 (1953)

[2] a. Pauling L., Corey R.B. and Branson H.R., "The structure of proteins: two hydrogen-bonded configurations of the polypeptide chain", Proc. U.S. Nat. Acad. Sci. **37,** 205-211 (1951). b. Pauling L. and Corey R.B., "The pleated sheet, a new layer configuration of polypeptide chains", Proc. U.S. Nat. Acad. Sci. **37,** 251-256 (1951). c. Pauling L. and Corey R.B., "The structure of synthetic polypeptides", Proc. U.S. Nat. Acad. Sci. **37,** 241-250 (1951).

[3] Cochran W., Crick F.H.C. and Vand V., "The structure of synthetic polypeptides. I. The transform of atoms on a helix", Acta Cryst. **5,** 581-585 (1952).

[4] Lucas A.A. and Lambin Ph., "Diffraction by DNA, carbon nanotubes and other helical nanostructures" Rep. Prog. Phys. **68,** 1181-1249 (2005).

[5] Lucas A.A., Lambin Ph., Mairesse R. and Mathot M., "Revealing the backbone structure of B-DNA from laser optical simulations of its X-ray diffraction diagram", J. Chem. Educ. **76,** 378-383 (1999).

[6] Olby R., "The Path to the Double Helix. The Discovery of DNA", *Dover Publications,* New York (1994).

[7] Polanyi M., "The X-ray Fibre diagram", Z. Phys. **7,** 149-180 (1921).

[8] Astbury W.T. and Street A., "X-ray studies of the structure of hair, wool and related fibres", Philosophical Transactions of the Royal Society London **A230,** 75-101 (1931); Astbury W.T. and Bell F.O., "X-Ray Data on the Structure of Natural Fibres and Bodies of High Molecular Weight" Tabul. Biol. **17**, 90-112 (1939).

[9] Pauling L. and Mirsky A., "On the structure of native, denatured and coagulated proteins", Proc. Nat. Acad. Sciences (USA) july (1936).

[10] Pauling L., Corey R.B. and Hayward R., "The structure of protein molecules", Sci. Amer. **191,** 51-59 (1954).

[11] Astbury W.T. and Bell F.O., "X-ray study of thymonucleic acid", Nature **141** 747-748 (1938).

[12] Horace Freeland Judson, "The Eighth Day of Creation", *Penguin Books,* London1979.

[13] Klug A., "Image analysis and reconstruction in the electron microscopy of biological macromolecules" Physica Scripta **14,** 245-256 (1978-79).

[14] Write the function $\exp(iq.r) = \exp[iq_\perp r\cos(\psi_q-\varphi)] \exp(iq_z z)$ as a Fourier series, that is as a power series of $\exp[i(\psi_q-\varphi)]$. The expansion coefficients are given, in the usual way, by an integral over φ of the original function multiplied by $\exp(in\varphi)$. This integral is just the integral representation of $J_n(x)$, modulo the factor $i^n = \exp(in\pi/2)$; see e.g. Arfken B. and Weber H., *Mathematical Methods for Physicists,* Academic Press, New York (1995).

[15] Bamford C.H., Hanby W.E. and Happey F., "The structure of synthetic polypeptides. I. X-ray investigation", Proc. Roy. Soc. **A205,** 30-47 (1951).

[16] Franklin R. and Gosling R.G., "The structure of sodium thymonucleate fibres. I. The influence of water content", Acta Cryst. **6,** 673-6771953. "II The cylindrically symmetrical patterson function" Acta Cryst. **6,** 678-685.

[17] Fuller W., "Who said « Helix »" , Nature **424,** 876-878 (2003).

[18] Watson J.D., *The double helix*, Athenaeum, New York (1968).

[19] Pauling L. and Corey R.B., "A proposed structure for the nucleic acids", Proc. U.S. Nat. Acad. Sci. **39,** 84-97 (1953).

[20] Crick F.H.C., *What Mad Pursuit*. Basic Books, Harper Collins, N.Y (1988).

[21] Chargaff E., "Chemical specificity of nucleic acids and the mechanism of their enzymatic degradation", Experientia **6**, 201-209 (1950).

[22] Fuller W., Wilkins M.H.F., Wilson H.R. and Hamilton L.D., "The molecular configuration of deoxyribonucleic acid IV. X-ray diffraction study of the A form", J. Mol. Biol. **12,** 60-80 (1963).

[23] Rich A., "Molecular Recognition between Protein and Nucleic Acids" in *The chemical bond. Structure and dynamics*, Edited by Ahmed Zewail, Academic Press N.Y., 31-86 (1992).

[24] To prepare the slide, each motif was first drawn repeatedly about 40 times in parallel on an A4 sheet of paper. For a paracrystalline (or gel) DNA fibre, the horizontal distances between the repeated motifs and their vertical positions were somewhat randomised, while for simulating crystalline fibres the motifs were set on a 2-D lattice. The 12 paper sheets were then assembled on a board in a 3 lines x 4 columns matrix and photographed together on a high-sensitivity film. The photographing distance to the board was such that the reduction produced about 5 motifs per mm on each of the film gratings. These specifications, the line thickness of the motifs, etc… need not be followed rigorously, as the high coherence of the light from a laser pointer is very forgiving of changes in these parameters. The film is mounted without cover on a standard projector slide.

Of the 12 motifs of the present simulation, 8 are the same as those of the original 9-gratings slide [5]; they are reproduced here for completeness and for convenience to the reader. Of the 4 new motifs, 2 aim at illustrating some fine but important details of the interpretation of the X-ray diagrams not covered in [5], and 2 are specific to the new A-DNA conformation.

[25] For the mathematically prepared reader, these extinctions are easy to understand: the total scattering amplitude of the two identical coaxial helices has the form $A(q)[1 + \exp(iq_z\Delta z)]$ where $A(q)$ is the amplitude of one single helix and where Δz is the axial distance between the two backbones. For $\Delta z = P/2$, as in the co-oriented helices of Fig.P8, and with $q_z = 2\pi l/P$, the factor in square brackets vanishes *for all odd integers l* by virtue of "the most beautiful formula of mathematics" $1 + \exp(il\pi) = 0$. For $\Delta z = P/8$, $3P/8$, $5P/8$ or $7P/8$, as in the counter-oriented

helices of Fig.P9, on the 4th layer line where $q_z = 4(2\pi/P)$, the amplitude again vanishes for the same reason. The axial shift of P/8 (or 7P/8) is unacceptable given the steric hindrances between adjacent nucleotides.

[26] Franklin R. and Gosling R.G., "Molecular configuration in sodium thymonucleate" Nature **171**, 740-741 (1953).

[27] Maddox B. Rosalind Franklin, *The dark lady of DNA*, Harper Collins, London (2002).

[28] Elkin Lynne Osman, "Rosalind Franklin and the double helix", Physics Today **56**, N°3, p.42-48 (2003).

Table I.

l	0	1	2	3	4	5	6	7	8	9	10
n	0 -10	1 -9	2 -8	3 -7	4 -6	5, -5	-4 6	-3 7	-2 8	-1 9	0 10
m	0 1	0 1	0 1	0 1	0 1	0, 1	1 0	1 0	1 0	1 0	1 0
$\mathbf{J_n}$	$\mathbf{J_0}$ J_{-10}	$\mathbf{J_1}$ J_{-9}	$\mathbf{J_2}$ J_{-8}	$\mathbf{J_3}$ J_{-7}	$\mathbf{J_4}$ J_{-6}	$\mathbf{J_5}$, $\mathbf{J_{-5}}$	$\mathbf{J_{-4}}$ J_6	$\mathbf{J_{-3}}$ J_7	$\mathbf{J_{-2}}$ J_8	$\mathbf{J_{-1}}$ J_9	$\mathbf{J_0}$ J_{10}

Caption of Table I.

Line 1 of the Table lists the l = 0 to l = 10 layer lines of the computed diffraction pattern of the Phosphorus helix of a single strand DNA (see Fig.3). Lines 2 and 3 lists the corresponding values of the integers n and m selected by the "selection rule" l = n + 10m and entering the CCV amplitude formula of Eq.(6) in the text. Line 4 lists the two leading Bessel terms (dominant in bold) contributing to the CCV amplitudes of the corresponding layer lines.

Figure Captions.

Fig.1 The atomic scattering factors of a few elements for Cu K_α X-rays (0.15 nm wavelength) in units of the classical electron radius $r_e = 2.8\ 10^{-15}$ m.

Fig.2 Principle of the geometrical construction of the hyperbolic layer lines in the X-ray diffraction pattern of a linear array of point scatterers (dark circles) with period P. ES: Ewald Sphere; LP: Layer Planes; S: observation Screen; LL: Layer Lines. The shaded regions are cones whose intersections by the screen generate the layer lines.

Fig.3 Computer simulated pattern of X-ray diffraction by a phosphorus helix of period P, radius r and atomic repeat p_a (see Table I for the leading Bessel contributions). Notice the central St Andrew's cross, the Bessel oscillations among the layer lines and the diamond pattern (highlighted by broken lines) with nearly empty North and South diamonds. The structural parameters of the helix (P, r, p_a) are readable from the geometrical parameters of the pattern ($2\pi/P$, α, $2\pi/p_a$).

Fig.4 The historic X-ray fibre diffraction patterns (a) of A-DNA and (b) of B-DNA plotted on the same scale. The 8^{th} layer line of pattern (a) occurs at about the 10^{th} layer line of pattern (b) reflecting a 20% increase in the DNA helix period. The A-DNA pattern shows crystalline spots on the first few layer lines.

Fig.5 High resolution X-ray fibre diffraction patterns of A-DNA showing the splitting of crystalline spots associated with the monoclinic symmetry of the crystallites [16].

Fig.6 The C-face-centred monoclinic unit cell of crystallites in a A-DNA crystal fibre [16]. The monoclinic cell, which is close to a hexagonal structure (one hexagon in the a-c face is highlighted), has the dimensions indicated [22]. The C2 symmetry axis demands that the DNA molecule must itself be dyadic and therefore have two anti parallel polar strands (b). The dyad excludes the possibility of a single strand (a) as well as a triple strand (c).

Fig.7 A schematic representation of the DNA double helix in its B conformation. Only the heterogeneous carbon-nitrogen rings of the flat bases (lower right) are shown, not their fringing chemical groups. A detailed representation of the 3'-5' oriented sugar-phosphate backbone is shown (upper right).

Fig.8 A photograph of the diffraction slide used in this paper for the optical simulations of X-ray diffraction by A- and B-DNA. The slide comprises 12 diffraction panels in 3 lines of 4 columns numbered P1 to P12. The motifs of the panels and their diffraction patterns are shown inset and in Figures P1 to P12.

Fig.9 (b): a molecular, space filling model of B-DNA. The backbones are in light grey and the bases in dark. The base pairs are perpendicular to the axis and are all seen edge on. (a): the diffraction motif used to simulate the B-DNA diffraction pattern at normal incidence in panel P12 of the slide. The backbone is represented by 10 black dots per period. The base pairs are represented as horizontal bars. The two sine backbones are separated by 3P/8.

Fig.10 (a) X-ray diffraction pattern of B-DNA (same as in Fig.4b). (b) optical diffraction simulation produced by panel P12 of the slide with the motif shown on Fig.9a.

Fig.11 (b) a molecular, space filling model of A-DNA. The backbones are in light grey and the bases in dark grey. Notice that a couple of base pairs define a see-through slit at each half period (dotted arrows). (a): the diffraction motif used to simulate the A-DNA diffraction pattern in panel P11 of the slide where the motifs are arranged on a 2-D lattice. The backbone is represented by 11 black dots per period. The base pairs are represented as bars inclined at 20° from the horizontal. Only the two base pairs seen edge on in each half-period are included in the model. The motif, as the molecule backbone, is C2 invariant.

Fig.12 (a) X-ray diffraction pattern of A-DNA (same as in Fig.4a). (b) optical diffraction simulation produced by panel P11 of the slide with the motif shown on Fig.11a. The broken lines in both patterns highlight the Saint Andrew cross produced by the zigzagging, edge-on base pairs of Fig.11. Note that the optical pattern in (b) is C2 invariant like the motif itself in Fig.11a.

P1. A set of equidistant, horizontal slits diffract light along regularly spaced layer lines. The band of maximal intensities is perpendicular to the slits.

P2. A set of equidistant slits rotated clockwise diffract light along regularly spaced layer lines. The band of maximal intensities is perpendicular to the slits (thin broken lines).

P3. A set of equidistant slits rotated anti clockwise diffract light along regularly spaced layer lines. The band of maximal intensities is perpendicular to the slits (thin broken lines).

P4. A zigzag motif diffract light to form a Saint Andrew cross. The arms of the cross are perpendicular to the zigzagging slits (thin broken lines).

P5. A zigzag double slit motif diffract light along regularly spaced layer lines. The motifs are arranged on a regular 2-D lattice on panel P5 of the slide, resulting in a spotty diffraction pattern. The maximal diffracted intensities define a Saint Andrew cross whose arms are perpendicular to the slits (thin dotted lines). The cross arms are clearly visible only at high-l layer lines (see text). The two reduced patterns on the right reproduce P4 (upper right) and P5 (lower right) for comparison. Notice the modulation of the arms of the Saint Andrew cross of the double slit motif (lower right).

P6. Diffraction by a sinusoidal motif. The band of maximal intensities is perpendicular to the zigzagging slope of the sine wave (thin dotted lines). There is no intensity in the meridian angle of the cross.

P7. Two coaxial sinusoidal motifs with different radii. The intensities diffracted by the small sine wave fall in the east and west angles of the Saint Andrew cross of the large sine wave. There is no intensity diffracted in the meridian angles of the large sine wave cross.

P8. A motif with two out-of-phase sine waves. Note that the true period of the motif is half the sine wave period. The layer line spacing is doubled as compared to the spacing of a single sine wave.

P9. A motif with two sine waves separated by 3P/8. Note the extinction of the 4^{th} layer lines.

P10. Diffraction by an atomic sinusoidal motif. Note the diamond pattern (broken lines) and the absence of intensity in the meridian diamonds.

P11. Optical simulation of the X-ray diffraction by an A-DNA fibre. The motif on the left is arranged on a 2-D lattice in panel P11 of the slide. The strong features on the 6 to 8^{th} layer lines of both the X-ray picture and the simulated pattern arise from the inclined base pairs seen edge on and forming a zigzagging double slit grating. (compare with P5).

P12. Optical simulation of the X-ray diffraction by a B-DNA fibre. The strong streaks on the 10^{th} layer lines of both the X-ray picture and the simulated pattern arise from the horizontal base pairs seen edge on.