X-RAY DIFFRACTION

AND

MOLECULAR MODELBUILDING STUDIES

ON THE

DEOXYRIBONUCLEIC ACID DOUBLE HELIX

by

Robert James Greenall, B.Sc.(Keele), M.Inst.P.

Volume I

A thesis submitted to the University of Keele for the degree of Doctor of Philosophy

1982

Department of Physics,
University of Keele,
Staffordshire. U.K.

IMAGING SERVICES NORTH
Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

Volume 1

Figure 2.1 between pages 34 & 35

Volume 2

Published journal article at the end of the volume

Excluded at the request of the university

ACKNOWLEDGEMENTS

I should like to thank my supervisor, Professor Watson Fuller, for his advice, and criticism and for the provision of research facilities. I am grateful to past and present members of his group, Drs. W.J. Pigram, G. Dougherty, C. Nave and D.C. Goodwin and Messrs. N.J. Rhodes, A. Mahendrasingam and V.T. Forsyth not only for discussions, suggestions and assistance but also for helping to provide a friendly atmosphere in which it has been pleasant to work.

It is also a pleasure to thank Professor R.A.J. Warren, who has been collaborating with this group for some years, for the provision of ϕw-14 DNA and for performing some of the measurements described in Chapter VI.

I am grateful to Professor M.H.F. Wilkins, F.R.S., who loaned the B-DNA diffraction pattern used in Chapter III and to Dr. M. Elder and his colleagues of the Densitometry Service, S.E.R.C., Daresbury Laboratory, who digitised the pattern.

Dr. G.A. Rodley of the University of Canterbury, New Zealand kindly provided some of his Axial Patterson calculations and the data on which they were based.

I am also indebted to my present colleagues, Dr. A. Miller and Mr. S.K. Burley of the Laboratory of Molecular Biophysics, Oxford, for discussions on coiled-coil diffraction theory and interpretation of the electron micrograph described in Chapter IV.

Mr. F. Rowerth and his technical staff have provided admirable service. In particular I should mention Mr. G. Dudley and his colleagues in the Mechanical Workshop and Mr. M. Daniels who prepared the photographic prints in this thesis.

I am grateful for the secretarial assistance I have received and I wish especially to thank Miss Helen Martin who not only produced a

most attractive typescript but also smiled bravely each time I presented her with just one more paper to photocopy.

Financial support was provided by a University of Keele Research Scholarship.

Finally, I have no doubt that no thesis is completed without considerable help in a multitude of ways from one's family and friends. To them all I express my gratitude and my incredulity that they manage to put up with me - especially in the mornings.

ABSTRACT

The known nucleic acid conformations and the methods by which they are determined using X-ray fibre diffraction are reviewed and discussed. A new stacking scheme for Watson-Crick base-pairs is described and a left-handed model of B-DNA which incorporates this scheme is presented and evaluated. This model is less successful than the conventional B-DNA model in explaining the observed diffraction pattern. The side-by-side of B-DNA is criticised in detail and its predicted diffraction pattern is found to compare unfavourably with that predicted by the double helix. Two forms of Patterson function have been applied to several data sets. The results suggest both that the accepted A-DNA indexing of Fuller et al (1965) is preferable to a new scheme proposed by Sasisekharan, Bansal and Gupta (1981) and also that a left-handed model of D-DNA may be in better agreement with the observed diffraction pattern than is the right-handed model of Arnott et al (1974) but neither function is found to be sufficiently robust to enable reliable conclusions to be drawn concerning molecular conformation. Expressions are derived which describe the effect on the Patterson functions of baseline errors in the measurement of intensities in diffuse patterns. The conformation and transitions of a bacteriophage DNA have been studied and a model is presented which explains the observed behaviour in terms of a groove-bridging putrescinyl linkage. Several models (incuding a preliminary coiled-coil model) of the structure of DNA under mechanical tension are described and compared with the observed diffraction patterns.

CONTENTS

CHAPTER I

THE STRUCTURE AND FUNCTION OF NUCLEIC ACIDS

1.1 The Biological Function of the Double Helix

Deoxyribonucleic acid (DNA) plays a central role in cellular biochemistry. The genetic information is encoded in the molecule, transmitted to the succeeding generation and translated into proteins which are ubiquitous in the structure and function of organisms.

DNA is an unbranched polymer. The monomer, or nucleotide, is formed by three groups: a sugar, a base and a phosphate. The sugar (2-deoxy-D-ribose) is a five membered ring. In ribonucleic acid (RNA), a closely related molecule, an additional hydroxyl group is bonded to C2' (Figure 1). The bases fall into two classes, the purines and the pyrimidines. The two major members of the purine class are adenine (A) and guanine (G). These are also the most common purines found in RNA. Thymine (T) and cytosine (C) are the major pyrimidines in DNA, but in RNA uracil (U) is present instead of thymine (Figure 3). In addition, a number of bases with minor chemical modifications are found in RNA. All the bases are flat, aromatic systems which are about 3.4 A thick. The bases are bonded (via N9 in purines and N3 in pyrimidines) to the sugar C1'. The nucleotide is completed by the addition of the phosphate group at C3'. The polymer is formed by a C3'-C5' phosphodiester bridge between successive nucleotides (Figure 2).

Early X-ray diffraction studies appeared to indicate that the secondary structure of DNA contained a high degree of regularity (Astbury 1947). This seemed paradoxical since it was clear that a completely regular structure would not be sufficiently flexible to encode the large amount of

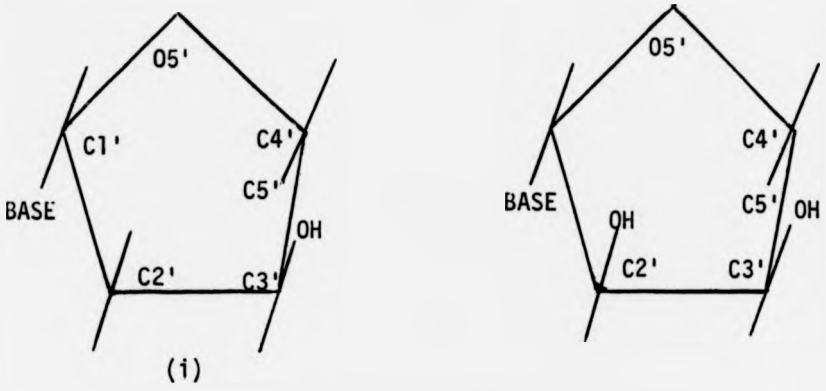


Figure 1.1 : The Common Pentose Sugars

(i) Deoxyribose; (ii) Ribose

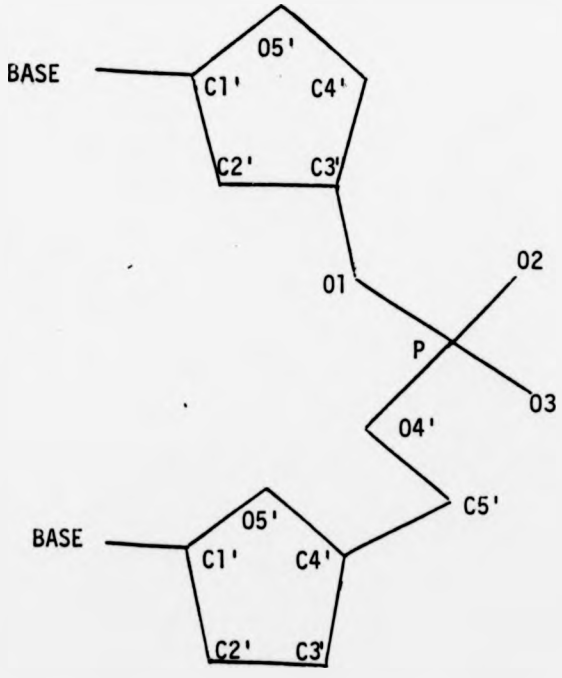


Figure 1.2 : The Phosphodiester Linkage

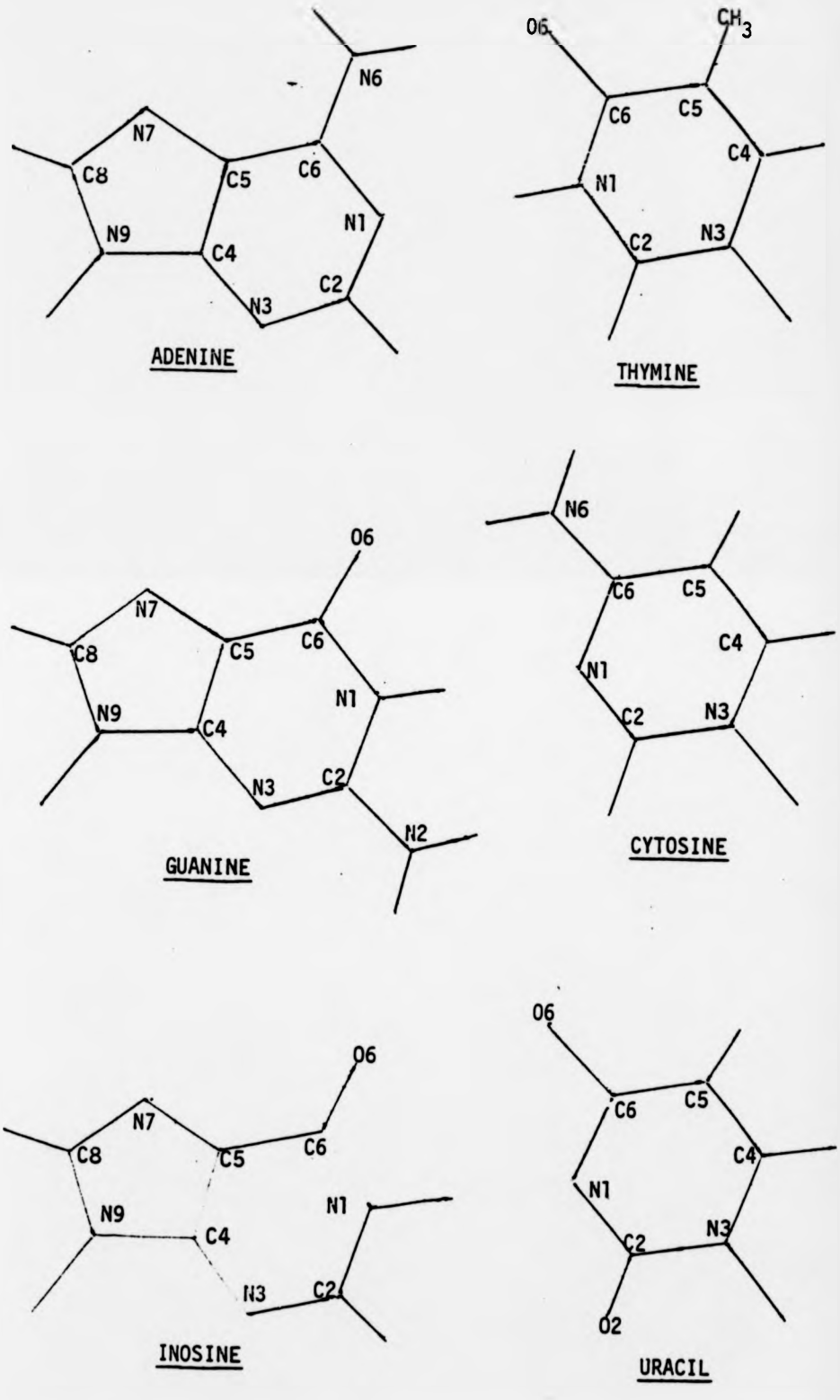


FIGURE 1.3 : THE STRUCTURES OF THE COMMON BASES

information needed in order to synthesise proteins. A particularly simple diffraction pattern such as those obtained by Franklin and Gosling (1953a) and Wilkins, Stokes and Wilson (1953) suggested a possible structure for DNA to Watson and Crick (1953a) which resolved this dilemma.

The structure consists of two polynucleotide chains joined together by base-pairing: adenine always pairs with thymine, and guanine always pairs with cytosine. The bases are held together by hydrogen-bonds which ensure the specificity of the interaction. The geometries of the two base-pairs are strikingly similar (Fig. 4) and so it is possible for the sugar-phosphate backbone to be wound into a perfectly regular double helix (Fig. 5). The sugar and phosphate attached at each side of a base-pair are related by a twofold rotation axis in the plane of the bases, so although each chain has a directional character defined by the C5'-C3' linkage, the double helix itself contains no directional marker. The helical symmetry of the molecule generates a further set of diad axes mid-way between the base planes. The model has the attractive chemical feature that the hydrophobic bases are hidden in the centre away from solvent molecules, whereas the negatively charged phosphate groups are easily accessible. In addition it explains naturally the discovery by Chargaff (1950) that the molar proportion of adenine is equal to that of thymine, and the proportion of guanine is equal to that of cytosine in DNA from most sources. DNA from sources wherein these rules are not obeyed, for example that from bacteriophage ϕx-174, is not double stranded.

The double helical model of DNA represents perhaps the best example of the central theme of molecular biology : that biological function is closely related to molecular structure. Watson and Crick (1953b) were immediately able to suggest a mechanism for DNA replication. They proposed that the strands of the molecule separated and that each strand then acted as a template upon which new nucleotides were added according to the base-

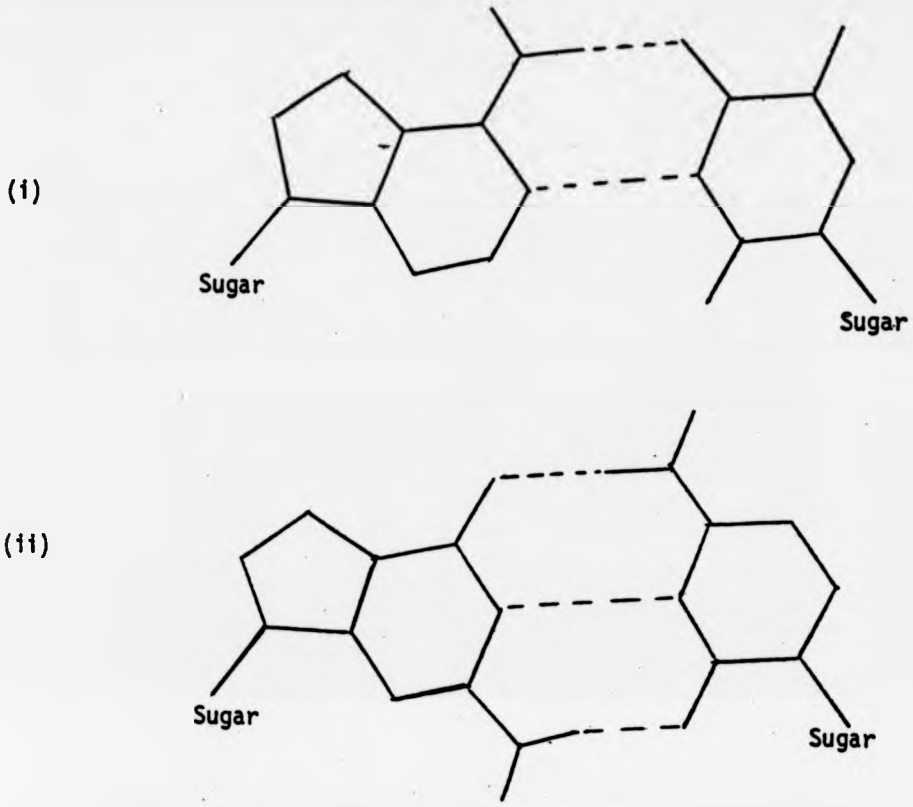


Figure 1.4 : The Geometry of the Watson-Crick Base-Pairs
(i) Adenine-Thymine; (ii) Guanine-Cytosine

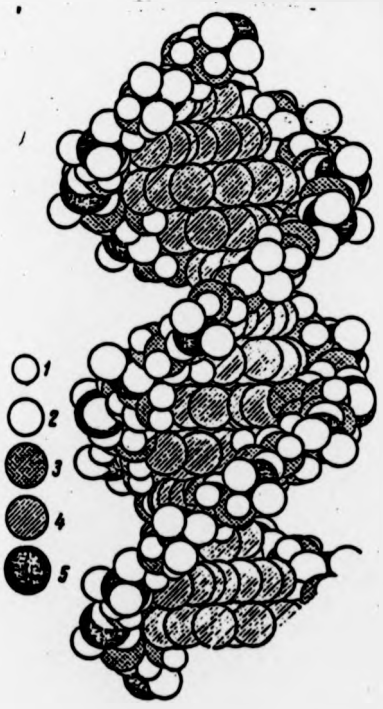


Figure 1.5 : The DNA Double Helix

Key

1. Hydrogen
2. Oxygen
3. Sugar Carbon
4. Base Carbon and Nitrogen
5. Phosphorous

pairing rules. In this way two identical daughter molecules would be formed, each of which would contain one strand from the parental DNA. Meselsohn and Stahl (1958) tested this hypothesis by growing bacteria in a medium containing ^{15}N which became incorporated into the DNA. When the sedimentation of DNA from daughter cells grown in ^{14}N was examined, it was found that the banding pattern was precisely that predicted by the semiconservative scheme of Watson and Crick.

The base-pairing hypothesis plays a major role not only in DNA replication but also in transcription and decoding. The proteins, whose synthesis is directed by DNA, are also unbranched polymers consisting of peptide monomers which make identical chemical links with each other. Variation is introduced into this primary structure via twenty different side-chains. A DNA molecule contains a number of genes each of which contains a message describing the sequence of side chains in a polypeptide. This message is recorded in the base sequence which is read in "words" of three bases, each word coding for a single peptide. When a protein is to be synthesised, a single-stranded messenger RNA (mRNA) molecule is transcribed upon the relevant gene using the base-pairing rules. The genetic information is then carried from the DNA, which in higher cells resides largely in the nucleus (although some cell organelles, such as the mitochondria, also contain DNA), to the ribosomes in the cytoplasm where proteins are manufactured. The ribosome grasps each word, or codon, of the mRNA in turn and the corresponding peptide is brought to the complex by a transfer RNA (tRNA) molecule which contains an exposed group of three bases, the anticodon, complementary to the codon. The new monomer is then added to the nascent polypeptide which detaches from the ribosome when growth is complete.

Whilst it is widely accepted that the replication and transcription

of genes follows this general framework, it is nonetheless recognised that we have only a poor understanding of these processes at the molecular level. Unresolved questions abound. For example, the cell will need to transcribe only a small proportion of the genes at any time, so the majority of the genes will be inactive. How are the required genes recognised? How is the switching between the dormant and active states controlled? At an early stage in the development of a higher organism, individual cells become specialised and utilise only a tiny fraction of the genome. How is the process of differentiation achieved? A DNA molecule may contain several thousand base-pairs. How are the intertwined strands unravelled during replication? Sexual reproduction is advantageous in selection because paternal and maternal genes are redistributed during recombination. How is this highly specific process organised? Added complexity is endowed upon these problems by the fact that DNA in higher organisms is complexed with other molecules, in particular, proteins.

If we admit that we are to a large extent ignorant of the processes of control and recognition, may we take solace from the fact that we have a detailed understanding of the genetic organisation of DNA? Unfortunately not, for the view of DNA as a string of genes is far too simple. It has been known for some time that even bacteria containing only one chromosome may undergo homologous recombination mediated by transduction, transformation or conjugation, but in the last ten years it has been discovered that bacterial cells, plants and animals may participate in "illegitimate" recombinational processes which join together non-homologous DNA segments. This recombination is effected by structurally and genetically discrete segments, known as transposons and insertion sequences, which may move around within the organismal DNA (Richmond, 1979; Cohen and Shapero, 1980). In addition, the amount of DNA contained by higher organisms greatly exceeds

that required to store the genetic information. Even if some of the excess may be assigned a controlling or structural role, the rest appears to be useless. This subject has recently been reviewed by Dawkins (1976), Doolittle and Sapienza (1980) and Orgel and Crick (1980). A final example serves to illustrate the complexity of DNA. The simple bacteriophage ϕx-174 has been found rather surprisingly to contain several stretches of DNA which are contained in more than one gene (Barrell et al, 1976; Brown and Smith, 1977; Smith et al, 1977). The advantage conferred upon the virus by such overlapping genes (apart from efficient packing of genetic information) is not yet clear.

This thesis is concerned primarily with the three-dimensional conformation of double-stranded DNA and a critical appraisal of the methods used in structure determination. The remainder of this chapter reviews the known conformations of nucleic acids and outlines the contents of the project.

1.2 Description of Double-Stranded Polynucleotides

In addition to the single-stranded nucleic acids such as mRNA and tRNA which exist within the cell, some viruses such as ϕx-174 contain single-stranded DNA and one-, three- and four-stranded polynucleotides have been observed in fibres. In this thesis we are concerned mainly with double-stranded nucleic acids and we will discuss here the way in which we may describe their three-dimensional structure. However in many respects the same methods may be used in describing helical polynucleotides with other than two strands.

The X-ray diffraction data from fibres typically give little information about spacings less than 3A. We are therefore compelled to use results from single crystal studies on nucleic acid components in order to determine the covalent stereochemistry. Arnott (1970) has presented details of a survey of the bond lengths and angles observed in such studies, and in

addition he has refined the structure of the base-pairs to agree with our current knowledge of hydrogen-bonding stereochemistry. His results (presented in figures 6 and 7) were used for all the structural models built by the author.

Whilst a set of three-dimensional co-ordinates for the atoms in DNA is useful, it is often more convenient and illuminating to characterise a polynucleotide in terms of the angles of rotation about the single bonds in the backbone. Figure 8 shows the nomenclature we will employ. We define the cis position to be zero and a positive torsion angle arises when (looking along the bond) the atom at the far end has to be rotated clockwise from zero. The angle χ, which describes the relative orientation of the sugar and base, is defined by the atoms C2', C1', N and C2 in the case of a pyrimidine or C4 in a purine. The values of χ which have been observed in both monomers and polymers may be divided into two classes separated by a rotational energy barrier. When $\chi \simeq 90^{o}$ the conformation is said to be anti; when $\chi \simeq 300^{o}$ the conformation is syn.

The backbone torsion angles fall naturally into three sectors and it will often be convenient to discuss conformations in terms of these general classes of angles rather than precise values. If τ is a torsion angle then it is said to be gauche $^{+}(g^{+})$ if $0 < \tau < 120^{o}$; gauche^{-} (g^{-}) if $-120^{o} < \tau < 0^{o}$, and trans if $120^{o} < \tau < 240^{o}$. (See figure 9).

Spencer (1959) pointed out that the sugar rings were unlikely to be flat since there would be steric interference between hydrogen atoms on adjacent carbons. It is now conventional to define the puckering of the ring by reference to the plane formed by C1', O5' and C4'. There are then four major conformations which can occur according to whether (i) C2' or C3' is further from the plane, and (ii) the further atom falls on the same side of the plane as C5' (endo) or on the opposite side (exo). Projections

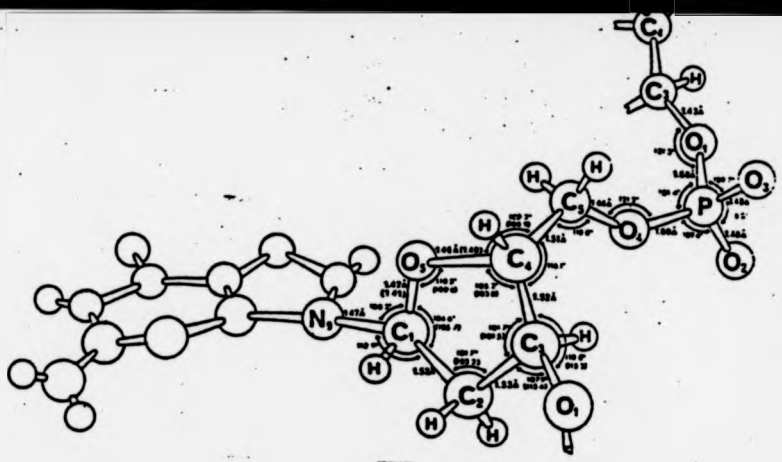


Figure 1.6 : Covalent Stereochemistry of the Sugar-Phosphate Backbone in the C3'-endo Conformation,

(Bracketed figures are for the C2'-endo Conformation). (From Arnott, 1967).

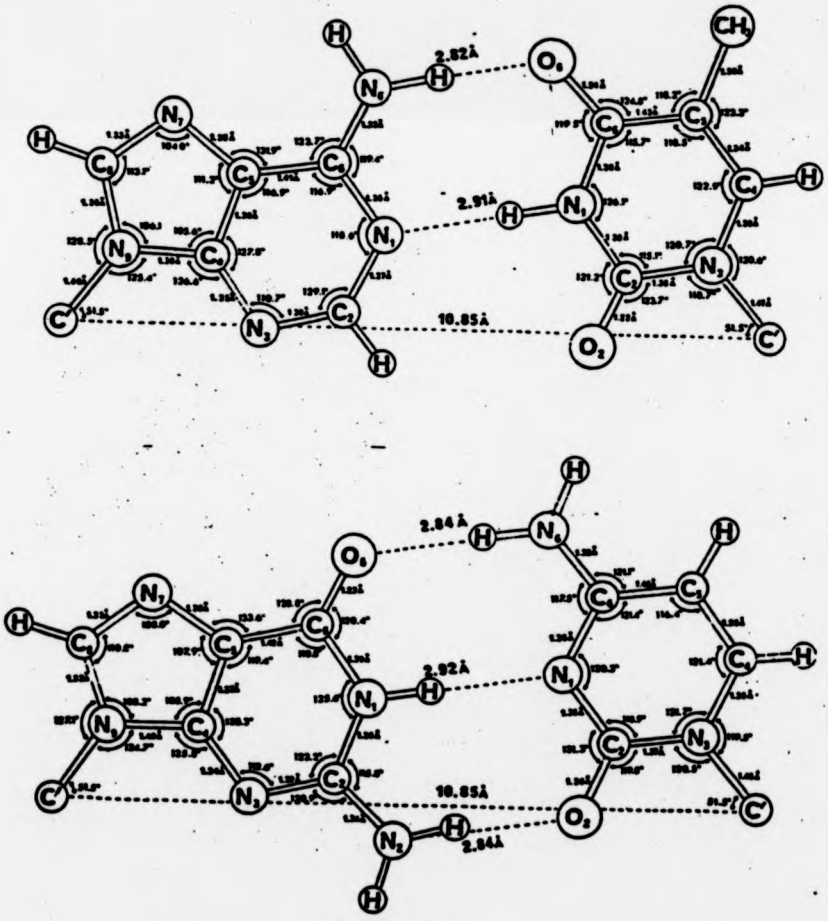


Figure 1.7 : Stereochemistry of the Base Pairs

(From Arnott, 1967)

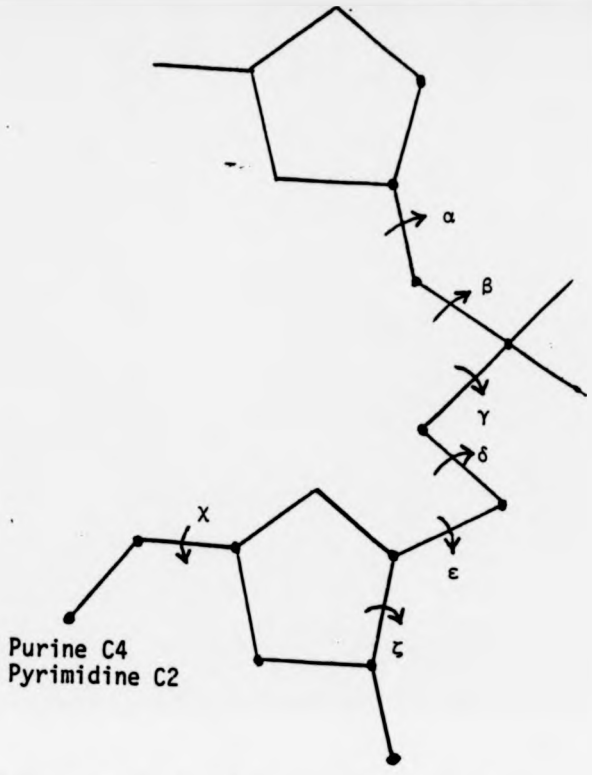


Figure 1.8 : Torsion Angle Nomenclature

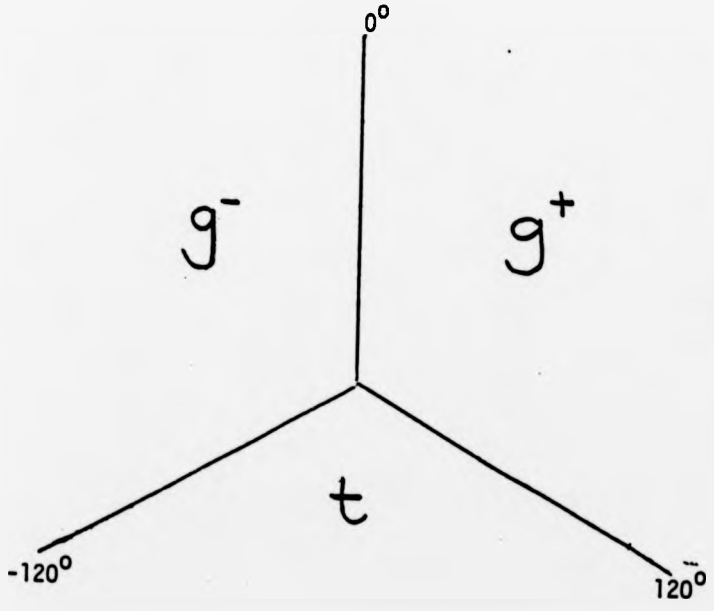


Figure 1.9 : Torsion Angle Sectors

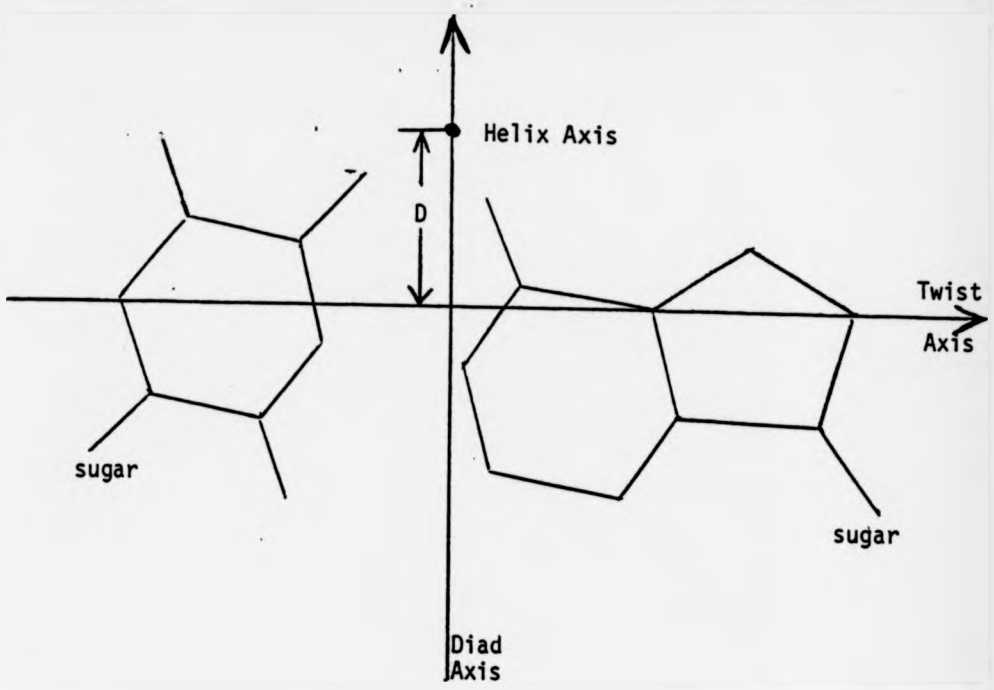


Figure 1.10: Parameters Defining the Base-Pair Positions

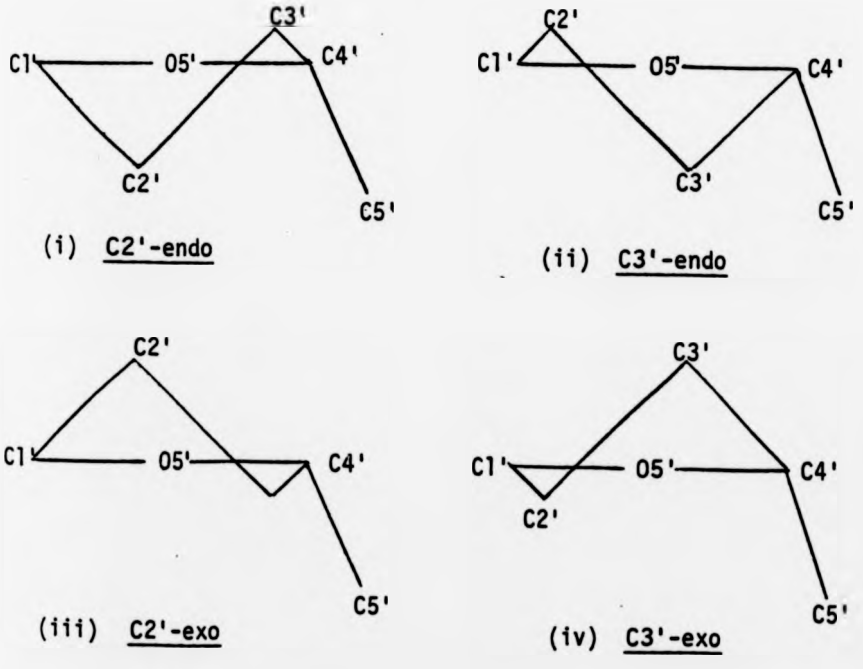


Figure 1.11 : Projections of the Four Major Sugar Puckers

of these conformations are shown in figure 11. Arnott and Hukins (1972a) have presented co-ordinates of standard sugars in the C2'-endo, C3'-exo and C3'-endo conformations based upon a survey of the furanose rings in β-nucleosides and β-nucleotides observed in single crystals.

Three parameters may be defined which describe the position and orientation of the bases. First, the base-pairs may be displaced by a distance D from the helix axis. Second, they may be tilted about the diad axis. Finally, the base-pairs may not be planar. It has been conventional in recent years to define the axis about which such twisting occurs to run from purine C8 to pyrimidine C2. Tilt and twist angles are both positive for anticlockwise rotation about their axes. D is positive when the twist axis is moved in the positive y-direction. (See figure 10).

1.3 Structural Studies on Nucleic Acids

1.3.1 Classical Double Helical Models

Polynucleotides have five backbone single-bonds about which rotation is possible, sugar rings which may adopt a variety of conformations and a glycosyl link between the base and sugar which may exhibit a wide degree of variation. It is not surprising therefore that they can adopt a number of distinctive conformations, characterised by the rise per residue (h) and the rotation per residue (t), which all conform to the Watson-Crick paradigm. These structures, which will be described briefly here, have been elucidated in the main by X-ray diffraction from fibres which characteristically give rather low resolution (≃ 3A) data. This technique and the confidence we may have in the results based thereon will be discussed in detail in subsequent chapters.

Until recently (1980) all double-stranded polynucleotide conformations could be assigned to one of two families named A and B. Members of the A-family (A-DNA, A-RNA, A'-RNA, A"-RNA and A*-RNA) are

characterised by C3'-endo sugars, positive base displacement and large, positive base tilts. In the B-family (B-, B'-, C-, C'-, C"-, D and E-DNA) the sugars are C3'-exo (or the minor variant of it, C2'-endo), the bases are near to the helix axis or slightly behind it and they are tilted in a negative sense.

1.3.1.1 The A-family

1.3.1.1.1 A-DNA

The first diffraction pattern from a crystalline fibre of DNA (designated A-DNA) was obtained with the sodium salt at low relative humidity and little excess salt (Franklin and Gosling, 1953b). Similar patterns are obtained with the potassium and rubidium salt but not with lithium. The A-form has been observed with a variety of naturally occurring DNA's with (A+T)/(G+C) varying from 0.42 to 1.85 (Hamilton et al, 1959). In an extensive study of synthetic polynucleotides, Leslie et al (1980) have found that neither base sequence nor base composition appears, in general, to prevent any DNA from adopting the A-form.

The conformation was determined in detail by Fuller et al (1965) who found it to be an 11_1 helix with a mononucleotide repeat unit and 28.15A pitch. The sugars are in the C3'-endo pucker. The bases are tilted and twisted 20^{o} and -8^{o} respectively and placed 4.25A in front of the helix axis. The space group, C2, requires four asymmetric units per cell but the cell dimensions indicate that only two molecules can be accomodated. These requirements are both satisfied if the molecule is oriented so that its diad axis points along the b-axis, and the crystal asymmetric unit is then one polynucleotide strand. Fuller et al (1965) showed that both the calculated structure factors and the intermolecular stereochemistry were most satisfactory if the diad pointed in the positive b direction.

Recently the model has been computer refined so as to impose

precise covalent stereochemistry (Arnott and Hukins, 1972b). They found A-DNA to be the least in need of modification of all the duplex conformations which had been determined by hand modelbuilding. The base tilt, twist and displacement in the refined version are 20.2^{o}, -1.2^{o} and 4.72 A respectively. Arnott and co-workers have recently produced a newly refined A-DNA model which appears to differ only marginally from their earlier one (reported in Arnott et al, 1980). The torsion angles of all the models are presented in table 2 for comparison.

It is naturally of interest to determine whether structures observed in fibres have any biological significance. Since A-DNA has always crystallised in a monoclinic lattice, it may be that lattice forces are predominant in stabilising the conformation. However, Arnott and co-workers (reported in Leslie et al, 1980) have determined the structure of two synthetic polynucleotides which maintain A-like geometry whilst crystallising in other systems. This indicates that the A-form is not merely an artefact of crystallisation.

1.3.1.1.2 A-RNA

Well oriented, crystalline diffraction patterns have been obtained from reovirus RNA (Langridge and Gomatos, 1963), rice dwarf virus RNA (Sato et al, 1966), the replicative form of MS2 virus RNA (Langridge et al, 1964) and wound tumour virus RNA (Tomita and Rich, 1964). The structure of the reovirus RNA was originally thought to be a 10_1 helix but Arnott et al (1966) showed that an 11_1 helix was also possible. The analysis of the patterns was more difficult than for those from DNA since the molecular packing was more complicated. The molecules crystallise in two different hexagonal systems named α and β (Langridge and Gomatos, 1963); Tomita and Rich, 1964; Arnott et al, 1966) but it was assumed that the molecular conformation was the same in both forms. Arnott et al (1967 a, b, c) showed that an elevenfold model based on reovirus data was slightly

preferable but it was not until the synthetic polyribonucleotide poly(A).poly(U) gave rise to a highly crystalline and exceptionally well oriented β-A-RNA pattern that the tenfold model could be confidently rejected (Arnott et al, 1973).

A-RNA is very similar to A-DNA. (See table 2 for the torsion angles). It contains antiparallel chains with Watson-Crick base-pairs displaced about 4A from the axis, tilted 16^{o} and twisted -6.9^{o}. The pitch is 30.9A.

1.3.1.1.3 A'-RNA

Arnott et al (1968) observed that addition of salt to a poly(I).poly(C) or poly(A).poly(U) fibre in the A-form gave rise to a new molecular structure, designated A'-RNA. The data from poly(I).poly(C) was used by Arnott et al (1973) in the characterisation and refinement of the molecule. A'-RNA is a 12_1 helix with a 36.2A pitch. Once again the torsion angles resemble those of A-DNA but the base tilt is halved to 10^{o}. Both A-RNA and A'-RNA have the rather large twist angle observed by Fuller et al (1965) in A-DNA but which was subsequently reduced upon refinement by Arnott and Hukins (1972b).

1.3.1.1.4 A"-RNA

A family of non-integral helices, called A"-RNA, was observed by Arnott et al (1968) in fibres of poly(A,U), poly(G,C) and poly(I,C) containing little salt. These structures have not been well characterised.

1.3.1.1.5 DNA-RNA Hybrids

Since at least short stretches of DNA-RNA hybrids are formed during transcription, it is of considerable interest to determine their structure. Milman et al (1967) have found that an RNA-DNA hybrid has a structure similar to A-DNA and O'Brien and MacEwan (1970) discovered that the synthetic hybrid poly(rI).poly(dC) is isostructural with A'-RNA. The

model of Milman et al (1967) is not well defined since no torsion angles nor co-ordinates were presented. However, O'Brien and MacEwan's model was computer refined and a full set of co-ordinates was published. The torsion angles have been calculated by the author and are recorded in table 2.

Initial results on the conformation of poly(dI).poly(rC) have recently been published (Chandrasekaran et al, 1980; Arnott, 1980). This tenfold helix (h = 3.1A, t = 36^{o}) is the least tightly wound member of the A-family. It is also distinguished by having a slightly different backbone conformation. Whereas all previous members fell in the $\alpha\beta\gamma\delta\varepsilon$ = $tg^{-}g^{-}tg^{+}$ class, this new structure (designated A^{*}-RNA) is $tg^{-}ttt$. Both the sugars, which are in the C3-endo pucker, and the bases, which have positive tilt and displacement, exhibit the conformation expected of the A-family. An interesting point is that A^{*}-RNA closely resembles the double helical model originally published by Crick and Watson (1954).

No RNA, synthetic or natural, nor any DNA-RNA hybrid has been observed in any form other than a member of the A-family. In addition, yeast phenylalanine tRNA, which is a "globular" molecule, contains short (somewhat irregular) helical stretches which are similar to A'-RNA in structure (Jack et al, 1976; Holbrook et al, 1978). It would be strange if this apparent preference were not exploited in biological systems, and indeed Arnott et al (1968) have suggested that DNA may be in the A-form during transcription.

1.3.1.2 The B-family

1.3.1.2.1 B-DNA

At high relative humidity the sodium salt of DNA gives diffraction patterns which indicate that the molecules are packed into a semi-crystalline array (Langridge et al, 1960a). Whilst the patterns are not sufficient to enable a precise determination to be made of the molecular conformation, they

do nonetheless exhibit rather strikingly the cross-shape which indicates that the structure is a helix (Cochran, Crick and Vand, 1952). It was a pattern of this type (Franklin and Gosling, 1953a) which led Watson and Crick (1953a) to postulate their model on the basis of modelbuilding studies. Langridge et al (1960a,b) subsequently refined the model using the more extensive data obtained from samples of the lithium salt which forms highly crystalline and oriented fibres. Comparison of the observed diffraction with that calculated from the co-ordinates published by Crick and Watson (1954) indicated that their model was not entirely satisfactory (Langridge et al, 1960b), however Langridge et al improved the agreement whilst retaining the characteristic features of the original structure.

Arnott and Hukins (1972b; 1973) have presented details of a computer-refined model which improves on the efforts of Langridge et al by imposing precise covalent stereochemistry. B-DNA is a 10_1 helix with 34A pitch. The sugars are in the C3'-exo pucker and the bases, which are in the anti orientation, are displaced 0.16A behind the helix axis, tilted -6^{o} and twisted -2.1^{o}.

Arnott and Hukins (1973) also built a B-DNA model with C2'-endo sugars but they were unable to discriminate between the two alternatives on the basis of the diffraction data alone. The stereochemistry marginally favoured the C3'-exo model. Arnott and Chandrasekaran (unpublished) have recently obtained improved diffraction data and they have re-refined the C2'-endo model which they now claim to be the better fit. The torsion angles have been presented in Arnott et al (1980). The conformational details of all the B models have been collected in table 2.

LiB-DNA packs into an orthorhombic lattice in space group $P2_12_12_1$. The molecular diad in the plane of the base points in the b direction. One polynucleotide strand forms the crystalline asymmetric unit so there are two molecules per unit cell, that in the centre being displaced 0.328c along

z relative to those at the corners (Langridge et al, 1960a). The crystal packing exploits the molecular symmetry in a remarkable way. The ratio a/b is equal to tan 36^{o} which maximises the number of equivalent contacts between the tenfold helices (Dover, 1977).

Although the diffraction patterns from the semi-crystalline sodium salt of B-DNA contain less information than those from the crystalline lithium salt, it is believed that the molecular conformation is essentially the same in both cases (Langridge et al, 1960a). Franklin and Gosling (1953b) have shown that upon varying the relative humidity around the fibre, the molecular conformation changes reversibly from the A-form (at low humidity) to the B-form. Cooper and Hamilton (1966) subsequently found that the salt concentration is an additional factor affecting the transition. At low salt concentrations (< 5% excess NaCℓ) the A-form was always observed at 92%, whereas high salt fibres (> 9% excess NaCℓ) gave the B-form even at 75% RH. These results were confirmed using infrared spectroscopy by Pilet and Brahms (1972, 1973), and using laser Raman spectroscopy by Erfurth et al (1975).

The apparent conservation of the B-form in both crystalline orthorhombic systems and semi-crystalline hexagonal systems seems to indicate that this conformation is not determined by lattice forces, but this conclusion is not straightforward since the data in the hexagonal case is sparse and changes in the molecular conformation may go undetected. However, Leslie et al (1980) have discovered that poly d(G-C).poly d(G-C) and poly d(A-C).poly d(G-T) maintain the classical B-form whilst crystallising in large unit cells which appear to contain four molecules, which indicates that B-DNA is a stable structure depending on intramolecular forces alone. This conclusion seems to be supported by the discovery of a crystalline sodium B-DNA from poly d(A-C).poly d(G-T) (Leslie et al, 1980).

In their study of DNA from a variety of natural sources, Hamilton et al (1959), working mainly with the sodium salt, concentrated on obtaining the A-form. However, they found that salmon sperm, calf thymus and normal human leucocytes give similar lithium B-DNA patterns which indicate that the conformations are isostructural. Wilkins and Randall (1953) showed that diffraction from DNA in oriented sperm heads gives rise to a B-like pattern. These findings suggest that the normal conformation for DNA in the cell is similar to B-DNA. Further support for this belief comes from neutron and X-ray diffraction studies on the DNA wrapped around nucleosome cores (Finch et al, 1981; Bentley, Finch and Lewitt-Bentley, 1981). The inter-base separation appears to be approximately 3.4 A. A wide variety of synthetic polynucleotides also give classical B-forms, and, except in a few special cases, neither base sequence nor base composition appears to prevent any polynucleotide attaining the B-form (Leslie et al, 1980).

Recent work by several groups has led to a re-investigation of the relationship between the structure of DNA in fibres, solutions and cells. Bram (1971a,b,c; 1972) used high angle X-ray scattering to show both that the fibre and solution structures of DNA differ and that the structure in solution varies with base content. Whilst such studies are certainly suggestive, it is rather difficult to draw definitive conclusions from the spherically averaged data obtained from solutions. Measurements on the length of DNA in electron micrographs suggested to Griffith (1978) that the molecule had 10.5 base-pairs per turn in solution. Energy minimisation studies by Levitt (1978) predicted a similar winding-up of the helix. In order to accommodate the change in the helical screw the sugar puckers changed quite dramatically and the angle between the normal to the base planes and the helix axis increased from the 6^{o} observed in B-DNA (Arnott and Hukins,

1972b) to 17^{o}. The bases are highly twisted giving rise to propellor-shaped base-pairs. The values of α, γ, δ and ε all move closer to those expected in staggered rotamers and χ drops from the high value of 82^{o} to 47^{o}. By interpolating into the calculated energy of 9_1, 10_1, 11_1 and 12_1 helices, Levitt concluded that the energetically most favourable conformation has 10.6 base-pairs per turn, but in a fibre diffraction study covering a wide range of humidity Zimmerman and Pheiffer (1979) confirmed the accepted 36^{o} rotation per residue. Support for the Levitt and Griffith conclusions came from the transient electric dichroism results of Hogan, Dattagupta and Crothers (1978) which suggested that the angle between the base normals and the helix axis is 17^{o}. Unfortunately this technique depends upon assumptions about the direction of the base transition moment and the mechanism of orientation in electric fields which may be questionable (Charney, 1978). However, Wang (1979), by inserting known lengths of oligonucleotide into covalently closed circular DNA and measuring the consequent change in electrophoretic mobility, has provided convincing evidence that the number of turns per residue in solution lies between 10.4 and 10.5 under physiological conditions.

A model explaining the discrepancy between fibre and solution studies has been put forward by Mandelkern, Dattagupta and Crothers (private communication). They suggest that in high salt, low DNA concentrations the molecules exist as single rod-like entities with structures like those found by Levitt, whereas in low salt, high DNA concentrations they conglomerate into bundles containing seven molecules which have conformations like B-DNA. The interaction between neighbouring molecules in the bundles is used to explain the conformational differences between each state. They have measured several hydrodynamic, thermodynamic and electro-optical parameters of DNA as a function of DNA and salt concentration and the results support their hypothesis. The conditions present in fibres are more like those in the

bundles than in solution so the difference between the number of base-pairs per pitch observed by diffraction and electrophoresis may reflect a slight change in the DNA structure which occurs as the molecules crystallise from solution.

1.3.1.2.2 B'-DNA

Poly(dA).poly(dT) exists as a 10_1 double-helix with Watson-Crick base-pairs (Arnott and Selsing, 1974). The molecule packs into both a hexagonal (α-B'-DNA) and orthorhombic (β-B'-DNA) lattice with no change in conformation. The pitch of the helix (32.4A) is only slightly less than that of B-DNA. A refinement using the C3'-exo B-DNA torsion angles of Arnott and Hukins (1973) as starting values produced a model which is very similar to the classical B-form (Arnott and Selsing, 1974). All the angles remained in the same range as those in B-DNA ($tg^-g^-tg^+$). The bases are situated on the helix axis and they are tilted and twisted -7.9^o and -1.0^o respectively. The B'-DNA conformation has also been observed recently by Leslie et al (1980) in poly(dI).poly(dC) and poly d(A-I).poly d(C-T).

1.3.1.2.3 C-DNA

At relative humidities of 66% or lower LiDNA may adopt the C-form. The molecules pack into both orthorhombic and hexagonal lattices depending upon the humidity and salt content (Marvin et al, 1961). Infrared dichroism studies by Brahms et al (1973) suggested that the sodium salt could also adopt the C-form. This has been confirmed by X-ray fibre diffraction studies of Arnott and Selsing (1975) using natural DNA, Leslie et al (1980) and N.J. Rhodes and A. Mahendrasingam (unpublished) in this laboratory using synthetic polynucleotides and A. Mahendrasingam (unpublished) using DNA from bacteriophage φw-14. The diffraction patterns obtained by Marvin et al (1961) indicated that the molecules were randomly screwed up and down the C-axis, but patterns obtained in this laboratory using poly d(A-C).poly d(G-T) have crystalline reflections even on higher layer-lines.

Marvin et al (1961) proposed that a 28_3 helix with h = 3.32A and t = 38.6^{o} is the representative member of a family of similar structures. The base-pairs are moved 1.5A further away from the helix axis than in the B-DNA model of Langridge et al (1960b). The tilt and twist are -6^{o} and -5^{o} respectively. The conformation is therefore a very close relative of B-DNA.

Arnott and Selsing (1975) pointed out that some of the torsion angles were highly distorted from the accepted B-DNA values. They used a computer method to derive the co-ordinates of a model with C3'-exo sugars which was more similar to B-DNA. They also produced a 9_1 model which still maintains torsion angles similar to those of B-DNA and which agrees reasonably well with the observed diffraction. Both models increase the base tilt slightly to about -8^{o} and move the bases slightly nearer to the axis. These models have torsion angles in the class $tg^{-}g^{-}tg^{+}$ which is the same as the C3'-exo B-DNA upon which they are based; whereas the Marvin et al angles are in the class $ttg^{-}tg^{+}$. The explanation for this appears to derive from the fact that the sugars in the Marvin et al model are much closer to C2'-endo and the chain angles then conform with those of C2'-endo B-DNA (see table 2).

Precise chemical repeat units can give rise to modifications of the C group of conformations whose symmetry reflects the primary structure (Leslie et al, 1980). The alternating dinucleotide poly d(A-G).poly d(C-T) has 9_2 symmetry (designated C"-DNA) which contains 9 dinucleotide-pairs per pitch, whereas poly d(A-G-C).poly d(G-C-T) and poly d(G-G-T).poly d(A-C-C) both have 9_1 helical symmetry (C'-DNA) which has 3x3 nucleotide pairs per pitch. Poly d(A-G-T).poly d(T-C-A), which gives only classical C-forms, is an exception to this rule.

In a fibre diffraction study on the conformation of DNA in various organic solvent/water mixtures, Zimmerman and Pheiffer (1980) have found that

DNA can exhibit C-like conformations containing from 8.0 ($t = 45^{o}$) to 9.6 ($t = 37.5^{o}$) residues per turn whose pitches are 26A (h = 3.25A) and 31.8A (h = 3.31A) respectively. This further extends the range of the C-family.

1.3.1.2.4 D-DNA

Low quality diffraction diagrams obtained by Davies and Baldwin (1963) indicated that poly d(A-T).poly d(A-T) assumed a conformation with axial periodicity about 24.5A which they named D-DNA. Similar but better quality patterns were obtained from poly d(I-C).poly d(I-C) by Mitsui et al (1970). They proposed a novel left-handed double helix with antiparallel strands and Watson-Crick base-pairs. This 8_7 helix contained sugars with an unusual O5'-endo pucker. The bases, which were in the anti orientation, were placed about 2A behind the helix axis.

Arnott et al (1974) prepared specimens of D-DNA using poly d(A-T). poly d(A-T) in salt conditions which would normally be expected to yield the A-form. (They also obtained poorly oriented patterns from poly d(G-C). poly d(G-C) which appeared to be D-DNA but they have now found that this polymer gives an A-form and not the D-form (Leslie et al, 1980)). These pack into tetragonal arrays (α-D-DNA). Subsequent experiments by Selsing, Arnott and Ratliff (1975) showed that poly d(A-T-T).poly d(A-A-T) adopts the D-form packed into hexagonal arrays (β-D-DNA). The molecular conformation appears to be unchanged. Arnott et al (1974) rejected the 8_7 helix of Mitsui et al (1970). Instead they claim the molecule has 8_1 symmetry with h = 3.0A and $t = 45^{o}$. The sugars are in the standard C3'-exo pucker and the bases are tilted -16^{o}. D-DNA is a member of the B-family therefore and with its high value of t it is the most tightly wound of the right-handed double helices for which precise co-ordinates are available.

The D-form has not been observed in any natural DNA, which calls into question its biological significance. However, some DNA's (e.g. crab

satellite DNA (Sueoka and Cheng, 1962a,b) contain stretches of the highly repetitious base sequences which seem to be required for its formation, and on this basis Arnott et al (1974) have suggested that it may play a structural role in such specimens.

1.3.1.2.5 E-DNA

A new conformation for poly d(I-I-T).poly d(A-C-C) at low relative humidity has recently been observed (Leslie et al, 1980). The molecule has 3_2 symmetry with a 48.7A pitch. The molecular asymmetric unit appears to contain five nucleotides with a mean rise per residue and rotation per residue per asymmetric unit of 3.25A and 48^{o} respectively. The structure approximates to a 15_2 helix. An initial model of the molecular conformation has been suggested in which all the backbone torsion angles are in the trans range (Chandrasekaran et al, 1980; Arnott, 1980).

1.3.2 Polynucleotides which do not form Watson-Crick double helices

The remarkable flexibility of the nucleic acids is further illustrated in those structures which do not conform to the Watson-Crick stereotype.

The widespread impression that base-stacking and satisfactory backbone stereochemistry are more important than base-base hydrogen bonding in structural stability was reinforced by the model of Arnott, Chandrasekaran and Leslie (1976b) for poly(C). This 6_1 helix is single-stranded with C3'-endo sugars and all but one of the backbone torsion angles are quite close to those observed in A-DNA. The bases, which are at an angle of 21^{o} to the helix axis, are stacked but in a manner somewhat different from those observed in other polynucleotides. Further evidence in favour of the importance of base interactions is furnished by the observation that homopolymers associate with the complementary monomers to form helical structures not unlike those adopted by the corresponding polymer duplex (Smith, 1978; Chandrasekaran et al, 1980). However,

Zimmerman, Davies and Navia (1977) have discovered a 2_1 structure for poly(A) in formamide (a denaturing solvent). The bases, which are on the outside of the helix, are involved in neither base-base electronic interactions nor base-base hydrogen bonds, but the helix is stabilised by formamide molecules which form hydrogen bonding bridges between adjacent bases.

Other polynucleotides have base-base hydrogen bonding which is not of the Watson-Crick kind. For example, poly(dT) poly(dA) poly(dT) forms a three-stranded 12_1 helix. Two of the strands form Watson-Crick base-pairs and this duplex is very similar to A-DNA. The third strand, which is wound inside the deep groove, forms non-standard hydrogen bonds to the bases on the other two. The triple-helical polypurine poly(I) poly(A) poly(I) obviously cannot form standard bonds, but it still winds into a 12_1 helix which falls clearly in the A-family. Both these structures, and those of poly(U) poly(A) poly(U) (two forms), poly(U) poly(A) poly(U) and poly(dC) poly(dI) poly(dC) have torsion angles which are remarkably similar to those in A-DNA (Arnott and Bond, 1973a,b; Arnott et al, 1976a). Poly (I) and poly(G) are possibly the only multistranded polynucleotides not to be members of the A-family. Arnott, Chandrasekaran and Martila (1974) proposed that these polymers (which are isostructural) formed four-stranded 23_2 helices with C2'-endo furanose rings. The bases make an angle of 5.6^{o} with the helix axis. This structure is therefore a member of the B-family. But Zimmerman, Cohen and Davies (1975) suggested a model with A-like stereochemistry. Both groups rejected triple-helical models. Diffraction patterns from most of the polynucleotides which are not double helical consist largely of diffuse scatter and few Bragg reflections so they contain rather less useful information than the crystalline patterns from double helical specimens. In addition, potential models contain more degrees of freedom than in the double helical case,

so refinement using the conventional linked-atom least-squares technique is problematic. Instead, more emphasis is placed upon modelbuilding studies and adjudication between competing models on the basis of stereochemistry alone is imprecise.

A final example is provided by poly(s^2U) (Mazumdar, Saenger and Scheit, 1974). It is a double-stranded structure with antiparallel chains, but the base-pairing scheme imposes non-equivalence upon the two strands. This helix is similar once again to A-DNA in having 11_1 symmetry and a 28.8A pitch, but one of the chains is more compressed than in A-DNA whereas the other is more extended.

1.3.3 Recent Developments

Whilst data from fibre specimens are relatively easy to obtain the analysis presents a number of disadvantages. The low resolution of the reflections enforces a heavy reliance upon modelbuilding methods which are inevitably largely empirical. In addition the sparseness of the data imposes assumptions about high molecular symmetry - for example, any sequence dependent effects are averaged out except possibly in those special polynucleotides with highly repetitious primary structures. Finally it is impossible to observe the structure of the ions and water in fibres, both of which are likely to have sterochemically important functions.

These problems may be ameliorated by high resolution diffraction studies on single crystals. A number of self-complemetary dinucleoside phosphates have been crystallised and solved to atomic resolution: uridylyl 3',5' - adenosine phosphate (UpA) (Seeman et al, 1971; Rubin et al, 1971, 1972; Sussman et al, 1972); GpC (Day et al, 1973; Rosenberg et al, 1976); ApU (Rosenberg et al, 1973; Seeman et al, 1976) and Ca^{2+}GpC (Hingerty et al, 1976). Whilst most of these molecules retain the conventional features of polynucleotides, the possibility of significant end-effects in such short segments renders uncertain how much confidence

we may have in extrapolating their structural details to the polymer case. These doubts will be diminshed as longer fragments are synthesised, crystallised and solved.

The oligomer d(CpGpCpGpCpG) (hereafter $d(CG)_3$) is now known to crystallise in an unexpected conformation (Wang et al, 1979). This so-called Z-helix has antiparallel strands held together by Watson-Crick base-pairs and it exhibits a considerable degree of internal regularity. It is novel in being the first left-handed helix to be observed. The molecule, which has been solved to 0.9A resolution, has a zig-zag backbone with alternating sugar pucker and sugar-base orientation. The cytosine-containing nucleotides are rather like those in B-DNA with an anti base-sugar orientation, C2'-endo sugars and the C4'-C5' bond is in the g^+ orientation. Nucleotides containing guanine are different: the sugar-base orientation is syn, the pucker is C3'-endo and the C4'-C5' bond is in the t conformation. The molecular asymmetric unit is therefore a dinucleotide so there are no diad axes in the planes of the bases. The base stacking is somewhat different from that observed in polymers: the base-pairs are sheared with respect to each other so that whilst the cytosines are stacked on the guanines, the guanines are unstacked and instead they interact with the O5' of an adjacent sugar. A final distinguishing feature is provided by the grooves. Since the bases are pulled away from the axis only one, deep, groove is observed which corresponds to the minor groove in B-DNA.

A closer study of the structure by Wang et al (1981) has revealed two different phosphate orientations. All phosphates in CpG sections and the majority in GpC sections are identical. But a minority of the phosphates are rotated about 1A away from this position to form a hydrogen bond with a magnesium ion. They have called the majority conformation Z_I and the minority one Z_{II}. The phosphodiester orientation is g^-t in the former

and g^+t in the latter. Wang et al (1981) proposed that a family of Z-helical structures can exist with varying proportions of Z_I and Z_{II}. An explanation for the more frequent observation of Z_I in $d(CG)_3$ is provided by a hydrogen bonding bridge from guanine to a nearby phosphate. The guanine amino group hydrogen which does not participate in base-pairing to the cytosine O2 hydrogen bonds instead to a water molecule which is bound to a phosphate group oxygen. A similar bridge occurs in the Z_{II} conformation but now two water molecules form a chain from the base to the phosphate. This latter structure is likely to be weaker and therefore less stable than the first. It is interesting that an AT base-pair could not form such a bridge at all and this might explain why the Z-helix has not been observed in AT oligomers.

The oligomers stack upon each other along the c-axis thereby approximating to an infinite molecule in which one in every six phosphate groups has been systematically removed. Whilst small distortions in the structure are evident from one nucleotide to the next, the most dramatic is that associated with the two phosphate conformations. Wang et al (1981) have produced idealised co-ordinates for infinite, regular Z_I and Z_{II} helices. The zig-zag structure in the Z_{II} helix is even more pronounced than that in Z_I. The torsion anglesof both helices are recorded in table 2. Since the molecular symmetry is 6_5 with a 44.6A pitch, the rotation and rise per dinucleotide are -60^o and 7.4A respectively. The base planes are tilted 7^o from the helix axis. The mean radius of the phosphate groups is about 9A so the molecule is slightly slimmer than B-DNA in which the phosphates are about 10A from the axis. The phosphate groups are also closer together in the Z-helices than in B-DNA so screening by cations may be important.

The similar molecule d(CpGpCpG) (hereafter $d(CG)_2$) crystallises in two forms (Drew, Dickerson and Itakura, 1978) with a reversible transition

between the low salt hexagonal lattice and the high salt orthorhombic lattice. The latter structure has been solved to 1.5A resolution by Drew et al (1980). The molecular conformation, which has been called Z'-DNA, is very similar to the Z-helix. The guanosines are in the syn orientation and the cytosines are anti so once again there is a zig-zag sugar-phosphate backbone wherein the rotation per dinucleotide is -60^{o}. The mean internucleotide rise is 3.8A (which is slightly higher than that in the Z-helix) and the bases are tilted 9^{o}. The major differences between the Z and Z' structures are in the sugar puckers. The Z'-helix contains C2'-endo sugars at cytosine (as in Z-DNA) but the C1'-exo pucker in guanosine is unusual. Since C1'-exo is a small structural variation of C2'-endo, the sugar puckers are relatively invariant along the backbone, in contrast to the situation in the Z-helix. The helical axes are not collinear with any of the crystallographic axes so no molecular stacking occurs.

The hexagonal crystals of $d(CG)_2$ have been studied by Crawford et al (1980) who obtained two different forms, one of which contained spermine ions and one without. They found that the molecular conformation was very similar to the Z-helix observed with the hexamer: the base orientation differed slightly and the cytosines were moved a little further from the helix axis. In both crystals the tetramers stacked to form quasi-continuous helices as in hexameric Z-DNA but unlike the orthorhombic Z'-DNA tetramers. Crawford et al (1980) proposed that the differences between the stacking of the two types of helix were due to the different ions found in the two crystals. Both Z_I and Z_{II} phosphate orientations were present in the $d(CG)_2$ molecules. In both the orthorhombic hexamer (which is in space group $P2_12_12_1$) and the orthorhombic tetramer (space group $C222_1$) there are two molecules per unit cell which are required by the crystal symmetry to be identical. But in the hexagonal $d(CG)_2$ there are three

molecules per unit cell, those at the corners requiring 6_5 screw axes whereas those at the centres require only 3_2 axes. Nonetheless the two types of molecule have very similar conformations.

Seven crystals containing d(CG) oligomers have now been solved at resolutions from 1.8A to 0.9A containing various species and concentrations of cations. The small differences between the family of molecules in these crystals appear to be due to the cations (Crawford et al, 1980).

The hexagonal $d(CG)_2$ crystals were the first of these oligomers to be studied. The surprising nature of the structure is illustrated by the fact that the solution was firstly attempted with molecular replacement methods (Blundell and Johnson, 1976) using A-DNA, B-DNA, C-DNA and D-DNA as trial molecules. When this method failed isomorphous derivatives were used in the solution.

It is obviously of interest to enquire whether the unusual behaviour of these oligomers can also be expressed in the polymer poly d(G-C). poly d(G-C). This synthetic polynucleotide has been observed usually in the A- and B-forms (Leslie et al, 1980) but Arnott et al (1980) have discovered that the sodium salt occasionally forms a statistically disordered crystal whose diffraction can be explained by a molecular conformation containing many of the characteristic features of the Z and Z'helices. The molecule (which has been called S-DNA) has 6_5 symmetry, a dinucleotide repeat unit and the pitch is 43.5A. The sugar-base orientation is syn at guanine and anti at cytosine. Whilst the cytosine nucleotides adopt a conformation not unlike those observed in A- and B-DNA, the guanosines are quite different with ε in the trans range (rather than gauche$^+$ in both A- and B-DNA) and β and γ in the gauche$^+$ range (rather than trans and gauche$^-$ in A- and B-DNA). The bases, which are positioned on the helix axis (unlike Z- and Z'-DNA) are tilted -5^o, are twisted by different amounts from the zero plane so the angle between the helix axis and the

normal to the base-planes is 18^{o} for guanine and 7^{o} for cytosine. A further feature which distinguishes S-DNA from the Z-helix is the fact that the molecule has two deep grooves which is a consequence of the base position and orientation. Arnott et al (1980) have also re-interpreted the patterns obtained from poly d(A-s^4T).poly d(A-s^4T) by Saenger et al (1973) in terms of a similar model with h = 7.6A and t = -51.4^{o}. Once again the purines have rather unorthodox stereochemistry.

Examination of table 2 reveals some interesting features about these four left-handed helices. All the conformation angles are in the same ranges in the CpG sections. In GpC stretches δ, ε, ζ and χ are in the same ranges but the conformations are distinguished by the angles α, β and γ. If we use the symbol "$\backsimeq$" to denote "in the same range" then the results may be summarised in the form:-

$$\alpha: \quad S \backsimeq Z_I \neq Z_{II}$$
$$\beta: \quad S \backsimeq Z_I \neq Z_{II}$$
$$\gamma: \quad S \neq Z_I \backsimeq Z_{II}$$

So the S helix is conformationally more similar to Z_I than to Z_{II}. Wang et al (1981) have calculated the Fourier transform of the Z_I helix and they find it to agree well with the diffraction pattern observed from poly d(G-C).poly d(G-C).

Since they also observed poly d(A-C).poly d(G-T) in the S-form, Arnott et al (1980) suggested that it was accessible to any DNA with alternating purine-pyrimidine sequences. However, poly d(A-T).poly d(A-T) which typically adopts the B and D conformations (Leslie et al, 1980) has not been reported in the S-form.

Poly d(G-C).poly d(G-C) was known to exhibit unusual optical properties in solution even before the crystallographic studies were undertaken. The experiments of Pohl and co-workers (Pohl, 1971; Pohl and

Jovin, 1972; Pohl et al, 1973; Pohl, 1976) using laser Raman spectroscopy, optical rotatory dispersion and ultraviolet circular dichroism indicated that a reversible and co-operative transition took place in the polymer in aqueous solution when the salt concentration was raised to 2.5M NaCℓ 1.8M $NaClO_4$ or 0.7M $MgCl_2$. No transition was observed with poly(dG).poly(dC), poly(G).poly(C) or poly d(I-C).poly d(I-C). Optical experiments give only rather indirect information concerning the detailed molecular structure, therefore Pohl and co-workers, in the absence of any models determined by X-ray diffraction which could explain the almost complete inversion of the circular dichroism signal during the transition, were content merely to refer suggestively to the low salt conformation as the R-form and the high salt conformation as the L-form.

Additional evidence for the Z-helix has recently come from ^{1}H and ^{31}P NMR spectroscopy. NMR signals from polymers tend to be rather broad so Patel et al (1979) concentrated on the oligomers $d(C-G)_n$ with n = 8, 10-15. The backbone conformation was monitored by ^{31}P NMR as a function of salt concentration. At low salt (approximately 1M NaCℓ) a single resonance was observed but at 4M NaCℓ two resonances appeared. NMR signals are sensitive to the chemical environment of the resonating nucleus so these results imply that there are two types of phosphate in the high salt solution whilst all phosphates are identical in low salt. Parallel ^{1}H NMR studies charted the chemical shift of the H_1' and H_3' resonances as the salt concentration was varied. These shifts are sensitive to the sugar-base orientation and the pucker respectively. The results indicated that both these parameters also fall into two distinct salt dependent classes. The circular dichroism results from the oligomers were also in agreement with those observed by Pohl's group from the polymer. These findings are clearly consistent with a B → Z transition as the salt concentration is increased.

Two peaks of approximately equal area are found in the ^{31}P NMR

spectrum from a 145 base-pair segment of oligo d(A-T) in low salt (<0.1M NaCℓ) solution whilst poly d(A-T).poly d(A-T) has only one, broad signal (Shindo, Simpson and Cohen, 1979). A similar experiment with 145 base-pairs of random sequence also revealed only one peak. This suggests that there might be a sequence-dependent local variation in conformation.

Solid state ^{31}P NMR signals from a poly d(A-T).poly d(A-T) fibre which gives B-DNA diffraction patterns also shows splitting when the magnetic field is parallel to the fibre whereas a calf thymus fibre under the same conditions exhibits only one, broad signal (Shindo and Zimmerman, 1980). The split signal contains two equal intensity peaks indicating that the phosphate groups have an equal probability of being in one of two orientations. These results, which are in contrast to those on GC oligomers, suggest that the high salt structure is regular, and therefore not Z-DNA, whereas an alternating conformation is present in low salts. The nature of this alternating structure has recently been visualised.

The tetramer d(ApTpApT) crystallises in a form quite unlike those observed in GC oligomers. The molecule contains Watson-Crick base-pairs but it does not form a mini-helix with four base-pairs. Instead, the first two nucleotides pair with a second molecule, but between the second and third nucleotides the backbone changes direction sharply to form two base-pairs with a third molecule (Viswamitra et al, 1978). Poly d(A-T).poly d(A-T) binds the *lac* repressor protein of *Escherichia coli* about 100-1000 times more strongly than does calf thymus DNA, (Riggs et al, 1972). In order to explain this discrimination and the NMR results, Klug et al (1979) have proposed an alternating structure for the polymer (named "alternating-B") based upon the tetramer structure. The sugar pucker in the adenosine residue is C3'-*endo* and that in the thymine residue is C2'-*endo*. The helical symmetry is therefore 5_1 with a dimer repeat. Thymine bases are well stacked upon adenines but not *vice*

versa. The structure was refined by the same procedure as that used by Levitt in the B-DNA studies described earlier and was found to be energetically no less favourable than regular B-DNA. The differences between the two structures are not sufficient to enable them to be discriminated with the quality of the X-ray fibre diffraction patterns generally observed. The torsion angles are given in table 2.

A most dramatic increase in our knowledge of B-DNA structure is likely to come from a detailed study of the complementary dodecamer d(CGCGAATTCGCG). Initial details of this dodecamer have been published by Wing et al (1980). This sequence is of particular interest since it contains the minimal recognition site of the EcoRI restriction endonuclease, d(GAATTC), flanked by CG sequences which have crystallised as left-handed helices. The cation content of the crystal was sufficient to produce such left-handed segments and indeed Dickerson's group originally expected to find a B-like central segment with left-handed ends but the molecule is actually completely right-handed with a structure similar to that in B-DNA. The helix, which was solved to 1.9A resolution, contains 10.1 base-pairs per turn with a mean internucleotide separation of 3.4A. No attempt was made to impose standard sugars but they all appear to be C2'-endo or O5'-endo with no sequence - dependent effects. However, assignment of the pucker at this relatively low resolution is uncertain. There are some departures from the classical B-DNA conformation. The bases, which are all anti, are propellor-twisted as predicted by Levitt (1978) implying that the molecule is perhaps similar to that which exists in solution. The helix axis is slightly curved, possibly due to lattice effects.

Whilst the structure of DNA is of immense interest, it must be remembered that it functions in the presence of proteins, often by specific recognition. The duplex fragment d(GAATTC) complexed with the EcoRI restriction endonuclease has now been crystallised (Young et al, 1981).

There appear to be four enzyme molecules and two DNA duplexes in each asymmetric unit of the cell which is in space group $P42_12$. This is the first example of a complex between a protein and a specific recognition sequence of DNA giving crystals of suitable quality for high resolution diffraction analysis. The enzyme has also been crystallised alone and a determination of its structural changes on binding may be of fundamental importance.

1.4 Outline of the Present Project

It is a propitious moment to re-examine the structure of DNA in fibres. The diffraction studies on oligonucleotides have provided the first examples of irregular backbones and left-handed helices. On the technical side, advances have been made not only in precise computer modelbuilding but also in the measurement of diffraction data. How confident can we be that the classical double helical models are correct? Could DNA be left-handed? These questions are discussed in chapter 3. Several left-handed Watson-Crick models have been built and compared with the observed diffraction. In the course of these studies an "inverted base-stacking" scheme has been discovered which increases the range of conformations accessible to the double helix. The stereochemical feasibility of transitions between these models is discussed.

A novel model for DNA which is dramatically different from the double helix has been proposed by workers in New Zealand (Rodley et al, 1976) and India (Sasisekharan et al, 1978a). The sugar-phosphate chains do not intertwine as in the Watson-Crick model but zig-zag along the molecular axis in short alternating stretches of left- and right-handed helix. The rationale behind this idea is to reduce the topological difficulties which are thought to be present during DNA replication. The stereochemistry and predicted diffraction from so-called side-by-side (SBS) models are examined in chapter 4.

The New Zealand group have proposed the use of the Patterson function and modifications of it to adjudicate between double helical and SBS models. This idea is examined in chapter 5.

The bacteriophage ϕw-14 contains chemically modified bases which may be of importance in the life-cycle of the virus or in the packing of the DNA into the phage head. Chapter 6 contains a discussion of the crystal and molecular structure of this DNA.

When they are stretched during drying DNA fibres sometimes give a diffraction pattern which has not been satisfactorily interpreted. Several models are considered in chapter 7.

Table 1.1 - Helical Parameters

	Symmetry	Number of nucleotides in asymmetric unit	Pitch (A)	Mean rise per nucleotide (A)	Mean turn per nucleotide (degrees)	Ref.
A -DNA	11_1	1	28.15	2.56	32.7	1
A -RNA	11_1	1	30.9	2.81	32.7	2
A'-RNA	12_1	1	36.2	3.02	30.0	2
A*-RNA	10_1	1	31.0	3.1	36.0	3
B -DNA	10_1	1	33.8	3.38	36.0	4
B'-DNA	10_1	1	32.9	3.29	36.0	5
C -DNA	28_3	1	31.0	3.32	38.6	6
C'-DNA	9_1	1	29.5	3.28	40.0	7
C"-DNA	9_2	1	58.2	3.23	40.0	7
D -DNA	8_1	1	24.3	3.04	45.0	8
E -DNA	3_2	5	48.7	3.25	48.0	7
S -DNA	6_5	2	43.5	3.63	-30.0	9
Z -DNA	6_5	2	44.6	3.70	-30.0	10
Z'-DNA	6_5	2	45.6	3.80	-30.0	11

References for Table 1.1

1. Fuller et al (1965) J. Mol. Biol., 12, 60.
2. Arnott et al (1973) J. Mol. Biol., 81, 107.
3. Chandrasekaran et al (1980) in Fibre Diffraction Methods, ACS141.
4. Langridge et al (1960) J. Mol. Biol, 2, 19.
5. Arnott and Selsing (1974) J. Mol. Biol., 88, 509.
6. Marvin et al (1961) J. Mol. Biol., 3, 547.
7. Leslie et al (1980) J. Mol. Biol., 143, 49.
8. Arnott et al (1974) J. Mol. Biol., 88, 523.
9. Arnott et al (1980) Nature, 283, 743.
10. Wang et al (1979) Nature, 282, 680.
11. Drew et al (1980) Nature, 286, 567.

TABLE 1.2 : Conformational Parameters of Polynucleotide Duplexes. (i) The A Family

	CLASS	α	β	γ	δ	ε	ζ	τ_{CN}	tilt	twist	D(Å)	Ref
Fuller A	$tg^-g^-tg^+$	-139.3	-81.2	-76.9	167.4	67.5	76.5	76.7	20.0	-8.0	4.25	1*
Arnott A1	$tg^-g^-tg^+$	178.1	-46.9	-84.7	-151.6	45.3	83.2	86.4	20.0	-1.2	4.72	2
Arnott A2	$tg^-g^-tg^+$	175	-45	-90	-149	47	83	-154^+	-	-	-	3
A-RNA	$tg^-g^-tg^+$	-151.1	-74.3	-62.1	179.8	48.1	83.2	74.9	16.0	-6.9	-	4
A'-RNA	$tg^-g^-tg^+$	-167.5	-59.9	-65.2	-167.0	43.6	83.2	79.5	10.0	-7.6	-	4
poly rI. poly dC	$tg^-g^-tg^+$	-166.1	-64.7	-75.5	-179.1	61.9	77.7	75.0	6.2	-7.0	4.85	5*
A^*-RNA	tg^-ttt	-	-	-	-	-	C3'-endo	-	0	-	0	6

TABLE 1.2 : Conformational Parameters of Polynucleotide Duplexes. (ii) The B Family

	CLASS	α	β	γ	δ	ε	ζ	τ_{CN}	tilt	twist	D(Å)	Ref
Crick-Watson B	tg^-ttt	-174.0	-64.3	-178.1	179.1	145.3	99.8	-	0	0	∿-3.5	7*
Langridge B3	$tg^-g^-tg^+$	147.2	-78.0	-79.3	-147.7	33.7	129.8	138.2	-2.0	0	-0.69	8*
Arnott B1	$tg^-g^-tg^+$	154.7	-95.6	-46.1	-146.5	36.4	156.5	142.5	-5.9	-2.1	-0.16	2,9
Arnott B2	ttg^-tg^+	-165.7	-135.9	-25.3	159.9	26.8	146.2	138.6	-0.1	-3.8	-0.14	9
Arnott B3	ttg^-tg^+	-133	-157	-41	136	38	139	-102^+	-	-	-	3
Levitt B	$tg^-g^-tg^+$	178	-85	-65	170	65	108	47^{++}	∿-17	∿-14	-	10
B'-DNA	$tg^-g^-tg^+$	145.3	-86.9	-52.1	-136.4	39.4	156.5	-144.5	-7.9	-1.0	-0.01	11
Marvin C	ttg^-tg^+	212.4	-148.1	-45.8	-216.7	48.0	140.4	139.2	-6.0	5.0	-2.24	12*
Arnott $2B_3$ C	$tg^-g^-tg^+$	160.6	-105.8	-38.7	-160.0	37.2	156.5	143.1	-8.0	1.0	-0.90	13

Cont.

Cont.

	CLASS	α	β	γ	δ	ε	ζ	τ_{CN}	tilt	twist	D(Å)	Ref
Arnott 9_1C	$tg^-g^-tg^+$	167.1	-116.3	-33.6	-172.1	37.9	156.5	143.6	-8.1	1.8	-1.50	13
D-DNA	$tg^-g^-tg^+$	141.4	-100.5	-61.8	-152.0	68.6	156.5	143.9	-16.0	5.6	-1.80	14
E-DNA	ttttt	-	-	-	-	-	-	-	-	-	-	6

TABLE 1.2 : Conformational parameters of polynucleotide duplexes (iii) Alternating B-DNA

Nucleotide Sequence	Class	α	β	γ	δ	ε	ζ	τ_{CN}	Ref.
5'TpA	$tg^+g^-tg^+$	-159.4	111.9	-60.1	151.2	58.9	99.1	90.5	16
5'ApT	$tg^-g^-tg^+$	-168.4	-81.9	-67.3	171.6	64.2	143.7	140.3	16

TABLE 1.2 : Conformational Parameters of Polynucleotide Duplexes. (iv) Left-handed Helices

	Nucleotide Sequence	Class	α	β	γ	δ	ε	ζ	τ_{CN}	Ref
S-DNA poly d(G-C). poly d(G-C)	5'CpG	$g^-g^+g^+tt$	-72	102	52	-153	178	76	89^+	3
	5'GpC	$g^-g^-g^-tg^+$	-103	-91	-110	-168	54	147	-159^+	3
poly d(As^4T). poly d(As^4T)	5's^4TpA	$g^-g^+g^+tt$	-92	99	60	-158	161	80	85^+	3
	5'Aps^{4}T	$g^-g^-g^-tg^+$	-106	-77	-96	-148	36	135	-154^+	3
Z_I	5'CpG	$g^-g^+g^+tt$	-94	80	47	179	-169	99	68^+	15
	5'GpC	$g^-g^-ttg^+$	-104	-69	-137	-139	56	138	-159^+	15
Z_{II}	5'CpG	$g^-g^+g^+tt$	-100	74	92	-167	157	94	62^+	15
	5'GpC	$g^-g^+ttg^+$	-179	55	146	164	66	147	-148^+	15

References for Table 1.2

1. Fuller et al (1965) J.Mol.Biol., 12, 60.
2. Arnott and Hukins (1972) Biochem. Biophys. Res. Commun., 47, 1504
3. Arnott et al (1980) Nature, 283, 743.
4. Arnott et al (1973) J.Mol.Biol., 81, 107.
5. O'Brien and MacEwan (1970) J.Mol.Biol., 48, 243.
6. Chandrasekaran et al (1980) in Fibre Diffraction Methods, ACS141.
7. Crick and Watson (1954) Proc. Roy. Soc., 223A, 80.
8. Langridge et al (1960) J.Mol.Biol., 2, 38.
9. Arnott and Hukins (1973) J.Mol.Biol., 81, 93.
10. Levitt (1978) Proc. Nat. Acad. Sci. U.S.A., 75, 641
11. Arnott and Selsing (1974) J.Mol.Biol., 88, 509.
12. Marvin et al (1961) J.Mol.Biol., 3, 547.
13. Arnott and Selsing (1975) J.Mol.Biol., 98, 265.
14. Arnott et al (1974) J.Mol.Biol., 88, 523.
15. Wang et al (1981) Science, 211, 171.
16. Klug et al (1979) J.Mol.Biol., 131, 669.

Key

The torsion angles of models marked with an asterisk were calculated by the author from the published co-ordinates.

The angle τ_{CN} is defined by the atoms C2', C1', N9 and C4 (for a purine base) except where marked + when it is defined by O5', C1', N9 and C4 or ++ when it is defined by O5', C1', N9, C8.

CHAPTER II

TECHNIQUES IN NUCLEIC ACID CRYSTALLOGRAPHY

This chapter collects together the techniques which will be of particular importance in this thesis. First the preparation and nature of DNA fibres are described. The following section outlines the theory of X-ray diffraction with particular emphasis on helical objects and the strategy for deducing molecular structure from a diffraction pattern. Finally, the experimental methods are discussed including a description of the design, writing and testing of a number of computer programmes, most of which were written by the author. A number of other programmes which were not of general significance were written and they will be described at the appropriate points in the text.

2.1 DNA Extraction and Purification

A large number of DNA's are now available from commercial sources. Calf thymus DNA (42% G-C base-pair content) was obtained from Sigma Chemical Company, Miles Research Laboratories and BDH Chemicals Limited. DNA from bacteriophage ϕw-14 was supplied by Professor R.A.J. Warren of the University of British Columbia, Vancouver.

DNA is generally found in the cell in conjunction with other molecules, particularly proteins, and the extraction processes commonly introduce various inorganic ions into the sample. It is desirable to remove as many impurities as possible using a standard technique. In this laboratory DNA is purified according to a modified version of Massie and Zimm's (1956) "hot phenol" extraction. Analar grade phenol was freshly distilled prior to the DNA purification and the distillate was dropped into a 0.1M NaCl solution. After shaking, the flask was allowed to stand and the phenol separated out into a layer at the bottom of the solution.

DNA dissolves rather slowly in high salt concentrations and can be damaged in low salt solutions. As a compromise the DNA was added to 0.002M NaCℓ. When the DNA had dissolved, 2M NaCℓ was added to raise the salt concentration to 0.1M. The DNA solution was then added to an equal volume of the fresh phenol and the mixture was shaken gently for twenty minutes before being centrifuged at 3000 r.p.m. for quarter of an hour. The upper layer was then carefully removed using a U-shaped Pasteur pipette to avoid drawing the phenolic phase through the interface and was centrifuged again if necessary. The DNA can then be precipitated by addition of an equal volume of cold propan-2-ol. The precipitate was wound onto a glass rod and washed in 80% ethanol, 95% ethanol and acetone.

2.2 Preparation of Samples for X-ray Analysis

DNA, in common with many other high polymers, does not form macroscopic single crystals of high quality. However, Wilkins (1962) discovered that it could be drawn into fibres which often exhibit a high degree of internal regularity.

The following procedure describes the preparation of a typical fibre of the sodium salt, but it may be desirable to alter the type and concentration of the counter ion.

Purified DNA was dissolved in 0.01M tris-HCℓ/0.01M NaCℓ, pH 7.6 buffer solution. The concentration of tris was kept sufficiently low to ensure that the major ionic component of the solution was NaCℓ whilst maintaining the pH at 7.6 $\pm$ 0.2 which is near physiological levels. The concentration of DNA was about 1mg/ml. The solution was spun for 12 hours at 40,000 r.p.m. in a 10 x 10ml angle rotor on an MSE 50 ultracentrifuge. A gel formed at the bottom of the tube and the supernatant was gently poured out. The ultraviolet absorption spectrum of the supernatant usually indicated that over 95% of the DNA had sedimented. Gels prepared in this way may be stored at 4^{o}C in sealed tubes for several weeks.

Fibres were pulled from gels using a frame (figure 1) described by Fuller et al (1967). Two glass rods of diameter about 150μ were prepared in a bunsen flame and mounted in plasticine as shown. The separation of the rods may be varied without altering their orientation by means of the knurled wheel. About 0.1mg of the gel was placed between the rods and was allowed to dry. The precise conditions required for making good fibres are rather difficult to describe since fibre-pulling is rather more an art than science - in general, if the fibre is stretched at all it should be done gradually and gently. The quality of the fibres can sometimes be improved by drying at 4°C or in the presence of a controlled atmosphere at a specific relative humidity.

Thick fibres tended not to be well oriented whilst thin ones required long exposure times to obtain a useful diffraction pattern. A reasonable compromise may be achieved if fibres are about 100-150μ in diameter.

2.3 The Nature of DNA Fibres

In the most highly ordered DNA fibres the molecules aggregate into microscopic crystals (crystallites) within which the arrangement of material is completely regular in three dimensions. Since any DNA molecule will be much larger than an individual crystallite, it seems likely that the molecules within a fibre are threaded from one crystallite to another via a relatively amorphous matrix. It is not possible to isolate the crystallites for study : the best that can be achieved is a fibre in which the long axes of the crystallites are parallel to the fibre axis but the azimuthal orientations are random. Both the sodium and lithium salts of DNA can give crystalline specimens (Fuller et al, 1965; Langridge et al, 1960a).

Many fibres exhibit disorder which may take a number of forms. Suppose for simplicity that the molecules are arranged on a regular two-

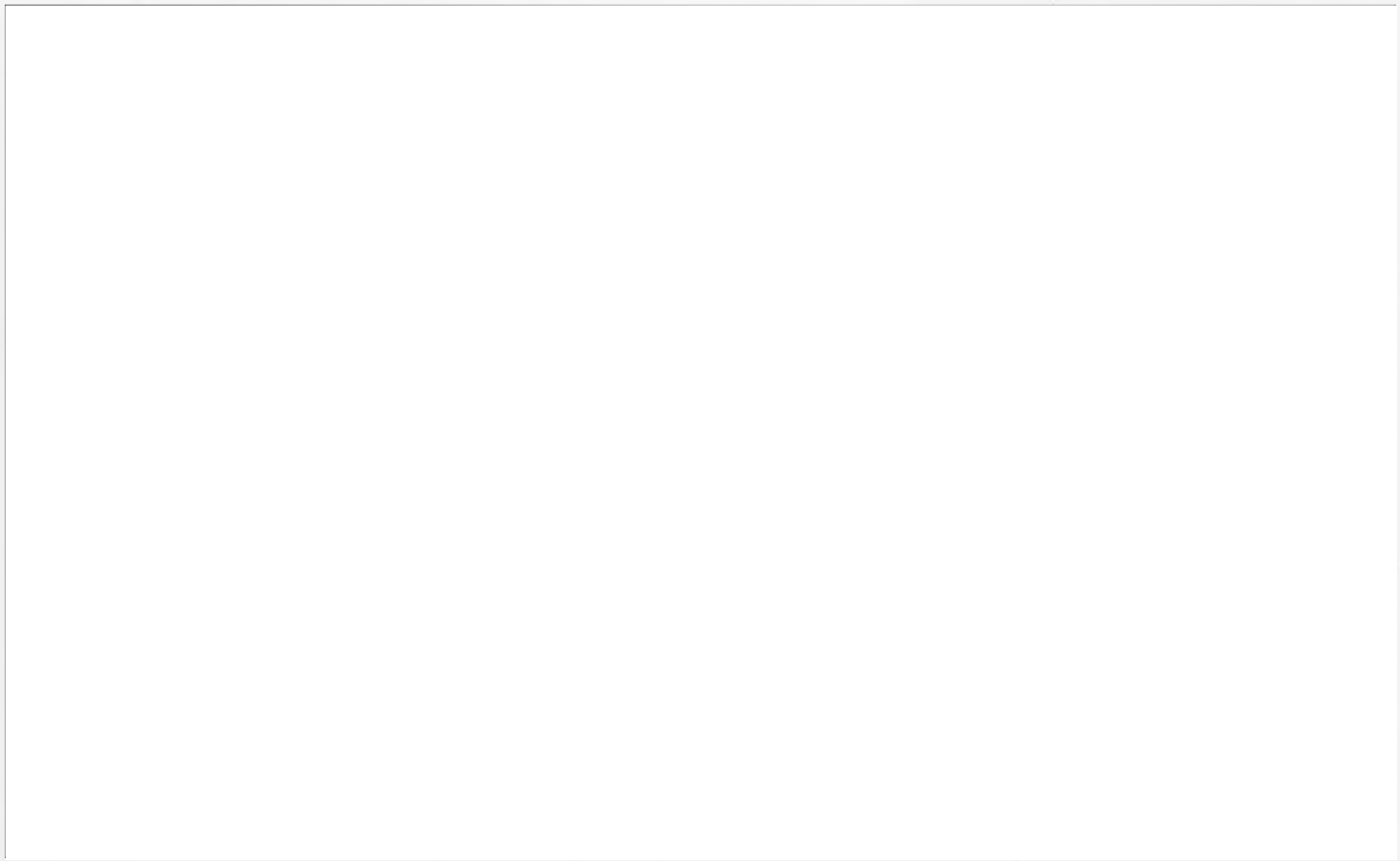


Figure 2.1 : Fibre Cell (from Fuller et al, 1967)

dimensional array perpendicular to the c-axis and that $\Delta\phi_i$ and Δz_i are the azimuthal orientation and displacement of the i th molecule with respect to some reference molecule. If $\Delta\phi_i = \Delta z_i = 0$ for all values of i then clearly the specimen is a crystalline one as described in the previous paragraph. Slippage disorder occurs if $\Delta\phi_i = 0$ for all i but the Δz_i have random values. In a specimen with rotation disorder $\Delta z_i = 0$ for all i and the $\Delta\phi_i$ have random values. Such a specimen, and also one in which $\Delta\phi_i$ and Δz_i are random and uncorrelated, is often said to be "oriented". If $\Delta\phi_i = \Delta z_i$ where the $\Delta\phi_i$ are random, the specimen is said to be screw-disordered. This is particularly likely to occur if the molecules are non-integral helices so they cannot arrange themselves to give an identical pattern of contacts with their neighbours in each pitch length. Instead they behave more like smooth interlocking helices any one of which could be screwed out of the array without disturbing the others (Klug and Franklin, 1958; Marvin et al, 1961). Naturally it is also possible to relax the assumptions that the molecules form regular two-dimensional arrays and that their c-axes are parallel; indeed even the best crystalline specimens show at least a small degree of disorientation about the long axis.

We must also consider disorder at the molecular level. We have implicitly assumed that all DNA helices are perfectly regular and this may be invalid particularly, for example, in the case of crystalline fibres wherein intermolecular interactions may produce small distortions. In addition, DNA fibres contain many water molecules and inorganic ions which may not have the same symmetry as the DNA. Finally, the atoms will all be in motion about their mean positions due to their thermal energy.

Each of these distortions will have concomitant effects upon the observed diffraction for which correction must be made during determination of the structure. A fuller treatment of this subject has been given by Vainstein (1966).

2.4 X-Ray Diffraction Theory

2.4.1 Introduction

Image formation is a two stage process. First radiation is scattered from an object, then the scattered waves are recombined to form the image. In order to examine the fine details of the object we need radiation whose wavelength, λ, is of the same order of magnitude, so molecular structure can only be determined using radiation with $\lambda \simeq 1A$. In the electromagnetic spectrum this corresponds to the X-ray region. But the refractive index with respect to X-rays in any medium differs from unity by no more than one part in 10^7 so the second stage of the image formation process is impossible since we cannot focus the scattered waves. We can however perform the recombination of the waves manually. In order to achieve this we need to find the relationship between the scattering material and the scattered field.

In figure 2 $\underline{S}_0$ is the wave-vector of an incident wave and $\underline{S}_1$ is the scattered wave. We set our origin at O and at P with position vector $\underline{r}$ is a small element of volume $d^3\underline{r}$ with scattering power $\rho(\underline{r})$. The phase difference between the incident and scattered waves is:-

$$\frac{2\pi}{\lambda}\ \underline{r}\ .(\underline{S}_0 - \underline{S}_1) = \frac{2\pi}{\lambda}\ \underline{r}\ .\underline{S} \qquad (1)$$

where $\underline{S} = \underline{S}_0 - \underline{S}_1$. (2)

$\underline{S}$ as defined in equation 1 is a dimensionless quantity measured in reciprocal lattice units (RLU). It will frequently be more convenient to subsume λ in $\underline{S}$ since this will simplify many of the equations we will use. It will be obvious from the context which definition is being used at any time. The amplitude of scattering from the element will be:-

$$dG(\underline{S}) = \rho(\underline{r})\ \exp(2\pi\ \underline{r}\ .\underline{S})\ d^3\underline{r} \qquad (3)$$

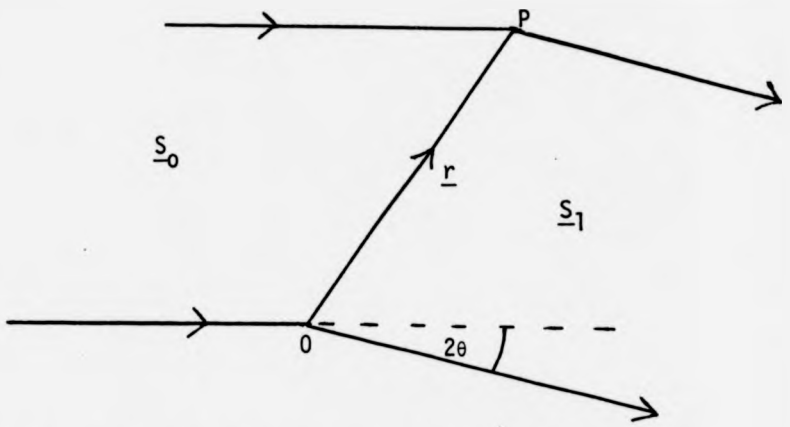


Figure 2.2 : The Geometry of the Scattering Process

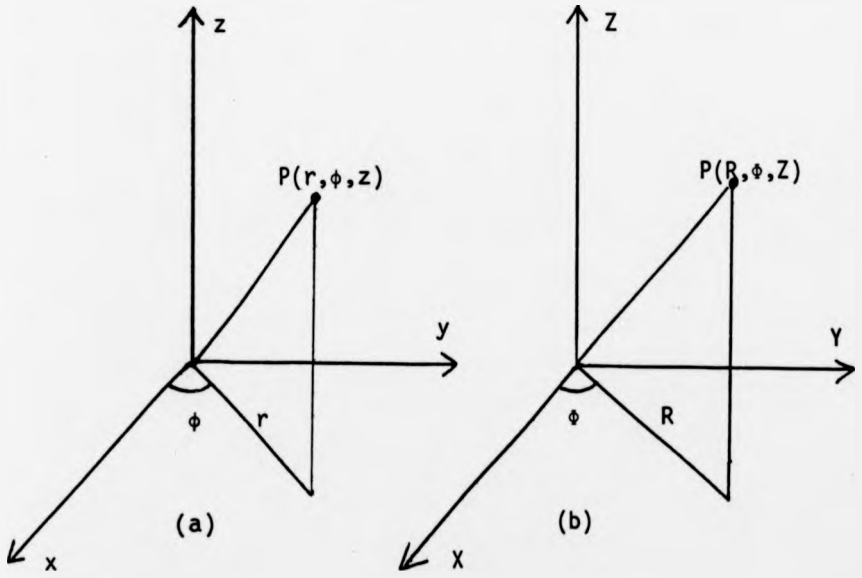


Figure 2.3 : Cylindrical Polar Co-ordinates in (a) Real Space and (b) Reciprocal Space

and so the total scattering due to the whole object is:-

$$G(\underline{S}) = \int_V \rho(\underline{r}) \exp(2\pi \ \underline{r} . \underline{S}) \, d^3\underline{r} \qquad (4)$$

where the integration is taken over the whole volume, V, of the object. This expression clearly shows why the wavelength of the radiation should be similar to the detail we wish to observe, for if λ were very large with respect to $\underline{r}$ then a big change in $\underline{S}$ would be required before $G(\underline{S})$ showed any variation.

It is clear from equation 4 that the amplitude of scattering is related to the scattering power by Fourier transformation. This important result, which means that the powerful methods of Fourier analysis are available to us, can be exploited immediately.

In X-ray scattering the scattering power is proportional to the electron density function. Now we may obtain certain advantages if we use a specimen which is highly ordered. We define a function P which is generated by the convolution of two functions f and g:-

$$P(\underline{u}) = \int f(\underline{r}) \, g(\underline{r} - \underline{u}) \, d^3\underline{r} \qquad (5)$$

where $\underline{u}$ is an arbitrary variable. If we replace f by $\rho(\underline{r})$, the electron density distribution of the unit cell, and g by a three-dimensional array of δ-functions whose value is unity at the lattice points and zero elsewhere, then $P(\underline{u})$ represents the electron density distribution of the entire crystal. But a general result of Fourier transform theory is that:-

$$T\,(f * g) = T\,(f)\,T(g) \qquad (6)$$

where T denotes the Fourier transform operation and * denotes the convolution operation.

So the Fourier transform of a crystal is the Fourier transform of the motif multiplied by the Fourier transform of the lattice. Now the

motif is a general arrangement of atoms and equation 4 shows that its Fourier transform will be a continuous function. The lattice function may be written in the form:-

$$g(\underline{r}) = \sum_{p,q,r} \delta(\underline{r} - (p\underline{a} + q\underline{b} + r\underline{c})) \qquad (7)$$

where $\underline{a},\underline{b},\underline{c}$ are the repeat distances along x,y and z and p,q and r are integers. Insertion of (7) into equation 4 gives:-

$$G(\underline{S}) = \sum_{h,k,\ell} \delta(\underline{a}.\underline{S}-h)\ \delta(\underline{b}.\underline{S}-k)\ \delta(\underline{c}.\underline{S}-\ell) \qquad (8)$$

which is non-zero only when:-

$$\underline{a}.\underline{S} = h, \qquad \underline{b}.\underline{S} = k, \qquad \underline{c}.\underline{S} = \ell \qquad (9)$$

These equations (known as the Laue equations) show that the diffracted amplitude from the crystal will only be non-zero when:-

$$\underline{S} = h\underline{a}^* + k\underline{b}^* + \ell\underline{c}^* \qquad (10)$$

where $\underline{a}^* = \underline{b} \times \underline{c}/\underline{a}.\underline{b} \times \underline{c}$

$\underline{b}^* = \underline{c} \times \underline{a}/\underline{b}.\underline{c} \times \underline{a}$

$\underline{c}^* = \underline{a} \times \underline{b}/\underline{c}.\underline{a} \times \underline{b}$

$\underline{a}^*$, $\underline{b}^*$ and $\underline{c}^*$ are the basis vectors of the reciprocal lattice. It may easily be shown that $\underline{S}$ is perpendicular to the plane in real space with Miller indices (hkℓ). So the continuous Fourier transform of the discrete motif is "sampled" at points in reciprocal space where $\underline{S}$ is a reciprocal lattice vector. But the total radiation scattered over a solid angle of 4π steradians depends only on the intensity of the main beam, not on the distribution of scattering material (neglecting absorption) so the advantage we obtain in using a crystalline specimen is that all the radiation is concentrated into spots which significantly reduces the required exposure time and enhances the signal-to-noise ratio. The disadvantage is that we

now see only a fraction of the molecular transform and so information is lost.

Combination of equations 4 and 10 gives:-

$$G(\underline{S}) = F(hk\ell) = V \int_0^1 \int_0^1 \int_0^1 \rho(xyz) \exp 2\pi i\ (hx+ky+\ell z).dxdydz \qquad (12)$$

where V is the volume of the unit cell. $F(hk\ell)$ is known as the structure factor or structure amplitude. It is generally inconvenient to deal with $\rho(xyz)$ which is a continuous function and which takes no explicit account of the fact that electrons tend to be localised around atomic centres. Instead we may write equation 12 in the form:-

$$F(hk\ell) = \sum_j f_j \exp 2\pi i\ (hx_j+ky_j+\ell z_j) \qquad (13)$$

where the sum is taken over all the atoms in the unit cell and f_j is the atomic scattering factor of the j th atom.

The atomic scattering factor describes the coherent scattering of X-rays from an atom. If relativistic effects are neglected,it is given by:-

$$f(\underline{S}) = \int |\psi(r)|^2 \exp(2\pi i\ \underline{r}.\underline{S})\ d^3\underline{r} \qquad (14)$$

(James, 1948), where $\psi(\underline{r})$ is the wave function of the atom. But $|\psi(\underline{r})|^2 = \rho(\underline{r})$, the electron density function. It is sufficient for our purposes to suppose that atoms are spherically symmetrical, in which case:-

$$f(S) = 4\pi \int_{r=0}^{\infty} r^2\rho(r)\ \frac{\sin Sr}{Sr}\ dr \qquad (15)$$

(James, 1948).

Now

$$f(0) = 4\pi \int_{r=0}^{\infty} r^2\rho(r)\ dr \qquad (16)$$

so the amplitude of scattering from an undeviated beam is equal to the number of electrons in the atom. As the scattering angle increases, phase differences between waves scattered from widely separated points within the atom become more significant and f decreases (figure 10).

A simple modification of equation 13 gives:-

$$F(\bar{h}\,\bar{k}\,\bar{\ell}) = \sum_j f_j \exp - 2\pi i\,(hx_j+ky_j+\ell z_j) \qquad (17)$$

Hence: $$|F(hk\ell)| = |F(\bar{h}\,\bar{k}\,\bar{\ell})| \qquad (18)$$

This result, known as Friedel's law, shows that the distribution of intensities in reciprocal space is centrosymmetric irrespective of the symmetry of the crystal.

2.4.2 Diffraction from Helical Objects

Equation 13 is generally applicable to crystalline specimens but the number of terms in the summation can often be reduced by taking account of the symmetry of the unit cell. Further simplification may be made for helical molecules which have much higher symmetry than the unit cell.

It is most convenient to describe scattering from helical objects in terms of the cylindrical polar co-ordinate system (Fig. 3). The Fourier transform of a general electron density distribution is then:-

$$F(R,\phi,Z) = \sum_n \int_0^\infty \int_0^{2\pi} \int_0^c \rho(r,\phi,z)\, J_n(2\pi Rr)$$

$$\times \exp(i[n\{\phi - \phi + \tfrac{\pi}{2}\} + 2\pi Zz])\, dr\, rd\phi dz \qquad (19)$$

where J_n is the nth order Bessel function of the first kind (Vainstein, 1966).

The electron density function in a helix may be expanded as a

two-dimensional Fourier series:-

$$\rho(r,\phi,z) = \frac{1}{c} \sum_{\ell} \sum_{n} g_{n\ell}(r) \exp i \left(n\phi - \frac{2\pi\ell z}{c}\right) \qquad (20)$$

where c is the repeat distance along the helical axis and $g_{n\ell}(r)$ is a member of a set of two-dimensional Fourier co-efficients. Cochran, Crick and Vand (1952) and Stokes (unpublished) have shown, by combining equations 19 and 20, that the Fourier transform on any layer-plane ℓ is given by:-

$$F_{\ell}(R,\phi) = \sum_{n} G_{n\ell}(R) \exp in(\phi + \tfrac{\pi}{2}) \qquad (21)$$

where:-

$$G_{n\ell}(R) = \sum_{j} f_j J_n(2\pi R r_j) \exp i\left(\frac{2\pi\ell z_j}{c} - n\phi_j\right) \qquad (22)$$

The summation in equation 22 is taken over all the atoms in the helical repeat unit. Two important points emerge from equation 21. First, due to the axial periodicity of the helix, the transform is confined to a set of layer-planes defined by:-

$$Z = \frac{\ell}{c} \qquad (23)$$

Second, the rotational periodicity of the helix is also present in the transform - this is embodied in the selection rule:-

$$n = \frac{\ell - Nm}{K} \qquad (24)$$

which determines those components included in the summation in equation 21. In equation 24, N is the number of repeat units contained in K turns of the helix and m is any integer.

If the helix contains more symmetry elements then the transform may be simplified further. For example, DNA molecules frequently contain a diad axis perpendicular to the helical axis so that for each atom at (r_j, ϕ_j, z_j) there is an equivalent one at $(r_j, -\phi_j, -z_j)$. Under these

circumstances equation 22 reduces to:-

$$G_{n\ell}(R) = \sum_j f_j J_n(2\pi R r_j) \cos\left(\frac{2\pi \ell z_j}{c} - n\phi_j\right) \qquad (25)$$

(e.g. Langridge et al (1960b)) and so $G_{n\ell}(R)$ is systematically real.

Although the axial periodicity of the molecule introduces discrete layer-planes into the molecular transform, there is no periodicity perpendicular to the axis and so the transform on the planes is continuous. It is useful to consider the intensity of scattering from a single molecule. Using the usual properties of waves, it is clear that the intensity on any layer-plane is given by:-

$$I_\ell(R,\Phi) = F_\ell(R,\Phi)\, F_\ell^*(R,\Phi) \qquad (26)$$

where the asterisk denotes the complex conjugate. So for a molecule such as that discussed in the previous paragraph:-

$$I_\ell(R,\Phi) = \sum_n G_{n\ell}^2(R) + 2 \sum_{\substack{n\ n' \\ n>n'}} \sum G_{n\ell}(R)\, G_{n'\ell}(R) \exp i(n-n')\left(\Phi + \frac{\pi}{2}\right) \qquad (27)$$

If the molecule were uniformly rotated about its axis whilst it was exposed to the X-rays, we would observe the cylindrically averaged intensity:-

$$\left\langle I_\ell(R,\Phi) \right\rangle_\Phi = \int_0^{2\pi} I_\ell(R,\Phi)\, d\Phi \Big/ \int_0^{2\pi} d\Phi \qquad (28)$$

which, ignoring a constant 2π is:-

$$\left\langle I_\ell(R,\Phi) \right\rangle_\Phi = \sum_n G^2{}_{n\ell}(R) \qquad (29)$$

(Franklin and Klug, 1955).

We are now in a position to describe the diffraction pattern from a helical object. Figure 4 showing the behaviour of a number of

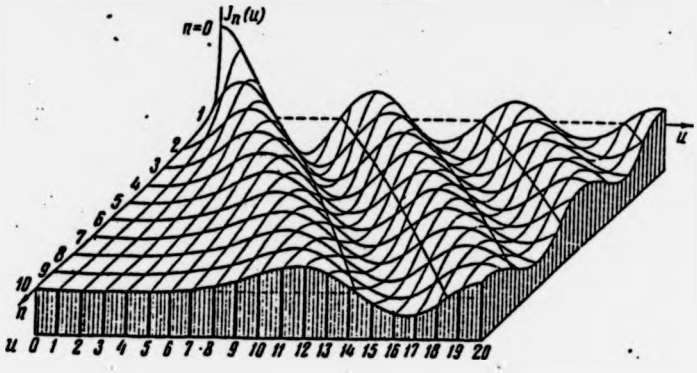


Figure 2.4 : $J_n(u)$ for n = 0 to 10

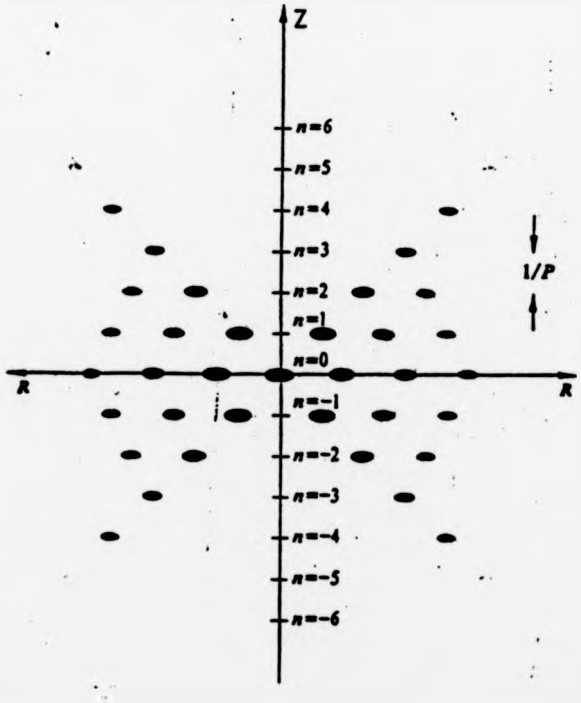


Figure 2.5 : Diffraction from a Continuous Helix

P is the helix pitch. The dark points represent peaks in the molecular transform.

Bessel functions illustrates three important points: (i) only J_0 is non-zero when the argument is zero; (ii) the argument of the first peak of each function increases with increasing order; (iii) the magnitude of the first peak of each function decreases with increasing order.

The diffraction pattern of a continuous helix contains only those Bessel functions for which m = 0 in equation 24 (Cochran, Crick and Vand, 1952), so J_ℓ is the only Bessel function which contributes to the ℓth layer-plane. The cylindrically averaged intensity (equation 29) then takes the particuarly simple form:-

$$I_\ell(R) = J_\ell^2 (2\pi Rr) \qquad (30)$$

This is illustrated in fig. 5. The cross-shaped pattern is characteristic of diffraction from a helix. It is conventional to call the line R = 0 the meridion and the plane z = 0 the equator.

If the helix is not continuous but discrete, m is allowed to take a continuum of values (Cochran, Crick and Vand, 1952). As a consequence, further cross-shaped patterns are set down with their centres along the meridion at points whose spacing is inversely proportional to the separation along the z-axis of the units in the helix (Fig. 6). (It can be seen that the continuous helix is a special case of this : the units in the helix are separated by an infinitesimal distance, so those parts of the transform with $m \neq 0$ are set at infinity). The figure and equation 21 show that more than one Bessel function now contributes to each layer-line. In principle the number is infinite but the properties of Bessel functions mentioned above ensure that only those with low order need be considered. Finally, it is clear that a continuous helix has cylindrical symmetry, but the projection of a discrete helix down the helix axis has rotational periodicity. As usual, periodicities in real space also manifest themselves in reciprocal space, so, for example, in the case of a 10-fold helix the Bessel function orders contributing to any layer-line are separated by ten.

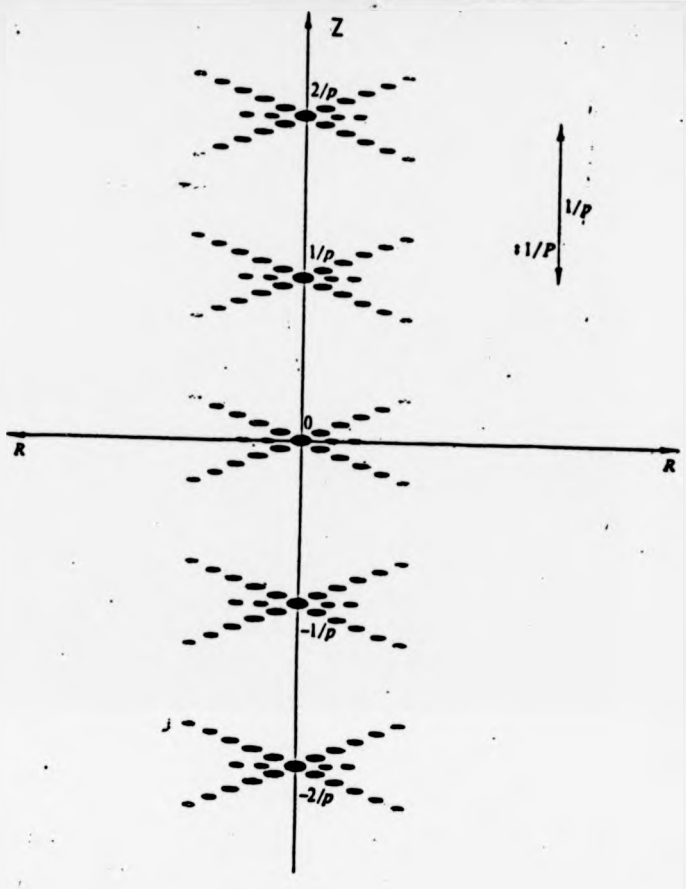


Figure 2.6 : Diffraction from a Discontinuous Helix

P is the helix pitch and p is the axial separation between the subunits.

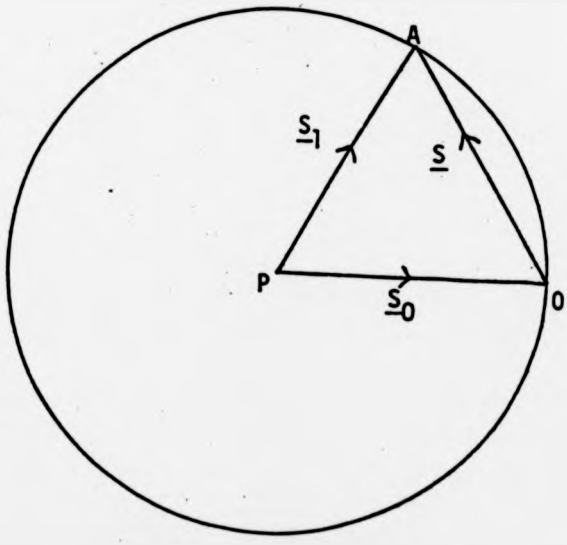


Figure 2.7 : The Ewald Construction

We have considered so far the diffraction pattern of an isolated helix, but as we have seen in section 2.3, DNA fibres consist of molecular arrays. In the case of a fully crystalline array of helices the molecular transform will be sampled at reciprocal lattice points as discussed in section 2.4.1. The structure factors are given by:-

$$F(h,k,\ell) = \sum_p \sum_n G_{n\ell}(R) \exp in\left(\Phi + \frac{\pi}{2} - \phi_p\right) \exp 2\pi i(hx_p + ky_p + \ell z_p) \qquad (31)$$

where (x_p, y_p, z_p, ϕ_p) are the fractional co-ordinates and the azimuthal orientation respectively of the pth molecule in the unit cell.

Arnott (1973) has derived the Fourier transforms of specimens exhibiting the types of disorder discussed in section 2.3. However we may also obtain a useful insight into scattering from such systems by the following more qualitative, intuitive approach. An X-ray beam incident upon a specimen selects what is regular in the specimen and concentrates the diffracted beams from the regular structure into sharp peaks (Bragg reflections). If a specimen is totally disordered then no consistent phase relationships exist between the diffracted beams and so no Bragg reflections occur. Instead the diffraction pattern consists of a continuum of scattered radiation. A gas of point scattering centres diffracts in this way (James, 1948). We might expect, therefore, that diffraction from a partially ordered system (such as a screw-disordered array of helices) would consist of Bragg reflections (arising from an ordered array of "average" unit cells) and diffuse scatter (arising from the disordered structure which results if one subtracts the average unit cell from the actual contents of each cell in the sample). Using this simple model we can predict which parts of the pattern will contain Bragg reflections.

Consider the screw-disordered system and suppose for simplicity that each unit cell contains only one molecule. If the array is truly

random then we may derive the average unit cell from any real cell in the array by uniformly smearing out each atom in the molecule along the helix on which it sits. The average cell therefore consists of a set of co-axial smooth helices the Fourier transform of which is simply the m = 0 branch of equation 21. So the Bragg reflections from a screw-disordered array of molecules will arise only from the m = 0 branch and will therefore be concentrated near the centre of the pattern. The disordered part of the array will give rise to diffuse reflections elsewhere, however, the one dimensional regularity of the molecule has not been destroyed so the diffuse scatter will still be concentrated onto layer-lines.

In an array exhibiting rotational disorder, the average unit cell is formed by smearing out each atom around the circumference of a circle whose centre is the helix axis. The Fourier transform of a circle involves only J_0 so the only portion of the diffraction pattern which shows Bragg reflections will be that due to this Bessel function. Once again, the diffuse scatter will be concentrated into layer lines.

The average unit cell of a sample with slippage disorder consists of an array of rods parallel to the helix axis. The projection down the helix axis retains the same order as would be exhibited by a crystalline array so the equatorial diffraction (which is due to this projection) consists of Bragg reflections and all the other layer lines show diffuse scatter.

2.4.3 The Geometry of Diffraction from Fibres

A crystal may be regarded as a three-dimensional diffraction grating. Diffraction is relatively easy to observe from one- and two-dimensional gratings since the elements of constant phase in S-space are planes and lines respectively. In the three-dimensional case the waves interfere constructively to give an intensity maximum only at points in S-space. It is clear therefore that if the recording device is held

stationary with respect to the crystal which in turn is stationary with respect to the incident beam, only a very small fraction, if any, of the available intensities will be observed. It is useful at this point to consider the conditions which must be satisfied if a reflection is to be observed. Figure 7 shows a section of S-space in which the origin is at O and the crystal is at P. When A is a reciprocal lattice point, $\underline{S}$ is a lattice vector and so, according to the theory developed in the previous sections, a diffracted beam will be emitted in the direction PA. To collect all the information available we must alter the relative orientation of the crystal and incident beam in order to bring the maximum number of lattice points into contact with the sphere of reflection.

As we rotate the crystal in real space, the Fourier transform rotates in reciprocal space in such a way as to conserve the relative orientation of the two co-ordinate systems. Now in a DNA fibre the crystallites are aligned approximately with their c-axes parallel to the fibre axis but their azimuthal orientations are random. Since there is no correlation between the crystallites, the diffraction pattern we observe is the same as we would obtain if we rotated a single crystallite uniformly about the fibre axis. Figure 8 illustrates this. It is clear that all reciprocal lattice points on a given layer-line with identical R values will cut the sphere of reflection at the same point and so diffraction from the crystal planes corresponding to these points will overlap on the diffraction pattern with consequent loss of information unless the Fourier transform systematically has an identical value at each of the points.

Figure 9 illustrates the effect of lack of parallelism of the crystallites on the diffraction pattern. The diffraction spots are drawn out into arcs of approximately constant angular width normal to the line drawn from each spot to the centre of the pattern.

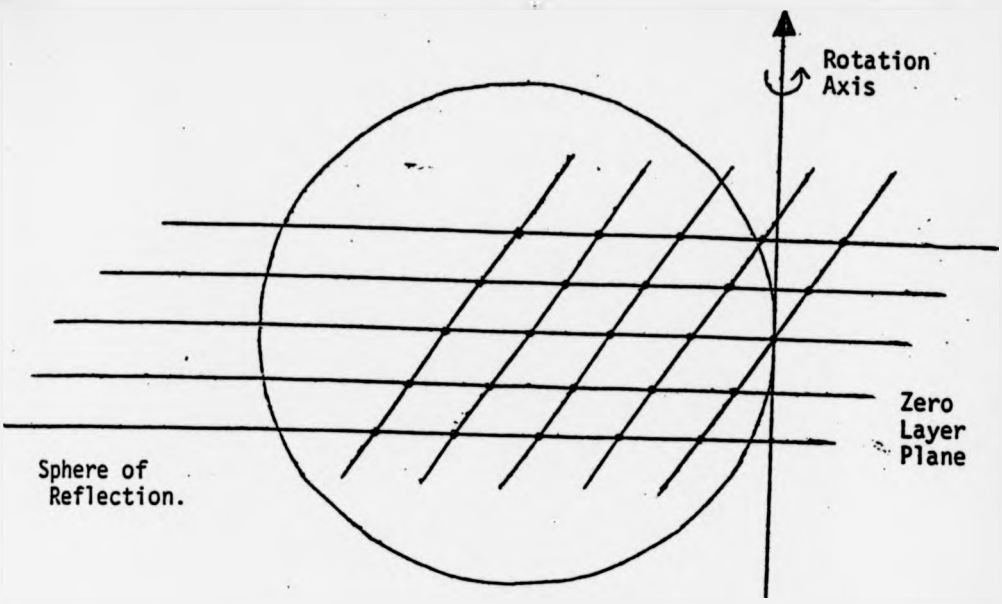


Figure 2.8 : Diffraction from a Rotating Crystal

Lattice points with the same values of ℓ and R will cut the sphere of reflection at the same point.

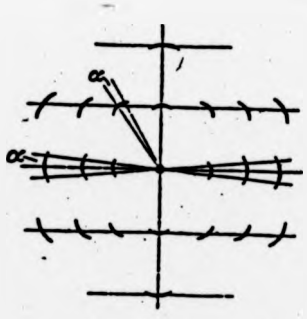


Figure 2.9 : Effect of Disorientation on the Diffraction Pattern

Reflections are drawn out into arcs of angular half-width α proportional to the disorientation of crystallites within the fibre. (From Vainstein, 1966).

2.4.4 The Phase Problem and the Strategy of Nucleic Acid Structure Solution

The ultimate objective of any crystallographic study is the production of an electron density map. We have seen that the electron density function of a helical molecule may be expanded as a two-dimensional Fourier series (equation 20). The Fourier transform of such a molecule is described in terms of the azimuthal harmonics, $G_{n\ell}(R)$ of equation 22. Klug, Crick and Wyckoff (1958) have shown that the co-efficients in equation 20 are related to the $G_{n\ell}$ via a Fourier-Bessel integral:-

$$g_{n\ell}(r) = \int_0^\infty G_{n\ell}(R)\ J_n(2\pi Rr)\ 2\pi R dR \qquad (32)$$

If we wish to compute $\rho(\underline{r})$ directly from these relationships we face several problems:-

(1) The $G_{n\ell}(R)$ values depend upon the molecular structure but in the observed diffraction pattern they are also modulated by the molecular packing which is present even in partially disordered specimens. We therefore need to propose a model for the packing which can then be used to correct to true $G_{n\ell}(R)$ values.

(2) Even if we remove the packing effect, several harmonics may be significant on each layer-line and in a specimen with relatively low rotational symmetry (e.g. a 10_1 helix) they may overlap which renders their measurement problematic.

(3) $G_{n\ell}(R)$ is, in general, a complex function. Whilst we may be fortunate enough to determine its amplitude from the intensity of the scattered waves, we cannot determine its phase. This is a classical problem of X-ray crystallography which, in the case of the globular proteins for example, may be solved by the method of multiple isomorphous replacement, or less frequently, by determining the position of a heavy atom in the structure followed by Fourier refinement (Blundell and Johnson, 1976). Marvin et al (1966)

suggested exploiting the latter method using phosphorous as the heavy atom. However, as the same authors pointed out, the resolution of the Fourier maps is limited by the smallest spacings present in the diffraction pattern and so an electron density map of DNA from fibre data would not be sufficient to determine atomic positions.

We are thus forced to adopt trial and error methods, building models (guided by experience, knowledge of acceptable stereochemistry and imagination), computing their transforms and comparing them with the observed data. The details of this procedure will become apparent in Chapter 3 when the building of models is discussed. Two quantative measures of the goodness of fit which are frequently used are R1 and R2:-

$$R1 = \sum_i \left| |F_o| - |F_c| \right|_i / \sum_i |F_o|_i \qquad (33)$$

$$R2 = \sum_i |I_o - I_c|_i / \sum_i I_{oi} \qquad (34)$$

where $(F_o)_i$ and $(F_c)_i$ are the ith observed and calculated structure amplitudes and $(I_o)_i$ and $(I_c)_i$ are the corresponding intensities.

2.5 Experimental X-Ray Methods

2.5.1 Equipment

Both Hilger and Watts microfocus generators and an Elliott rotating anode generator were used as a source of X-rays. A nickel filter was used to obtain the monochromatic CuK_α line (λ = 1.5418A). The Hilger and Watts sets were operated at 35kV and 3mA tube current. The typical exposure time for a 100μ thick fibre and a 3cm specimen-to-film distance was two days. The rotating anode was run at 35kV and 60mA tube current. The exposure time was between four and twelve hours.

Two types of camera were used: (i) pinhole cameras (such as those

described by Langridge et al, 1960a); and (ii) Elliott cameras with toroidal optics (Elliott, 1965). Both types of camera were continuously flushed with helium gas to reduce atomospheric scattering. The humidity was controlled by flushing the gas through an appropriate saturated salt solution (O'Brien, 1948).

2.5.2 Lattice determination

2.5.2.1 Introduction

The co-ordinates of reflections on an X-ray film were measured using a two-dimensional travelling microscope. These values were then converted to reciprocal space co-ordinates using the program FILM written by Dr. W.J. Pigram. Utilising the real-space co-ordinates of several uniformly distributed points on a calcite calibration ring, the program derives the least-squares best fit to the centre of the film and calculates the specimen to film distance. It then calculates the co-ordinates (R,Z) of each spot and finds ρ, the distance of the reflection from the origin of reciprocal space, and d, its Bragg spacing.

Miller indices (hkℓ) may then be assigned to each spot. The lattice parameters can be calculated from a set of simultaneous equations relating the observed ρ-values of a number of spots to their Miller indices and the reciprocal lattice parameters a^*, b^*, c^*, α^*, β^* and γ^*. Clearly there will be errors of observation in the measurement of ρ-values and so it is desirable to find lattice parameters which represent the best fit to the full set of observed ρ. This is generally achieved by the method of least-squares.

2.5.2.2 Optimisation by the Method of Least Squares

The method of least-squares is a classical optimisation algorithm which is often used by crystallographers, in particular for the refinement of lattice parameters and atomic co-ordinates in molecular models. Since the principles of the method are the same in both cases, they will be

described in general here.

The least-squares procedure minimises a function of the form:-

$$\phi = \sum_{j=1}^{M} w_j^2 (Q_{oj} - Q_{cj})^2 = \sum_{j=1}^{M} w_j^2 \Delta Q_j^2 \qquad (35)$$

where Q_{oj} and Q_{cj} are the observed and calculated values of some physical parameter of the system under investigation; w_j is a weighting factor which will generally quantify our confidence in the jth observation; M is the number of observations. We suppose the Q_c is calculated from a model system whose state is defined by a set of N parameters, $\{P_k\}$. We wish to find the shifts, Δp, which will minimise ϕ. This immediately suggests that we should expand ϕ as a Taylor series about p_k. If we ignore terms of $O(\Delta p_k^2)$ and higher, this becomes:-

$$\phi = \sum_{j=1}^{M} w_j^2 \Delta Q_j - \sum_{i=1}^{N} \Delta p_k \frac{\partial \Delta Q_i}{\partial p_k} \qquad (36)$$

ϕ will be minimised when $\partial\phi/\partial\Delta p_k = 0$, $(k = 0,1 \ldots N)$ which can be expressed analytically as:-

$$-2 \sum_{j=1}^{M} w_j^2 \left[\Delta Q_j - \sum_{i=1}^{N} \Delta p_i \frac{\partial \Delta Q_i}{\partial p_j} \right] \frac{\partial \Delta Q_j}{\partial p_k} = 0 \qquad (37)$$

In matrix terms:-

$$\underline{Q}.\underline{M} - \underline{P}.\underline{M}^T.\underline{M} = 0 \qquad (38)$$

so $$\underline{P} = \underline{Q}.\underline{M}.(\underline{M}^T.\underline{M})^{-1} \qquad (39)$$

where $$\underline{Q} = [W_1 \Delta Q_1 \ldots\ldots\ldots W_M \Delta Q_M], \qquad (40)$$

$$\underline{M} = \begin{bmatrix} W_1 \frac{\partial Q_1}{\partial p_1} & \cdots\cdots\cdots & W_1 \frac{\partial Q_1}{\partial p_N} \\ \vdots & & \vdots \\ W_M \frac{\partial Q_M}{\partial p_1} & \cdots\cdots\cdots & W_M \frac{\partial Q_M}{\partial p_N} \end{bmatrix} \qquad (41)$$

$$\underline{P} = \begin{bmatrix} \Delta p_1 & \cdots\cdots\cdots & \Delta p_N \end{bmatrix} \qquad (42)$$

Since the method calculates only shifts, not absolute values of p_k, it is necessary to furnish the algorithm with a set of starting parameters. Frequently these values will determine the nature of the final solution and so it is particularly important that care be exercised in designing a model for refinement. The aim of optimisation is, after all, to find the best fit to the data of a model like the initial one - the algorithm should not be expected to find new types of model.

In general the new values of p_k which we obtain by adding on the calculated shifts will not minimise ϕ because the Taylor expansion from which equation 37 was derived is only an approximation. Therefore, it is usual to repeat the procedure using the new parameter values from each successive cycle until the change in ϕ from one cycle to the next falls below some specified value.

2.5.2.3 Lattice Refinement

The theory in the previous section may easily be adapted to the task of refining lattice parameters. The author has written programs in Algol for use on the CDC7600 at UMRCC which refine monoclinic, hexagonal, tetragonal and orthorhombic lattices. Since these programs are all very similar in form, only one (the hexagonal case) will be described here.

In a hexagonal lattice:-

$$\rho_c^2 = (h^2 + hk + k^2)\, a^{*2} + \ell^2 c^{*2} \qquad (43)$$

so in the matrix terms of the previous section:-

$$\underline{Q} = \left[(\rho_o - \rho_c)_1 \ \ldots\ldots\ (\rho_o - \rho_c)_N\right] \qquad (44)$$

$$\underline{M} = \begin{bmatrix} \partial\rho_1/\partial a^* & \ldots\ldots & \partial\rho_N/\partial a^* \\ \partial\rho_1/\partial c^* & \ldots\ldots & \partial\rho_N/\partial c^* \end{bmatrix} \qquad (45)$$

where N is the number of observed reflections.

The data may be input to the program in one of two forms: (i) the Miller indices and observed ρ values of the spots are read in; (ii) the specimen to film distance is read in together with the Miller indices and observed real space co-ordinates of the spots. In the latter case, the program calculates the ρ values. A maximum of ten cycles is performed but execution is terminated before this stage if the calculated shifts vary by an amount less than a figure specified by the user. Finally the root-mean-square discrepancy between the observed and calculated ρ values is calculated.

2.5.3 Intensity Measurements

The measurement of intensities is an area in which technical advances have recently been made. These methods will be discussed in Chapter 3 during an assessment of the quality of the X-ray data upon which double helical models are based.

In this laboratory intensities have been measured by taking layer-line traces through Bragg reflections using a microdensitometer in a manner similar to that discussed by Langridge et al (1960a) and Marvin et al (1961). Since the only part of this project which involved measurement

of intensities is that concerned with ϕw-14 DNA, the detailed description of the method is presented in Chapter 6.

2.5.4 Modelbuilding

Initial models were built with wire skeletal models similar to those described by Langridge et al(1960b). The scale was 4 cm to 1A. The bases were represented by flat plates whose tilt and displacement from the helix axis could be varied.

Although wire models can be used with a reasonable degree of accuracy it is nonetheless useful to be able to check the detailed stereochemistry of a set of co-ordinates. Three programs were written by the author for this purpose. They were all in Algol and were run on the CDC 7600 at UMRCC.

The first program, BONANG, was very similar to one described by Pigram (1968). In a typical run, the co-ordinates of three repeat units of a helix were read in and were then sorted into ascending order along the z-axis - this renders more efficient the computation to be described next since all the models considered by the author were helical and extended more in the z-direction than the x- or y-directions. The maximum length of a covalent bond, d_{cov}, was set at 2A, and the maximum distance for a van der Waals' contact, d_{vdw}, was set at 3.5A. The program then took each atom in the ordered list in turn and calculated the distance between it and all atoms whose z-co-ordinates were greater until the discrepancy in z exceeded d_{vdw}. (Clearly any atom further along the list could not make a bad contact with the atom under consideration). If the distance between any two atoms was less than d_{cov} then that contact was recorded as a covalent bond. Any contacts greater than d_{cov} and less than d_{vdw} were recorded as van der Waals' interactions if the two atoms involved were not covalently linked to a common atom. Finally, the program calculated the angles between all covalently linked atoms. The use of this program immediately showed

whether the model under consideration was feasible.

The program IBC calculated distances between atoms in adjacent base-pairs. The input parameters were the rotation per residue, rise per residue, tilt angle and twist angle. In addition, the position of the twist axis could be varied. The geometry involved in these calculations has been discussed in detail by Fuller (1961). Any contacts less than some distance input as data were printed out. It was also possible for the user to specify pairs of atoms whose interatomic distances were calculated. This facility was useful in considering hydrogen bonding geometry.

The program IHC calculated distances between the atoms in two adjacent molecules. The co-ordinates of equivalent parts of the molecules and their relative orientation were required as data. In addition the co-ordinates of those atoms furthest from the helix axis were read in. Any contacts between the molecules less than some distance specified by the user were printed out. This computation has also been described by Fuller (1961).

2.5.5 Fourier Transform Calculations

2.5.5.1 Computing Equipment

Fourier transform calculations were performed on an ITT 2020 microcomputer with 48K of RAM core store. Back-up storage was provided by twin mini-floppy disk-drive units. Each disk held 116K kilobytes of information. All data were typed in via a keyboard. Programs were written in APPLESOFT II BASIC (BASIC Programming Reference Manual, APPLE Computer Inc.).

The use of such a mini-floppy file-based system presents special problems requiring house-keeping programs, a suite of which was written by the author and will be described below.

2.5.5.2 The Program HELIX 1

Fourier transform programs were written for the cylindrical systems described earlier in this chapter. The main program, HELIX 1, calculated and stored (i) the azimuthal harmonics, $G_{n\ell}(R)$; and (ii) the cylindrically averaged squared transform of a set of scattering centres. The program was partly based on an Algol program written and described by Fuller (1961) and so it will not be discussed in detail here.

The program resides on a disk and is read into the core store at the beginning of a run. It then opens and reads a disk file which contains the parameters which are used in calculating the atomic scattering factors (to be described in the next section). Next it opens and reads a disk file which contains all the remaining data:-

(i) the helical parameters N,K;

(ii) a flag which indicates whether the structure has a diad perpendicular to the helix axis;

(iii) the initial and final values of R and the interval in R at which the transform is to be computed;

(iv) the components of the transform to be computed;

(v) the atomic data: r,ϕ,z,wt,code,wwt for each atom where wt is simply a weighting factor and code and wwt are parameters which describe the type of the scatterer.

The program then calculates the transform of each atom and adds the result into a cumulative total. Finally, the user is asked whether he wishes to calculate the cylindrically averaged squared transform. The results may be sent to a printer and disk storage.

In a typical run, a real transform of 33 atoms with the m = -1 family contributing to ℓ = 0 - 3 and the m = 0 and +1 families contributing to ℓ = 0 - 10 could be calculated along layer lines from R = 0 - 0.4A^{-1} in

steps of $0.01A^{-1}$ in about five hours.

The program was tested by calculating a number of known Fourier transforms. For example, equation 22 shows that the components of the Fourier transform of a discrete helix with atoms of unit and constant scattering power at $(\pm \frac{1}{2\pi}, 0, 0)$ are:-

$$G_{nn}(R) = J_n(R) \qquad (m = 0) \tag{46}$$

When these results were compared with a table of Bessel functions they were found to agree to at least 1 part in 10^4 which is satisfactory.

2.5.5.3 Calculation of Atomic Scattering Factors

Cromer and Waber (1974) have used a function of the form:-

$$f_{cw}\left[\frac{\sin\theta}{\lambda}\right] = \sum_{i=1}^{4} a_i \exp\left[\frac{-b_i}{\lambda^2}\sin^2\theta\right] + C \tag{47}$$

to fit the scattering curves derived from various types of wavefunction. Vand, Eiland and Pepinsky (1957) have used a function of the form:-

$$f_{VEP}(\sin\theta) = A\exp(-a\sin^2\theta) + B\exp(-b\sin^2\theta) \tag{48}$$

In these expressions a_i, b_i, c, A, a, B and B are constants which depend upon the type of atom; a list of these parameters for those atoms of interest to us is given in table 1.

The original version of the program Helix 1 was written prior to 1974 so the parameters for f_{cw} were not available and therefore, f_{VEP} was used to evaluate scattering factors. The author has compared the form of the scattering curves computed using these two methods in order to determine whether a change to Cromer and Waber's method was desirable. The results are plotted in fig. 10. Clearly there is no significant difference between the two approximations for carbon, oxygen and nitrogen.

Atom	f_{VEP}				f_{CW}								
	A	a	B	b	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4	c
C	2.240	0.413	3.760	9.91	2.31000	20.8439	1.02000	10.2075	1.58860	0.568700	0.865000	51.6512	0.215600
N	2.528	0.444	4.472	7.47	12.2126	0.005700	3.13220	9.89330	2.01250	28.9975	1.16630	0.582600	-11.529
O	3.093	0.529	4.907	6.22	3.04850	13.2771	2.28680	5.70110	1.54630	0.323900	0.867000	32.9089	0.250800
P	8.816	0.6295	6.184	13.72	6.43450	1.90670	4.17910	27.1570	1.78000	0.526000	1.49080	68.1645	1.11490

$$f_{VEP}(\sin\theta) = A\exp(-a\sin^2\theta) + B\exp(-b\sin^2\theta) \quad : \quad \text{Vand, Eiland and Pepinsky, (1957)}$$

$$f_{CW}\left(\frac{\sin\theta}{\lambda}\right) = \sum_{i=1}^{4} a_i \exp\left(\frac{-b_i}{\lambda^2}\sin^2\theta\right) + C \quad : \quad \text{Cromer and Waber, (1974)}$$

TABLE 2.1 - Scattering Factor Parameters

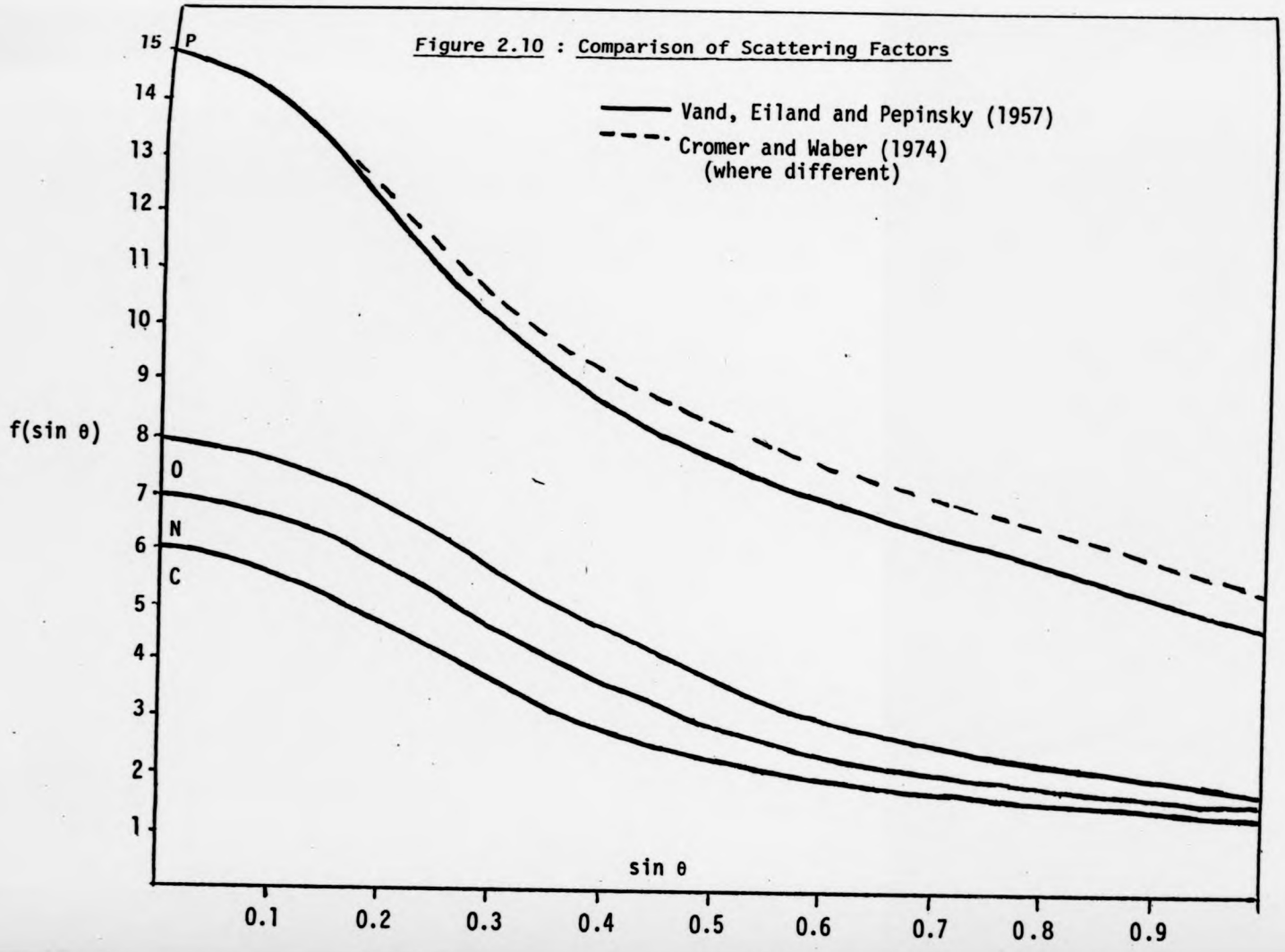


Figure 2.10 : Comparison of Scattering Factors
Vand, Eiland and Pepinsky (1957)
Cromer and Waber (1974)
(where different)
P
O
N
C
f(sin θ)
sin θ
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

The curves for phosphorous do diverge appreciably growing to a maximum discrepancy of about 10% when Sin θ = 1. However, the smallest useful spacings observed from DNA fibres correspond to $\rho \simeq 0.4A^{-1}$ for which Sin θ = 0.31. Hence the difference between the two curves can be ignored without introducing significant error into the Fourier transform calculations. The physical explanation of the discrepancy between these two curves provides another example of the reciprocal relationship between real- and diffraction-space. The Vand, Eiland and Pepinsky scattering curve for phosphorous was calculated using wavefunctions which took no account of exchange interactions. But the inclusion of exchange effects tends to shrink the calculated electron density function - i.e. the Gaussian sphere which is used as an approximate model of the atom has a larger radius with exchange than without. The Fourier transform of the Cromer and Waber sphere therefore tends more slowly to zero.

The computation of the Cromer and Waber function is the more expensive because it contains more terms. Therefore, since the results do not decisively favour either function, the Vand, Eiland and Pepinsky parameters have been used throughout the present work.

The scattering factors thus calculated must be amended to take account of the effect of the scattering from water (Langridge et al, 1960b). Fuller (1961) has described an empirical method for achieving this but it is not entirely satisfactory. A number of alternative methods will be discussed in detail in Chapter 3.

2.5.5.4 A Suite of 'Housekeeping' Programs for HELIX 1

These programs were written largely in order to achieve two desirable objectives:-

(1) to make efficient use of storage space;

(2) to keep the input and output files for each transform on the same disk.

It is not always possible to predict how large the output file from HELIX 1 will be so it is usually best to direct output from the program to an empty disk. When the computation is complete the file can then be moved to the same disk as the input files. If it transpires that there is no room then the input files are transferred to the new disk. The housekeeping programs therefore consist almost entirely of routines for moving files of the types SCAT (scattering factor parameters), FT DATA, TRANS (output file containg $G_{n\ell}(R)$'s) and CAST (cylindrically averaged squared transforms) from one disk to another and clearly they are relatively trivial.

In addition it is necessary to write input files. The program WRITE SCAT simply asks for the number of scattering types to be input followed by A,a,B,b,wwt for each type (where wwt is a parameter which describes how the water in the fibre affects scattering from that type of atom). The program WRITE FT DATA prompts the user to give as data the other parameters to be used in HELIX 1 (described in section 2.5.5.2). Both these programs are made particularly convenient to use by the interactive nature of the system - in, for example, a card based system, it is necessary for the user to remember the structure of input files.

Since FT DATA files are usually quite large, it is almost inevitable that errors will occur when writing them. Therefore the author has written EDIT FT DATA. The program opens the file to be edited and asks what type of editing is to be done. If data other than the co-ordinates are to be changed then all such data must be re-typed. This is relatively easy since there is not a great deal of it and the program prompts the user in the same way as WRITE FT DATA. It would generally be very inconvenient to re-type the atomic co-ordinates so if an error occurs in this section the user simply types the letter

R,F,Z,W or C to denote the type of parameter, the number of the atom in the list and the new value. The program has been written flexibly so that each of these facilities may easily be used repetitively.

It is useful to be able to add together two transforms (for example, of a base and phosphate) and so the program ADD TRANSFORMS was written. The cylindrically averaged squared transform of the new transform may then be calculated using the program CALCULATE CAST which simply evaluates equation 29.

Finally, it is a good idea to have programs which will read the contents of a file and print it onto the screen or the printer in a convenient format.

The full list of utility programs written by the author to aid the calculation of Fourier transforms is therefore as follows:-

WRITE SCAT	WRITE FT DATA	
READ SCAT	EDIT FT DATA	
MOVE SCAT	MOVE FT DATA	
READ TRANS	READ CAST	ADD TRANSFORMS
MOVE TRANS	MOVE CAST	CALCULATE CAST

2.5.6 Structure Factor Calculations

The structure factor program HELIX 2 is a straightforward application of equation 31. A file of type TRANS is given as data, and the crystal lattice parameters and molecular orientations are input via the keyboard. There are two ways of selecting which structure factors are to be calculated. First, initial and final values of h, k and ℓ may be typed in and all structure factors with indices between these values are computed. Second, a file containing a list of the structure factors required is attached to the program. A third possibility (which has not yet been added) is to give a maximum ρ or R value as data and

then calculate all the structure factors which fall within the sphere or cylinder defined thereby.

In the second case, if the file contains the observed structure factors then the program scales the calculated set and computes the residuals R1 and R2.

In a slight modification of this program (Helix 3) the number of molecules per unit cell was fixed at two. The first molecule was placed at the origin with azimuthal orientation ϕ. The second molecule was situated at $(\frac{1}{2}, \frac{1}{2}, z, \phi)$. Both z and ϕ could be systematically varied between any user-specified limits. Apart from this the program proceeded exactly as Helix 2 so the structure factors and residuals could be evaluated as a function of the orientation and the relative heights of the two molecules. This facility was useful in the work to be described in Chapter 3.

2.6 Refinement of Molecular Models

During the manual model-building described in section 2.5.4 care is taken to ensure that models are stereochemically satisfactory. However, it is inevitable that there will be discrepancies between the bond lengths and angles calculated from the model co-ordinates and those observed in single crystal studies. There is no guarantee that such anomolies can be ameliorated without radically changing the model, so it is always desirable at the end of a structure analysis to attempt to build a model with precise stereochemistry. The model-building program used in this laboratory is based on the linked-atom least-squares (LALS) technique (Arnott and Wonacott, 1966). It was written by W.J. Pigram and was substantially modified by D.C. Goodwin. The program was important in the present work and so it will be described in general here; however, detailed accounts have already been presented (Pigram,

1968; Goodwin, 1977).

A schematic flow-chart of the program is shown in figure 11.

A molecule may be described in terms of the co-ordinates of each atom with respect to some set of axes. However, such a description makes no explicit use of the information we have from single crystal studies of the covalent stereochemistry. An alternative method which exploits this information has been described by Eyring (1932). In the idealised molecule shown in fig. 12 we set up axes initially at atom 2 with the x-axis pointing from atom 2 to atom 1; the y-axis in the plane formed by atoms 1, 2 and 3, and the z-axis completes the right-handed set. The co-ordinates of atom 1 in this frame are clearly $(L_1, 0, 0)$. If we now set the origin at atom 3 in a similar manner the co-ordinates of atom 2 are $(L_2, 0, 0)$ whereas those of atom 1 are given by:-

$$\underline{X}' = \underline{A}_{23}.\underline{X}_1 + \underline{L}_{23} \qquad (49)$$

where $\underline{X}_1$ are the co-ordinates of atom 1 in the reference frame at atom 2; $\underline{A}_{23}$ is a matrix which describes the anticlockwise rotation of $\pi - \theta_1$ about z followed by the rotation of τ_1 about the new x-axis (which is required to bring the axes into the correct orientation at atom 3); $\underline{L}_{23}$ is the position vector from atom 2 to atom 3. The dihedral angle defined by atoms 1, 2, 3 and 4 is denoted τ_1. During refinement the covalent stereochemistry is maintained but the dihedral angles are allowed to vary. We therefore require the derivatives of the atomic co-ordinates with respect to each of these parameters. They may be calculated analytically in a manner which is analogous to the derivation of the co-ordinates.

The procedure described above would be relatively inflexible since all the atoms must lie on one chain. Three extra facilities have been incorporated into the program: first, pendant atoms may be

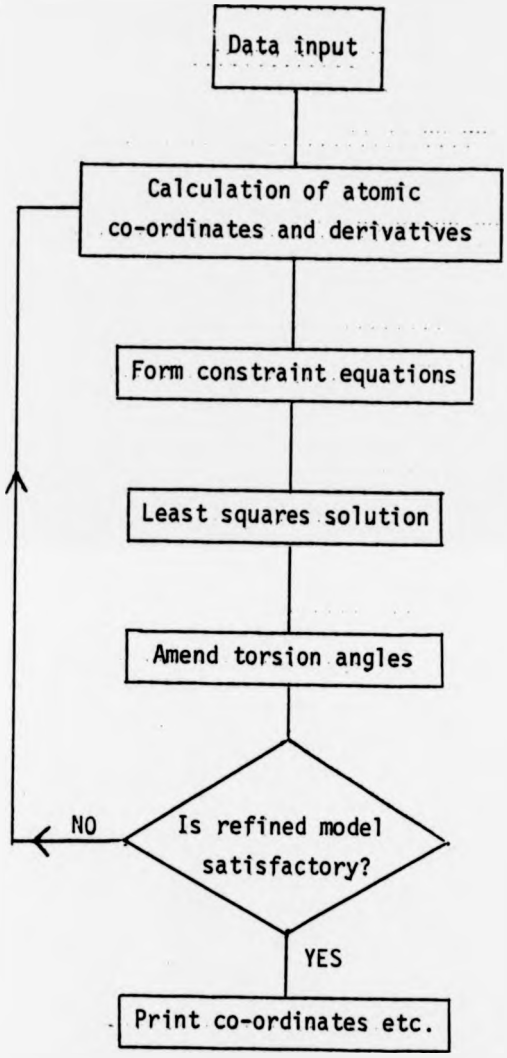


Figure 2.11 : Flow chart of the modelbuilding program

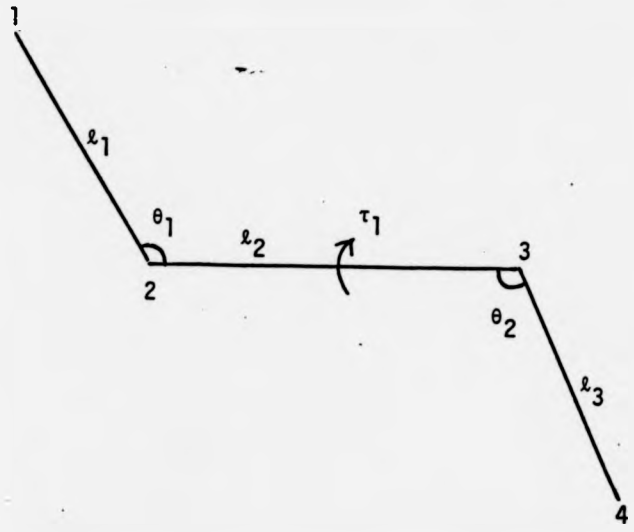


Figure 2.12 : Modelbuilding Nomenclature for an Idealised Molecule

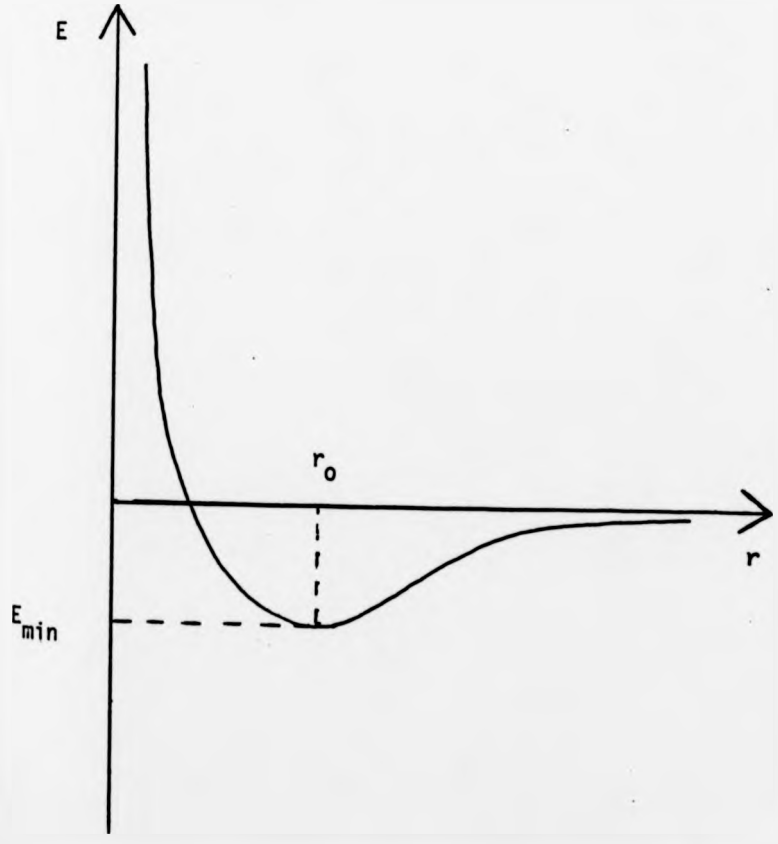


Figure 2.13 : The Lennard-Jones Potential Function

introduced at any point in the main chain; second, branch chains of any length may be added at any chain atom; third, atoms may be introduced which are in fixed positions with respect to the final co-ordinate system.

Clearly the initial set of dihedral angles (derived from the co-ordinates of the wire models) together with the covalent stereochemistry will give a set of computed co-ordinates in quite close agreement with the wire model. But this model will generally still not be satisfactory - for example, errors of measurement and inaccuracies in the design of the wire model may lead to co-ordinates whose helical parameters differ slightly from those desired - and so it must be refined. During the refinement procedure two types of constraints are used: (i) energy constraints; and (ii) geometrical constraints.

Once the atomic co-ordinates have been calculated the distances between those atoms which are not covalently connected are calculated and a run-time specified number of contacts with the worst energies are recorded. The energy, E, of interaction is calculated using the Lennard-Jones function:-

$$E = \frac{B}{r^{12}} - \frac{A}{r^{6}} \qquad (50)$$

where A and B are constants which depend upon the two atoms involved and r is the interatomic distance. This may be written in the form:-

$$E = A\left(\frac{r_o^{6}}{2r^{12}} - \frac{1}{r^{6}}\right) \qquad (51)$$

where r_o is the equilibrium separation. Values of A for various pairs of atoms are given in table 2. The form of the potential function is illustrated in Fig. 13. Since A is similar for all contacts not involving hydrogen it was set at the same value for all atom pairs. When

TABLE 2.2

Values for the constant, A, of the Lennard-Jones potential as used by Scott and Sheraga (1966) to give energy values in K cal/mole. The equilibrium separation is r_o.

Atom Pair	A	r_o(Å)
C-C	370	3.2
C-N	366	3.1
C-O	367	3.0
C-H	128	2.8
N-N	363	3.0
N-O	365	2.9
N-H	125	2.7
O-O	367	2.8
O-H	124	2.6
H-H	47	2.4

A = 366 the units of the energy are approximately k cal/mole (Goodwin, 1977). The value of A was divided by three for any hydrogen atom involved in a contact. The minimum energy, E_{min}, calculated from equation 51 is:-

$$E_{min} = -\frac{A}{2r_o^6} \qquad (52)$$

The discrepancy between the calculated and the desired energy for each of the contacts recorded is then added into a function of the sort defined in equation 35 for subsequent refinement.

The energy constraints will not generally be sufficient to ensure the retention of the correct helical symmetry in the refined model so additional, geometrical, constraints must be applied. Those used in the present work were relatively straightforward. For example, the C4' atoms in successive residues of a B-DNA model would be constrained so that h = 3.4A and t = 36^o. Three such pairs of atoms need to be so constrained in a dinucleotide model. The equations for this and a number of other constraints have been discussed in detail by Pigram (1968) and Goodwin (1977).

The energy calculated by the method just described is not a true measure of the total molecular energy since we have ignored, for example, electrostatic, solvent and base-base electronic interactions. In this sense it is immaterial what value is assigned to A. In practice it is convenient to use A as a simple way of altering the relative weight of the energetic and geometrical constraints.

When the constraint equations are set up the refinement proceeds in a manner analogous to that described in section 2.5.2.2. At the end of each cyle the total energy of the non-bonded interactions is compared with that obtained in the previous cycle and if the modulus of the discrepancy between the two figures is less than some specified

value, the refinement finishes and the atomic co-ordinates are printed, otherwise the procedure is repeated.

The present author discovered an error in the programme written by Goodwin (1977) which was the current version in use in this laboratory. In the coding of equation 52 the A had been omitted so whilst the calculated energy was correct, the routine was attempting to refine towards an optimum energy which was wrong.

When this fault had been corrected, the program was tested. Pigram (1968) tested the original version by refining the structures of cyclohexane, paraffin, poly-L-proline and poly-L-alanine. The latter molecule was selected for the present test.

The stereochemistry of the monomer is shown in fig. 14. This molecule can fold into an α-helix with the hydrogen-bonding arrangement shown in fig. 15. Since the torsion angle about the C'-N bond is 180^{o} (Pauling, Corey and Branson, 1951) there are only three variable parameters in the repeating unit: ϕ, ψ and χ (fig. 14). The α-helix is in a potential well (Nemethy and Scheraga, 1965) so the structure can be, and was, refined without the specific inclusion of helical constraints or hydrogen bonding interactions. The energy gained by hydrogen bonding enables the atoms involved to approach closer than the sum of their van der Waals' radii, so if the helix is to be refined using van der Waals' interactions alone, the atoms which would have been involved in the hydrogen bonding must be excluded from the search for van der Waals' contacts. Six residues were sufficient to include all significant contacts. Both right and left-handed helices were built using both the corrected and uncorrected programs for comparison and with a range of values for A. The results, and those obtained by Pigram (1968), are presented in table 3.

It is clear that Goodwin's program gives rise to wild final

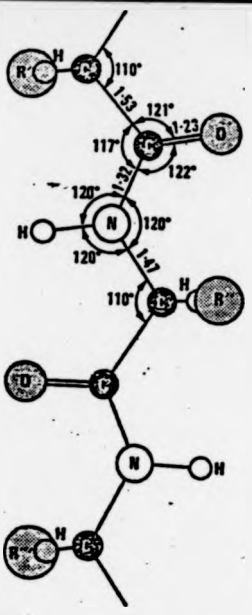


Figure 2.14 : Stereochemistry of the Peptide Monomer

(From Bragg, 1975)

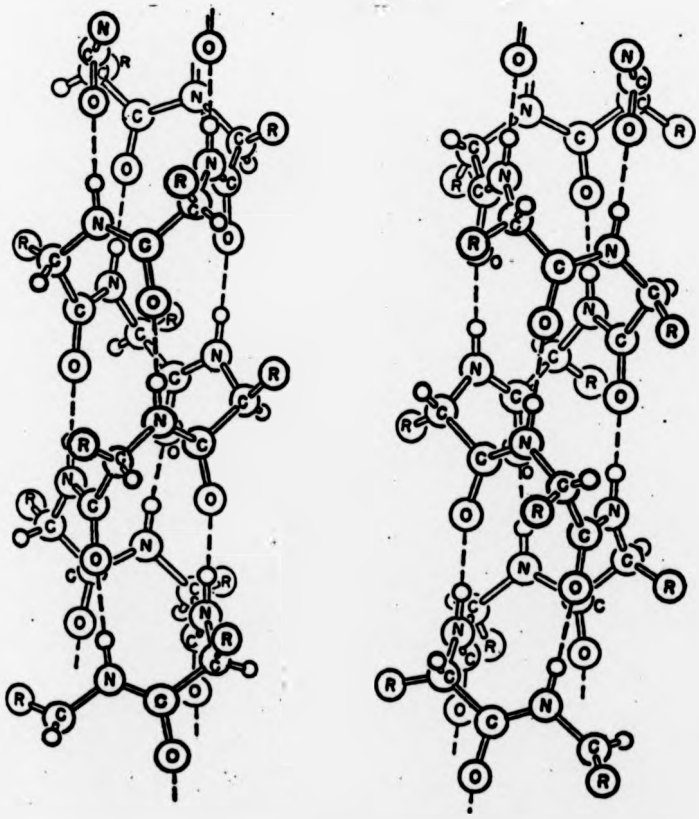


Figure 2.15 : Right and Left-handed α Helices

(From Bragg, 1975)

TABLE 2.3 - Torsion Angles of Right- and Left-handed α-helices derived using the modelbuilding program

			A					
		PROG	0.1	1.0	10	100	1000	WP
LEFT	χ	DG RG	2521.09 201.88	-168.46 201.92	214.89 201.91	219.06 201.91	219.43 201.91	200.92
	φ	DG RG	11.97 57.49	68.96 57.54	49.84 57.53	49.05 57.52	48.98 57.52	57.31
	ψ	DG RG	7542.62 51.20	35.20 51.21	48.29 51.21	48.55 51.21	48.58 51.21	51.13
RIGHT	χ	DG RG	-73137.71 -182.03	-87.73 -182.03	-166.52 -181.94	-168.52 -181.92	-168.70 -181.92	-181.85
	φ	DG RG	-2306.06 -61.30	53.69 -61.30	-49.46 -61.30	-48.24 -61.29	-48.12 -61.29	-61.37
	ψ	DG RG	-41.51 -48.84	-91.85 -48.84	-48.65 -48.84	-49.40 -48.84	-49.48 -48.84	-48.74

A is the parameter in equations 51 and 52.

The angles χ, φ and ψ are defined in figure 14.

Values labelled DG were produced by the incorrect version of the program described by Goodwin (1977). Those labelled RG were produced by a version corrected by the author. The column labelled WP shows the values obtained by Pigram (1968) using A = 100.

values whereas the corrected version refines for all values of A to the final torsion angles found by Pigram (1968). We may conclude that the program is now correct. The stereochemistry of the final models has been discussed by Pigram (1968).

A further program, PREP, was written in Basic for the ITT 2020 by the author based on an earlier version described by Pigram (1968). Given the atomic co-ordinates of a model this program calculates various parameters which are useful in setting up the data for the modelbuilding program, for example, bond lengths and angles, torsion angles and direction cosines of pendant atoms.

CHAPTER III

STUDIES ON THE CONFORMATION OF DOUBLE HELICAL DNA

3.1 Introduction

This chapter falls into two sections. In the first section we examine critically the data against which polynucleotide models have been designed. In particular, we examine how closely two independent data sets for B-DNA agree and discuss methods which may improve the accuracy with which intensities may be measured in future. A number of slightly different approaches have been made to the problem of correcting calculated Fourier transforms of polynucleotides for the effect of water in the fibre. These methods are discussed and compared and their effect on polynucleotide refineme is examined. The linked-atom least-squares procedure for refining polynucleotides is criticised. In particular, the theoretical soundness of the method is questioned.

In the second half of the chapter various polynucleotide modesl are presented. Thes e include both right- and left-handed regular double helical conformations. These are compared with models devised by other workers. An inverted base-stacking scheme is described and models incorporating this scheme are presented. Transitions between right- and left-handed models with normal and inverted stacking are discussed.

3.2 Criticism of Methods and Data

3.2.1 Intensity measurements, corrections and scaling

Structure factor amplitudes are commonly presented with a degree of precision which they scarcely deserve. For example, Arnott and Hukins (1973) quote the observed amplitudes for B-DNA to four significant figures. Whilst computerised methods have enhanced the reproducibility and precision of modelbuilding studies, estimation of the errors in intensity measurements

has advanced slowly. Yet the aceptability of models depends ultimately upon the accuracy of the intensities. In order to decide between competing models we need to quantify our confidence in the data upon which they are based. This is particularly so in the case of DNA where for example the alternative models may share the same gross features (e.g. Watson-Crick base-pairing, anti-parallel strands and right-handed helical symmetry) but differ in relatively minor ways (e.g. sugar pucker). If the accuracy of intensity measurement is low then we may not be justified in posing such esoteric questions. In the case of cellulose, for example, the problem is even more acute. The residual between data sets collected by different authors from the same photograph differed by as much as 49%. Nonetheless, each author presented a model which gave a residual of less than 20% when compared with his own data. These models differed on such fundamental points as chain polarity. (Example quoted by Miller and Brannon (1980)).

Crystalline fibres of DNA yield more data than can be obtained from cellulose and so the effect of random errors in the intensity measurements may be less likely to produce such dramatically different data sets. Two sets of data for B-DNA have been published (Langridge et al, 1960a; Arnott and Hukins, 1973) and it is of interest to calculate their residual. Two problems must be overcome before this can be done. First, Langridge et al (1960a) presented their data in terms of the quantity f_m where

$$I(hk\ell) = f_m^2\left[1 + \exp 2\pi i \left(\frac{h}{2} + \frac{k}{2} + \frac{\ell d}{c}\right)\right] \qquad (1)$$

The fractional displacement along z of the molecule at the centre of the cell is given by d and f_m is the transform of a single molecule. Langridge et al (1960a) set d at $^1/_3$ for convenience. This value was used in equation 1 in order to convert their data to structure factor intensities. Second, there is no one-to-one correspondence between the sets of reflections observed by the two groups. Therefore those reflections which are common to both sets have

been selected. The two groups of authors have indexed the patterns slightly differently. In particular they sometimes disagree where two composite spots are adjacent to each and it is difficult to decide precisely which reflections with similar radial co-ordinates in reciprocal space are in each of the spots. In such cases the intensities of the two spots have been added together and treated as one composite reflection. The data has been scaled so that the sum of the intensities in each set is the same. The reflections considered in this analysis are shown in table 1. The residual R has been calculated where:-

$$R = \frac{\sum | F_A - F_L |}{\sum F_L} \qquad (2)$$

and F_A and F_L are the observed amplitudes in the Arnott and Hukins (1973) and Langridge et al (1960a) data sets respectively. The residual calculated in this manner is 31.8%.

Both these data sets were collected using essentially the same technique. The variation of optical density through a reflection was measured along a line from the centre of the pattern. The area under the densitometer trace was then measured using a planimeter. The major problem in fibre diffraction measurements tends to be in the assessment of the background level which depends upon such factors as helium scattering and diffuse diffraction from amorphous regions within the fibre. The background was corrected in both cases according to the method described by Langridge et al (1960a). The area A of each spot must then be corrected for such factors as the Lorentz and polarisation effects and the arcing of the spots due to disorientation of crystallites within the fibre. Both sets of authors have simply multiplied A by R (the radial co-ordinate of the reflection in reciprocal space) in order to effect the Lorentz correction. Arnott and Hukins (1973) have multiplied by $\tan 2\theta/(1 + \cos^2\theta)$ to effect the polarisation correction. No justification is given for this procedure. The normal polarisation correction requires

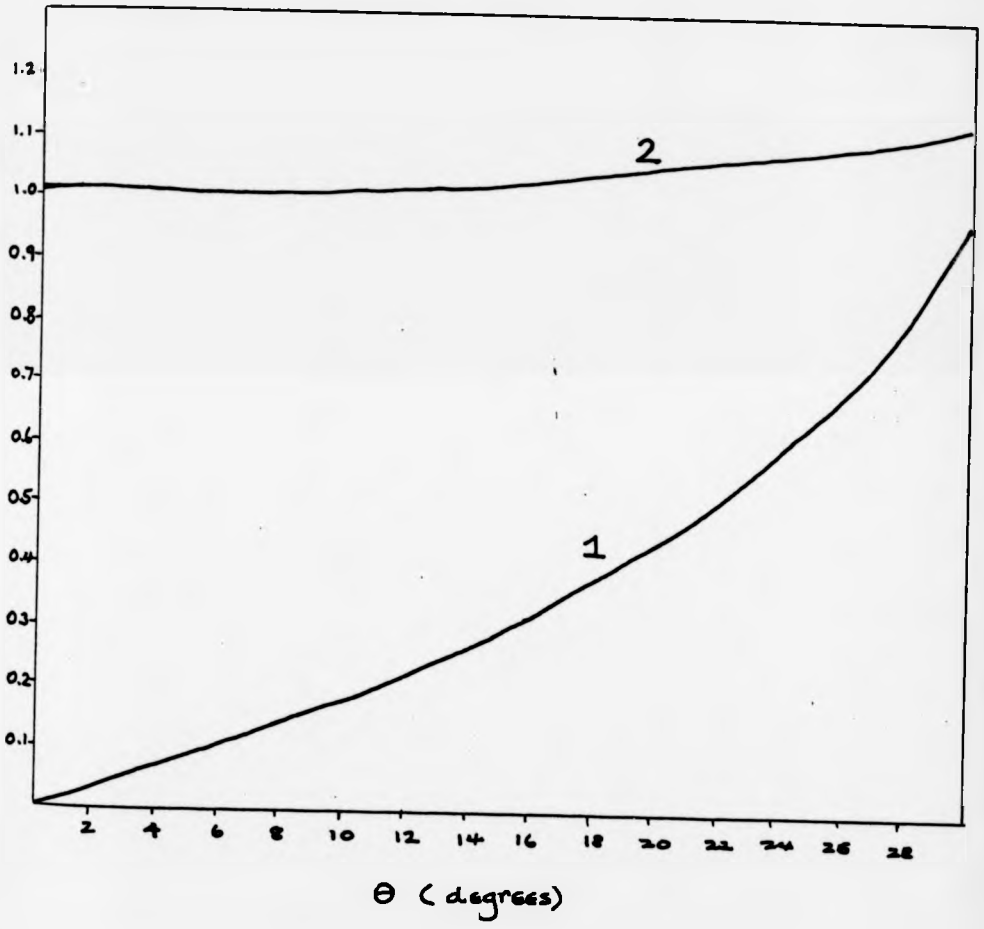


Figure 3.1 : Comparison of the polarisation corrections quoted by (1) Arnott and Hukins (1973) and (2) Langridge et al (1960a).

In case (1) $P = \tan 2\theta/(1+\cos^2\theta)$

and in case (2) $P = 2/(1+\cos^2 2\theta)$

TABLE 3.1 : The observed structure factor intensities common to Langridge et al (1960a) and Arnott and Hukins (1973) on the same (arbitrary) scale

h	k	ℓ	I_L	I_A
1	1	0	880	58998
2	0	0	0	79
3	1	0	160	130
2	2	0	360	305
4	0	0	160	181
1	3	0	80	41
4	2	0	160	99
3	3	0		
5	1	0	80	20
0	4	0	0	15
2	4	0	1000	725
6	0	0	1440	956
5	3	0	320	181
4	4	0	160	89
7	1	0	0	12
3	5	0	0	130
8	0	0	800	1253
7	3	0		
6	4	0	800	358
0	6	0	800	751
8	2	0		
2	6	0		
5	5	0	1120	2586
1	0	1	300	156
0	1	1	360	322
1	1	1	35	25
2	0	1	0	5
2	1	1	0	2
1	2	1		
3	0	1	30	23
4	0	1		
3	2	1		
0	3	1		

TABLE 3.1 Cont.

h	k	ℓ	I_L	I_A
4	1	1		
1	3	1	22	30
4	2	1		
5	0	1		
3	3	1	15	30
1	4	1		
5	2	1		
4	3	1	90	169
6	1	1		
3	4	1	30	89
1	0	2	450	253
0	1	2	1020	525
1	1	2	310	181
2	0	2	370	253
2	1	2	690	458
1	2	2		
3	0	2	180	179
4	0	2		
3	2	2		
0	3	2		
4	1	2		
1	3	2	40	46
1	4	2		
5	2	2		
4	3	2	150	420
6	1	2		
3	4	2	90	223
7	1	2		
2	5	2		
6	3	2		
5	4	2		
3	5	2		
7	2	2	104	551
1	1	3	420	322
2	0	3	800	502
0	2	3	600	417
3	1	3		

TABLE 3,1 Cont.

h	k	ℓ	I_L	I_A
2	2	3	200	258
3	3	3		
5	1	3	160	253
2	4	3		
6	0	3	120	266
5	3	3	520	223
6	2	3	520	238
7	1	3	560	340
8	2	3		
2	6	3		
5	5	3	120	287
2	1	4	60	15
1	2	4		
3	0	4	30	20
2	1	5	300	7
1	2	5		
3	0	5		
3	1	5		
2	2	5	887	938
4	0	5		
3	2	5		
0	3	5		
4	1	5		
1	3	5	462	1358
2	3	5	330	79
5	1	5	100	35
3	3	5		
0	4	5		
1	4	5		
5	2	5		
4	3	5	462	1402
6	0	5		
6	1	5		
3	4	5	210	30
1	1	6	40	12
2	0	6	80	48
0	2	6	160	79

TABLE 3,1 Cont.

h	k	ℓ	I_L	I_A
2	2	6	40	56
4	0	6	1000	571
4	1	6		
1	3	6	880	287
4	2	6	1520	620
2	1	7	60	56
1	2	7		
3	0	7	30	48
4	0	7		
3	2	7		
0	3	7		
4	1	7		
1	3	7	88	120
2	3	7	210	56
2	1	8	750	458
1	2	8		
3	0	8	1110	1345
3	1	8		
2	2	8		
4	0	8		
3	2	8		
0	3	8		
4	1	8		
1	3	8	684	3547
0	2	9	240	143
3	1	9		
2	2	9	160	181
4	1	9		
1	3	9	280	99
0	0	10		
1	0	10		
0	1	10		
1	1	10		
2	0	10		
2	1	10		
0	2	10		
1	2	10		
3	0	10	10147	9458

TABLE 3.1 Cont.

I_L and I_A are the observed intensities quoted by Langridge et al (1960a) and Arnott and Hukins (1973) respectively. They have been scaled so that $\sum I_L / \sum I_A = 1$. For reasons outlined in the text the (110) reflection was not included when the scale factor was calculated. However, it has been placed on the same scale as the other reflections and it is shown in the above list.

multiplication of A by $2/(1 + \cos^2 2\theta)$. Inspection of figure 1 shows that the latter correction has little effect on the intensities in the region in which DNA diffracts whereas the Arnott and Hukins correction shows a significan variation in this range. If this is the form of the correction which had been applied then one would expect to find that the low angle intensities in the Arnott and Hukins data set were appreciably reduced with respect to the higher angle reflections. In fact comparison of the two sets shows no such systematic variation. It would appear therefore that the polarisation term has been incorrectly recorded in Arnott and Hukins (1973). Langridge et al (1960a) have accounted for crystallite disorientation by multiplying A by ℓ, the arc length of a reflection on the photograph. Arnott and Hukins (1973) appear to have ignored this factor. This is unlikely to be significant if the pattern was obtained from a well oriented fibre.

If two equally good data sets show a disagreement index R = 31.8% then this indicates that attempting to modify a model so that its residual is significantly less than 30% may be futile since the accuracy of the data does not merit this degree of refinement. The assumption that the data sets are equally good is not necessarily true. The data published by Arnott and Hukins is more extensive than that observed by Langridge et al. However Arnott and Hukins have made no attempt to quantify the error in their measurements (although their paper concerns the precision with which models may be built). Langridge and co-workers assess that the intense reflections are measured accurately to within 20% whereas others may be in error by 40%. It is obviously difficult to decide which set of observations is more reliable under these circumstances.

A number of other methods for determining intensities are available and it will be useful to consider these here in order to see if we may reduce or quantify the errors. Two rather unreliable methods may be mentioned first for completeness. Instead of measuring the integrated intensity from layer line traces it is also possible simply to measure the peak height. Originally

this was done by eye with the aid of a graded strip of film but more recently it has been accomplished with densitometers. In both cases it is necessary to make assumptions about the spot size and shape in order to convert the peak height to an integrated spot intensity. These and a number of other corrections have been discussed in detail by Franklin and Gosling (1953c). It is unlikely that these methods will yield more accurate results than that discussed by Langridge et al (1960a).

Two dimensional scanning microdensitometers are advantageous in several ways. In these devices a light spot rasters across the film automatically and produces a file containing a quasi-continuous map of the optical density. Figure 2 shows a small section of such a map produced from the crystalline B-DNA pattern shown in Plate 1. An Optronics P-1000 scanner was used with 100 μm raster. The output file contains an array of 600 x 600 numbers in the range 0-255 representing optical densities in the range 0-3. One of the major advantages of this system is that a whole film may be measured in seconds rather than weeks. However the reduction of the optical density file to a list of structure factors is not trivial and a considerable amount of software is required. The suite of programs known as GENS (M. Pickering, P.A. Machin and M. Elder, SRC Daresbury Laboratory, unpublished) contains a number of useful routines, in particular, CENTRE AND THETA. The first finds the centre of the film in raster co-ordinates from either a diffracted salt ring or from symmetry related elements in the pattern. The second routine determines the angle of tilt of the film relative to the scanner co-ordinate axes.

A third routine, BACK, attempts to calculate the background correction required at each spot by defining a function which depends upon the sum of the lowest numbers appearing in the vicinity of a spot weighted by the number of appearances. The mean of this sum is then subtracted from each of the elements in the vicinity of the spot. The program AXIS (Meader

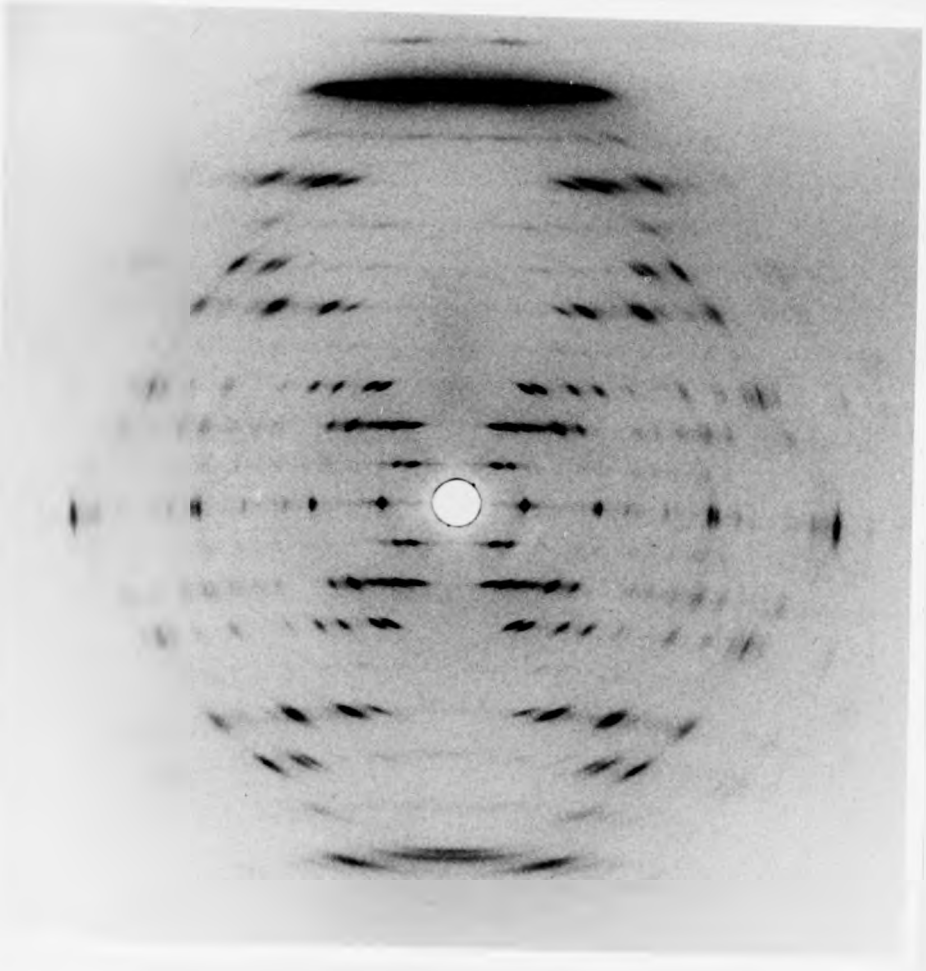


Plate 3.1 : X-ray Diffraction Pattern of Li B-DNA (courtesy of Professor M.H.F. Wilkins)

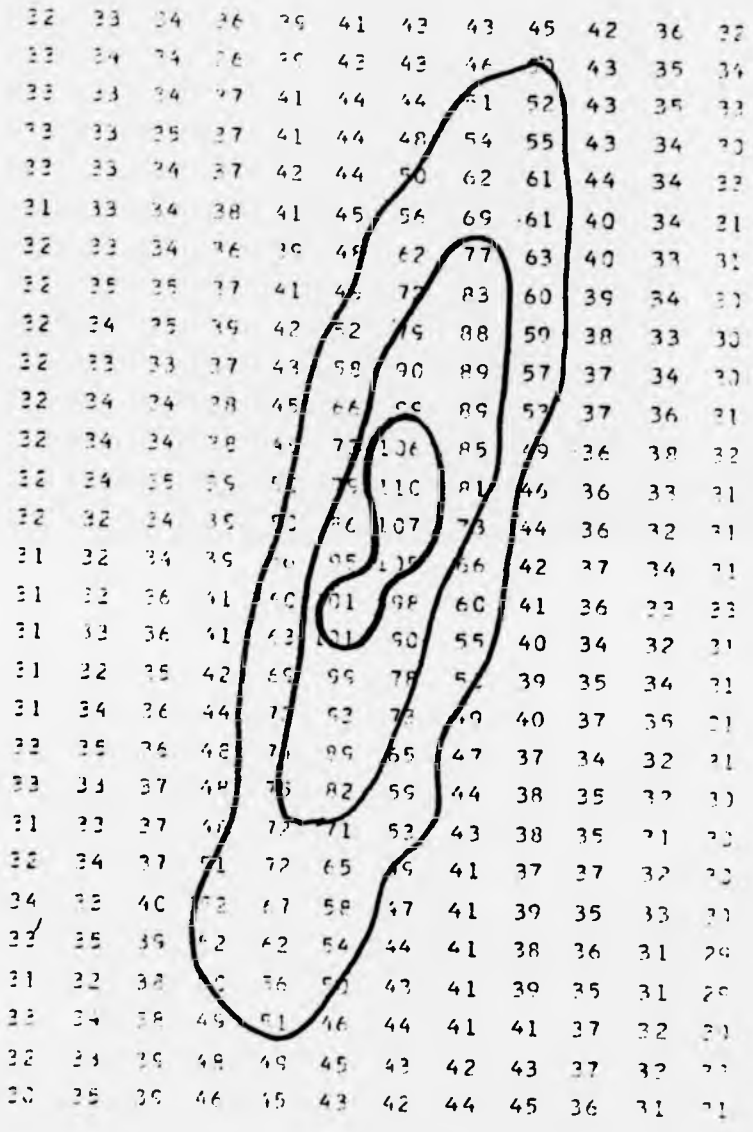


Figure 3.2 : Section of an Optronics map of the B-DNA diffraction pattern in Plate I showing the intensity contours of one reflection

et al, 1980) utilises a cubic spline fitting procedure to simulate the variation of background near a reflection. Whilst this procedure may produce accurate results with well resolved spots, it may give rise to a false background surface when spots are close together. Under such circumstances the knots to which the splines are tied are forced close together so as to avoid impinging on adjacent spots so small errors in the height of the knots can yield a surface which seriously overestimates the background contribution (Fraser, 1981). Fraser et al (1976) have described how the coherent particle length and the disorientation function of the crystallites within a fibre may be used in determining a background correction.

When background corrections have been applied, Lorentz and polarisation corrections may be applied on a point basis rather than over a whole reflection. This is particularly useful for reflections near the meridion where correction functions often break down. However, the number of such reflections is low in DNA diffraction patterns. Once all corrections are complete it is necessary to determine the integrated reflection intensity. When the file represented in figure 2 was created only Neisser (1980) had attempted to solve this problem. His method was specifically designed to measur intensities from the terephthalate series of polymers. Unfortunately he assumed that reflections are elliptical and this is unsatisfactory with DNA diffraction patterns wherein reflections tend to be crescent-shaped. Therefore integrated intensities could not be calculated from the B-DNA file mentioned earlier. Fraser (1981) has written more general programs which are available at the Daresbury Laboratory. These techniques are now being investigated in this laboratory. Despite the considerable problems involved in utilising optical density files there is no doubt that these methods will be increasingly used in the future.

A further significant advance in this area may be expected from the development of position-sensitive detectors. (See Schelten and Hendricks (1978) for a review of these instruments). Both one- and two-dimensional

detectors have now been constructed for the measurement of both X-ray and neutron intensities and they have mainly been utilised in small-angle scattering experiments. In the case of a two-dimensional detector (which would be the more useful type for fibre diffraction experiments) the sensitive area is divided into a number of energy "bins" (typically 256 x 512) and the number of counts registered in each bin is recorded by an on-line computer. An enormous advantage of such a system over photographic film results from the large dynamic range of the detector: the highest intensity which may be measured is limited only the amount of computer store available. Therefore both high and low intensities may be recorded during the same exposure thereby eliminating the need to perform error-prone scaling of film packs. In addition, the problem (which arises from film blackening) of measuring adjacent weak and strong reflections will be considerably reduced. By performing electronic pulse height analysis it will also be possible to discriminate between the energies of incident photons thus allowing the rejection of, for example, fluorescence and Compton Scattering signals and thermal diffuse scattering which merely add noise to the observed diffraction pattern. When used in conjunction with a high flux source, such as a synchrotron or storage ring, counting statistics may be employed to assess the error in the measurement of each reflection. Unfortunately the spatial resolution of area detectors (typically about 1 mm full-width half-maximum in X-ray detectors) is at least an order of magnitude lower than that of films. However this problem will be reduced when high flux sources and well-collimated, focussed beams are available which will make large specimen-to-detector distances feasible. A two-dimensional multi-wire proportional detector is currently being commissioned for use at the SRC Daresbury Laboratory storage ring which should be available for general use in 1982.

Whilst area detectors may improve the accuracy of measurements by removing some of the background noise, they do not resolve the fundamental

problem of determining the background surface due to fibre disorder. However, when used in conjunction with automated routines for mapping the background they will at least enable the intensities to be measured objectively and reproducibly. Competing models emanating from different laboratories may then be compared with more confidence against an agreed data set.

We turn now to the problem of scaling the observed and calculated diffraction. Wilson (1942) has shown that the mean value of the structure factor intensities is related to the scattering factors of the N atoms in the unit cell by the expression:-

$$\left\langle |F(hk\ell)|^2 \right\rangle = \sum_{j=1}^{N} f_j^2 \qquad (3)$$

This expression may be used in single crystal studies to determine the absolute values of the structure factors. Unfortunately fibres tend not to yield a sufficient number of observed intensities for a statistical analysis of this kind and so some empirical method is required.

The problem is to find a number, K, which, when multiplied by the calculated amplitudes, puts the two sets of structure factors on the same scale. The simplest solution is to let:-

$$K = \frac{\sum F_o}{\sum F_c} \qquad (4)$$

However the sum of the intensities of the observed reflections is a measure of the energy diffracted into the Bragg spots. Therefore a simple application of the law of conservation of energy suggests that we should use:-

$$K = \frac{\sum I_o}{\sum I_c} \qquad (5)$$

This appears to be the procedure adopted in the early studies on DNA (Langridge et al, 1960a; Fuller et al, 1965). Certain assumptions must be made if this factor is to be used. Not all the energy diffracted by a

crystal is directed into Bragg reflections: diffuse background arises from both static and dynamic disorder within the crystal. If we are able to correct the observed intensities in order to remove the effect of the underlying background then equation 5 should be used as the scale factor. Should it be thought that the corrected observed structure factors still differ significantly from their true values then an alternative approach is to include K as a parameter in a least squares refinement routine (Arnott and Selsing, 1974). It is of interest to compare the agreement between the observed and calculated diffraction as the scale factor is varied. The author has calculated the quantities:-

$$R1 = \frac{\sum |F_o - K^{\frac{1}{2}}F_c|}{\sum F_o} \qquad (6)$$

$$R2 = \frac{\sum |I_o - KI_c|}{\sum I_o} \qquad (7)$$

using the published observed and calculated structure amplitudes for B-DNA (Arnott and Hukins, 1973) and D-DNA (Arnott et al, 1974). In both cases only K was varied (so K = 1 corresponds to the published scale). The results for B-DNA are presented in figure 3 and those for D-DNA are shown in figure 4. In the case of B-DNA two sets of curves are shown. In one set, all the published data was used in the calculation whereas in the other set the suspect (110) reflection was omitted. Arnott and Hukins (1973) claim that R1 = 31% for their model. Inspection of the curves shows conclusively that the published observed amplitude of the (110) reflection is incorrect since the residual is 39% (at k = 1) when this spot is included and 31% when it is ignored. The correct curve for R1 reaches a minimum of 30% when K = 1.2 which is not significantly different from the value given by Arnott and Hukins. However these results indicate that their refinement routine does not find the optimum scale factor. Fortunately the curves for both B-DNA

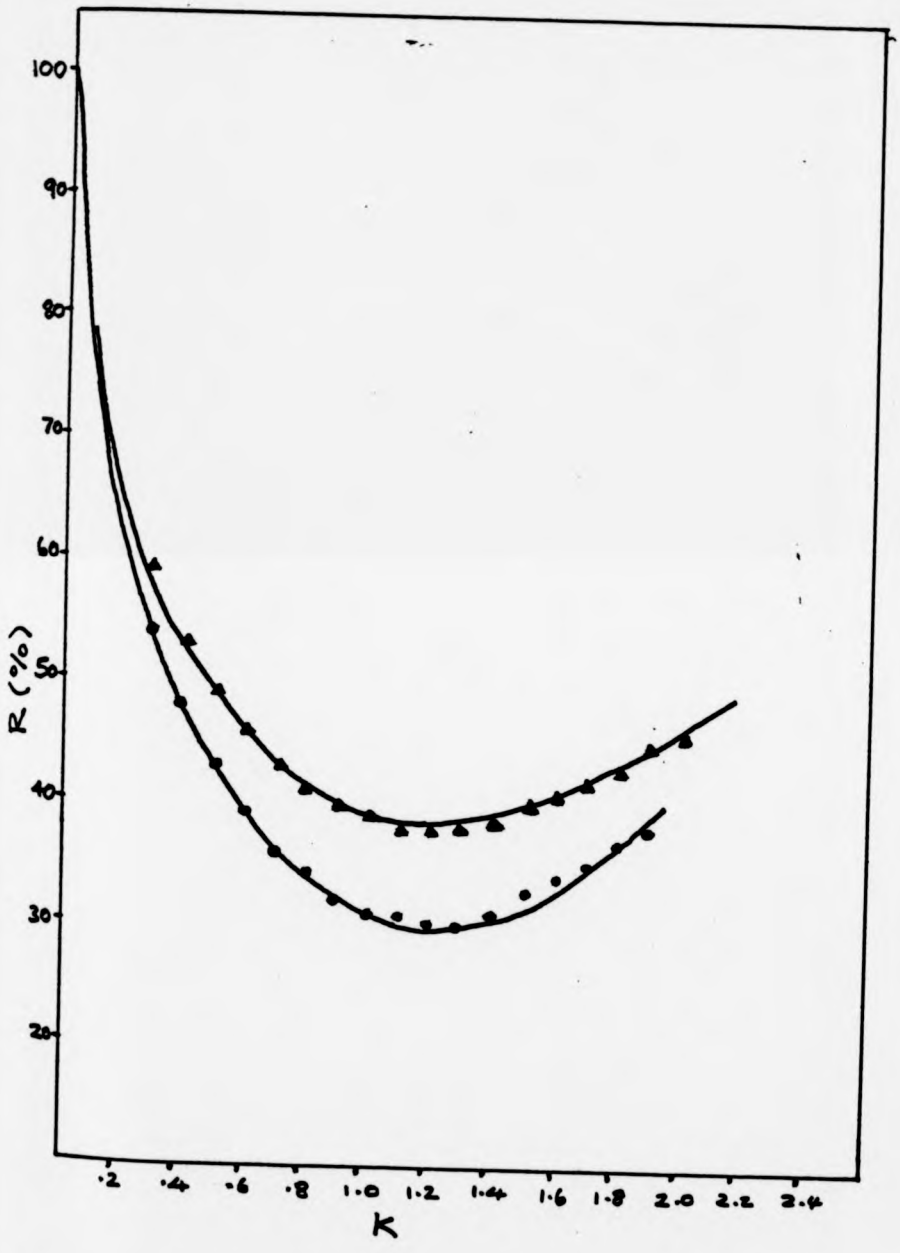


Figure 3.3 : Variation of the residual (R) of the B-DNA data of Arnott and Hukins (1973) as a function of scale factor (k)

▲ All reflections included

● (110) reflection omitted

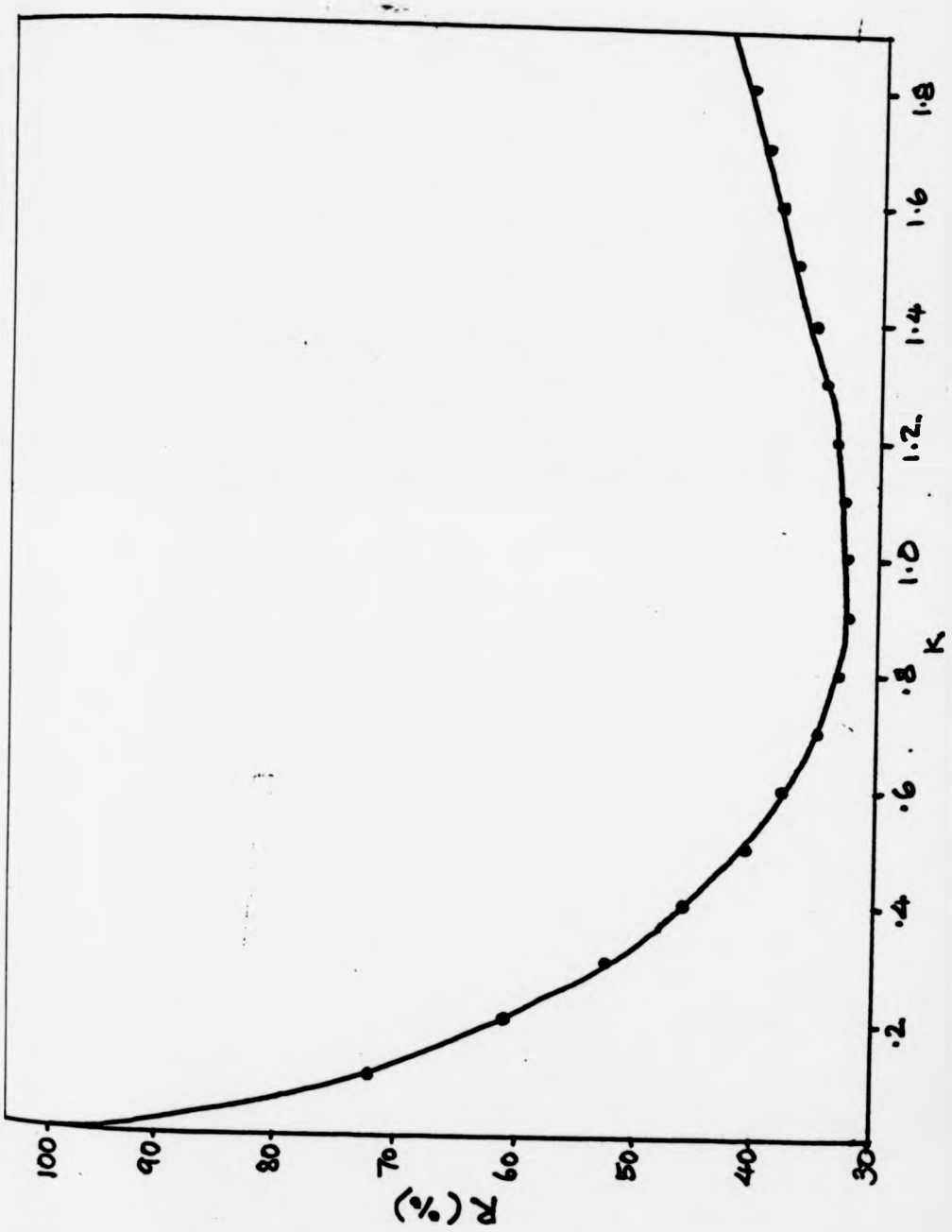


Figure 3.4 : Variation of the residual of the D-DNA data of Arnott et al (1974) as a function of scale factor (k)

and D-DNA show only a shallow minimum and so the scaling factor does not appear to be critical in determining the value of R1. Unless stated otherwise, all the structure factors quoted in the present work will have been scaled so that:-

$$\sum I_o / \sum I_c = 1$$

We are now in a position to make a detailed comparison of the B-DNA diffraction data obtained by Langridge et al (1960a) and Arnott and Hukins (1973). The scale factor calculated earlier for the common reflections in the two sets has also been applied to the reflections which are not common. The results are plotted in figure 5.

On the equatorial plane both sets of data are in quite close agreement from R = 0 to $0.15A^{-1}$. At higher values of R the disagreement is more marked. In particular, Arnott and Hukins find the intensity of the triplet formed by the (820), (260) and (550) reflections to be almost double that measured by Langridge et al.

There is no significant difference between the data sets on the first layer-plane. However the data of Arnott and Hukins extends to higher values of R. There are also some small discrepancies in the indexing of the two patterns. For example, Arnott and Hukins have resolved a doublet at $R \simeq 0.15A^{-1}$ which they have assigned to the (231) and (421) reflections whereas Langridge et al have recorded a spot in the same region which they have indexed as (501) and (331).

In the region $R < 0.1A^{-1}$ on the second layer-plane there are significant differences between the two data sets. The intensities observed by Langridge et al are systematically higher than those of Arnott and Hukins. On the remainder of this plane the differences are insignificant, however, the data of Arnott and Hukins is once again the more extensive.

On the third layer-plane the intensities of Langridge et al are again systematically higher than those of Arnott and Hukins. Both the fourth and

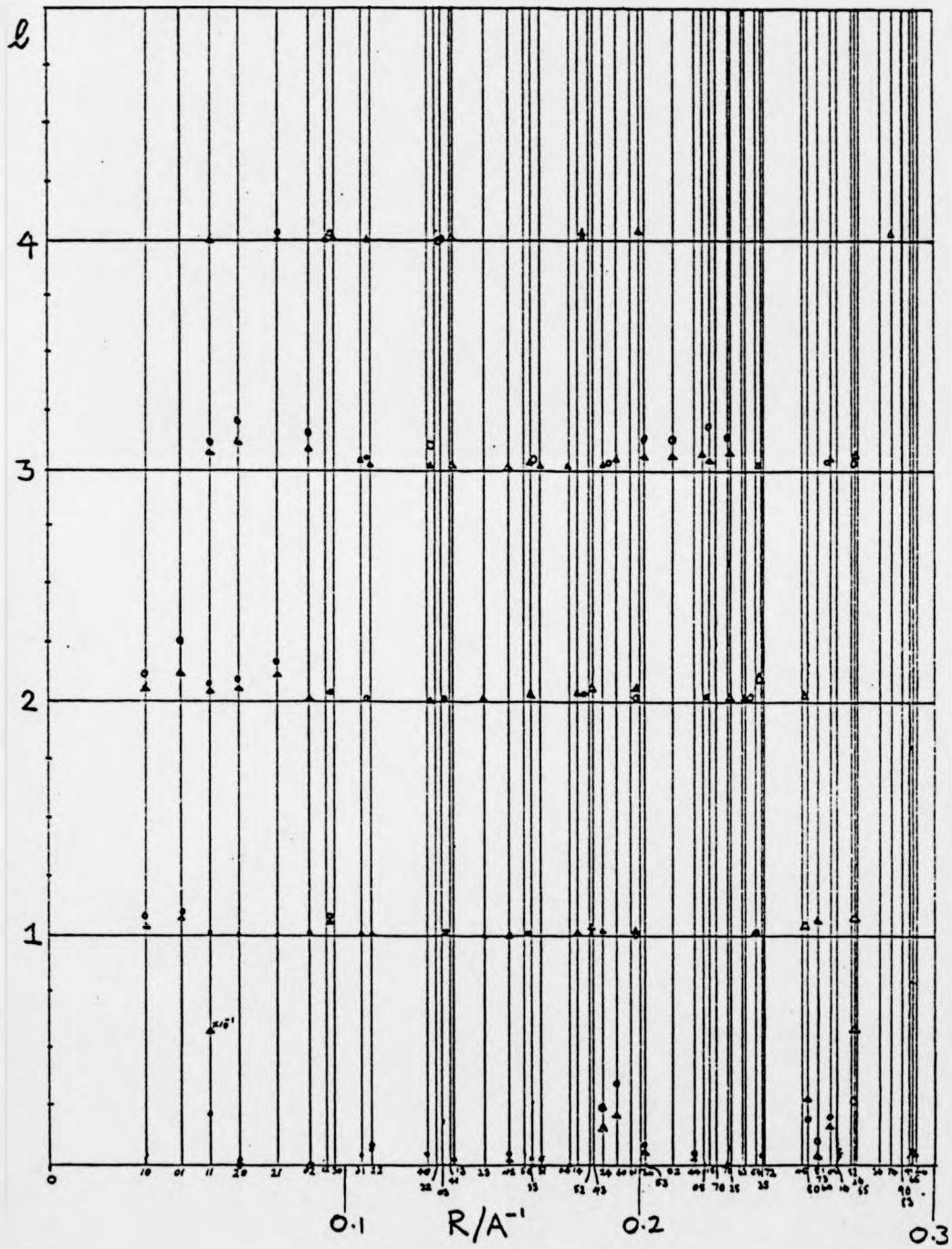
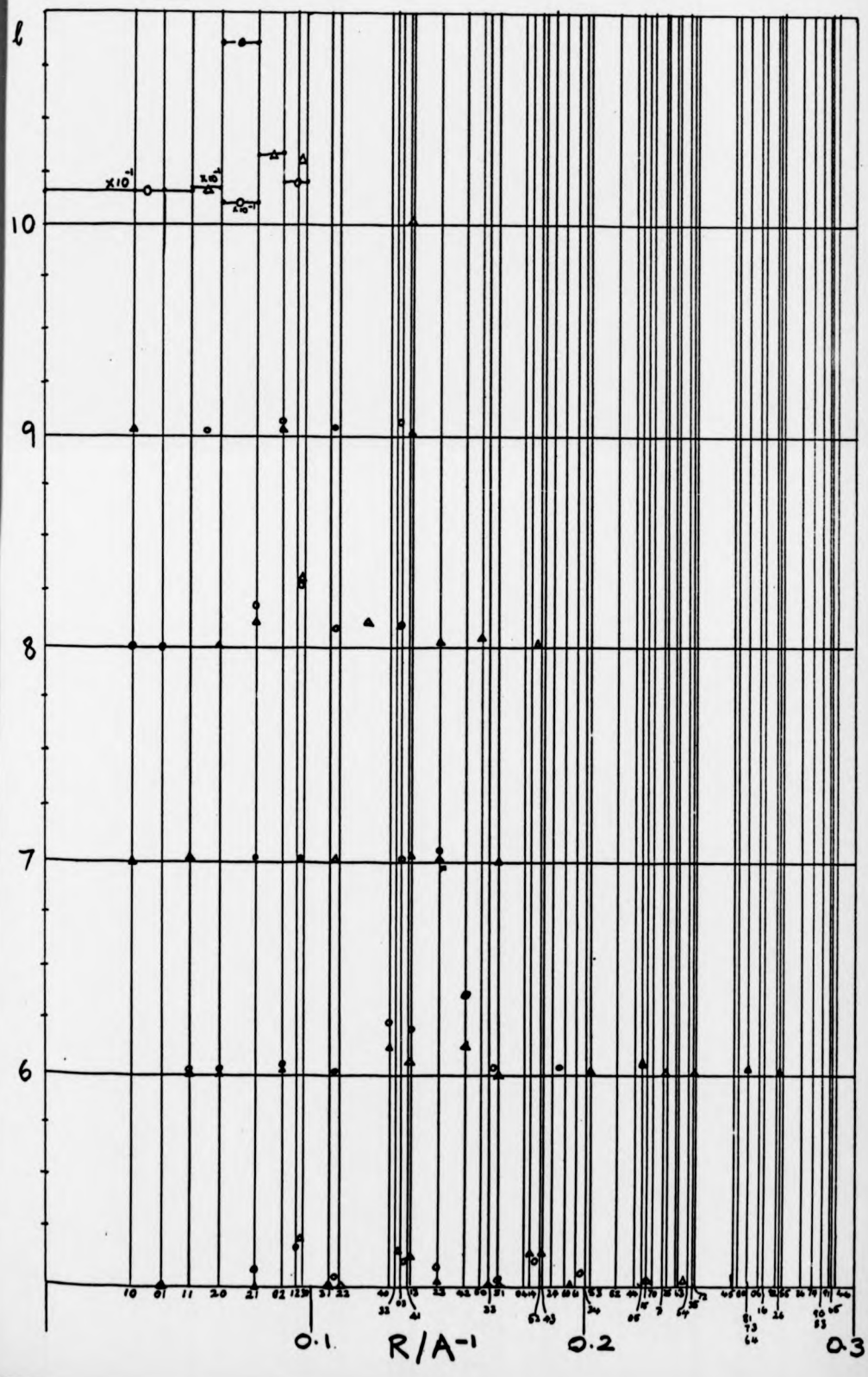


Figure 3.5 : Comparison of the observed intensities from B-DNA obtained by Arnott and Hukins (Δ) and Langridge et al (O)


l
10
9
8
7
6
0·1
R/A-1
0·2
0·3

fifth layer-planes show good agreement and there are only small discrepancies on the sixth layer-plane except in the region $R \approx 0.15A^{-1}$. The seventh, eighth and ninth layer-planes are also good but Arnott and Hukins have collected more data.

The comparison of intensities on the tenth layer-plane is made difficult by two factors. First, the arcing of reflections which arises from crystallite disorientation becomes more serious as the diffraction angle increases. Second, the high intensities recorded on this layer-plane from B-DNA tend to cause excessive film blackening which leads to an artificial broadening of the observed reflections. This may be reduced by recording these intensities on an underfilm but the scaling of the films which this necessitates also gives rise to errors. As a result of these two factors, not only is the accurate measurement of intensities hard to achieve, but also the indexing of the reflections is not unambiguous. Figure 5 shows that Arnott and Hukins and Langridge et al have indexed this layer-plane quite differently. A detailed comparison is not possible therefore; however, the total intensity observed by Arnott and Hukins is the higher of the two sets.

3.2.2 Diffraction from Water in Polynucleotide Fibres

The total amplitude of radiation scattered from water within DNA fibres is about 50% of that scattered by the DNA (Langridge et al 1960b). Although the water is largely disordered it nonetheless contributes to the observed diffraction pattern and its effect is particularly significant at small angles (Bragg and Perutz, 1952). An approximate correction may be made to the calculated diffraction by assuming that the solvent behaves as an electron gas whose electron density is the mean of that of the solvent and which occupies the space around the DNA molecules (Wrinch, 1950). We may then calculate the contribution of the solvent to the diffraction by

Babinet's principle (Langridge et al, 1960b). Then the corrected Fourier transform, F', of the entire system is related to that of the DNA, F, by the expression:-

$$F' = F - V\sigma F_V \qquad (8)$$

where σ is the mean electron density of the solvent; F_V is the Fourier transform of the solid of uniform density from which the solvent has been displaced and V is its volume.

Fraser, MacRae and Miller (1965) have used this expression to determine the scattering from fibrous proteins wound into coiled-coils. In their treatment they made simplifying assumptions which enabled them to calculate the transform F_V of the continuous molecular volume. Calculation of F_V is not straightforward in the case of DNA (Langridge et al 1960b), however O'Brien and MacEwan (1970) obviated the need for the calculation by populating the DNA molecular volume with a random array of water molecules, calculating the transform of the array by normal methods and scaling the result so that the predicted diffraction was consistent with an electron density of 0.33 electron/A^3 (the electron density of free water of unit density). An even simpler approach is possible and this has been adopted in most studies on polynucleotides. It involves the modification of the atomic scattering factors rather than applying the correction at the molecular transform level. The procedure is exactly analogous to that described above. The general correction is given by:-

$$f'(\sin\theta) = f(\sin\theta) - V\sigma\,\phi\,(\sin\theta) \qquad (9)$$

Here f' and f are the modified and uncorrected atomic scattering factors respectively; ϕ is the Fourier transform of the shape of the solid from which solvent has been excluded by the atom under consideration and V is the volume of that shape. A number of studies have been reported in which this expression, or a similar one, has been used: however they differ in

detail. In particular the values of the volume V and the form of ϕ have varied. These methods and their effects on polynucleotide transforms will be compared here so that we may discover whether the choice of method is likely to have a significant effect during a polynucleotide conformational refinement.

Arnott and co-workers have taken V in equation 9 as the volume derived from the van der Waals radius of each atom (Campbell-Smith and Arnott, 1978). This approach has the disadvantage that it takes no account of the variation in an atom's volume as it becomes charged or participates in hydrogen-bonding. An alternative approach used by all other workers (Langridge et al, 1960b; Fuller, 1961; Fraser, MacRae and Suzuki, 1978) is to use partial atomic volumes, V_s, derived from molecular volumes in solution. This method has been described in detail by Langridge et al (1960b). A comparison of the van der Waals and solvent volumes is shown in table 3.3.

Langridge et al (1960b) have assumed that atoms such as those in the sugar rings are approximately spherical with a diameter of 3A whereas base atoms form a flat plate 3A thick. Fuller (1961) has suggested that the Fourier transforms of these two objects differ insignificantly, therefore he assumed that all atoms were spherical. Since this procedure is computationally simpler than that of Langridge et al, it will be adopted here in the calculation of the scattering factors of Langridge et al, Fuller and Arnott. Both Langridge and Fuller used a sphere of constant radius so that ϕ may be calculated once only at various values of sin θ prior to the calculation of any f' and then used in an interpolation routine. This was desirable when computers were relatively slow however today we may calculate ϕ many times with a different spherical radius appropriate to the atom under consideration. Arnott uses the van der Waals radius in the calculation of ϕ. We will also compare Langridge's scattering factors when calculated with a sphere of constant radius 2A and a variable radius

TABLE 3.2

Values of parameters used in
evaluating Arnott's scattering factors

(See text for explanation of the symbols)

	r_A(A)	b(A)	α	β	V_H(A^3)
C	1.7	1.09	44.6^o	95.8^o	2.0
N	1.5	1.01	52.8^o	85.0^o	2.7
O	1.4	0.96	57.6^o	80.0^o	3.0

TABLE 3.3

Comparison of solvent volume (V_s) and van der Waals' volume (V_{vdw}) of polynucleotide scattering centres

Scattering Centre	V_s $(A^3)^1$	V_{rdw} $(A^3)^2$
P	23.6	28.7
-O-	9.1	11.5
$O^{\frac{1}{2}-}$	6.9	11.5
C_1, C_3, C_4	19.5	22.6
C_2, C_5	24.7	24.6
-O- (sugar)	7.1	11.5
N (ring)	2.5	14.1
NH_2	12.8	19.5
C	16.4	20.6
CH	21.6	22.6
=O	9.1	11.5
CH_3	32.0	26.6
-OH	-	14.5

1. Langridge et al (1960b)
2. Calculated by the author from values given by Campbell-Smith and

derived from the appropriate solvent volume. The Fourier transform ϕ of a hard sphere of radius r_o is given by:-

$$\phi(\rho) = \frac{3}{X^3} (\sin X - X \cos X) \quad (10)$$

where $X = 2\pi\rho r_o$ (11)

and $\rho = 2 \sin \theta/\lambda$ (12)

Fraser, MacRae and Suzuki (1978) have suggested that the use of a hard sphere in the calculation of ϕ is undesirable since ϕ oscillates with ρ. Instead they used a Gaussian sphere of radius r_o given by:-

$$\rho(r) = \exp\left[-(r/r_o)^2\right] \quad (13)$$

where $V_s = \frac{4}{3}\pi r_o^3$ (14)

The normalised expression for ϕ is then:-

$$\phi(\rho) = \exp\left(-\pi V_s^{2/3} \rho^2\right) \quad (15)$$

This expression tends asymptotically to zero as ρ increases and so the oscillatory component is eliminated.

Fuller (1961) has not used equation 9. Instead he calculated f' from the expression:-

$$f'(\sin\theta) = f(\sin\theta)\left[1 - \frac{V\sigma}{Z}\phi(\sin\theta)\right] \quad (16)$$

where Z is the number of electrons in the atom. If we write this in the form:-

$$f' = \frac{f}{Z}(Z - V\sigma\phi) \quad (17)$$

then we may see that f' is the unitary scattering factor f/Z multiplied by the bracketed term which represents the effective scattering power of the atom reduced from the <u>in vacuo</u> value by the scattering from the solvent

background. Whilst this expression may not have the sound theoretical basis of equation 9 it could nonetheless be a good empirical approximation to f'.

Each of the expressions described so far is suitable for the calculation of f' of individual atoms. However we also need to find f' of ionised groups (such as $O^{\frac{1}{2}-}$) and atoms with covalently bonded hydrogens. Langridge et al (1960b) have given values of V_s for $O^{\frac{1}{2}-}$, CH, CH_2 etc. and these are shown in table 3.3. Campbell-Smith and Arnott (1978) have devised a correction to V which allows for the volume of n attached hydrogen atoms of volumes V_H given by:-

$$V_H = \frac{1}{3}\pi r_H^3 \left[2 + \cos\beta\,(\sin^2\beta + 2)\right] - \frac{1}{3}\pi r_A^3 \left[2 - \cos\alpha\,(\sin^2\alpha + 2)\right] \qquad (18)$$

where

$$\alpha = \arccos\left[\frac{r_A^2 + b^2 - r_H^2}{2r_A b}\right] \qquad (19)$$

$$\beta = \arccos\left[\frac{r_H^2 + b^2 - r_A^2}{2r_H b}\right] \qquad (20)$$

r_A is the radius of the atom to which the hydrogen is bonded, r_H is the radius of the hydrogen atom and b is the covalent bond length. The values of these parameters have been calculated by the author and they are recorded in table 3.2.

Although none of the authors mention it, it seems necessary to take account of the increased number of electrons in non-atomic scattering centres. If Z' is the number of electrons in the scatterer and Z is the atomic number of the major atom then a reasonable correction is effected by multiplying f(sin θ) by Z'/Z. For example in $O^{\frac{1}{2}-}$, Z'/Z = 8.5/8 and

Table 3.4 : Summary of Features in Scattering Factor Calculations

Method	Sphere	Radius of sphere	Z'/Z correction
Fraser	Gaussian	$\left(\frac{3}{4\pi} V_s\right)^{1/3}$	Yes
Fuller	Hard	2A	No
Arnott	Hard	$\left(\frac{3}{4\pi} V_{vdw}\right)^{1/3}$	Yes
Langridge	Hard	2A	Yes

in $-CH_3$, $Z'/Z = 9/6$ etc. Inspection of the scattering factor curves published by Langridge et al (1960b) indicates that a correction of this kind was used by these authors.

We now compare scattering curves calculated with these methods. Table 4 summarises the characteristic features of the methods for convenience. Curves for each of the scattering groups listed by Langridge et al (1960b) are shown in figures 6(a-ℓ). In figures 7(a-ℓ) we compare scattering curves derived using Langridge's method both with a fixed sphere radius (2A) and with the radius set equal to that of a sphere whose volume is V_s. Several points are worth noting:-

(1) The value of V_s for phosphorous recorded by Langridge et al (47 A^3) appears to be incorrect. If this value is used then $f'(0) = -0.51$ whereas they claim it should be 7.2. The value of V_s in table 3 has been derived using $f'(0) = 7.2$. The calculated curve (figure 6ℓ) then agrees well with that shown in figure 7 of Langridge et al.

(2) The van der Waals volumes used by Arnott are usually greater than the solvent volumes used by all the other authors. The discrepancy is particularly marked in the case of ring nitrogen (Table 3). The only exception is the CH_3 group and the values for C_2 and C_5 are almost equal in both methods.

(3) The objection to using V_{vdw} instead of V_s is demonstrated by the curves for oxygen (figures 6 h-k). All Arnott's curves are identical irrespective of the nature of the oxygen whereas the other methods give rise to differences albeit small ones. It is physically realistic to expect that O = (phosphorous) and - O - (sugar) will scatter differently since the former is more likely to be closely surrounded by water.

(4) The effect of the Z'/Z correction factor may be observed by comparing the curves calculated using the Fuller method with the others. At high values of ρ the curves for f' tend towards those for f except in the Fuller method. However, when Z' = Z, and so no correction is made, the Fuller and Langridge curves are very similar.

(5) The oscillatory nature of the Fourier transform of a hard sphere mentioned by Fraser et al (1978) is not manifest in the corrected scattering factors in the region of reciprocal space in which DNA scattering is significant since the first zero in ϕ occurs at $\rho = 0.36\ A^{-1}$ for a sphere with 2 A radius. The radii of the variable spheres are less than 2 A so their transforms will pass through zero at even higher scattering angles (figure 8).

(6) The peaks in Arnott's curves tend to occur at higher values of ρ than those from the other three methods and his curves are generally the lowest in magnitude as a result of the effect which the larger volumes he uses have on both the value of ϕ and the weight attributed to it.

(7) Fraser's curves are generally the largest in magnitude since the Fourier transform of a Gaussian sphere falls to zero quicker than that of a hard sphere with the same radius (figure 8).

(8) The effect on the curves of using a variable rather than fixed radius is most significant at high scattering angles (figure 7). The low angle discrepancy between the variable radius method of Arnott and the fixed radius methods of Langridge and Fuller is therefore due mainly to the weight attributed to ϕ by the sphere's volume.

Since these correction factors are largely empirical it is difficult to decide which is best so it is important to compare transforms

Figure 3.6 : Comparison of scattering factor curves of atoms found in polynucleotides with those obtained after correction for the effect of water

a) C_1, C_3 and C_4 (i.e. C_H)

b) C_2 and C_5 (i.e. CH_2)

c) C with no hydrogen atom

d) CH (base)

e) CH_3

f) N (ring)

g) NH_2

h) -o- (phosphate)

i) $O^{\frac{1}{2}-}$

j) -o- (sugar)

k) =O

l) P

Key

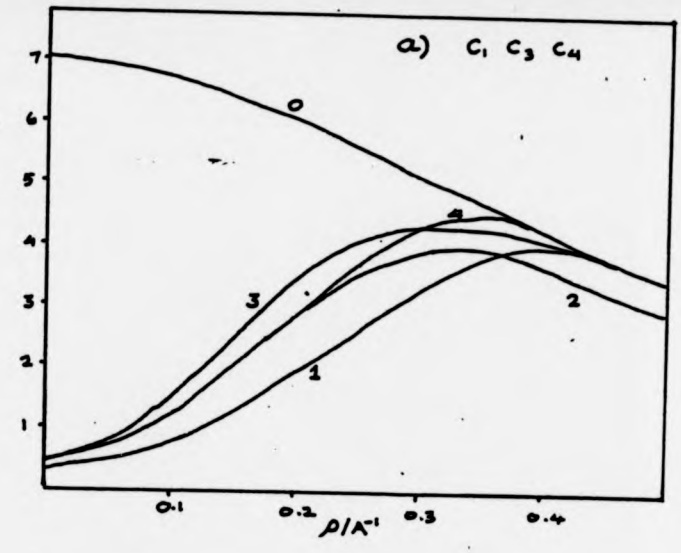
0) Uncorrected scattering factor

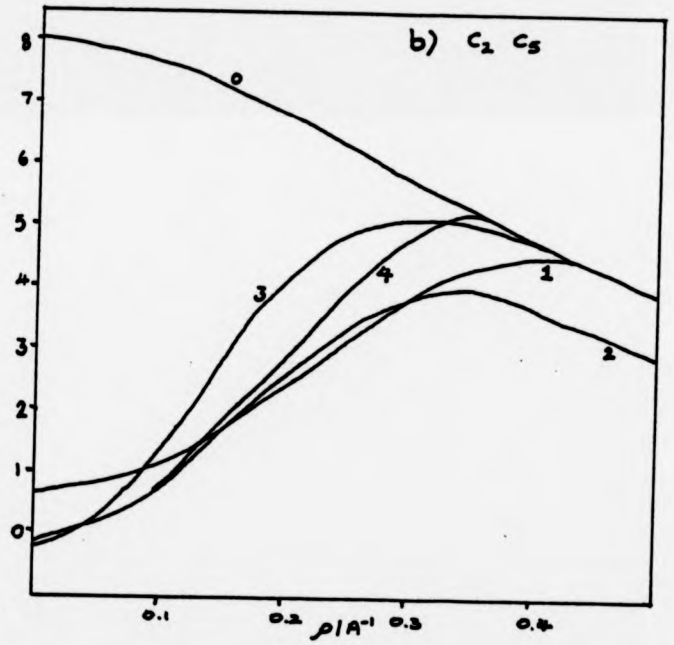
1) With Arnott correction

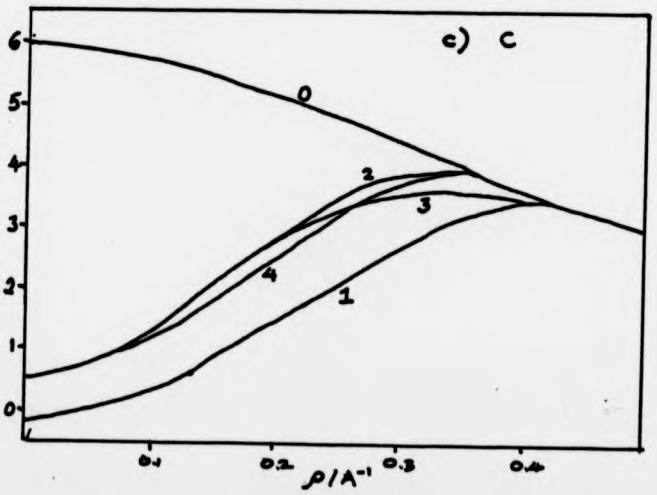
2) With Fuller correction

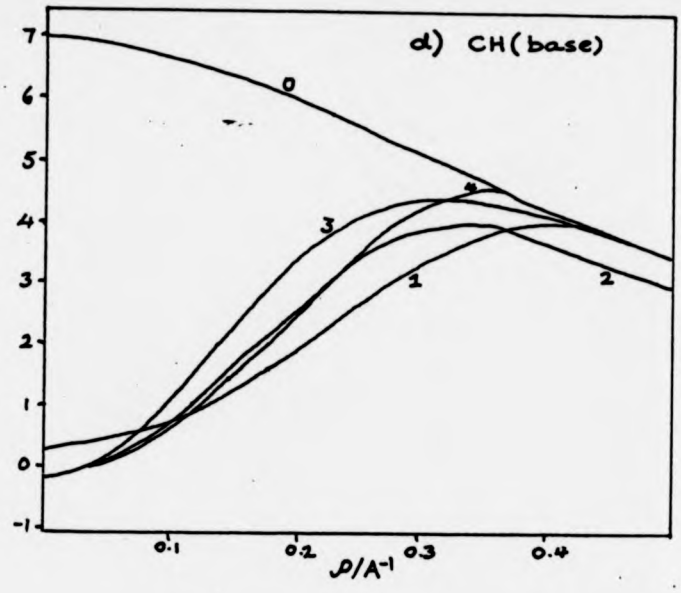
3) With Fraser correction

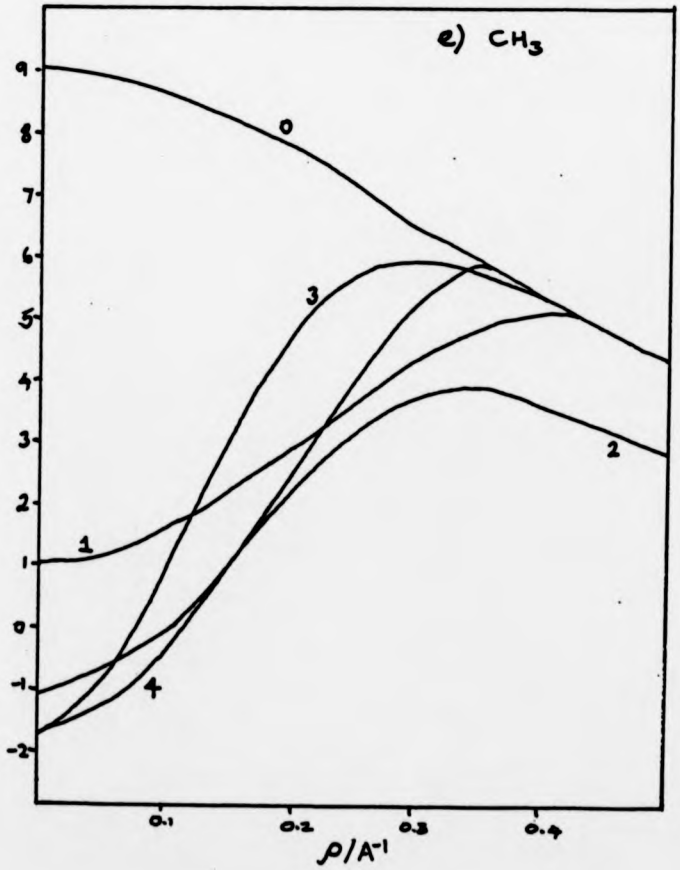
4) With Langridge correction

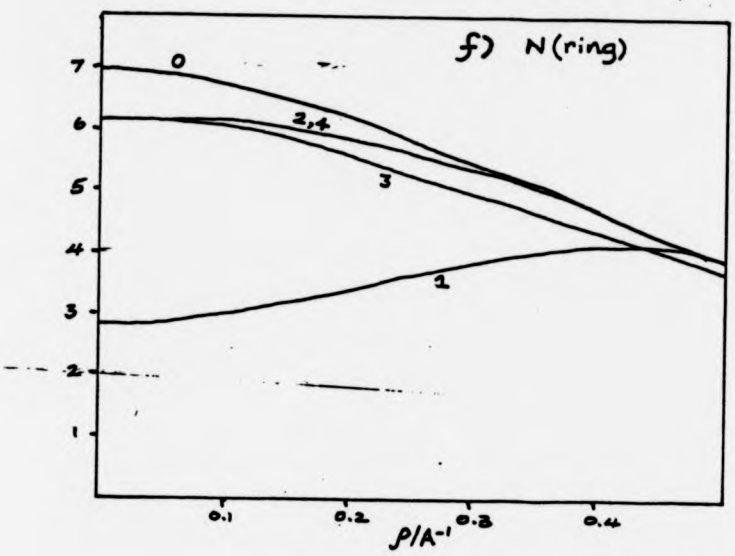

a) C_1 C_3 C_4
0
1
2
3
4
0.1
0.2
0.3
0.4
ρ/A^{-1}

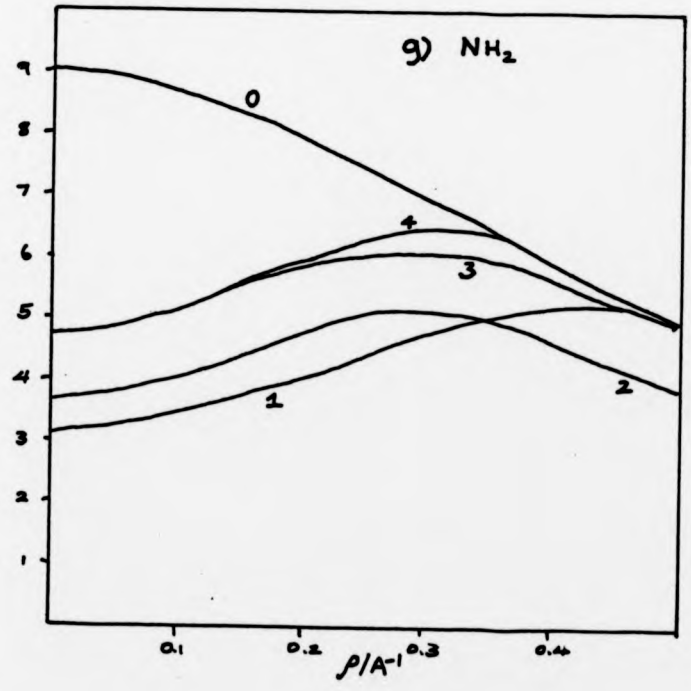

b) C_2 C_5
0
1
2
3
4
0.1
0.2
0.3
0.4
ρ/A^{-1}

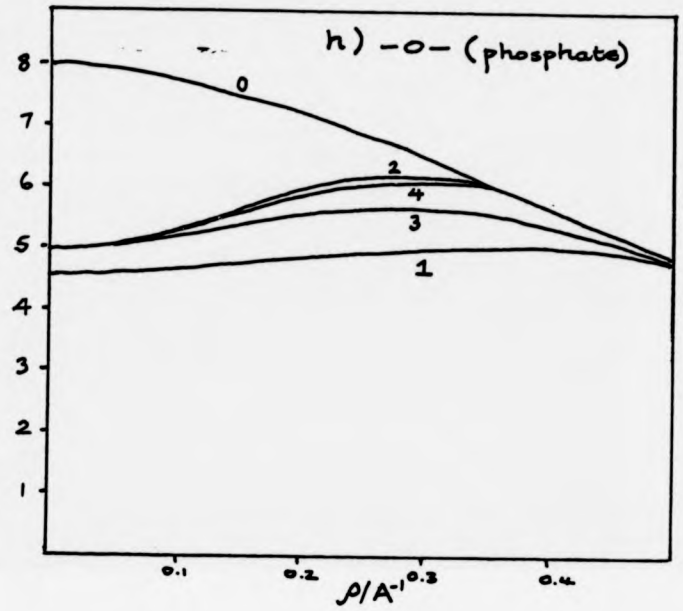

c) C
0
1
2
3
4
0.1
0.2
0.3
0.4
ρ/A^{-1}

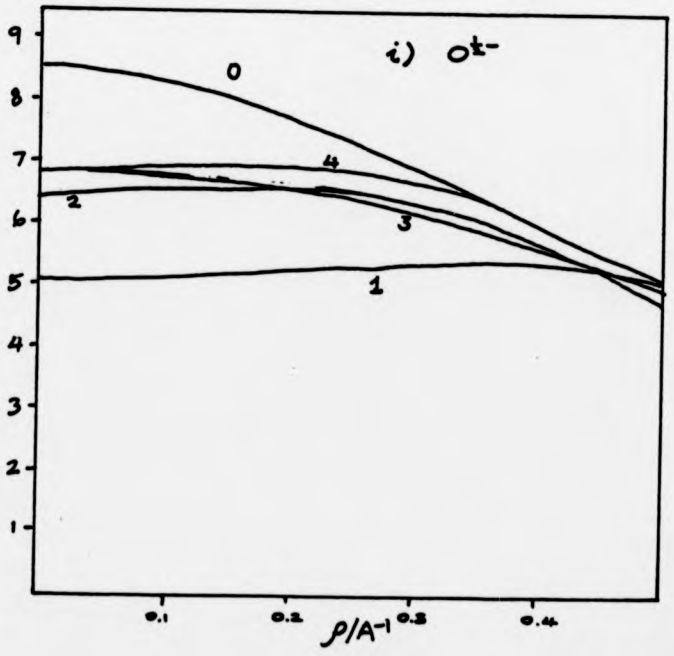

d) CH(base)
0
1
2
3
4
7
6
5
4
3
2
1
0
-1
0.1
0.2
0.3
0.4
ρ/A⁻¹

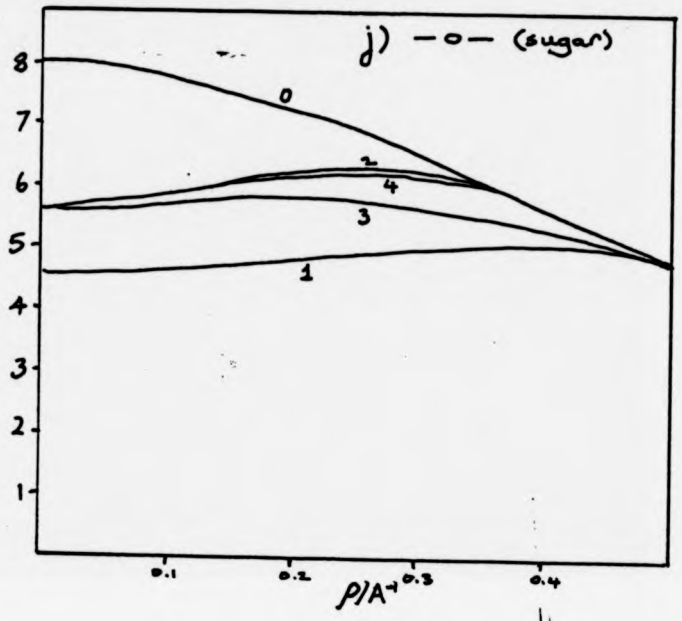

e) CH3
0
1
2
3
4
9
8
7
6
5
4
3
2
1
0
-1
-2
0.1
0.2
0.3
0.4
ρ/A⁻¹

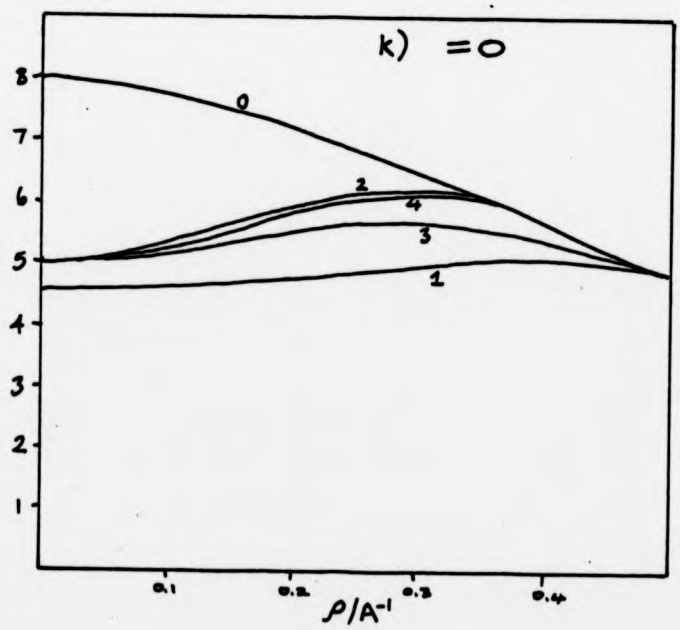

f) N(ring)
0
2,4
3
1
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/A⁻¹


g) NH₂
0
4
3
1
2
9
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/A⁻¹


h) -o- (phosphate)
0
2
4
3
1
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/A⁻¹


i) O²⁻
0
4
2
3
1
9
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/A⁻¹


j) —o— (sugar)
8
7
6
5
4
3
2
1
0
2
4
3
1
0.1
0.2
0.3
0.4
ρ/A⁻¹


k) =o
8
7
6
5
4
3
2
1
0
2
4
3
1
0.1
0.2
0.3
0.4
ρ/A⁻¹

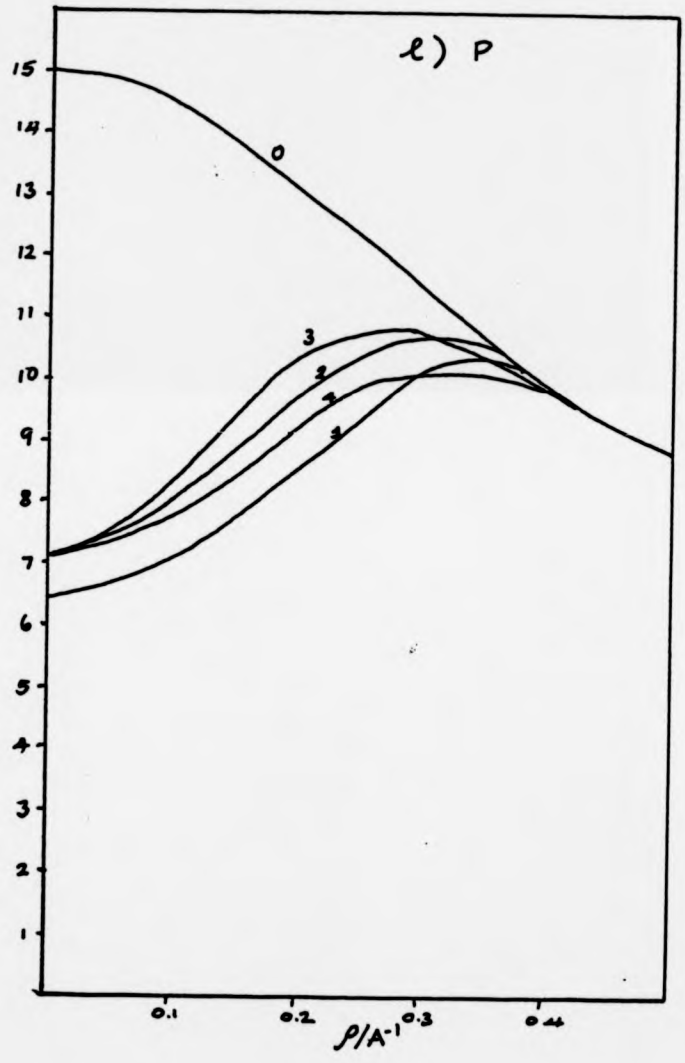

ℓ) P
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
3
2
1
0.1
0.2
0.3
0.4
ρ/A⁻¹

Figure 3.7 : Water-weighted scattering factor curves obtained using the Langridge method with the sphere radius set to 1) 2A and 2) the van der Waals radius (R) of the scattering group. Uncorrected scattering factors (0) are shown for comparison.

a) C_1, C_3 and C_4

b) C_2 and C_5

c) C

d) CH

e) CH_3

f) N(ring)

g) NH_2

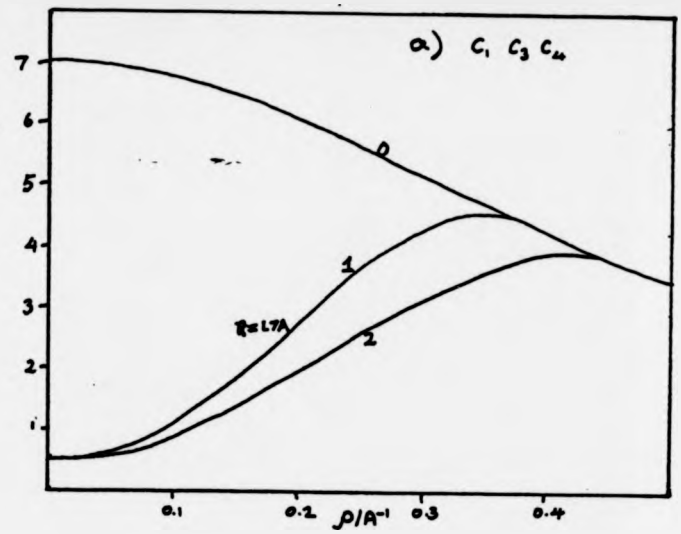
h) -o- (phosphate)

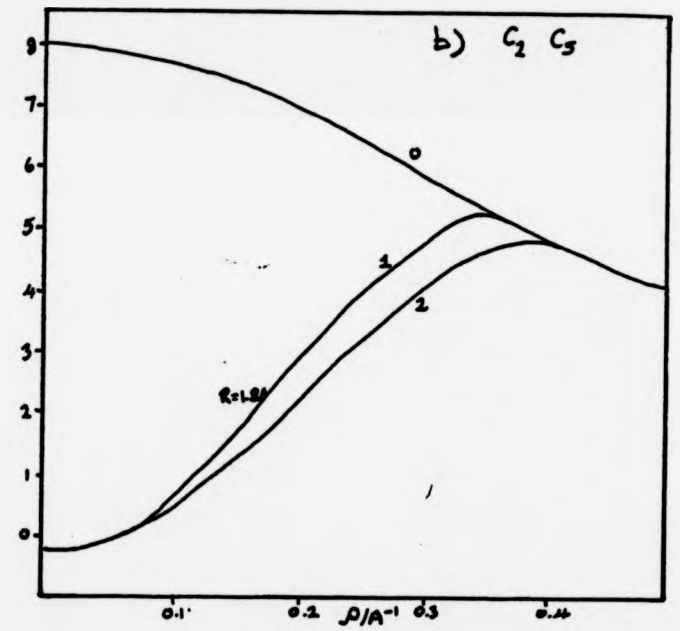
i) $O^{\frac{1}{2}-}$

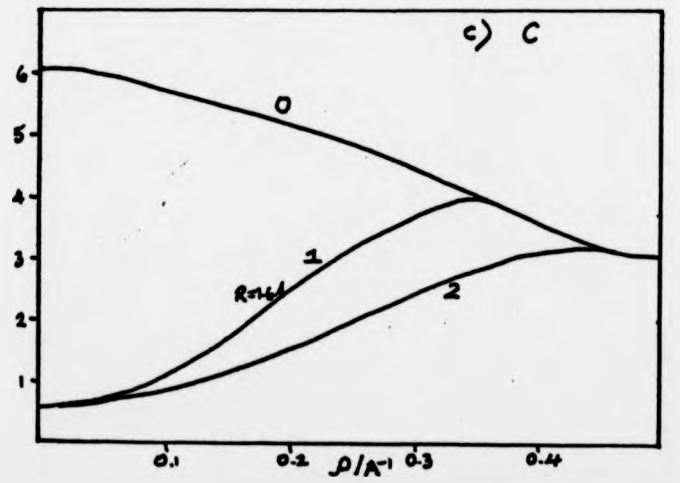
j) -o- (sugar)

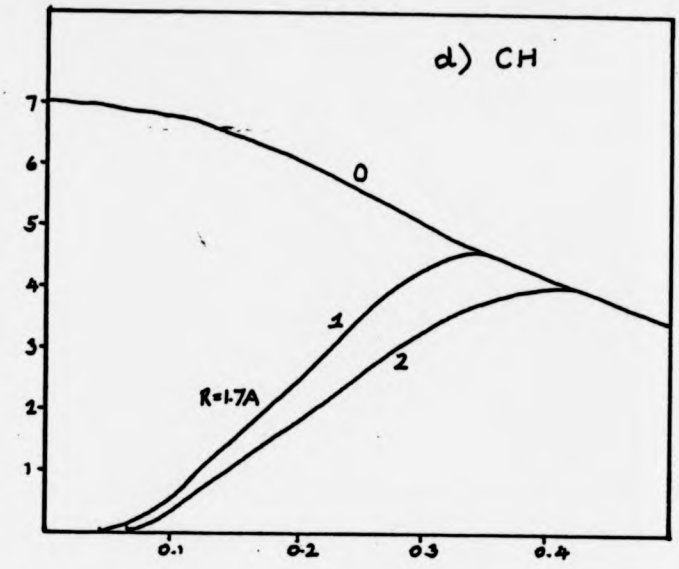
k) =O

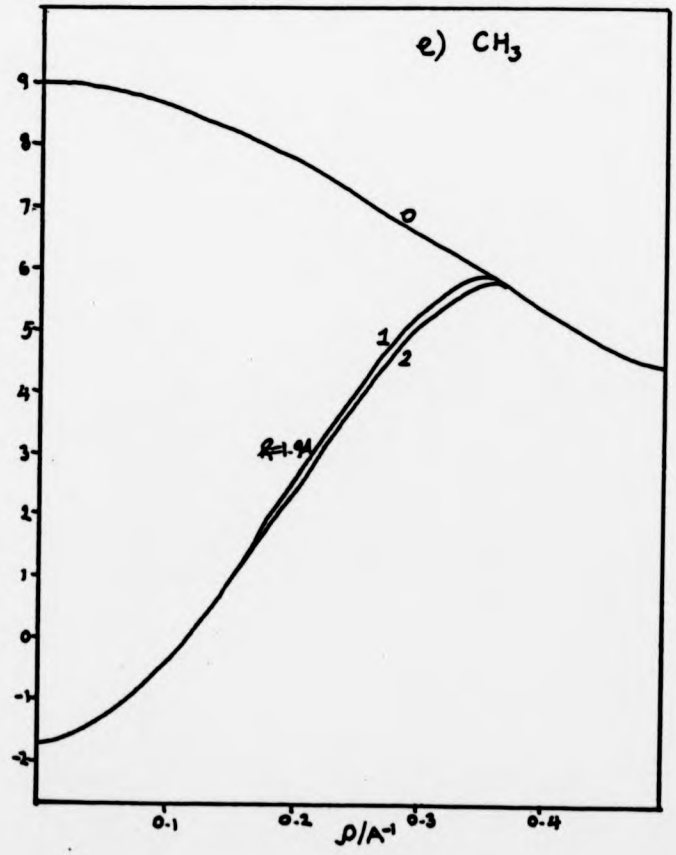
l) P

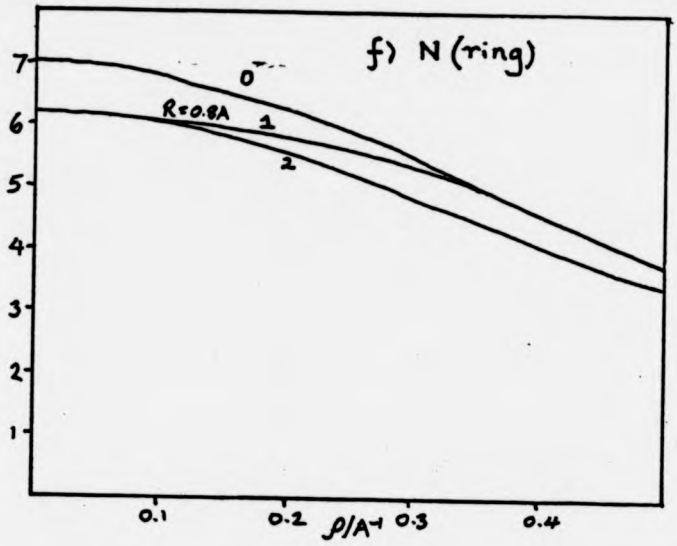

a) C_1 C_3 C_4
7
6
5
4
3
2
1
0
1
2
R=1.7Å
0.1
0.2
$\rho/Å^{-1}$
0.3
0.4

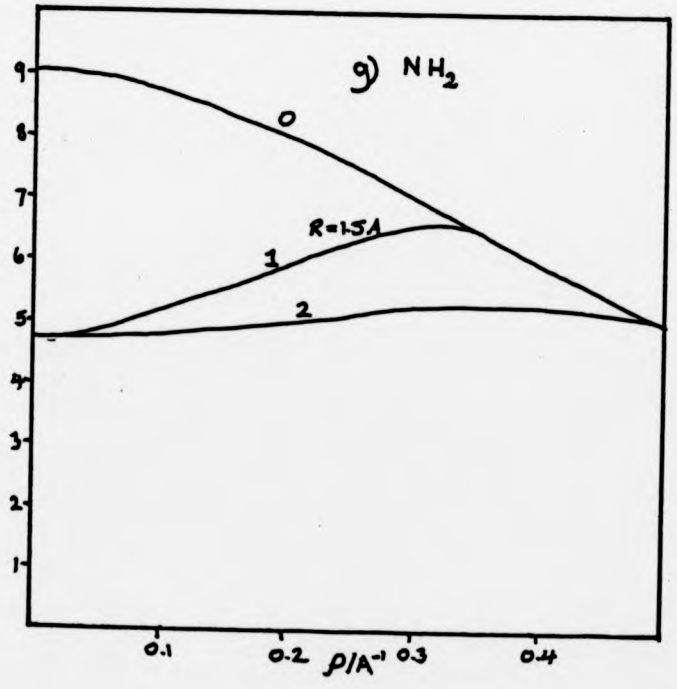

b) C_2 C_5
8
7
6
5
4
3
2
1
0
0
1
2
R=1.8Å
0.1
0.2
$\rho/Å^{-1}$
0.3
0.4

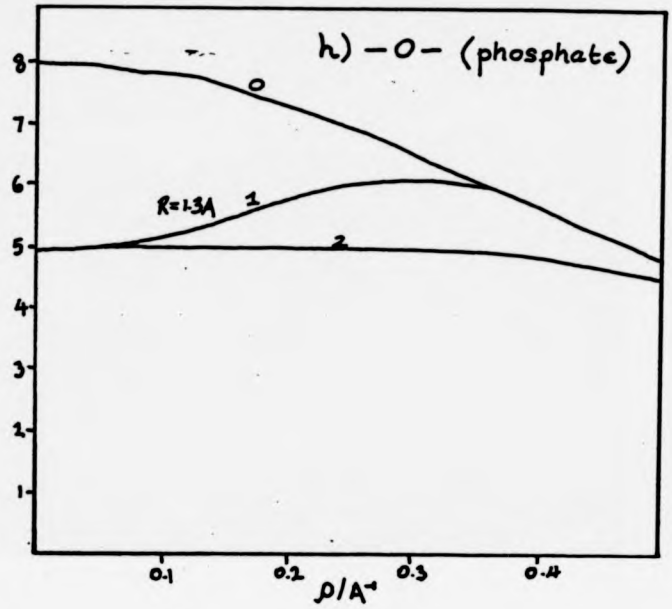

c) C
6
5
4
3
2
1
0
1
2
R=1.6Å
0.1
0.2
$\rho/Å^{-1}$
0.3
0.4

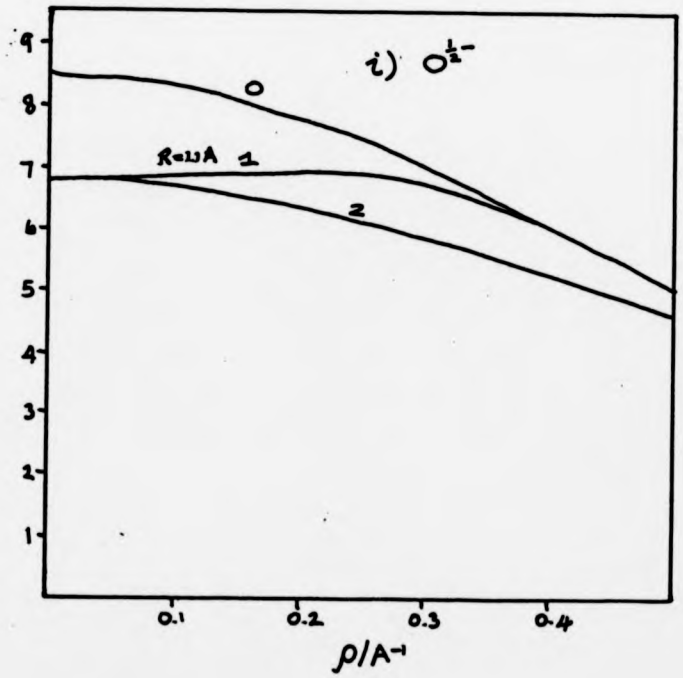

d) CH
7
6
5
4
3
2
1
0
1
2
R=1.7A
0.1
0.2
0.3
0.4

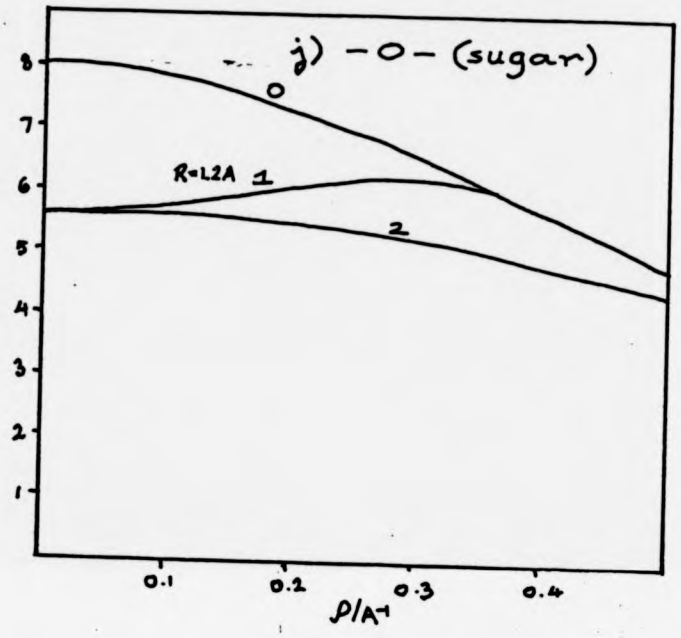

e) CH_3
9
8
7
6
5
4
3
2
1
0
-1
-2
0
1
2
R=1.9A
0.1
0.2
0.3
0.4
p/A^{-1}

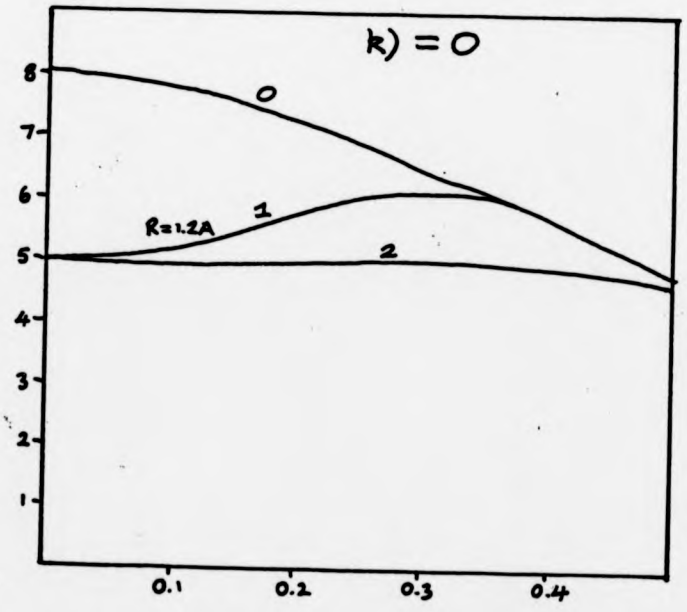

f) N (ring)
0
R=0.8A
1
2
7
6
5
4
3
2
1
0.1
0.2
ρ/A⁻¹
0.3
0.4

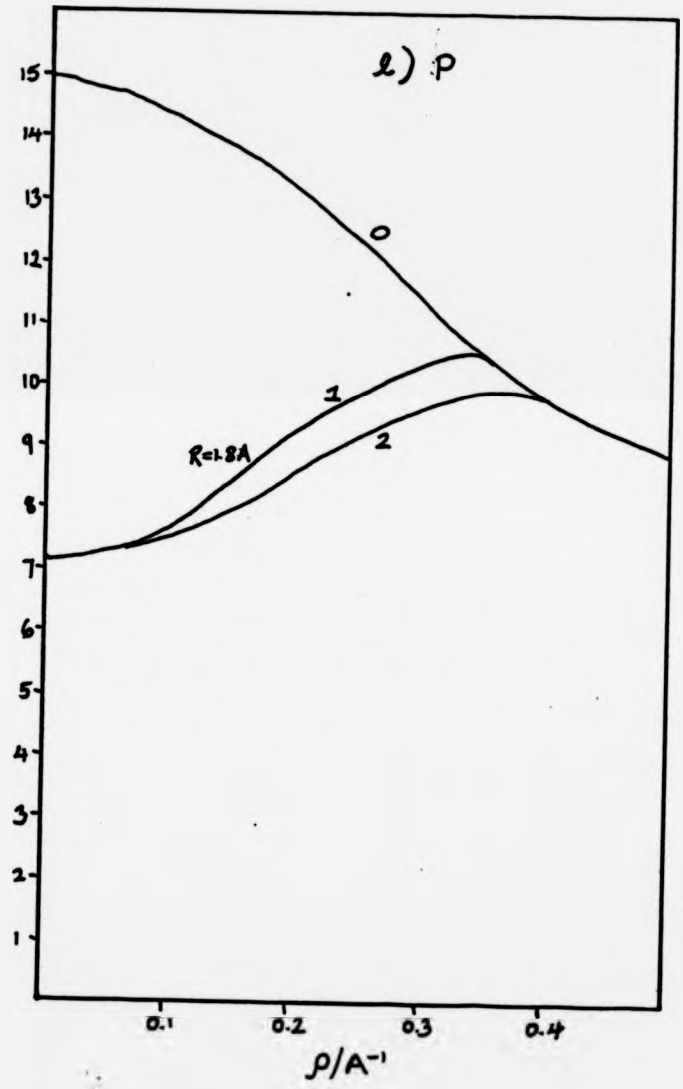

g) NH2
0
R=1.5A
1
2
9
8
7
6
5
4
3
2
1
0.1
0.2
ρ/A⁻¹
0.3
0.4


h) —O— (phosphate)
0
R=1.3Å 1
2
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/Å⁻¹


i) O½⁻
0
R=1.3Å 1
2
9
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/Å⁻¹


j) -O- (sugar)
0
R=1.2A 1
2
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4
ρ/A⁻¹


k) = 0
0
R=1.2A
1
2
8
7
6
5
4
3
2
1
0.1
0.2
0.3
0.4


2) P
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
1
2
R=1.8A
0.1
0.2
0.3
0.4
ρ/A⁻¹

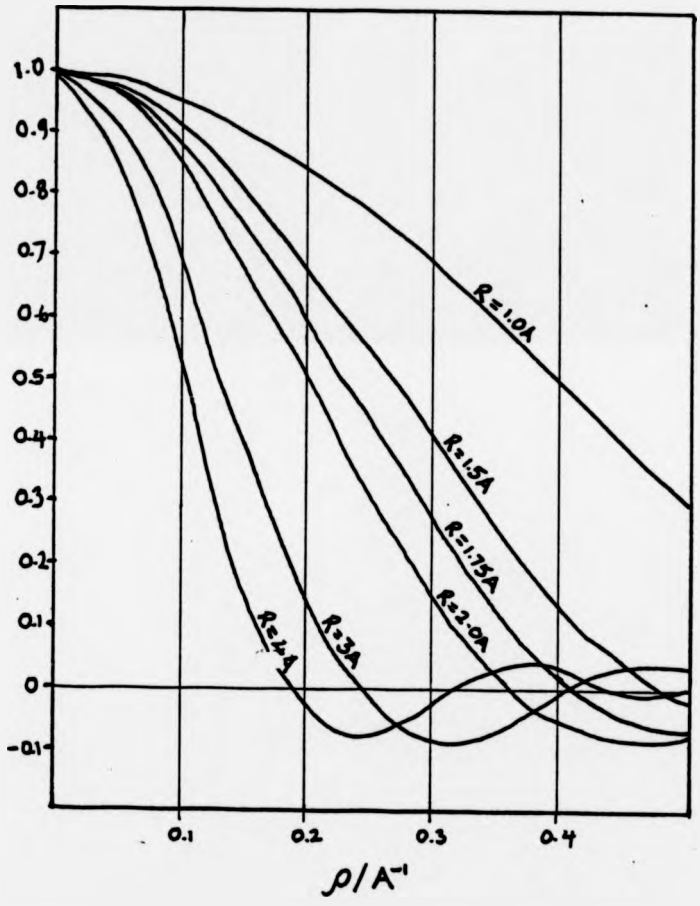
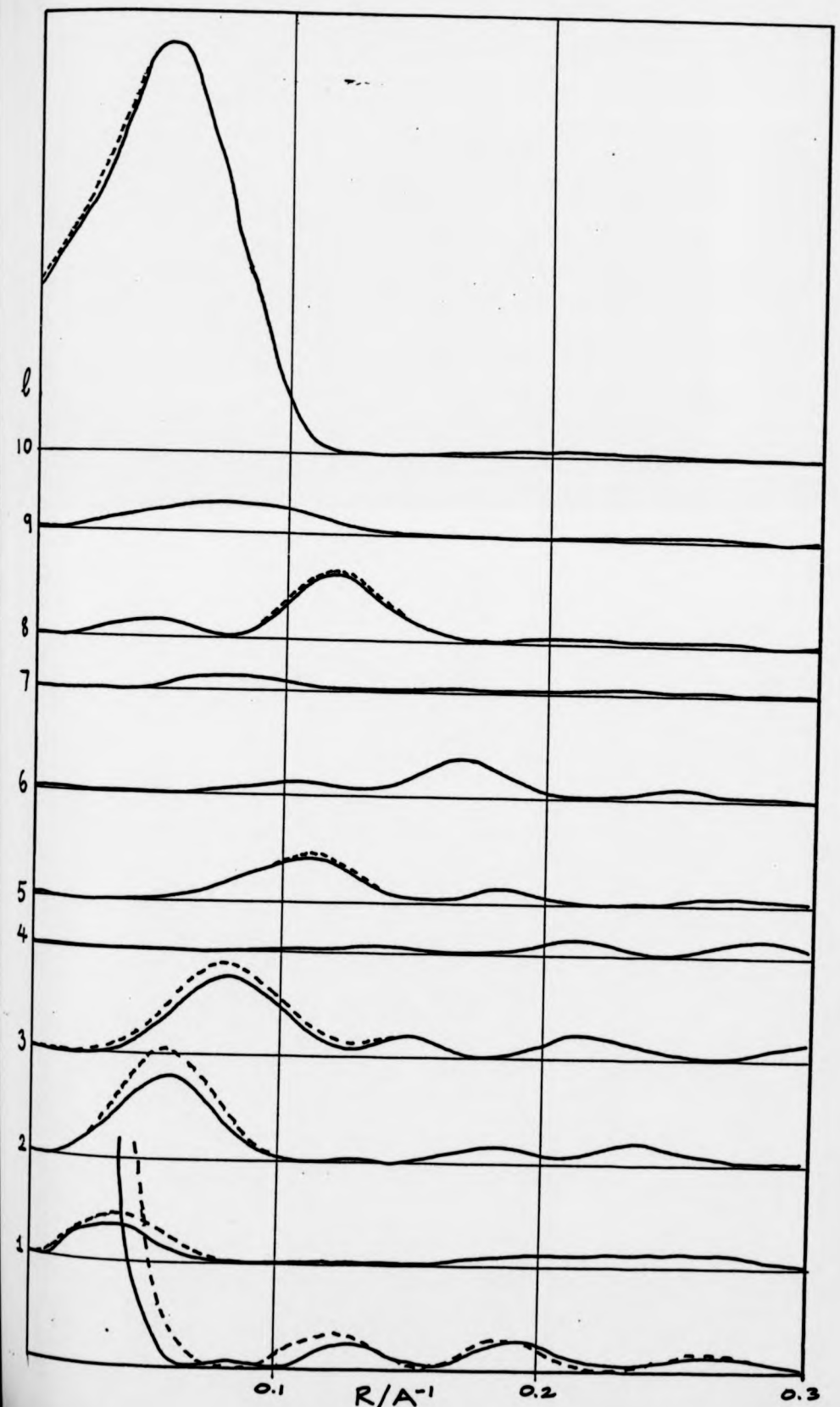


Figure 3.8 : Comparison of the scattering amplitudes of hard spheres with various radii R.

Figure 3.9 : Comparison of the cylindrically averaged squared Fourier transforms of B-DNA calculated using Arnott's scattering factors (——) and Langridge's scattering factors (---). In each case the temperature factor was set at $4A^2$.


l
10
9
8
7
6
5
4
3
2
1
0.1
R/A⁻¹
0.2
0.3

calculated with them in order to see if any differences are significant. Figure 9 shows a comparison of the cylindrically averaged squared transforms of B-DNA (Arnott and Hukins, 1972b) calculated using Arnott's and Langridge's scattering factors. The largest discrepancies occur at low values of ρ as expected. In particular the first zero on the equator occurs at a smaller value of ρ using Arnott's method and in addition the peak heights on the first three layer-lines are lower. These features, which may be significant when comparing models published by different authors, can change the value of R1 by about five percentage points. We will return to this point when discussing the models later in this chapter. Unless otherwise stated all transforms in this thesis have been calculated using Langridge's method.

3.2.3 Criticism of the Linked-Atom Least-Squares Technique

The linked-atom least-squares refinement method, which has been described in Chapter 2, is now widely used in the analysis of polynucleotides, polysaccharides and fibrous polypeptides. Here we consider two aspects which may have important implications for the accuracy of the results.

Consider the idealised six atom molecule shown in figure 10. The chain runs from atom 1 to atom 6 at the origin and its conformation is defined by the torsion angles τ_1, τ_2, τ_3 and τ_4. We wish to minimise a function Φ such as that defined in equation 2.35. Now if the discrepancies between observed and calculated structure factors or non-bonded contacts are the properties of the model which are to be minimised then Q_c depends explicitly upon the atomic co-ordinates. However, the torsion angles are the explicit parameters in the LALS refinement algorithm and the co-ordinates are merely implicit variables. The rather complex relationship between the torsion angles and the co-ordinates introduces a problem which has not been discussed elsewhere. Suppose we concentrate upon the minimisation of the

of the discrepancies between observed and calculated structure factors. Let G_i be the amplitude of the transform of atom i. Then G_i is a function of $\underline{S}$ (the reciprocal space vector) and $\underline{r}_i$ (the position vector of the atom). Now $\underline{r}_i$ is a function of τ_1, τ_2, τ_3 and τ_4; $\underline{r}_2$ is a function of only τ_2, τ_3 and τ_4 and so on. We may therefore write the calculated amplitude of the molecular transform at $\underline{S}$ in the form:-

$$G_c(\tau_1, \tau_2, \tau_3, \tau_4, \underline{S}) = G_1(\tau_1, \tau_2, \tau_3, \tau_4, \underline{S}) + G_2(\tau_1, \tau_3, \tau_4, \underline{S}) + G_3(\tau_3, \tau_4, \underline{S}) + G_4(\tau_4, \underline{S}) \qquad (21)$$

The discrepancy between the observed and calculated transform at $\underline{S}$ may be expressed in terms of the familiar first order Taylor expansion:-

$$G_o - G_c(\underline{S}) = \left(\frac{\partial G_1}{\partial \tau_1}\right) \Delta\tau_1 + \left(\frac{\partial G_1}{\partial \tau_2} + \frac{\partial G_2}{\partial \tau_2}\right) \Delta\tau_2 + \left(\frac{\partial G_1}{\partial \tau_3} + \frac{\partial G_2}{\partial \tau_3} + \frac{\partial G_3}{\partial \tau_3}\right) \Delta\tau_3 + \left(\frac{\partial G_1}{\partial \tau_4} + \frac{\partial G_2}{\partial \tau_4} + \frac{\partial G_3}{\partial \tau_4} + \frac{\partial G_4}{\partial \tau_4}\right) \Delta\tau_4 \qquad (22)$$

where $\Delta\tau$ are the shifts which we are seeking in the torsion angles. The obvious point which arises from this equation is that the torsion angles are not treated equally. Instead, whereas τ_1 affects only the contribution of atom 1, τ_4 affects the contributions of all the atoms. The physical significance of this is clear. If all the atoms have equal scattering power then a given change in τ_4 will generally have a more significant effect than the same change in τ_1. Thus the usual assumption of least-squares analysis, that the parameters are independent and have equal weight, is not satisfied.

It is important to determine whether this has any effect on the final value of the parameters of the model. It is difficult to devise an

adequate test of any physical significance, however, the author, in conjunction with Mr. A. Mahendrasingam, has investigated two methods of building the same dinucleotide. It is clearly desirable that the refinement, which was carried out using the energy function described in the previous chapter, should give the same result irrespective of the building method. In the first method (figure 11a) the main chain starts at C1' of the upper sugar (atom 1), traverses the sugar via a dummy bond to C3' (atom 2) and then runs down the backbone, to the C4' of the lower sugar, crosses the sugar via a second virtual bond to C1' and thence to N9 and C8 of a purine (atoms 9 and 10) and on to a dummy atom (number 11) situated on the diad axis. When built in this manner, τ_9 represents the twist angle between the bases and τ_{10} is the tilt angle. The angles τ_6 (which is determined by the sugar pucker) and τ_8 were kept fixed. The base tilt and twist were also fixed. Helical constraints were applied between the C1', C2' and C3' atoms of the two residues in order to preserve the correct rise and turn per residue. Those atoms shown on the diagram which do not form part of the main chain were added as pendants at appropriate points.

The second method of building the molecule progresses in the C3' to C5' direction as shown in figure 11b. The chain starts at C1' of the lower sugar and runs across the ring to C4', along the backbone to C3' of the upper sugar, across to the C1' and then via the base to the final reference frame in the same manner as described before. Once again the sugar pucker and base planarity were preserved by fixing the appropriate torsion angles (τ_6 and τ_8). It should be noted that these two methods do not strictly satisfy our requirement that the same torsion angles should be present in each case, but in a different order since, for example, the glycosyl angle τ_7 is defined differently. However, they are sufficiently similar for us to perform at least a preliminary test. In model 1, for

example, the torsion angle formed by the atoms 1C1', 1C3', O1 and P is the least significant since it is furthest from the origin, but in model 2 the same torsion angle is in the middle of the chain and would therefore be expected to be more significant.

The precise nature of the model we chose to build need not concern us here; in fact it is a right-handed C2'-endo B-DNA model which will be discussed in more detail later in this chapter. The torsion angles were derived using the program PREP from an initial set of co-ordinates. The co-ordinates of the two models calculated by the modelbuilding program prior to the first cycle of refinment were compared and they were found to agree within 0.01 A. The torsion angles which we may compare are shown in table 5. The first column contains the initial values used in both models. The final values of model 1 are shown in the second column and the final values of model 2 are in the third column. None of the angles differs by more than 6^{o} between the models and the final atomic co-ordinates are quite similar. The transforms of the models have not been calculated since it is very unlikely that such small differences will give rise to observable discrepancies. A difference is apparent, however, between the final Φ values of the models. Table 6 shows the initial and final value of Φ in both cases. The discrepancy between the initial values of this parameter is due to the precision with which the helical constraints are satisfied before refinement commences. This discrepancy arises from the accumulation of errors which is a familiar problem in computer procedures which utilise repeated multiplication. In fact the absolute value of Φ is of little importance since it depends on the weights which have been assigned to the constraints. The helical constraints are given a high weight and thus they tend to dominate Φ. The discrepancy between the initial Φ values of the models is therefore insignificant. However the final Φ of Model 1 is two orders of magnitude lower than that of Model 2.

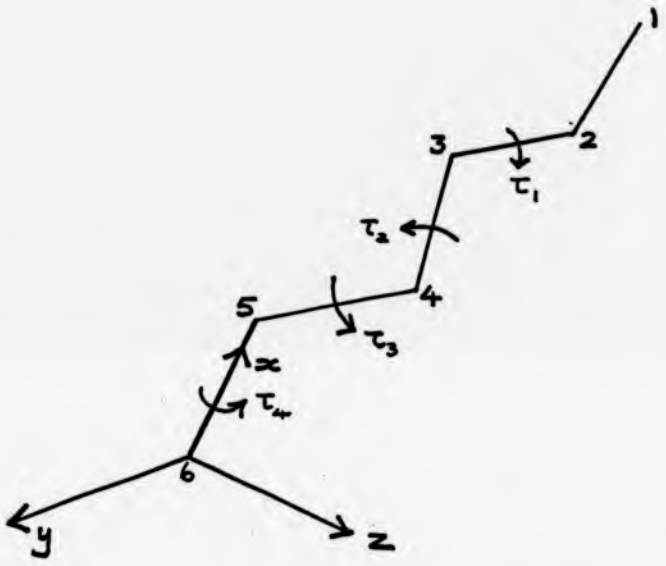


Figure 3.10 : The idealised molecule used in the discussion of the linked-atom least-squares technique

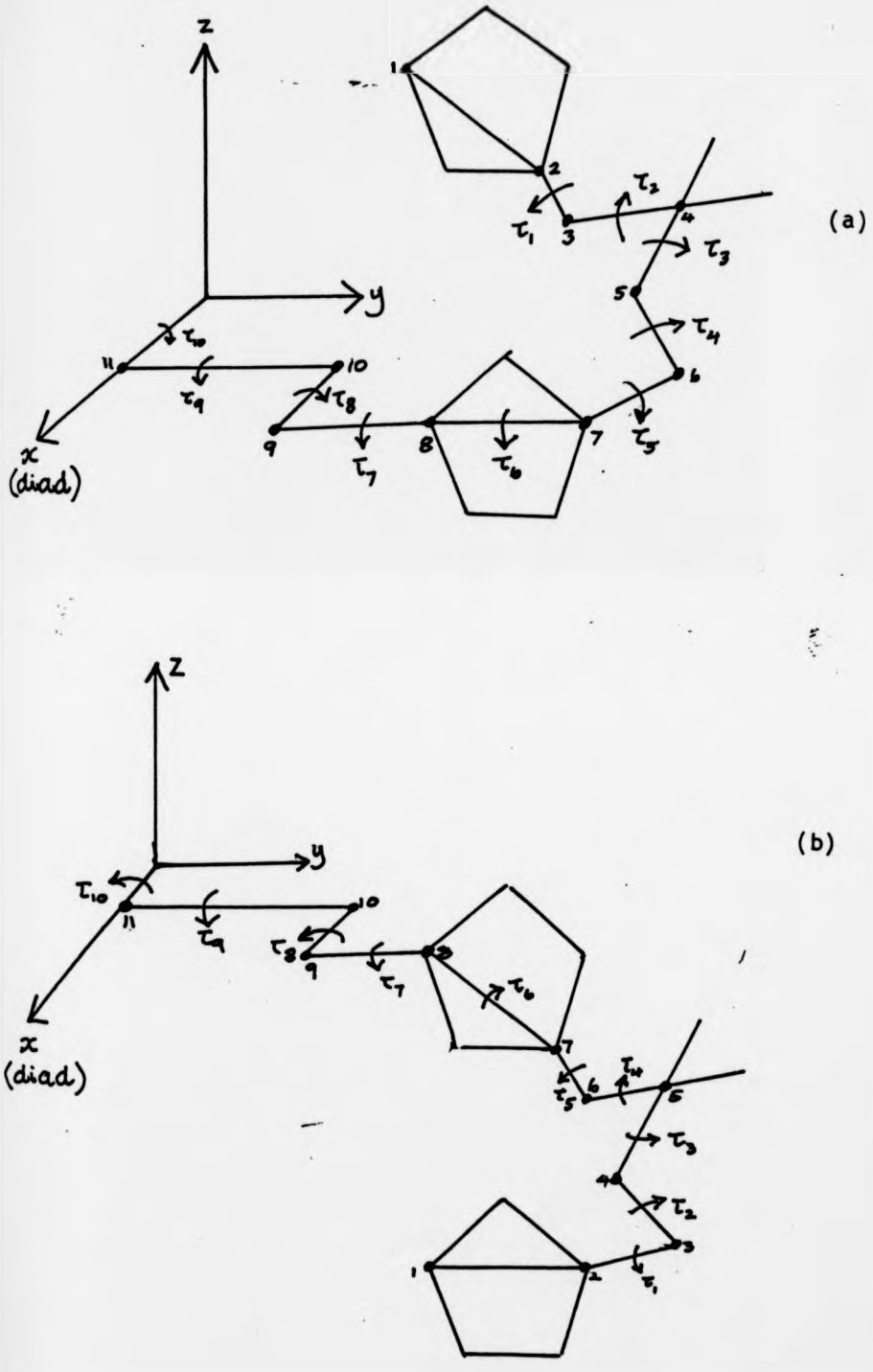


Figure 3.11 : The two dinucleotide modelbuilding methods used in the test of the linked-atom least-squares technique

Table 3.5 : Comparison of Initial and Final Torsion Angles of the LALS Test Models

Torsion angle	Initial Value	Final Value	
		Model 1	Model 2
τ(C1'-C3'-O1-P)	108.0	114.5	120.0
τ(C3'-O1-P-O4)	-115.0	-121.5	-124.4
τ(O1-P-O4-C5')	-54.0	-44.4	- 47.3
τ(P-O4-C5'-C4')	180.0	171.6	167.8
τ(O4-C5'-C4'-C1')	-33.0	-41.3	-40.0

Table 3.6 : Initial and Final 'Figures of Merit' (Φ) of the LALS Test Models

	Initial Φ	Final Φ
Model 1	51.0885	0.0067
Model 2	68.1849	0.2618

Table 3.7 : Final Values of the Helical Constraints in the LALS Test Models

Atom	Model 1			Model 2		
	Δr(A)	Δφ(deg)	Δz(A)	Δr(A)	Δφ(deg)	Δz(A)
C1'	0.000	0.00	0.000	-0.002	0.00	0.000
C2'	0.000	0.00	0.000	0.002	0.11	-0.001
C3'	0.000	0.00	0.000	0.000	0.05	0.000

Δr, Δφ and Δz are the discrepancies between the final values of r, φ and z for each atom and the corresponding values demanded by the helical constraints.

Once again the difference is due to the helical constraints which are satisfied better in Model 1 than in Model 2 (Table 7). Nonetheless the final discrepancies in Model 2 are still satisfactory.

The conclusion we may draw is that the direction in which the chain is built gives rise to only an insignificant difference in the final model. This may be an artefact of the modelbuilding procedure used. Firstly, the initial model we have studied is already nearly satisfactory. It would be more testing to study the behaviour of the program when the initial model is a greater distance from a minimum in Φ-space. Further work is currently in progress on this aspect. Second, the geometrical constraints on the chain may be the limiting factor in the determination of the final conformation. We may conclude that the linked-atom least-square technique is at least a practicable method of producing stereochemically acceptable models which will give a reasonable fit to experimental data. However, the efficacy of the method should not be allowed to obscure its unproven theoretical basis. Until the soundness of the method has been confirmed, it is doubtful whether this technique can be claimed to give the "best" possible model or whether it is justifiable to calculate standard deviations of the parameters as a measure of its accuracy (Campbell-Smith and Arnott, 1978; Arnott and Hukins, 1973). This implies that the attempt by Arnott (1980b) to discredit the B-DNA model of Levitt (1978) may be unsound. Levitt's model was derived using his own program (Levitt and Lifson, 1969) which refines against both X-ray data and stereochemical constraints. In the latter case terms are included which take into account not only non-bonded contacts but also covalent bond stretching and bending. Since helical constraints are not applied between one nucleotide and the next, irregularity may be present within the molecule. Arnott has re-refined this model using his own program and then used Hamilton's test (1965) to suggest that the B-DNA model of Arnott and Hukins (1972b) is superior. This procedure of course is open to the

objection that Levitt's model is not being used in the comparison since Arnott's program modifies it. But an interesting point which arises is that the backbone torsion angles of Levitt's original model are left largely unchanged by the LALS refinement whereas the base tilt and twist are significantly changed. This is precisely what one would expect if the unequal parameter weights are important since these are the two torsion angles closest to the origin and therefore the most significant in the refinement procedure. It is noteworthy that the propellor-like base twists predicted by Levitt and criticised by Arnott as stereochemically unreasonable have since beeen observed in the single crystal analysis of the B-DNA dodecamer (Wing et al, 1980).

We wish to suggest that the method of Levitt (1978) has several advantages over the LALS technique which make it desirable for use in future polynucleotide refinements. First, the parameters, which in this case are the atomic co-ordinates, are independent and equally weighted. Therefore it should be easier to devise quantitative statistical tests for use in the comparison of models. Second, since the parameters are independent it is less likely to produce biased results, which in the LALS method might be dependent upon the way in which the model is built. Third, if the relative positions of the sugar atoms, for example, are refined at all in the LALS method then the means of doing so are rather contrived whereas in Levitt's method the positions of all the atoms are refined in a more natural manner. In addition the disposition of the bases would no longer be defined rather artificially by the tilt, twist and displacement and small deviations from a 'perfect' Watson-Crick base-pair may readily be incorporated. Fourth, irregularity within the molecule, perhaps as a function of base-sequence, would be possible. Such effects have been observed in single crystal studies on oligonucleotides (Wing et al, 1980; Drew et al, 1981) but cannot easily be accommodated into the LALS procedure. Finally, the assumption of strict helicity in LALS modelbuilding may obscure the

the existence of many models with slight deviations from perfect symmetry. Hingerty (1979) has pointed out that only small variations in the covalent angles of the doublehelical loops of tRNA significantly altered the helical parameters of the model. Naturally the Levitt method is not without disadvantages the most obvious of which is the need for a set of "spring constants" in the energy terms for bond stretching and bending. However, a consistent set of such constants should be attainable from single crystal studies on short nucleic acid fragments and in particular from the B-DNA dodecamer and any future molecules of equal or greater length from which end effects should largely be eliminated.

3.3 Modelbuilding Studies

3.3.1 Base and Sugar Co-ordinates

Although many papers on polynucleotide conformations contain the co-ordinates of the bases, they are not immediately of use to anyone wishing to build models since they are presented with respect to a reference system in which the base planes are tilted and twisted. It is useful therefore to derive the co-ordinates in a system in which these rotations have been removed. The base co-ordinates used in the model-building studies in this thesis were derived from those of B-DNA (Arnott and Hukins, 1972b). A single cycle of the modelbuilding program was used to build the bases with the tilt and twist removed and with the line from purine N9 to pyrimidine N3 passing through the origin. The final co-ordinates are presented in table 8.

The three standard sugars derived by Arnott and Hukins (1972a) do not contain hydrogen atoms. Whilst these atoms do not contribute a great deal to the diffraction pattern and may therefore be ignored in Fourier transform calculations, nonetheless they may be involved in non-bonded interactions with other atoms in a model and their inclusion in modelbuilding

Table 3.8 : Atomic Co-ordinates of Watson-Crick Base-Pairs

Purine	x(A)	y(A)	z(A)	r(A)	ϕ(deg)
N1	-0.47	0.55	0	0.68	126.8
C2	0.80	1.09	0	1.36	53.8
N3	1.14	2.36	0	2.62	64.2
C4	0.04	3.12	0	3.12	89.3
C5	-1.26	2.72	0	2.99	114.9
C6	-1.49	1.34	0	2.01	138.0
N7	-2.12	3.79	0	4.34	119.2
C8	-1.32	4.80	0	4.98	105.4
N9	0.00	4.48	0	4.48	90.0
Adenine					
N6	-2.71	0.80	0	2.82	163.6
O6	-2.61	0.73	0	2.71	164.6
Guanine					
N2	1.93	0.22	0	1.94	6.5
Pyrimidine					
N1	-0.82	2.29	0	2.43	109.7
C2	0.24	3.12	0	3.12	85.6
O2	1.40	2.73	0	3.06	62.8
N3	0.00	4.48	0	4.48	90.0
C4	-1.26	4.96	0	5.12	104.2
C5	-2.34	4.14	0	4.75	119.5
C6	-2.08	2.73	0	3.44	127.1
Cytosine					
N6	-3.08	1.89	0	3.61	148.4
Thymine					
O6	-3.11	1.94	0	3.66	148.0
CH_3	-3.75	4.67	0	5.98	128.8

Table 3.9 : Atomic Co-ordinates of Sugars in the Four Major Puckers

The first three sugars are in the same reference frame as those of Arnott and Hukins (1972a). The co-ordinates of the non-hydrogen atoms in the C2'-exo pucker are from Pigram (1968).

C3'-exo	x(A)	y(A)	z(A)
H1	-1.79	-0.43	-0.91
H21	-2.75	1.65	-0.49
H22	-2.06	1.76	1.11
H3	-0.60	3.20	-0.08
H4	1.33	1.45	-0.67
C3'-endo			
H1	-1.79	-0.54	-0.85
H21	-2.04	1.79	-1.08
H22	-2.76	1.65	0.50
H3	-0.68	2.17	1.60
H4	0.77	1.65	-1.00
C2'-endo			
H1	-1.81	-0.09	-0.99
H21	-2.79	1.67	0.26
H22	-1.78	1.37	1.66
H3	-0.59	3.14	0.63
H4	1.09	1.59	-0.87
C2'-exo			
C1	1.42	0.00	0.00
H1	1.80	-0.78	0.63
C2	1.84	1.35	0.55
H21	1.80	1.41	1.62
H22	2.81	1.67	0.21
C3	0.76	2.24	-0.06
H3	0.97	2.49	-1.07
C4	-0.49	1.37	0.00
H4	-1.05	1.55	0.89
O5	0.00	0.00	0.00
C5	-1.38	1.60	-1.21

studies is therefore sometimes desirable. Their co-ordinates were calculated using one cycle of the modelbuilding program. The chain was built in the same manner as that described by Arnott and Hukins (1972a) and the hydrogen atoms were added at appropriate points. It was assumed that the hydrogen to carbon bond length was 1.07 A and that the arrangement was tetrahedral in each case. Although the latter assumption is not strictly correct, the co-ordinates so derived are sufficiently accurate for our purposes since hydrogen atoms are relatively small and flexible and so their exact position is not critical. The co-ordinates for the hydrogens in sugars in the C2'-exo pucker were also derived in this manner. This pucker was not considered by Arnott and Hukins (1972a) since it had not been observed in any of the furanose rings which had been studied by single crystal techniques. However, Pigram (1968) has described a refined sugar in this pucker which he derived from modelbuilding studies and his ring co-ordinates have been used in the present case. Only the hydrogen atom co-ordinates of the first three sugars are presented in table 9 since the reference frame is the same as that used by Arnott and Hukins (1972a). The co-ordinates are given of all the atoms in the C2'-exo sugar since they are not easily accessible elsewhere.

3.3.2 Description of an Inverted Base-stacking Scheme

DNA consists of two chains of opposite polarity. As one looks into the minor groove of B-DNA the right-hand chain proceeds down the molecule from C5' to C3! Whilst building left handed B models we found that it was possible to build the chains in the opposite direction so that the right-handed chain proceeds from C3' to C5' down the molecule. This novel base-stacking scheme had apparently never been considered before and so all the published double-stranded polynucleotides and oligonucleotides previously published contained what we will call α-stacking. Subsequent to our discovery the second scheme, which we will call β-stacking, was observed in

the Z-DNA tetranucleotide (Wang et al, 1979). This structure contains alternating purine and pyrimidine bases along the molecule but our model-building studies have shown that β-stacking is accessible to DNA with any base-sequence.

3.3.3 The Design of Molecular Models for B-DNA and a Preliminary Comparison with the Diffraction Data

For 25 years all the published models for B-DNA have contained a right-handed screw axis. The possibility of left-handed B-DNA has been examined by Wilkins and co-workers although no details of the models have been published. Interest in left-handed helices declined when Fuller et al (1965) claimed that only right-handed models for A-DNA were acceptable and argued that B-DNA was probably of the same handedness since the A → B transition occurred with facility within fibres. However strong evidence in favour of left-handed Z-DNA helices has been obtained by Wang et al (1979) and Drew et al (1980). Left-handed sections were also proposed within the side-by-side model for B-DNA (Rodley et al, 1976; Sasisekharan and Pattabiraman, 1976). Therefore we undertook this study in order to evaluate left-handed models for B-DNA. Gupta et al (1980a, b) have published details of such a model which they claim is both stereochemically satisfactory and in good agreement with the B-DNA diffraction pattern. However, their model may be obtained simply by twisting the nucleotides of a conventional model about the helix axis : that is, it is topologically equivalent to the right-handed models for B-DNA published by Langridge et al (1960b) and Arnott and Hukins (1972b). We refer to such models as α-stacked. Whilst examining left-handed models we discovered the β-stacking described in the previous section which gives rise to conformations which are topologically distinct. This stacking was subsequently proposed by Hopkins (1981) who refers to it as chain configuration II.

He built several CPK models of A- and B-DNA which incorporated β-stacking but he gave no co-ordinates. Since this novel idea has not been explored in detail elsewhere we concentrated on building left-handed models of this type. The procedure adopted was essentially that described by Langridge et al (1960b) except that the final stage of refinement utilised the modelbuilding program described in Chapter II in order to impose the precise stereochemistry observed in single crystal studies on nucleic acid fragments. The procedure is illustrated by describing several models and their transforms.

The aim of the first stage was simply to obtain a plausible wire model (βLHB1) whose transform was consistent with the major features of the diffraction pattern of B-DNA. The sugar, which was held in the C3'-exo pucker, was maintained in the anti orientation and the bases, which were untilted and untwisted, were set 1A behind the helix axis. With these constraints it was possible to obtain a stereochemically acceptable sugar-phosphate chain conformation with the phosphorous atom about 8.5 A from the helix axis as indicated by the diffraction pattern. The cylindrically averaged intensity transform was promising as a first approximation (figure 12a). Superimposed upon this figure is the transform of the B-DNA model proposed by Arnott and Hukins (1972b) for comparison.

Both transforms pass through zero on the equator at R ≈ 0.08 A^{-1} but the amplitudes of the secondary peaks do not agree. Care must be exercised at high scattering angles since more than one Bessel function may be significant. The cylindrical transform takes no account of interference between such terms whereas structure factor calculations do include interference. However, inspection of the amplitudes published by Arnott and Hukins (1973) indicated that the predicted intensities were rather low at R = 0.2 A^{-1} so the large value of the transform of βLHB1 at this point was not considered to be a serious discrepancy. The transforms are in good agreement on the first layer-line except at R = 0.1 A^{-1} where an extra

peak is apparent in the βLHB1 curve. This peak, which was found to be a characteristic feature of left-handed B-DNA transforms, arises from constructive interference of the base and phosphate components. In right-handed models only a small peak arises since the contribution of the base transform tends to cancel those of the sugar and phosphate (compare figure 13a with figure 3 of Langridge et al, 1960b). The observed structure factors at this point are much weaker than those at lower scattering angles so the presence of the 0.1 A^{-1} peak will be an important discrepancy. On the second and third layer-lines the major βLHB1 peaks are too low whereas the transform is too high on the fourth layer-line where the observed intensity is very weak. The major contribution to the discrepancies on ℓ=2 and 4 arises from the position of the phosphorous which is situated such that $\theta+\phi = 45^{o}$ ($\theta = 2\pi z/c$; see Fuller (1961) and Fuller et al (1967)). Therefore on ℓ=2 the phosphate transform is modulated by $\cos 2(\theta + \phi) = 0$ and on $\ell = 4$ it is modulated by $\cos 4(\theta + \phi) = -1$. Agreement with the Arnott and Hukins transform is good on $\ell = 5$ and 7, but on the sixth layer-line the major peak of the βLHB1 transform occurs in the wrong position. The higher layer-lines arise from fine detail within the molecule which need only be considered later in the refinement process.

In a later model (βLHB3) the bases were maintained in the same position as in βLHB1 but the phosphate group was moved so that $\theta + \phi \simeq 70^{o}$ which was adjudged to be necessary to correct the discrepancies on layer-lines two and four. This improved the transform on $\ell = 2$, and to a lesser extent on $\ell = 3$ and $\ell = 0$, but it left $\ell = 4$ largely unchanged (figure 12b). In addition it reduced the extraneous peak on $\ell = 1$ but at the expense of reducing the inner peak also.

Consideration of the components of the transform and wire model building indicated that little further improvement could be achieved unless the base parameters were altered, in particular the tilt. It is necessary to consider the stereochemical consequences of this. In A-DNA where the

base displacement is high and positive the tilt must also be positive or short contacts arise between adjacent bases. If the displacement is negative the tilt must also be negative. But if the helix is left-handed the reverse is true : high positive tilts may only be coupled with negative displacements and vice versa. The average values of θ and ϕ with the base disposition employed hitherto were 0^{o} and 108^{o} respectively. Thus the amplitude of the modulation of the base transform on $\ell = 1$ was $\cos(\theta + \phi) = -0.31$. The $m = 0$ components of the transforms are shown in figure 13a. The extraneous peak at $R = 0.1\ A^{-1}$ might be reduced and the peak at $R = 0.03\ A^{-1}$ might be enhanced, as required, if the base transform were reduced without affecting the other two components. If we assume that tilting the base has a negligible effect on the average value of ϕ then the mean z co-ordinate required to reduce the base component by a factor of two is given by $\cos(\theta + 108^{o}) = -0.15$, or $\theta \simeq -10^{o}$ which corresponds to $z = -0.94$ A. Since the mean radial co-ordinate of the base atoms is 3.4 A this gives a tilt whose magnitude is $\arctan(0.94/3.4) \simeq 15^{o}$. Since the rotation about the tilt axis tends to reduce the z coordinates the tilt is positive according to our convention. Several models of this type with slightly different base tilt and displacement were built and their transforms were calculated. It was often possible to obtain good agreement on the lower layer-lines but all the transforms contained a very high peak at $R \simeq 0.08\ A^{-1}$ on $\ell = 9$. This is a serious discrepancy since no intense reflections are observed at this point.

Small negative base tilts were also examined. Although this is tilting against the sense of the helix, no serious inter-base contacts occur if the magnitude of tilting remains low. A range of models was built with tilt = -4^{o}, -8^{o} and -10^{o}. One such model (βLHB12) which was found to be promising had bases tilted by -4^{o} and displaced by -1A and the phosphate was situated such that $\theta+\phi = 60^{o}$. The cylindrical transform is given in figure 12c. Since the agreement was good on most layer-lines except the

eighth, this model was rebuilt using the refinement program described earlier. This has the advantage that precise stereochemistry may be imposed. In the resulting model (βLHB14) the phosphate group had moved considerably, which caused the transform to deteriorate, and short contacts had been introduced between O2 and C4' (2.3 A) and P and C4' (2.9 A). This possibility had been evident from the wire models and steps had always to be taken to remove these contacts. It would appear that this had introduced some other undesirable feature, for example eclipsed conformations, which according to the criteria of the modelbuilding algorithm were more serious than the two short contacts. Subsequent attempts to increase these distances (especially between O2 and C4') required the introduction of artificial constraints : (i) the van der Waals radius of O2 and O3 was increased to 2.0 A in order to push the phosphate group to a better position; (ii) the weight of the helical constraints was reduced relative to that of the non-bonded interactions; and (iii) the weight of the worst short contact was increased by a factor of ten. An acceptable model (βLHB16(3)) was finally obtained as a result of these changes.

On the equator (figure 12d) agreement is acceptable but the problem of the first layer-line has not been solved. Agreement is good on ℓ = 4,5,6 and 7 but the modifications required in order to move the major peak on the eight to a higher radial co-ordinate have caused the agreement on the second and third lines to deteriorate (compare figures 12c and 12d). This has also introduced an unwanted peak on the ninth layer-line but it is less serious than that produced by tilting the bases in the opposite direction. The m=0 and m=1 components of the transform are displayed in figure 13b and c.

The model obtained, whilst not entirely satisfactory, is no worse in accounting for B-DNA diffraction pattern than is the α-stacked left-handed model BIV of Gupta, Bansal and Sasisekharan (1980). They

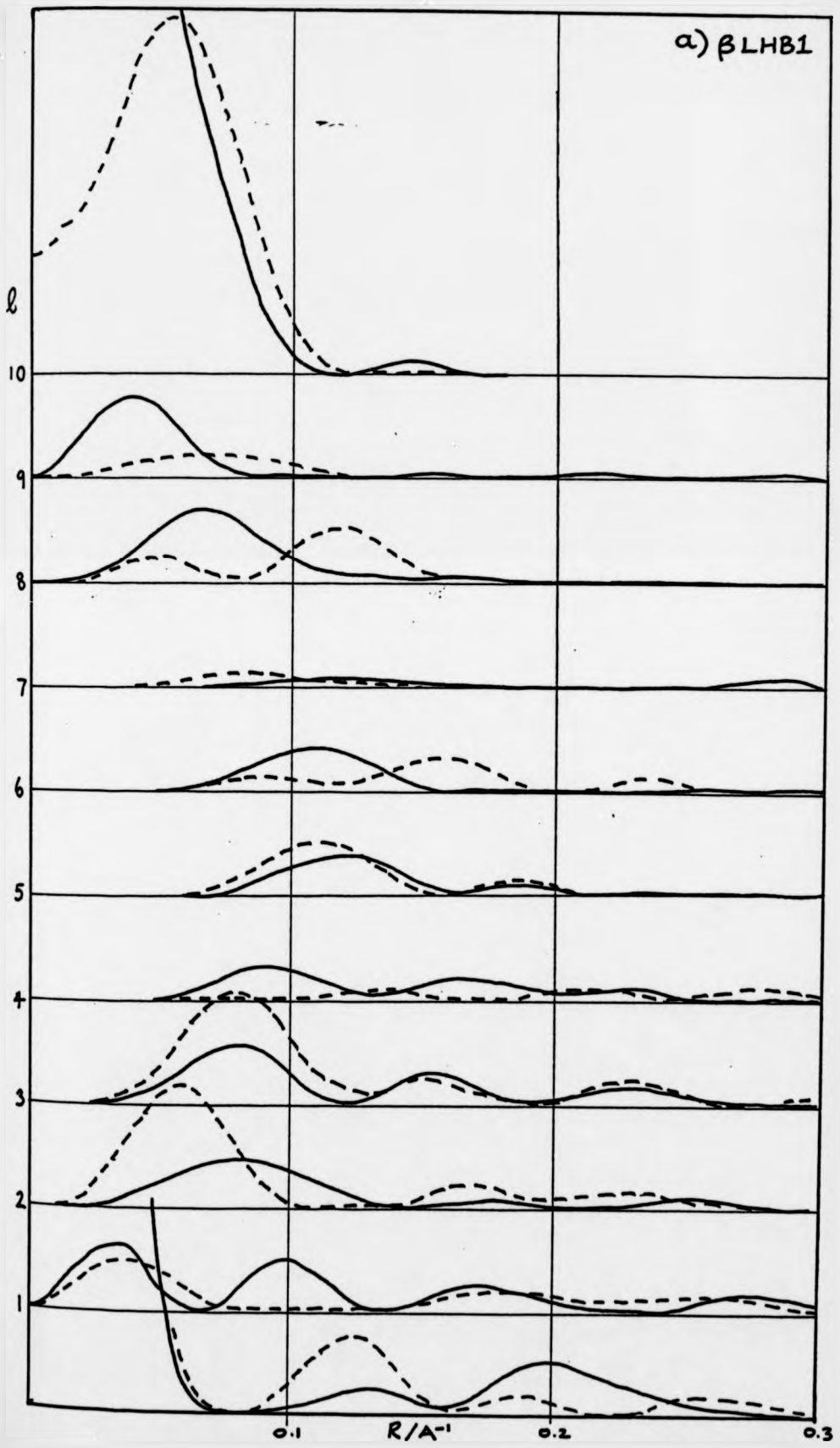
Figure 3.12 : Cylindrically averaged squared fourier transforms of β-stacked, left-handed B-DNA models (——) compared with the Arnott and Hukins model (---)

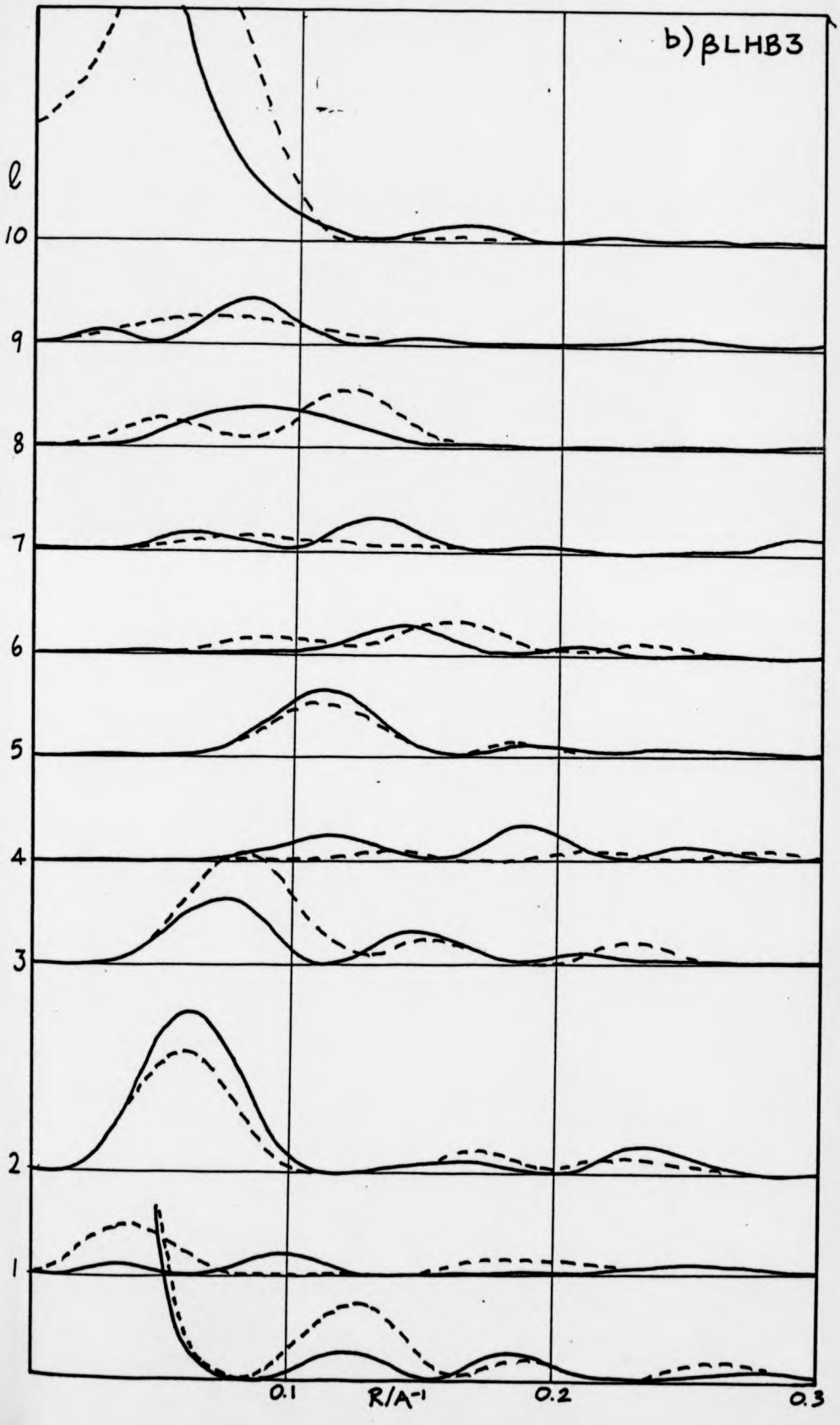
a) βLHB1

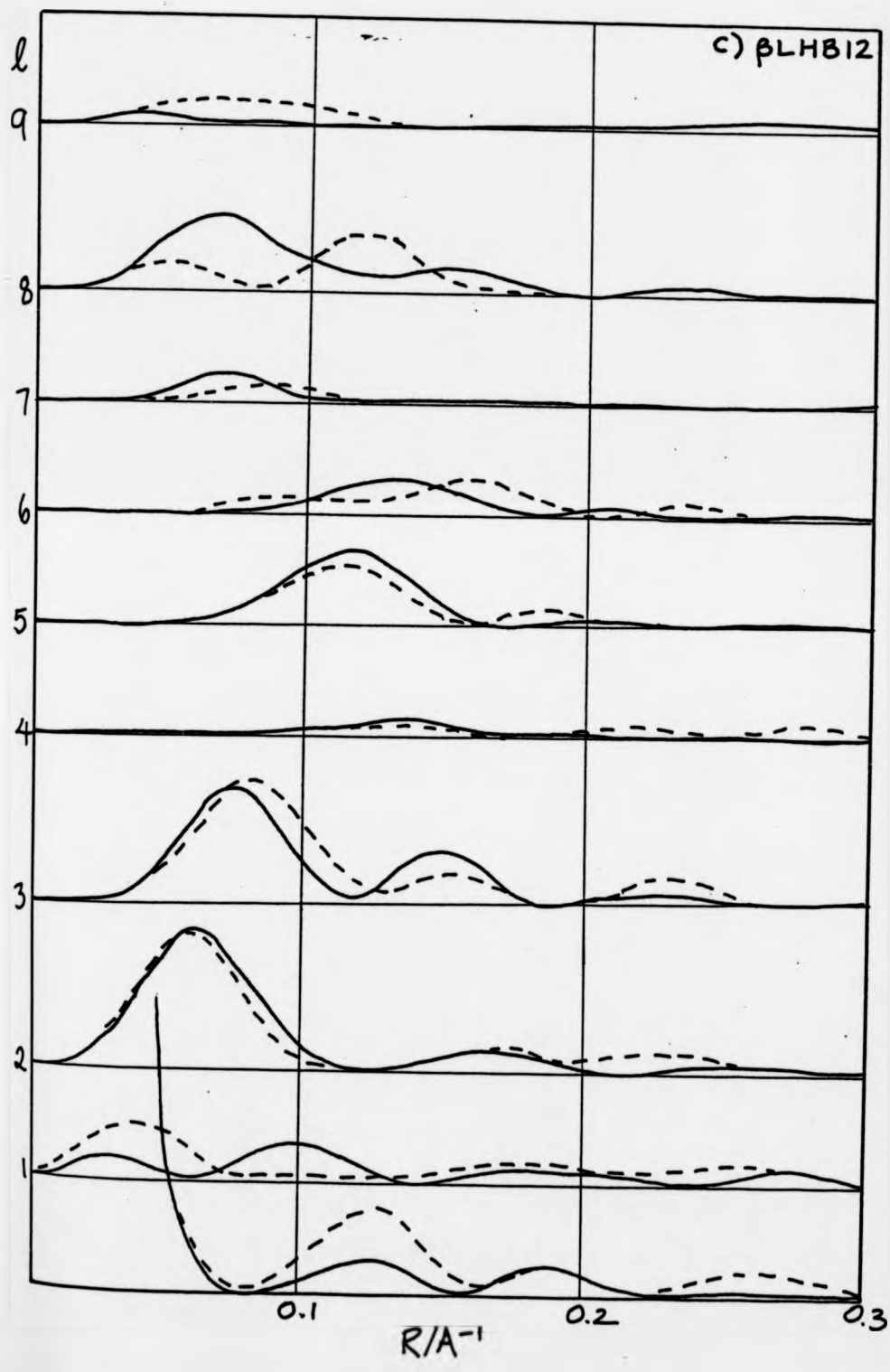
b) βLHB3

c) βLHB12

d) βLHB16 (3)


a) βLHB1
ℓ
10
9
8
7
6
5
4
3
2
1
0.1
R/A⁻¹
0.2
0.3


b) βLHB3
ℓ
10
9
8
7
6
5
4
3
2
1
0.1
R/A⁻¹
0.2
0.3


c) βLHB12
l
9
8
7
6
5
4
3
2
1
0.1
0.2
0.3
R/A⁻¹

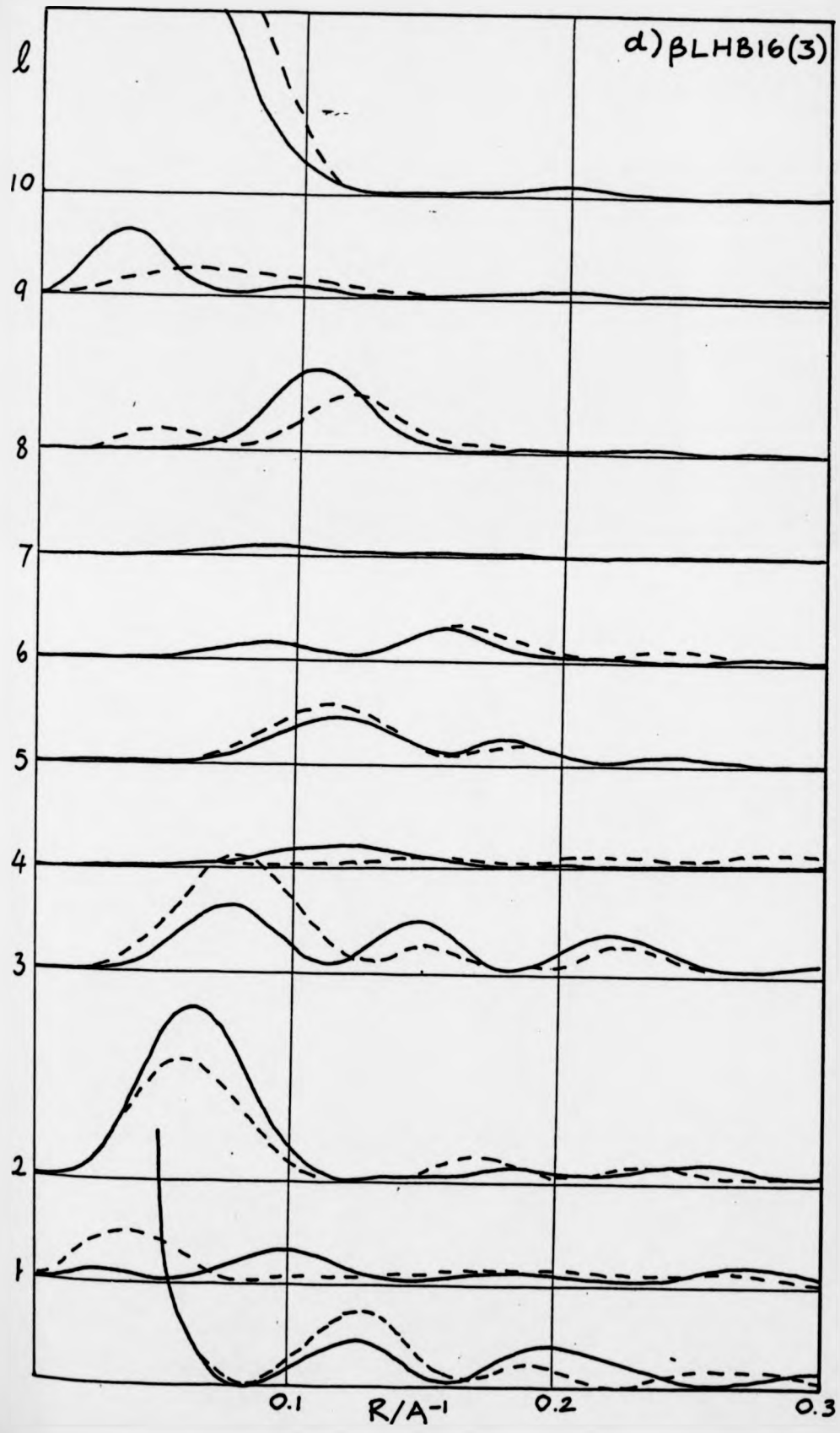

d) βLHB16(3)
l
10
9
8
7
6
5
4
3
2
1
0.1
R/A-1
0.2
0.3

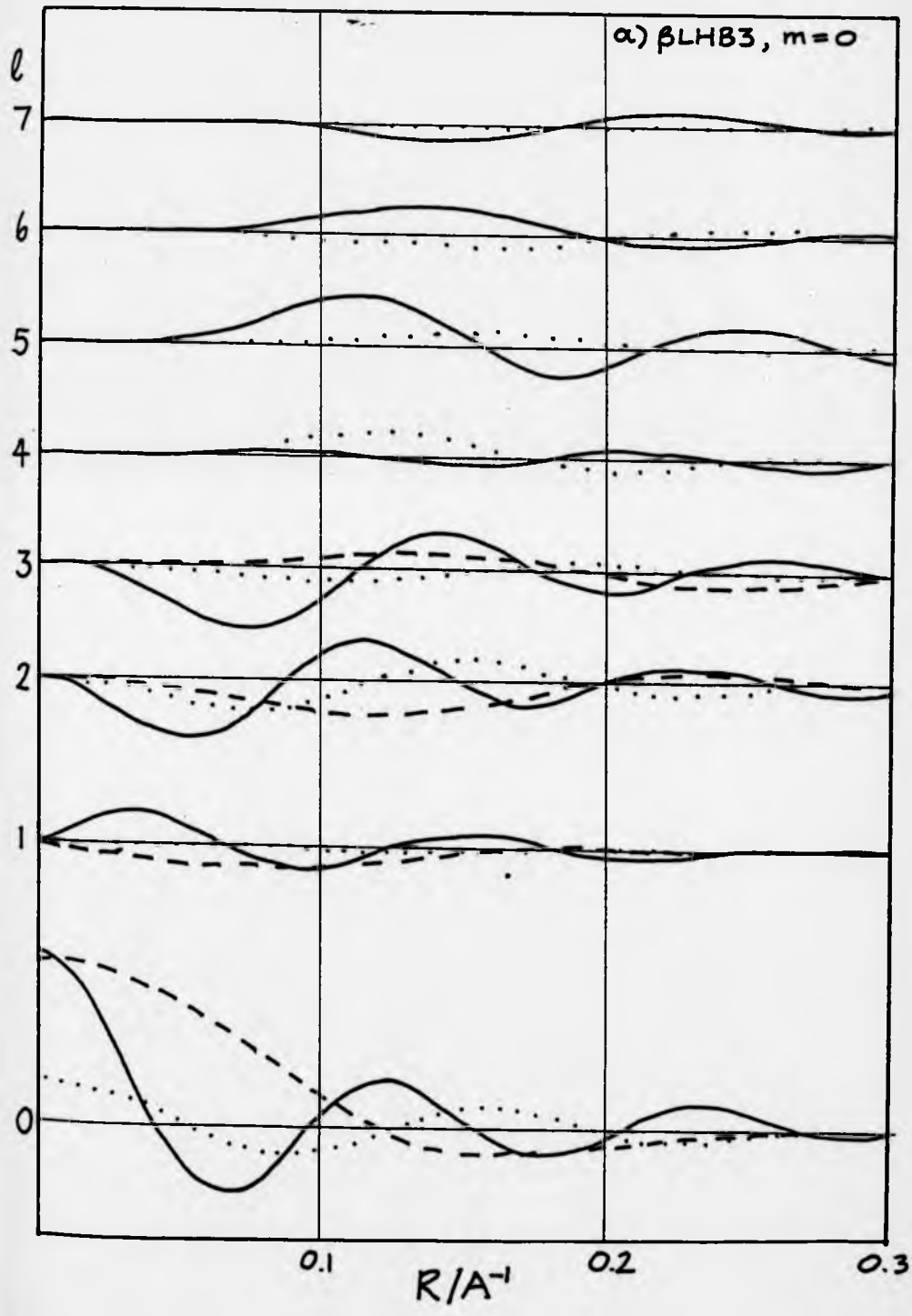
Figure 3.13 : Components of the Fourier transforms of β-stacked, left-handed B-DNA models.

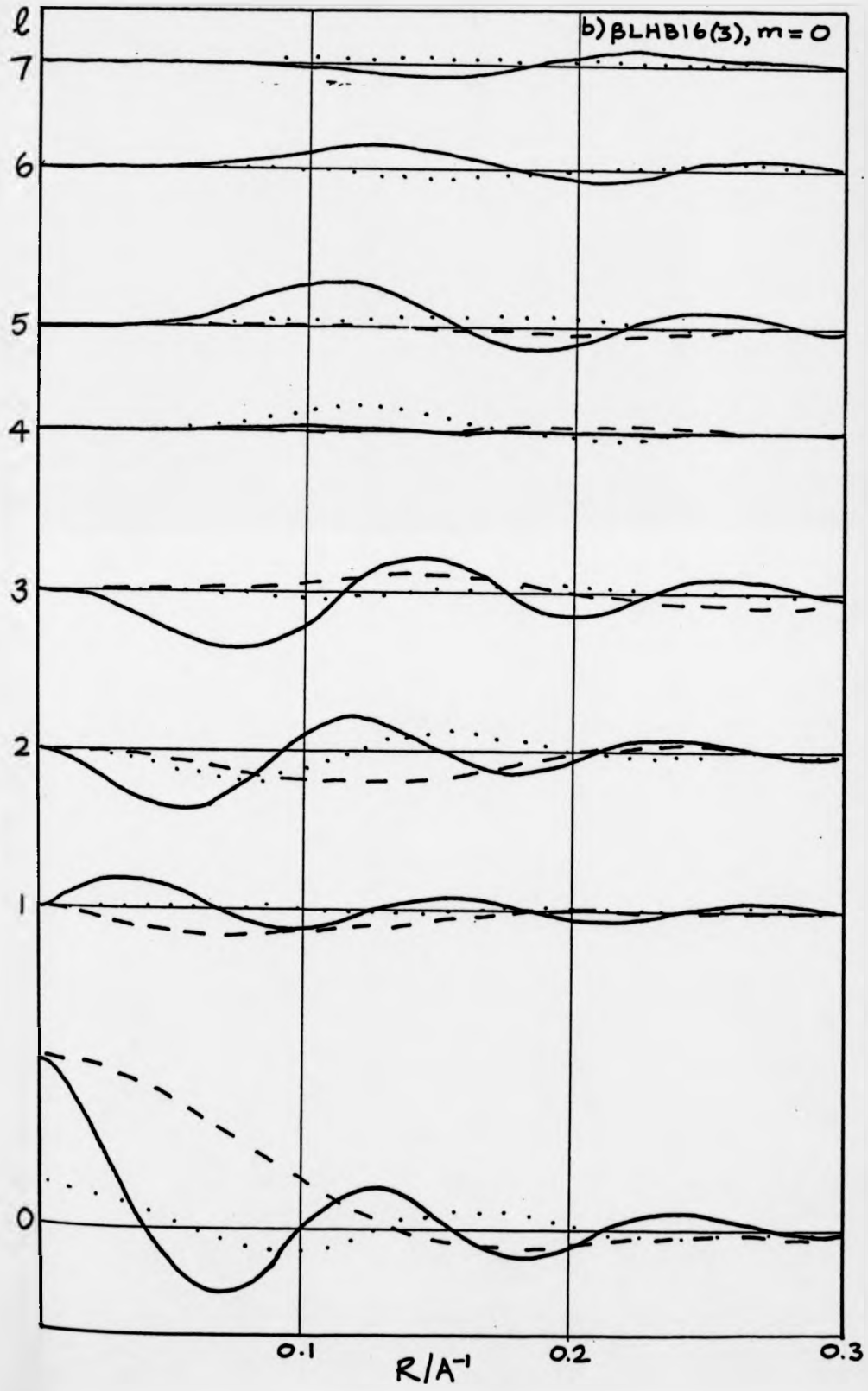
a) m = 0 components of βLHB3

b) m = 0 components of βLHB16 (3)

c) m = 1 components of βLHB16 (3)

In each case phosphate (——), base (---) and sugar (···) components are shown.


a) βLHB3, m=0
ℓ
7
6
5
4
3
2
1
0
0.1
0.2
0.3
R/A⁻¹


b) βLHB16(3), m = 0
ℓ
7
6
5
4
3
2
1
0
0.1
0.2
0.3
R/A⁻¹

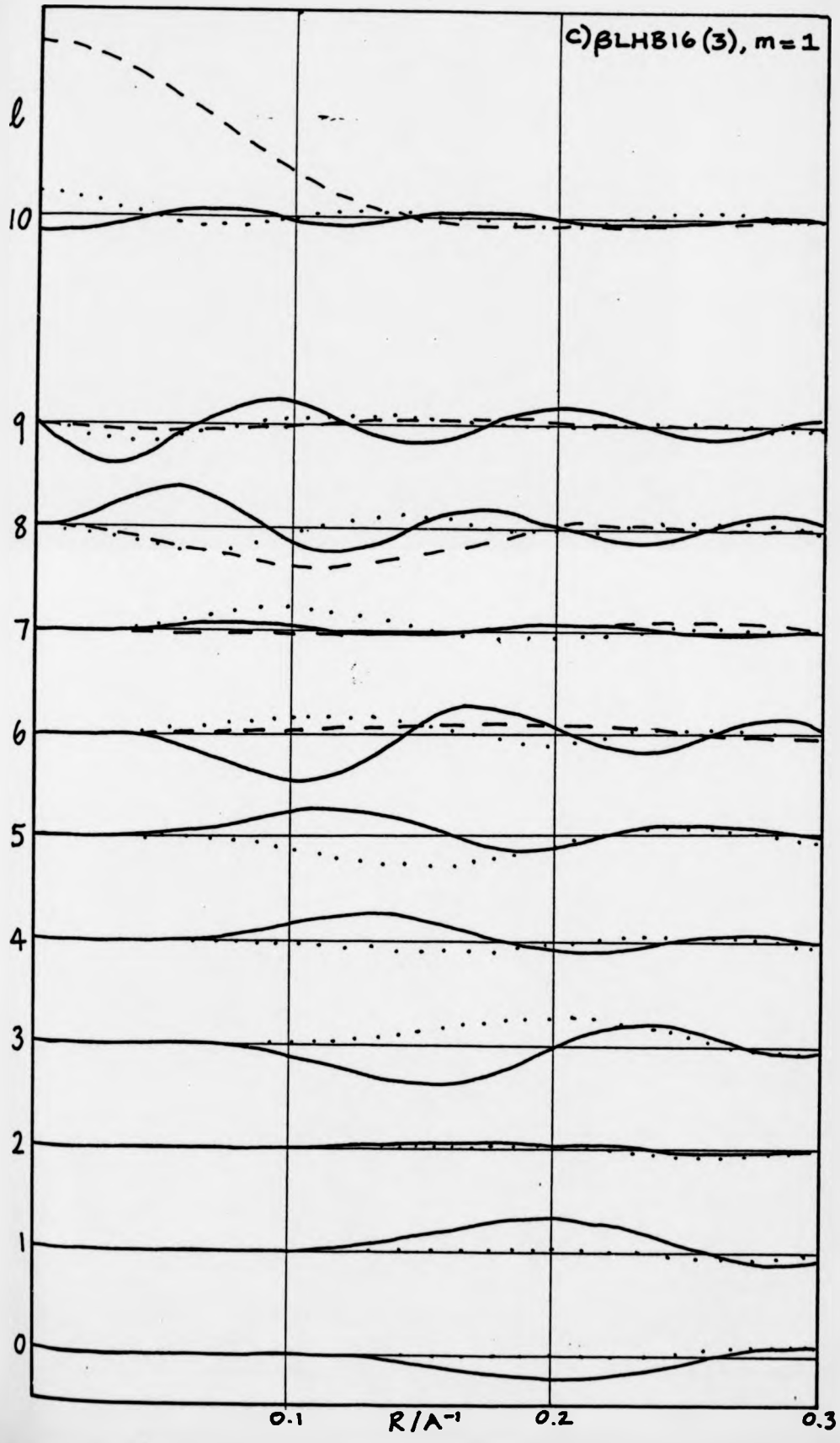
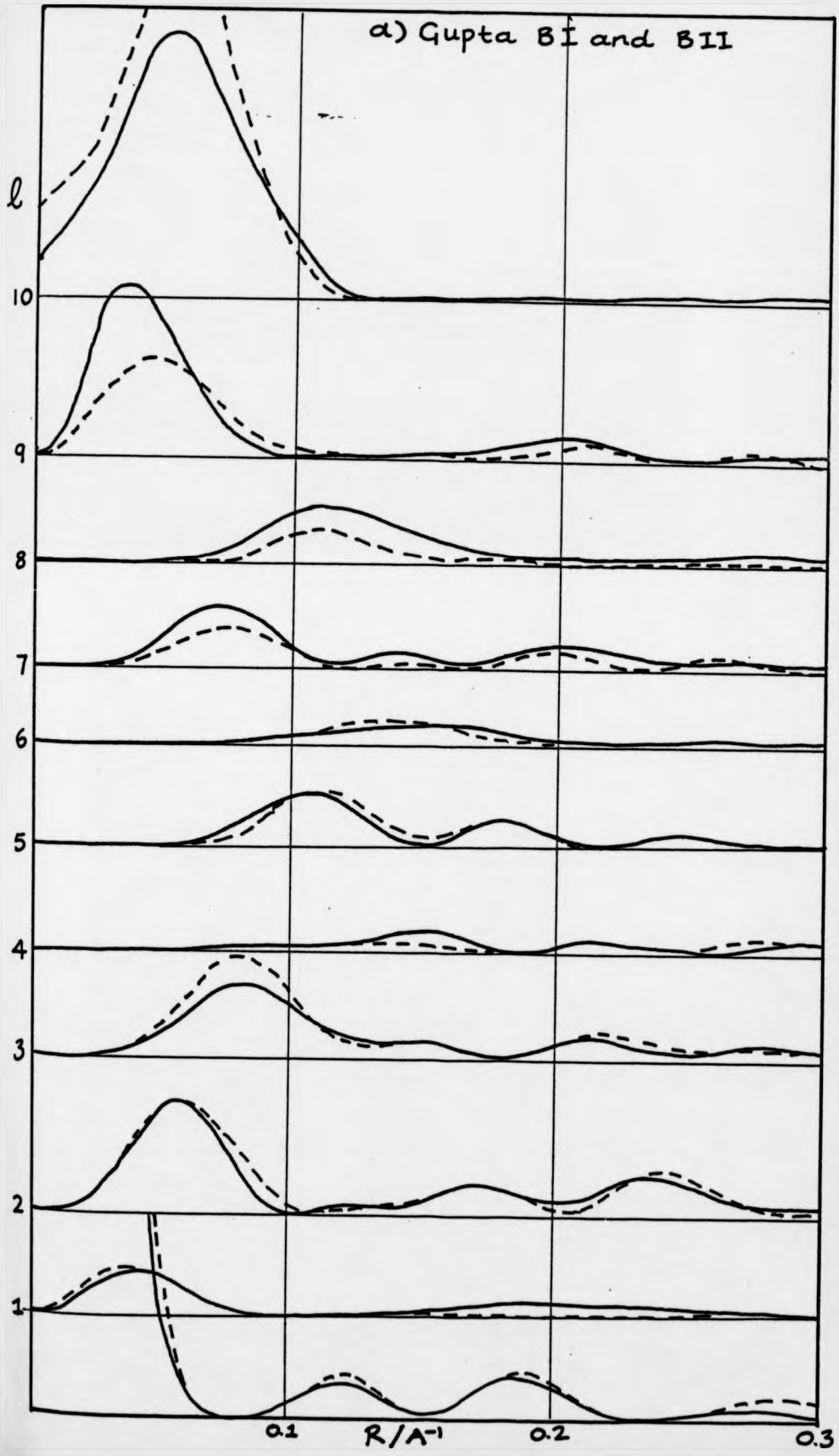

c) βLHB16 (3), m=1
ℓ
10
9
8
7
6
5
4
3
2
1
0
0.1
R/A⁻¹
0.2
0.3

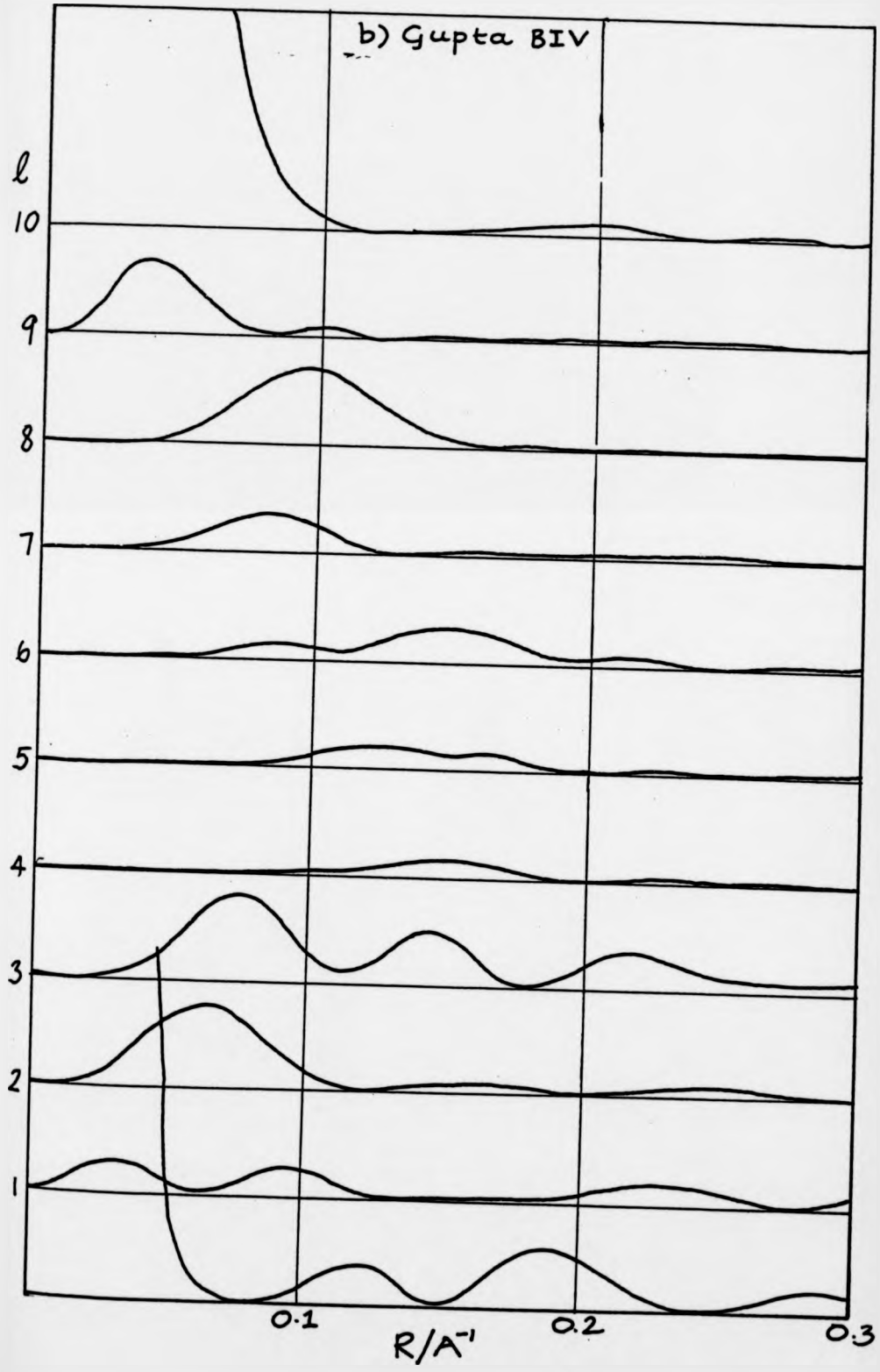
Figure 3.14 : Cylindrically averaged squared Fourier transforms of the B-DNA models of Gupta et al (1980b)

a) Models BI(——) and BII(---)

b) Model BIV

The transforms were calculated using Langridge's scattering factors and with the temperature factor set at $4A^2$.


a) Gupta BI and BII
l
10
9
8
7
6
5
4
3
2
1
0.1
R/A⁻¹
0.2
0.3


b) Gupta BIV
l
10
9
8
7
6
5
4
3
2
1
0.1
0.2
0.3
R/A⁻¹

published transforms of their model calculated using the scattering factors of Arnott and Hukins (1973). We have repeated this calculation with the preferable Langridge scattering factors discussed earlier (figure 14b). Their model gives rise to the same discrepancies as βLHB16 on ℓ=1 and ℓ=9. It is in slightly superior agreement on ℓ=2 and 3, but it is worse on ℓ=5 and 7.

Whilst refining the structure of the model of Arnott and Hukins as a preliminary exercise to the development of the left-handed models, difficulty was experienced in attaining convergence. The structure was modelled in the conventional 5' → 3' direction described earlier. The plot of ϕ (equation 2.35) versus cycle number (figure 15) shows that the refinement was converging for the first seven cycles but it then became unstable. The refinable parameters all underwent a large shift in cycle 7 (figure 16). No serious short contacts are present in the published model so the behaviour during cycle 7 probably arises from some ill-conditioning in the normal equations. An attempt was made to distort the initial model slightly in the hope that this would enable the algorithm to avoid the observed singularity but all models with C3'-*exo* puckering behaved in a similar manner whereas C2'-*endo* models refined with no difficulty (figures 15 and 16). The transform of the best model (RHB1), which is shown in figure 17, is only slightly different from that of the accepted model. The transforms of the models BI and BII of Gupta, Bansal and Sasisekharan calculated with Langridge scattering factors are shown for comparison (figure 14a).

3.3.4 Determination of Molecular Packing from the X-ray Data

As pointed out earlier, the cylindrically averaged intensity transform of a model is only an approximate representation of its diffraction pattern. In this section we compute structure factors of the models.

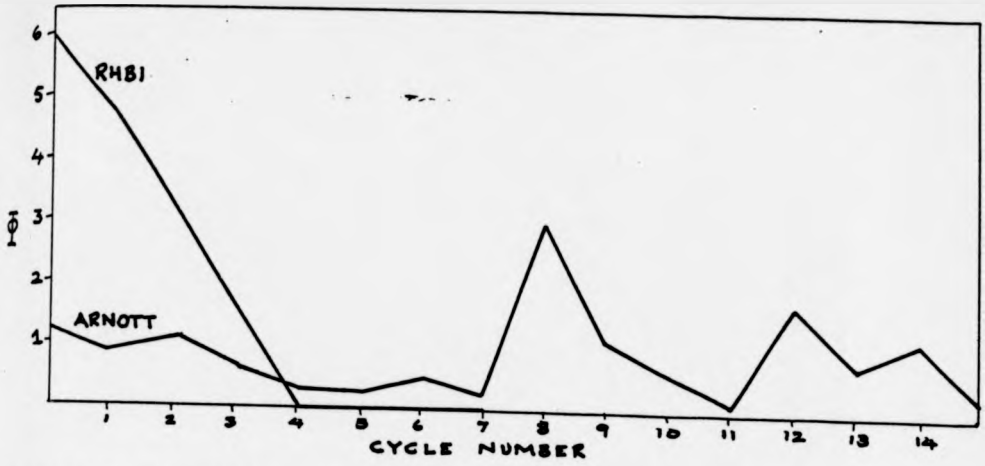


Figure 3.15 : Comparison of the values of the modelbuilding "figure of merit" Φ of the Arnott and RHBI models as a function of refinement cycle number

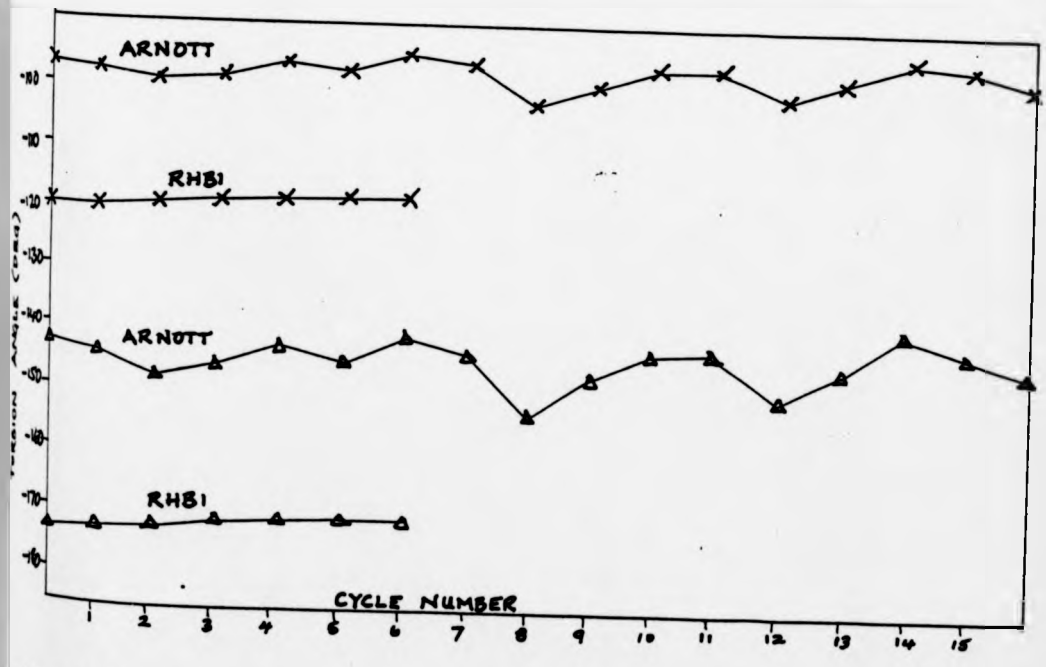


Figure 3.16 : Comparison of the values of the torsion angles β (x) and δ (Δ) in the Arnott and RHBI models as a function of refinement cycle number. The behaviour of the other angles is similar

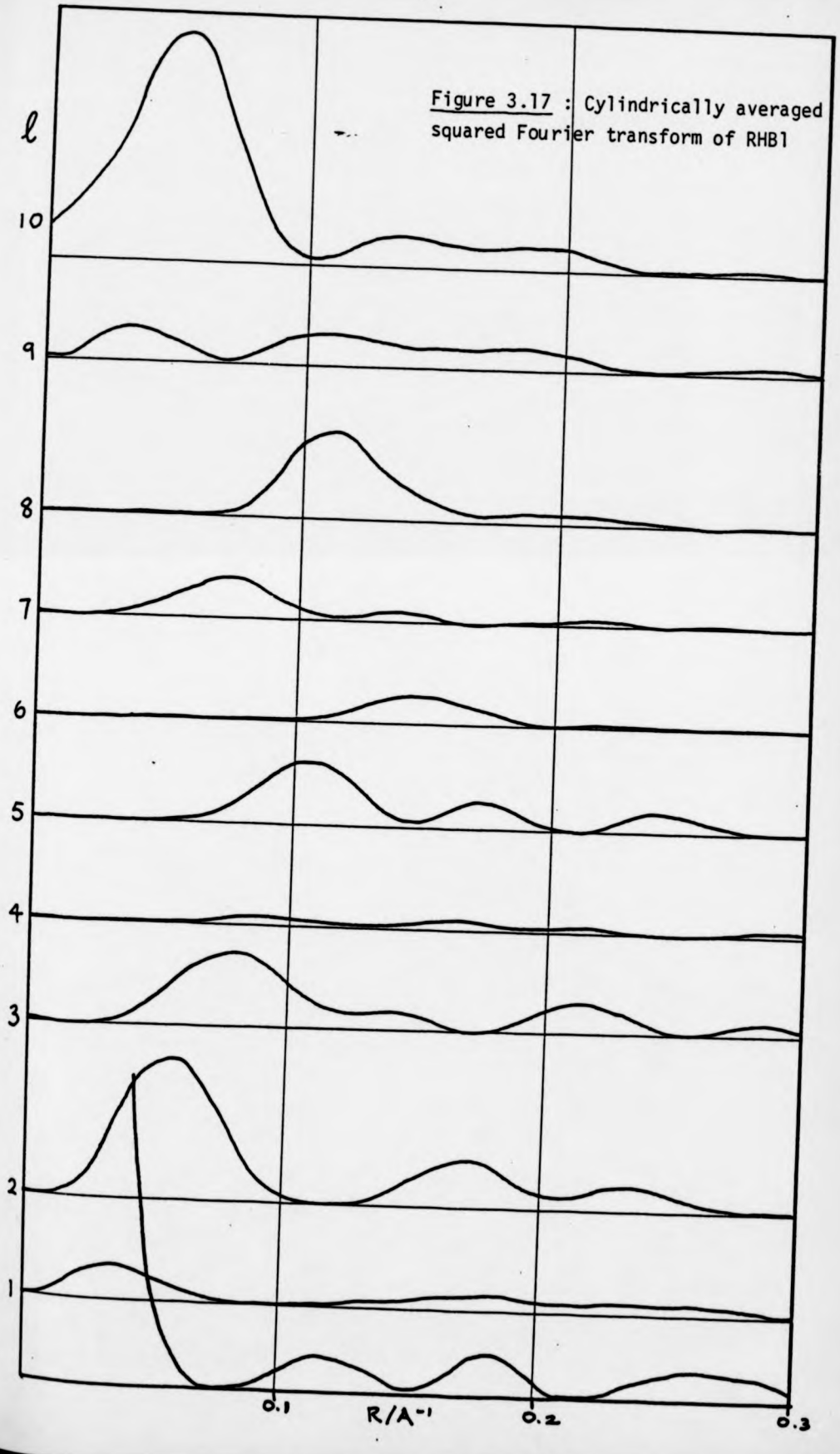


Figure 3.17 : Cylindrically averaged squared Fourier transform of RHB1
l
10
9
8
7
6
5
4
3
2
1
0.1
R/A⁻¹
0.2
0.3

This is an appropriate point at which to consider the effect of the various scattering factors described earlier on the structure factors and residuals. Table 10 shows residuals of the Arnott and Langridge models as a function of the data set and scattering factors employed and the molecular azimuthal orientation. It is unequivocally clear that the molecular orientation cannot be $\phi=0^{o}$ (which corresponds to a molecular diad oriented along the a axis) but more important is the fact that R varies from 32.5% to 40.2% when $\phi=90^{o}$. The difference between the residuals of two different models calculated with the same data and scattering factors is generally not much greater than the variation between the residuals for any one model calculated with different scattering factors or data. Therefore it is essential to adopt a standard procedure if rival models are to be evaluated so all residuals presented here have been calculated with Langridge scattering factors and the observed intensities of Arnott and Hukins (1973). A standard temperature factor, $B = 4\ A^{2}$, has also been imposed. This value was used by Langridge et al (1960b). In fibre diffraction the temperature factor accounts for the attenuation of the observed intensities as a function of scattering angle which arises from both thermal and static disorder within the specimen. The choice of a specific value is necessarily somewhat arbitrary but the inclusion of B as a refinable parameter has led to negative values (Arnott and Hukins, 1973; Gupta, Bansal and Sasisekharan, 1980) which are physically meaningless. The effect of such values, which lead to an increase in attenuation as the scattering angle decreases, is apparent in the cylindrical transforms published by Gupta et al.

It is worth examining the value of R as the relative molecular displacements along z and the azimuthal orientation, ϕ, are varied. Langridge et al (1960a) found that a relative displacement $z_f = 0.327$ (which corresponds to a shift $\Delta z = 11.1$ A of the central molecule relative to

Table 3.10 : Variation of Crystallographic Residual (R) as a Function of Molecular Model, Data Set, Scattering Factors and Azimuthal Orientation

Model	Data	f'	R(%)	
			$\phi=0^o$	$\phi=90^o$
Langridge[1]	Langridge[2]	Langridge[1]	47.6	32.7
Langridge	Langridge	Fuller[3]	48.1	32.5
Langridge	Arnott[4]	Langridge	46.9	34.1
Langridge	Arnott	Fuller	48.3	35.3
Arnott[5]	Langridge	Langridge	49.2	36.2
Arnott	Langridge	Fuller	49.6	36.7
Arnott	Arnott	Langridge	48.5	38.1
Arnott	Arnott	Fuller	49.5	40.2
Arnott	Arnott	Arnott[4]	-	35.1

In all calculations $B = 4A^2$ and $z_f = 0.327$

1) Langridge et al (1960b)
2) Langridge et al (1960a)
3) Fuller (1961)
4) Arnott and Hukins (1973)
5) Arnott and Hukins (1972b)

those at the corners) and $\phi=90^{o}$ accounted adequately for the diffraction pattern and gave satisfactory packing. Arnott and Hukins (1973) agreed that $\phi=90^{o}$ but they preferred Δz = 10.8A. With this arrangement of molecules the space group is $P2_12_12_1$ which is consistent with the observed systematic absences although Donohue (1969, 1971) has suggested that the crystal symmetry may be no higher than triclinic. Gupta et al have claimed that for their three models with mononucleotide repeats the best fit with the X-ray data is obtained when z_f = 0.328 and ϕ =92^{o} giving R = 36% for BI; when z_f = 0.328, ϕ_1 = 90^{o} and ϕ_2 = 87^{o} giving R = 35% for B II; and when z_f = 0.32 and ϕ = 92^{o} giving R = 37% for B IV. In B I and B IV the constraint is maintained that both molecules have the same orientation. Although the space group is no longer $P2_12_12_1$ in either case, Dover (1977) has pointed out that the tenfold screw symmetry of the helix combined with the particular lattice parameters observed for B-DNA guarantees the maximum number of favourable intermolecular contacts if the orientations are identical. We have maintained this constraint in our calculation of $R(\phi,z)$.

Table 11 shows $R(\phi,z_f)$ of the Arnott and Hukins model. The lowest R (35.6%) was obtained when ϕ = 91^{o} and z_f = 0.313 which differs insignificantly from the value of 35.7% obtained when ϕ = 90^{o} hence there is no evidence from this that the space group is not $P2_12_12_1$. However, the value obtained when Arnott and Hukins packing parameters were used was 38.1% in contrast to the 31% quoted by them. The residual of the Langridge model was found to be 34.1% so Arnott's model certainly does not represent an improved fit with the diffraction data (but its stereochemistry is superior). The residual of RHB1 attains a minimum of 36.7% at $\phi=92^{o}$ and z_f = 0.31 so it is a competitor to the accepted model if X-ray constraints alone are imposed (table 12). If the orientations of the two molecules are constrained to be identical then the best residual for BI of Gupta et al is 40.1% at ϕ = 90^{o} and z_f = 0.31 and the

Table 3.11 : $R(\phi, z_f)$ of the Arnott and Hukins Model of B-DNA

z_f \ ϕ	85	86	87	88	89	90	91	92	93	94	95
0.30	42.4	40.8	39.4	38.3	37.5	37.0	36.9	37.0	37.4	38.2	39.3
0.31	40.6	39.1	37.8	36.8	36.1	35.7	35.6	35.7	36.1	36.9	38.0
0.32	41.0	39.3	38.1	37.3	36.6	36.0	35.9	35.9	36.3	37.1	38.1
0.33	44.3	42.5	41.4	40.5	39.8	39.2	38.9	39.0	39.2	40.2	41.2
0.34	49.2	47.4	46.2	45.3	44.5	43.9	43.6	43.6	43.9	44.8	45.8

Table 3.12 : Residual as a function of azimuthal orientation, ϕ, and fractional displacement, z_f, of RHB1 using the observed intensities of Arnott and Hukins (1973)

z_f \ ϕ	85	86	87	88	89	90	91	92	93	94	95
0.30	45.2	43.4	41.7	40.3	39.1	38.1	37.5	37.3	37.8	39.1	40.9
0.31	44.4	42.6	41.1	39.7	38.5	37.5	36.9	36.7	37.1	38.5	40.3
0.32	45.5	43.5	42.0	40.7	39.5	38.4	37.8	37.4	37.9	39.3	41.2
0.33	47.6	45.5	43.9	42.5	41.3	40.2	39.7	39.4	39.9	41.3	43.5
0.34	50.7	48.5	46.7	45.1	43.8	42.7	42.2	42.0	42.5	44.0	46.4

Table 3.13 : $R(\phi_1\ z_f)$ of βLHB16

z_f \ ϕ	85	86	87	88	89	90	91	92	93	94	95
0.30	49.0	48.5	48.2	48.4	48.8	49.0	49.0	49.0	48.3	48.3	49.1
0.31	47.2	46.6	46.3	46.5	46.9	47.1	47.0	47.1	46.4	46.3	47.0
0.32	46.9	46.4	46.1	46.1	46.5	46.7	46.6	46.6	46.0	45.7	46.4
0.33	50.0	49.5	49.3	49.1	49.4	49.5	49.5	49.5	48.9	48.6	49.2
0.34	53.6	53.1	52.7	52.3	52.3	52.6	52.6	52.7	52.3	52.1	52.7

Table 3.14a : $R(\phi, z_f)$ of Model I of Gupta et al

z_f \ ϕ	85	86	87	88	89	90	91	92	93	94	95
0.30	46.3	45.1	44.1	43.3	42.7	42.6	42.8	42.9	43.3	43.8	44.5
0.31	44.1	42.9	42.0	41.1	40.6	40.5	40.7	40.8	41.2	41.7	42.4
0.32	43.6	42.5	41.6	40.9	40.2	40.1	40.3	40.3	40.5	41.0	41.7
0.33	45.6	44.4	43.4	42.7	42.2	42.2	42.3	42.2	42.4	43.0	43.9
0.34	49.1	47.8	46.8	46.0	45.5	45.6	45.6	45.5	46.0	46.7	47.8

Table 3.14b : $R(\phi, z_f)$ of Model II of Gupta et al

z_f \ ϕ	85	86	87	88	89	90	91	92	93	94	95
0.30	45.7	44.3	42.9	42.1	41.8	41.3	41.4	41.8	42.5	43.6	45.0
0.31	43.2	41.8	40.5	39.9	39.6	39.2	39.3	39.7	40.3	41.2	42.7
0.32	42.8	41.4	40.2	39.5	39.0	38.7	38.8	39.1	39.6	40.5	42.1
0.33	45.3	43.8	42.5	41.7	41.3	41.1	41.3	41.6	42.1	43.1	44.8
0.34	49.1	47.5	46.2	45.3	44.8	44.7	44.8	45.3	46.0	47.0	48.6

Table 14c : $R(\phi, z_f)$ of Model IV of Gupta et al

z_f \ ϕ	85	86	87	88	89	90	91	92	93	94	95
0.30	47.5	46.9	46.2	45.8	46.2	46.7	47.0	47.1	47.1	47.3	47.4
0.31	46.4	45.8	45.0	44.8	45.0	45.4	45.6	45.7	45.6	45.8	46.1
0.32	45.9	45.4	44.7	44.5	44.7	45.2	45.2	45.2	45.0	45.2	45.5
0.33	48.8	48.2	47.5	47.4	47.8	48.1	48.1	48.1	47.8	48.2	48.5
0.34	52.5	51.4	50.7	50.5	50.8	51.0	50.8	50.8	50.6	51.2	51.7

and the best for BII is 38.7% at $\phi = 90^{o}$ and $z_f = 0.32$ (table 14a, b). If the packing parameters of Gupta et al are used these values rise to 41.6% and 41.5% respectively. Thus both models are significantly less successful than the Arnott model or RHB1 in accounting for the X-ray pattern.

The models considered so far have all been right-handed. The residuals of the left-handed βLHB16 and B IV are shown in tables 13 and 14c respectively. The behaviour of the residual of βLHB16 is different from those of all the other models in having two minima. One (R = 45.9%) occurs at $\phi = 87.5^{o}$ and $z_f = 0.313$ and the other (R = 45.7%) is at $\phi = 94^{o}$, $z_f = 0.314$. The best residual if ϕ is constrained to 90^{o} is 48.5% at $z_f = 0.314$. The minimum residual of B IV is 44.5% at $\phi = 88^{o}$ and $z_f = 0.32$ which increases to 45.2% at $z_f = 0.32$ if $\phi = 90^{o}$. The residual is also 45.2% when calculated with the packing parameters quoted by Gupta et al in contrast to the 37% they claim. Thus both the α-stacked model B IV of Gupta et al and the β-stacked LHB16 described here give significantly worse agreement than RHB1 or the Arnott model with the diffraction data.

3.3.5 The Stereochemistry of the Models

Gupta et al (1980b) have undertaken a survey of the backbone torsion angles found in single crystals of dinucleoside monophosphates and fibres of polymeric nucleic acids in which they found a correlation between the sugar pucker and the torsion angles about the P-O1 and P-O4 bonds (β and γ). In particular, in those molecules adopting the C2'-endo pucker ($130^{o} \lesssim \zeta \lesssim 160^{o}$) these angles were always in the tg^{-} domain in contrast to those molecules adopting the C3'-endo pucker ($70^{o} \lesssim \zeta \lesssim 100^{o}$) in which the same angles fell in the $g^{-}g^{-}$ domain. They refer to these combinations of ζ, β and γ as the "preferred correlation". In addition they found restricted sets of values for the remaining backbone angles: $200^{o} \lesssim \delta \lesssim 200^{o}$ and $40^{o} \lesssim \varepsilon \lesssim 70^{o}$. Sasisekharan, Gupta and Bansal (1981) and Gupta et al (1980b) have imposed these values in constructing a variety of models for B-DNA. We shall

Table 3.15 : Co-ordinates of the Asymmetric Unit of RHB1

Successive nucleotides may be generated by adding 36^{o} and 3.4 A to ϕ and z respectively. The base atom co-ordinates are the same as those of Arnott and Hukins (1972b)

Phosphate	r(A)	ϕ(deg)	z(A)
O1	8.70	95.5	3.11
O2	10.70	92.2	1.70
O3	8.97	101.6	0.80
P	9.24	94.0	1.62
O4	8.47	85.3	1.21
Sugar			
C1	5.90	66.5	0.47
C2	6.88	71.4	-0.56
C3	8.20	68.3	0.08
C4	7.96	69.0	1.58
C5	8.53	77.9	2.14
O5	6.52	68.0	1.75

Table 3.16 : Co-ordinates of the Asymmetric Unit of βLHB16

Successive nucleotides may be generated by adding -36^{o} and 3.34 A to ϕ and z respectively.

Phosphate	r(A)	ϕ(deg)	z(A)
O1	7.80	95.9	-3.53
O2	9.85	100.9	-2.32
O3	9.57	86.2	-2.52
P	8.85	94.2	-2.35
O4	7.89	94.3	-1.07
Sugar			
C1	5.48	77.6	0.38
C2	6.14	75.1	-0.97
C3	7.61	74.1	-0.59
C4	7.78	81.4	0.58
C5	8.35	91.0	0.21
O5	6.50	82.1	1.25
Purine			
N1	0.75	134.2	0.04
C2	1.27	56.5	0.07
N3	2.55	65.1	0.16
C4	3.12	90.1	0.22
C5	3.04	115.7	0.19
C6	2.21	138.5	0.10
N7	4.40	119.3	0.27
C8	5.01	105.6	0.34
N9	4.49	90.1	0.31

Cont.

Table 3.16 (Cont.)

Purine (Cont.)	r(A)	ϕ(deg)	z(A)
GUN2	1.82	4.51	0.01
GUO6	2.83	164.2	0.05
ADN6	2.94	163.3	0.06
Pyrimidine			
N1	2.46	111.3	0.16
C2	3.12	86.5	0.22
N3	4.49	90.1	0.31
C4	5.16	104.3	0.35
C5	4.82	119.4	0.29
C6	2.32	128.0	0.13
O2	3.00	63.3	0.19
CYN6	2.06	169.1	0.03
THO6	3.76	148.0	0.14
THME	6.08	128.4	0.33

concentrate here only on those models in which the asymmetric unit is a mononucleotide. Sasisekharan and co-workers criticised the C3'-exo model of Arnott and Hukins (1972b) since the β and γ angles fell within the g^-g^- range and therefore the nucleotide does not adopt the preferred correlation. They also pointed out that the angle α has a higher value (155^o) in the Arnott and Hukins model than is observed in any single crystals which leads to short contacts between O2 and C2'. When the sugar pucker is C2'-endo and β and γ are in the tg^- range such contacts are not present. Since their model B II adopts this preferred correlation, Gupta et al claim that it is sterochemically superior to the Arnott and Hukins model. However, Arnott and Hukins also presented details of a C2'-endo conforming to this criterion which they felt was only marginally inferior to the one with the C3'-exo pucker. Arnott et al (1980) have refined the C2'-endo model once again and they claim that it is superior to all previous attempts.

In table 17 we compare the backbone torsion angles of these models with those of RHB1 and BLHB16 and also the angles within those residues of the B-DNA dodecamer which contain C2'-endo sugars (Dickerson and Drew, 1981). It is apparent that a wide range of torsion angles are adopted by corresponding angles within dodecamer residues. In addition, it is noteworthy that only one residue (C11) conforms to the preferred correlation principle whereas in all other residues the angles β and γ are in the g^-g^- domain. The RHB1 model, whilst distinctly different from the C2'-endo models of Gupta et al and Arnott and co-workers is nonetheless unexceptional in falling within the ttg^-tg^+ domain. The values of α and δ which it adopts are closer than those in any of these models to the average values found within the dodecamer. Similarly the value of β is in better agreement but the conformation about this bond is eclipsed. The value of γ in the Gupta model is closest to the dodecamer values and there is little discrimination between the rather low values of ε in all

Table 3.17 : Backbone Torsion Angles of B-DNA Nucleotides Containing C2'-endo and C3'-exo Sugars

	α	β	γ	δ	ε	ζ	Domain	Reference
Dodecamer residues								
G4	-177	-88	-63	180	57	156	tg^-g^-tg+	
G10	-157	-94	-67	169	47	143	$tg^-g^-tg^+$	
C11	-103	150	-74	139	56	136	$t\ t\ g^-\ tg^+$	
G16	-185	-86	-69	171	73	136	$t\ g^-g^-\ tg^+$	1
A17	-186	-98	-57	190	54	147	$t\ g^-g^-tg^+$	
A18	-183	-97	-57	186	48	130	$t\ g^-g^-tg^+$	
G22	-177	-86	-67	179	50	150	$t\ g^-g^-tg^+$	
Arnott B1	155	-96	-46	-147	36	157	$tg^-g^-tg^+$	2
Arnott B2	-166	-136	-25	160	27	146	$t\ tg^-tg^+$	2
Arnott B3	-133	-157	-41	136	38	139	$t\ tg^-tg^+$	3
Levitt	178	-85	-65	170	65	108	$tg^-g^-tg^+$	4
RHB1	-172	-120	-43	173	39	143	$t\ tg^-tg^+$	5

Cont.

Table 3.17 Cont.

	α	β	γ	δ	ε	ζ	Domain	Reference
Gupta B I	-176	-68	-91	179	75	97	$tg^-g^-tg^+$	6
Gupta B II	-132	-158	-58	144	41	149	$t\ tg^-tg^+$	6
Gupta B IV	-119	-156	-90	135	36	137	$g^-tg^-tg^+$	6
βLHB16	39	86	166	-120	40	157	$g^-g^+tg^-g^+$	5

References

1) Dickerson and Drew (1981)

2) Arnott and Hukins (1972b)

3) Arnott et al (1980)

4) Levitt (1978)

5) This work

6) Gupta et al (1980b)

the fibre models. Since RHB1 is nearly within the $tg^-g^-tg^+$ domain and is gnerally close to the dodecamer structure it is likely that only a small modification of it may produce a fibre model which is superior to all previous attempts. In particular one could investigate the effect of incorporating base-pairs with the propellor twists observed in the single crystal structure.

The left-handed model B IV of Gupta et al falls within the $g^-tg^-tg^+$ domain so the change in handedness arises mainly from the α and β angles. However, when the bases are inverted as in βLHB16 a quite distinct conformation results which is even unlike those which occur in the Z-DNA helices (cf. table 1,2). Only the value of ϵ is close to those observed within right-handed B-DNA in contrast to the α-stacked model B IV where the change in handed ness is accommodated without a major modification of the backbone.

We now examine the intermolecular contacts between the RHB1 and βLHB16 models packed in the B-DNA unit cell in order to determine whether a correlation exists between the optimum packing and the best packing parameters suggested by the crystallographic residuals. The RHB1 model explains the diffraction pattern best when $\phi = 92^o$ and z = 10.14 A. When arranged in this manner two short contacts occur between O2 atoms on adjacent helices; the distances are 2.39 A and 2.60 A compared with the optimum value of 2.80 A. A similar problem has arisen with previous models for B-DNA and a slight distortion of the molecular structure was proposed to alleviate it (Langridge et al, 1960b; Arnott, Dover and Wonacott, 1969). Although the X-ray data was insufficient to preclude the possibility that $\phi = 90^o$ when z = 10.14 A, the stereochemistry deteriorates markedly since the distances between the two O2 pairs falls to 2.11 A and 2.48 A.

The X-ray data suggested that z = 10.2 A and $\phi = 87.5^o$ or 94.0^o were the best packing parameters for βLHB16. In both cases one short contact of 2.30 A between O2 atoms occurs so neither the X-ray data nor the stereochemistry may be used to distinguish between these two arrangements. The most satisfactory stereo-

chemistry to be obtained without distorting the molecular structure was when z = 11.2 A, $\phi = 94^{o}$ (when the O2 separation increased to 2.56 A) and z = 11.6 A, $\phi = 88^{o}$ (when the separation increased to 2.78 A). However, the respective residual values (48.6% and 52.3%) indicated a deterioration in the agreement between the observed and predicted diffraction patterns.

3.4 Discussion and Conclusions

Although the results of the work described in this chapter suggest that β-stacked, left-handed models do not appear to explain the structure of B-DNA in fibres, such structures may nonetheless have a biological role. Nordheim et al (1981) have proposed that Z-DNA may act as a regulator by producing a dramatic change in the local environment of a particular gene or in a more long-range fashion by the propagation of the effect of a B → Z transition via supercoiling. Such modifications might for example affect the transcription rate of a gene. Clearly any modification such as β-stacked structures which produce significant changes has the potential for exploitation in recognition or control processes.

Until comparatively recently it was generally regarded as inconceivable that a transition affecting the helical sense of a polynucleotide could occur in a fibre without considerable stereochemical difficulty. However experiments largely stimulated by the discovery of Z-DNA have demonstrated that such transitions are possible. For example Arnott et al (1980) and Leslie et al (1980) have observed both the B and Z conformations in the same fibre of poly d(G-C).poly d(G-C) and Behe et al (1981) report the existence of both in poly d(G-m^5C).poly d(G-m^5C). In addition, Klysik et al (1981) have found evidence for the occurrence of B → Z transitions in stretches of d(G-C) inserted into plasmids. Further support for the transition in solution has come from deuterium exchange (Ramstein and Leng; 1980), nmr (Patel et al, 1979) and optical experiments

(Pohl and Jovin, 1972). It is of interest to determine the mechanism of such transitions and as a preliminary exercise we built CPK models of the eight distinct structures containing α- or β-stacked bases, left- or right-handed symmetry and syn or anti glycosidic link orientation in order to determine which were feasible for consideration in future, more extensive modelbuilding studies. No attempt was made to impose any specific number of residues per turn or to ensure precise helical symmetry : instead our aim was to determine which structures could be excluded on general grounds. We refer to an α-stacked, right-handed model with anti sugar-base orientation as αRa and so on. We found αRs and βRa models exceedingly difficult to build due to steric clashes and we consider that no polynucleotide conformation will be found which contains these characteristics. Although αLa and αLs models were difficult to build these characteristics could probably be incorporated into polymers. All other types of model αRa, βRs, βLa and βLs could be built with ease with a variety of base tilts and displacements. The acceptability of αRa models is of course well known since all members of the A and B families of polynucleotides have these characteristics, and the βLHB16 model falls in the βLa class but no βRs or βLs models have been proposed so there is scope for investigating the possibility that models with these novel characteristics might be capable of explaining the currently known diffraction patterns of polynucleotides. We did not examine models with dinucleotide repeats but Z-DNA for example is a combination of βLa and βLs which is stereochemically acceptable. This conformation is apparently only accessible to molecules with alternating purine and pyrmidine sequences whereas the structures we examined were not restricted in this way. Most transitions between such structures involve only a re-arrangement of the sugar-phosphate chain, for example to accomodate a change in handedness, or a rotation about the sugar-base link. However, transitions between α and β structures require a break in the hydrogen bonds joining the base-pairs followed by a flipping

over of the bases and rejoining of the bonds since these classes of conformation are topologically distinct. Ivanov and Minyat (1981) have found that transitions between B and A forms of DNA show fast kinetics (~ 10s) whereas those between B and Z forms are slow (between ten and sixty minutes) as might be expected from the major conformational change which is required. Further work is in progress in this laboratory to investigate the stereochemistry of the transitions.

The work described here has highlighted some of the technical problems of fibre diffraction analysis. Although these are less acute for the crystalline patterns which may be obtained from DNA fibres than for most other fibrous systems which give more diffuse diffraction patterns, they are nontheless significant. For example the different methods of accounting for diffraction from water have been shown to be capable of changing the value of the residual of a model by about five percentage points. Nor is this the most important effect since we have shown (section 3.4) that varying the method affects the best value which can be obtained for the packing parameters. This raises questions about the correctness of refined models for DNA. For example, Arnott et al (1969) claim that their computer refinement of the B-DNA model of Langridge et al (1960b) resulted in a superior model. However, we have shown that if the residuals of the Arnott and the Langridge models are calculated using the same data and the same method, Langridge's model is better (table 10). Thus the superiority claimed by Arnott et al for their model appears to arise mainly from the use of incorrect scattering factors and negative temperature factors. (It should be admitted however that the residual of the Arnott model might be expected to be a little higher since the stereochemical constraints were much more severe - thus we are not suggesting that the Arnott model is wrong, merely that the support it receives from the X-ray data is less impressive than Arnott et al claim). The similarly impressive residuals using Arnott's scattering factors quoted by Gupta et

al (1980) for their B-DNA models also deteriorate when Langridge's scattering factors and a positive temperature factor are imposed. These results underline the need for a uniform procedure to be adopted when models are to be compared.

When we adopted such a procedure we found that none of the models of B-DNA published by Gupta et al (1980) is a serious rival to the established model. However the RHB1 model presented here is attractive since it contains torsion angles which are similar to those observed in the B-DNA dodecamer and its residual is only slightly higher than that of the Arnott and Hukins model. Left-handed models are less successful. Neither the model of Gupta et al nor βLHB16 accounts well for the B-DNA diffraction pattern but β-stacked, left-handed models have been found to be stereochemically plausible with few problems concerning intermolecular or intramolecular contacts. Indeed a β-stacked, left-handed model for D-DNA has been devised in this laboratory which is superior to the standard model of Arnott et al (1974) in both stereochemistry and agreement with the diffraction pattern (A. Mahendrasingam, unpublished results). In addition to the comparison of observed and predicted intensities a further crucial test of the βLHB model may come from intercalation of drugs and dyes into the structure. Such chromophores are known to intercalate between the base-pairs of DNA giving rise to a change in the helix pitch and concomitant re-arrangement of the sugar-phosphate chain (see Pigram (1968) for example). Preliminary modelbuilding suggested that this would not be easily explained by βLHB since the sugar-phosphate chain is significantly more taut and there is little scope for the base-pairs to move apart by 3.4 A.

Finally, it should be added that recent studies have underlined some of the more fundamental problems of fibre diffraction analysis. In particular the structure of the B-DNA dodecamer has shown that the average nucleotide principle, generally assumed to hold in models based on fibre

data, is not correct. Future studies should therefore attempt to adopt a somewhat more sophisticated approach by incorporating for example the base roll, propellor twists and principle of anticorrelation of sugar puckers which have been found within the dodecamer (Dickerson and Drew, 1981). Fibre diffraction seems likely to remain the most important technique for investigating polynucleotide polymorphism but single crystal analysis may make more incisive contributions in the determination of the fine details of conformation. However it remains to be seen whether the full range of structures adopted in fibres will also occur in crystals.

Appendix to Chapter III

Derivation of the Co-ordinates of S-DNA and a General Method for the Calculation of the Co-ordinates of a Helical Polymer from the Chain Torsion Angles

It has become fashionable to present preliminary models for polynucleotides in terms of the torsion angles rather than the atomic co-ordinates. These angles are often of interest in themselves and in some cases, for example helical polypeptides, they may be easily used to derive the co-ordinates. However the polypeptides contain only two backbone torsion angles per monomer. The polynucleotides, by contrast, contain six. In addition these angles define the positions of only a relatively small proportion of the atoms. If a standard sugar pucker is not used then derivation of the co-ordinates of the bases may prove difficult. One example of this problem is the preliminary report of the structure of S-DNA (Arnott et al, 1980). Presenting a model in this way gives rise to difficulties for any other workers who may wish, for example, to calculate the diffraction pattern of the molecule or to examine its stereochemistry or to design similar models.

Two distinct attempts were made to calculate the co-ordinates of S-DNA but both failed. The first method illustrates the problems which are encountered when information such as the base disposition is unavailable. The second method was unsuccesful for a different reason; however, it will be described here since it may prove useful in future work despite its failure when applied to the problem for which it was devised.

Method 1

Most of the computer models described in this thesis have been built according to the method shown in figure 3.11a . However an alternative

method is available (figure A3.1). The model consists of two chains. The first starts at the C3' atom of the upper sugar (atom 1) and runs down the backbone and across the sugar to the base in the same way as in the models described earlier (figure 3.11a) But the main chain runs from the C5' of the upper sugar (atom 11), across the sugar and base to atom 17 on the helix axis. The chain then runs to atom 18 at the origin of co-ordinates at which point the first chain atoms are added. Two dummy atoms, 1' and 2', are added as pendants to the upper sugar in the C3' and O1 positions. In order to preserve chain closure the distances between atoms 1 and 1' and also 2 and 2' are constrained to be zero. Whilst this method is rather more cumbersome than the one commonly used, it has the advantage that the rise per reside ($\ell_{17,18}$) and the rotation per residue (τ_{16}) may be changed with facility and also given values which are precisely maintained.

The S-DNA model was built using this method from the data provided by Arnott et al (1980). Unfortunately, the co-ordinates of the chain-closure atoms showed large discrepancies (about 2A). Wire modelbuilding studies suggested that there was no large error in the published torsion angles since structures similar to those described by Arnott et al (1980) could be built. The discrepancies appear to arise from the relatively large amount of information which must be included in the building procedure but which has not been published:-

(1) the base tilt and twist;

(2) the base displacement;

(3) the base shear parallel to the twist axis;

(4) the positions of the sugar atoms.

The first of these should not have a very large effect. The base tilt is given as "approximately 7^{o}" and the twist is probably not more than a few degrees. However, the base displacement had to be estimated from figure 2 of Arnott et al (1980). In addition, since the repeating unit is a dinucleotide, there is no requirement that the bases be in helically

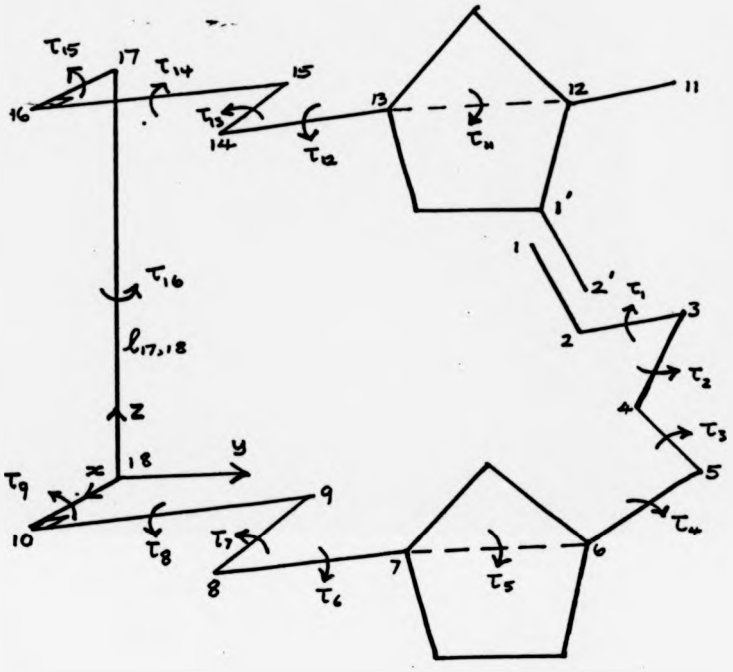


Figure A3.1 : Building a dinucleotide using the two branch method

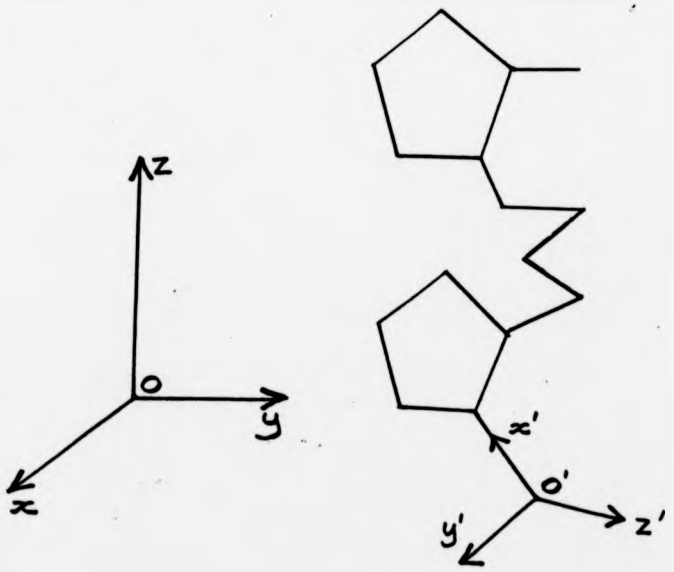


Figure A3.2 : Building a dinucleotide in an arbitrary reference frame O'x'y'z'. The helix axis is along Oz and the diad axis is along Ox.

equivalent positions. Indeed successive bases are sheared with respect to each other parallel to the twist axis. The degree of shear also had to be estimated from the diagram. Both of these quantities need to be accurately determined since the positions of the bases affect the co-ordinates of all the other atoms in the backbone. The wire modelbuilding studies indicated that the base shear is very important in allowing chain closure. If the shear was zero it was not found to be possible to close the chain, (especially in 5'CpG regions). It is unlikely that the values determined from the diagram were sufficiently accurate. Finally, the positions of the sugar atoms were derived from standard C3'-endo and C2'-endo sugars whereas the ζ angles indicate that the puckers have been distorted (table A3.1). This is not a criticism of the S-DNA model since it is perfectly feasible that polynucleotide sugar puckers might differ somewhat from the ideal values. However, it contributes further uncertainty to the information needed to build the model. It should also be noted that the ring covalent geometry is correlated with its pucker (Arnott and Hukins, 1972a).

In view of these difficulties, no further progress was made with this method.

Method 2

The second method consists of several steps. First the modelbuilding program was used to derive the co-ordinates in an arbitrary reference frame of the backbone atoms from the known covalent stereochemistry and the torsion angles. In figure A3.2, for example, two residues have been built and the final reference frame is O'x'y'z'. We now wish to derive the co-ordinates of the atoms in the reference frame Oxyz where z is the helix axis. In general therefore we require to perform the matrix operations:-

$$\underline{X} = \underline{R}.\underline{X}' + \underline{L} \qquad (A1)$$

where $\underline{R}$ is a 3 x 3 matrix which effects the three rotations necessary to

orient O'x'y'z' and Oxyz so that corresponding axes are parallel and $\underline{L}$ is a column vector which translates the origin of O'x'y'z' to that of Oxyz. Thus we need to find twelve unknowns - the nine elements of $\underline{R}$ and the three of $\underline{L}$.

We know Δz and $\Delta\phi$, the rise and rotation per residue respectively. We can use this information to derive the radial co-ordinates of atoms in Oxyz as follows. Consider atoms 1 and 2 in figure A3.3 which are corresponding atoms in successive residues of the helix. We can calculate d_{12}, the distance between 1 and 2, from the co-ordinates of these atoms in the arbitrary reference frame:-

$$d_{12} = \left[(x_1' - x_2')^2 + (y_1' - y_2')^2 + (z_1' - z_2')^2\right]^{\frac{1}{2}}$$

Therefore, $\Delta\ell$, the distance between the projections of 1 and 2 onto the xy-plane is given by:-

$$\Delta\ell = (d_{12}^2 - \Delta z^2)^{\frac{1}{2}} \qquad \text{(A2)}$$

Looking in projection down the helix axis (figure A3.4) we see that:-

$$\sin \tfrac{1}{2}\Delta\phi = \frac{\frac{1}{2}\Delta\ell}{R} \qquad \text{(A3)}$$

$$\text{so } R = \tfrac{1}{2}\Delta\ell \operatorname{cosec} \tfrac{1}{2}\Delta\phi \qquad \text{(A4)}$$

where R is the radial co-ordinate of atoms 1 and 2. Using this procedure we may therefore find the radial co-ordinates of any atoms in the backbone.

We now arbitrarily assume that in our helical reference frame, Oxyz, the co-ordinates of atom 1 are $(x_1, 0, 0) \equiv (R_1, 0, 0)$. The co-ordinates of atom 2 are then $(R_1, \Delta\phi, \Delta z)$. Similarly the co-ordinates of corresponding atoms further along the chain are $(R_1, 2\Delta\phi, 2\Delta z)$ ... $(R_1, n\Delta\phi, n\Delta z)$ etc.

In equation A1 we have twelve unknowns, four for each space co-ordinate. Thus in order to derive the elements of $\underline{R}$ and $\underline{L}$ we must use the co-ordinates of four atoms in the chain. Consider just the x co-ordinates.

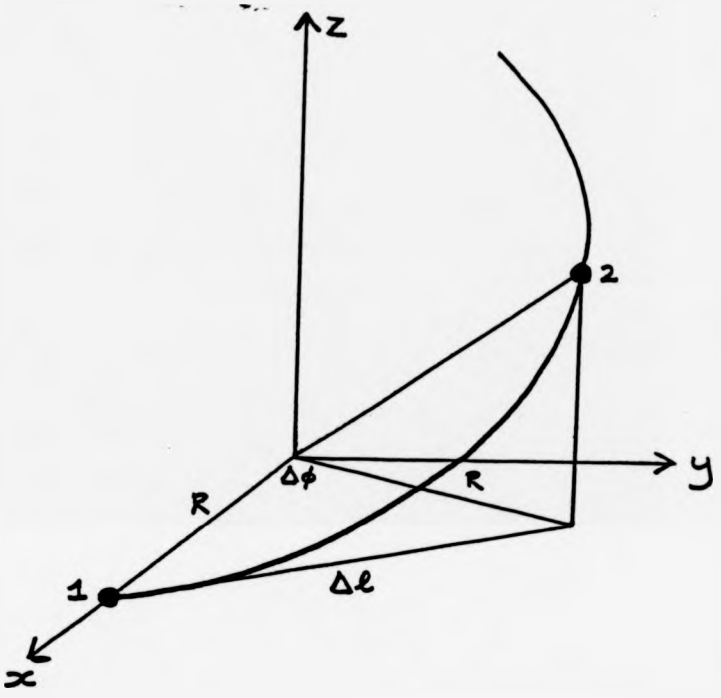


Figure A3.3 : Construction for the derivation of the radial co-ordinates (R) of two successive atoms on a helix

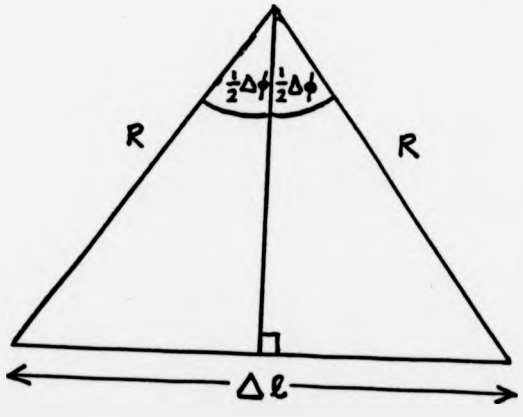


Figure A3.4 : Projection down the z-axis of the previous diagram

Equation A1 may be written as:-

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{pmatrix} \qquad (A5)$$

Therefore $x_1 = R_{11}x'_1 + R_{12}y'_1 + R_{13}z'_1 + \ell_1$ (A6)

and similar equations may be written for the x co-ordinates of each of the four atoms. These four equations may be written in the form:-

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x'_1 & y'_1 & z'_1 & 1 \\ x'_2 & y'_2 & z'_2 & 1 \\ x'_3 & y'_3 & z'_3 & 1 \\ x'_4 & y'_4 & z'_4 & 1 \end{pmatrix} \begin{pmatrix} R_{11} \\ R_{12} \\ R_{13} \\ \ell_1 \end{pmatrix} \qquad (A7)$$

or $\underline{A} = \underline{B}.\underline{C}$ (A8)

We know all the elements of $\underline{A}$ and $\underline{B}$ and we wish to find $\underline{C}$:-

$$\underline{C} = \underline{B}^{-1}.\underline{A} \qquad (A9)$$

We perform a similar procedure using the y and z co-ordinates which eventually gives us all the elements of $\underline{R}$ and $\underline{L}$. Equation A1 may then be used to derive the co-ordinates of all the backbone atoms in the helical reference frame. In general this will not be the most convenient frame since the x-axis passes through atom 1 rather than being aligned along the diad axis for example. However, this may be remedied by a simple rotation about z and translation along z.

The method was tested by building several known polynucleotides. Although the first section of the procedure, finding the radial co-ordinates of the atoms, gave quite encouraging results (table A3.2) the final helical

parameters for A-DNA, B-DNA and S-DNA were widely different from their actual values (table A3.3). This is probably due to the errors which accumulate when performing repeated multiplication operations such as in the linked-atom method. Finding an average set of co-ordinates for the helically related atoms by building more residues in the arbitrary frame did not improve the results since the errors grow larger as the chain length is increased.

The conclusion to be drawn from these studies is that presentation of polynucleotide models in terms of their torsion angles alone does not enable other workers to test the model.

TABLE A3.1 : Comparison of the angle ζ in standard sugar rings and S-DNA

Pucker	Standard sugars[1]	S-DNA[2]
C3'-endo	83.2	76
C2'-endo	146.2	147

1. Arnott and Hukins (1972a)
2. Arnott et al (1980)

TABLE A3.2 : Comparison of the radial co-ordinates of two atoms in B-DNA with those derived using method 2

Atom	B-DNA[1] (A)	Derived (A)
C3'	8.24	8.20
P	9.02	8.91

1. Arnott and Hukins (1972b)

TABLE A3.3 : Helical parameters of some test models observed using method 2

Model	t(deg)	h(A)	N	Pitch(A)
A-DNA	34.7 (32.7)[1]	2.57 (2.56)	10.4 (11.0)	26.7 (28.15)
B-DNA	38.1 (36.0)[2]	3.31 (3.38)	9.5 (10.0)	31.5 (33.8)
S-DNA	51.5 (60.0)[3]	5.66 (7.26)	7.0 (6.0)	36.6 (43.5)

t and h are the rotation and rise per asymmetric unit respectively. N is the number of asymmetric units in one turn of the helix. The bracketed figures are the values of the parameters observed in fibres.

1. Fuller et al (1965)
2. Langridge et al (1960a)
3. Arnott et al (1980)