# Keele 8 U N IVERSITY 

This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, noncommercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

## X-RAM DIFFRACTION

ANO

MOLECULAR MODELBUILDing studies

ON THE
deoxyribonucleic acid couble helix
by

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

Volume II

A thesis submitied to the University of Keele for the degree of ioctor of Philosophy

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

## CHAPTER IV

THE SIDE-BY-SIDE MODEL : IS DNA A DOUBLE HELIX?

### 4.1 Introduction

The lack of an objective solution of the phase problem combined with the low resolution of diffraction data from polynucleotide fibres has provided an opportunity for dissenters from the double-helix hypothesis to criticise the details of the Watson-Crick model and sometimes to present alternative conformations. For example, the fibre diffraction evidence for Watson-Crick base-pairs has been a source of considerable debate (Donohue, 1969, 1970; Wilkins et al, 1970; Crick, 1970; Arnott, 1970). Recently attention has focussed on the sugar-phosphate backbone. A number of workers have described models for DNA in which the two polynucleotide strands are in side-by-side association rather than intertwined in a double-helical structure (Rodley et al, 1976; Sasisekharan and Pattabiraman, 1976, 1978; Sasisekharan et al, 1977, 1978; Cyriax and Gäth, 1978; Pohl and Roberts, 1978). In all these side-by-side (SBS) models the two polynucleotide strands are antiparallel and are linked through complementary base-pairing of the Watson-Crick type. In the SBS models which have been described in most detail, a region of five nucleotidepairs in a right-handed double-helical conformation similar to that in the Watson-Crick B model is followed by five nucleotide-pairs in a lefthanded conformation which is in turn followed by another right-handed region and so on throughout the length of the molecule. Both groups of workers who have proposed detailed models of the SBS type have made the point that the model they have described should be seen as one member of a
whole family of possible models differing for example in the number of nucleotide-pairs in the left- and right-handed regions. The initial stimulus for the construction of these models was concern over the topological difficulties of strand separation for the Watson-Crick model during replication, transcription and recombination. Both groups have argued that these processes may be more easily explained by the SBS model wherein strand separation would not involve unwinding of a helix.

Two distinct types of SBS structure have been proposed and designated type I and type II (Sasisekharan and Pattabiraman, 1977). They are distinguished according to the relative orientation of the sugar rings with respect to the helix axis in the right- and left-handed regions along one polynucleotide chain. In type I structures sugar rings in left- and right-handed regions "point in an approximately the same direction" whereas in type II structures sugar rings in right-handed regions "point in an approximately opposite direction" to those in lefthanded regions. The direction in which a sugar ring points is defined by the direction of the $C_{2}^{1} \rightarrow O_{5}^{\prime}$ vector.

The SBS model proposed by Rodley et al (1976) is of type I and models of both type I and II have been described by Sasisekharan and Pattabiraman (1977). There is some dispute between these authors on the stereochemical feasibility of type I models and a suggestion that type II models are energetically more favourable (Sasisekharan et al, 1978). However experience would suggest that it is extremely difficult to exclude all polymer conformations of a particular type on the basis of energy calculations, and, in the modelbuilding studies described in this chapter, we have found it rather easier to build the type I model, although the construction of both types of model presented stereochemical difficulties.

The only SBS model for which atomic co-ordinates have been published is the type I model described by Rodley et al (1976). In
presenting this model the authors expressed the hope that specialists in vartous areas of polynucleotide research would consider the possibility of its occurrence under various physical condttions. In a recent discussion of the evidence for the dquble-helical structure of DNA, Crick et al (1979) emphasised the need for calculations of the $X$-ray diffraction which would be expected from structures of the SBS type. In the following sections we describe such calculations for the SBS model published by Rodley et al (1976). In addition to reporting precise calculations for this model we also describe similar calculations for a simple modification of this model which might be expected to improve the agreement with the observed X-ray diffraction. These are the first detailed calculations which have been published for models of the SBS type (Greenall, Pigram and Fuller, 1979; Appendix), although Arnott (1979) has quoted an overall reliability index for the agreement between the SBS model of Rodley et al (1976) and the observed diffraction data. In contrast to the claim by Rodley and coworkers that the SBS model accounts satisfactorily for the B-DNA diffraction pattern, the reliability index quoted by Arnott suggests that the SBS model is in substantially worse agreement with the observed $X$-ray data than is the best model of the Watson-Crick type. However, the soundness of Arnott's calculation (which will be discussed later) is open to doubt and in any case it is not clear that a single parameter such as the reliability index is the best way of comparing models of this kind and in the following we describe the full molecular Fourier transforms which will form the basis for any consideration of possible structural refinement of the SBS model. In addition we will discuss the constraints imposed upon future SBS models by the extensive polymorphism of DNA and crystal packing. The New Zealand group have suggested that Patterson functions may be the most straight-forward way of deciding between SBS and double-helical models. These arguments will be considered in Chapter 5. In the present chapter we will also describe briefly additional biological and physical arguments
adyanced in favour of the SBS model,

### 4.2 Detailed Description of the Model

Atomic cQ-ordinates of two SBS models, both of type $I$, have been published (Rodley et al, 1976; Millane and Rodley, 1981). No details of type II models have yet been reported and so these conformations will not be discussed.

Rodley et al (1976) have designated the two strands of the molecule $A$ and $B$, and the nucleotides have been numbered as shown in figure 1. This nomenclature will be employed in this chapter.

The model contains Watson-Crick base-pairs stacked upon each other with an axial separation of approximately 3.45A. The gross character of the model is defined by five nucleotide pairs in a right-handed helix (nucleotides 9, 10, 1 and 2) followed by another five pairs forming a lefthanded helix (nucleotides 4, 5, 6 and 7) and so on. There are, therefore, two bends or folds in the polynucleotide backbone every ten nucleotidepairs. These bends are not identical : as one goes down the 5'-3' chain the bend between a right- and left-handed segment (nucleotides B2, B3 and B4) is called a q-bend whilst that between a left- and right-handed segment (nucleotides B7, B8 and B9) is called a p-bend. In p-bends the C4'-C5' conformation changes from gg used in right-handed regions to gt used in left-handed regions. In this sense it is similar to the kinks proposed by Crick and Klug (1975) to explain the folding of ONA in nucleosomes. The sequence of conformers at $q$ bends is $g g-t g-g t$. All the nucleotides are in the anti conformation. All but one of the sugars is in the C3'-endo pucker. The exception, which is $\mathrm{C} 3^{\prime}$-exo, is in the q -bend region. The phosphate groups in the left-handed regions tend to have higher radial co-ordinates than those in right-handed sections so the maximum diameters of the two regions differ by $1-2 A$. The molecule contains two diad axes perpendicular to the molecular axis in each asymmetric unit. The distance
A strand B strand


Figure 4.1 : Nomenclature for SBS

- p-bend
- q-bend
- diad axis

The box contains the molecular asymmetric unit (From Rodley et al, 1976).
between the N3 and N9 atoms in each nucleotide-pair was maintained at 9.2A and the centres of the lines joining these two atoms in each basepair were constrained to lie as closely as possible along a line parallel to the molecular axis.

The model contains a nett $35^{\circ}$ right-handed twist every ten nucleotide-pairs since the right-handed regions have a greater magnitude of rotation. This model is not therefore strictly side-by-side in the sense originally intended by Rodley et al (1976). If there were no nett twist then the transverse cross-section of the molecule would be roughtly heart-shaped with all the p-bends at one side and all the q-bends at the other. In addition the $3^{\prime}$ ends of the two chains are on the opposite side of the molecule from the 5 ' ends.

The more recent model (Millane and Rodley, 1981) retains the same general features, However, the long-range right-handed twist has been increased from $35^{\circ}$ to $46.8^{\circ}$ per ten nucleotide-pairs. The backbone conformation has been changed from tg to gt at residue A8 and the C3'-endo pucker originally at residue 810 has been altered to C2'-endo. The constraint that the bases should be precisely stacked as described earlier has been removed.

The stereochemistry of the SBS model will be considered in detail later in this chapter.

### 4.3 Methods

### 4.3.1 Derivation of the Co-ordinates of 02 and 03

The SBS model published by Rod ley et al (1976) does not contain the phosphate oxygens 02 and 03 or any base atoms besides N3 or N9. Since it is clear that these atoms must be included in any comprehensive study of the nolecular stereochemistry or diffraction pattern, their co-ordinates were calculated. These derivations will be described in this and the
following section,
In figure 2 the co-ordinate axes are aligned in the manner defined by Rodley et al (1976), 0 is the position of a chosen phosphorous atom and 1, 2, 3 and 4 are the positions of the associated 01, 02, 03 and 04 atoms respectively. The position vectors of 0,1 and 4 are $\underline{r}_{0}, \underline{r}_{1}$ and $\underline{r}_{4}$ respectively. The vector from 0 to 1 may be written in the form:-

$$
\begin{align*}
r_{01} & =\left(x_{1}-x_{0}\right) \underline{i}+\left(y_{1}-y_{0}\right) \underline{j}+\left(z_{1}-z_{0}\right) \underline{k}  \tag{1}\\
& =\Delta x_{10} \underline{1}+\Delta y_{10} \underline{j}+\Delta z_{10} \underline{k} \tag{2}
\end{align*}
$$

Similarly:-

$$
\begin{align*}
& \underline{r}_{04}=\Delta x_{40} \underline{i}+\Delta y_{40} \underline{j}+\Delta z_{40} k  \tag{3}\\
& \underline{r}_{14}=\Delta x_{41} \underline{i}+\Delta y_{41} \underline{j}+\Delta z_{41} \underline{k} \tag{4}
\end{align*}
$$

Each of these vectors may be calculated immediately since we know the co-ordinates of 0,1 and 4. We now define the vectors $\underline{a}=\underline{r}_{01} \times \underline{r}_{04}$ and $\underline{b}=-\underline{r}_{14} \times \underline{a}$ which have the directions shown in the diagram. From the definition of $\underline{a}$ and $\underline{b}$ it follows that $\underline{b}$ is the bisector of the angle formed by 02, P and 03 and $\underline{a}$ (which is perpendicular to $\underline{b}$ ) lies in the plane formed by these atoms. The position vectors of 2 and 3 are then given by the expressions:-

$$
\begin{align*}
\underline{r}_{2} & =\underline{r}_{0}+\underline{r}_{02}  \tag{5}\\
& =\underline{r}_{0}+(\alpha \underline{\hat{a}}+\beta \hat{b}) \tag{6}
\end{align*}
$$

and

$$
\begin{align*}
r_{3} & =\underline{r}_{0}+\underline{r}_{03}  \tag{7}\\
& =\underline{r}_{0}+(-\alpha \underline{a}+\beta \underline{b}) \tag{8}
\end{align*}
$$



Figure 4.2 : Construction used in the derivation of the co-ordinates of 02 and 03.


Figure 4.3 : Projection perpendicular to the $\mathfrak{a}-\underline{b}$ plane of the construction in the previous diagram.
where $\underline{\hat{a}}$ and $\underline{\hat{b}}$ are unit vectors along $\underline{a}$ and $\underline{b}$ respectively and $\alpha$ and $\beta$ are constants whose values depend upon the stereochemistry of the phosphate group. The parameters $\alpha$ and $B$ are given by:-

$$
\begin{align*}
& \alpha=\left|\overrightarrow{\mathrm{PO}}_{2}\right| \sin (\theta / 2)  \tag{9}\\
& \beta=\left|\overrightarrow{\mathrm{PO}}_{2}\right| \cos (\theta / 2) \tag{10}
\end{align*}
$$

where $\theta$ is the angle formed by 02, $P$ and 03 (ftgure 3). Now:-

$$
\begin{align*}
\underline{a}= & \underline{r}_{01} \times \underline{r}_{04} \quad \text { (by definition) }  \tag{11}\\
= & \left|\begin{array}{lll}
\underline{i} & \underline{j} & \underline{k} \\
\Delta x_{10} & \Delta y_{10} & \Delta z_{10} \\
\Delta x_{40} & \Delta y_{40} & \Delta z_{40}
\end{array}\right|  \tag{12}\\
= & \left(\Delta y_{40} \Delta z_{40}-\Delta z_{10} \Delta y_{40}\right) \underline{i} \\
& +\left(\Delta z_{10} \Delta x_{40}-\Delta x_{10} \Delta z_{40}\right) \underline{j}  \tag{13}\\
& +\left(\Delta x_{10} \Delta y_{40}-\Delta y_{10} \Delta x_{40}\right) \underline{k} \\
= & a_{x} \underline{i}+a_{y \underline{j}}+a_{z-}(s a y) \tag{14}
\end{align*}
$$

$$
\text { Hence:- } \hat{\underline{a}}=\frac{\underline{a}}{a}=\frac{a_{x} \underline{i}+a_{y} \underline{j}+a_{z} \underline{k}}{\left(a_{x}^{2}+a_{y}^{2}+a_{z}^{2}\right)^{\frac{1}{2}}}
$$

and similarly:-

$$
\begin{equation*}
\underline{\underline{b}}=\frac{b_{x} \underline{i}+b_{y} j+b_{z} \underline{k}}{\left(b_{x}^{2}+b_{y}^{2}+b_{z}^{2}\right)^{\frac{1}{2}}} \tag{16}
\end{equation*}
$$

The values of $\alpha$ and $\beta$ were calculated from the co-ordinates of the phosphate group in B-DNA (Arnott and Hukins, 1972b). We may now calculate
$\hat{a}$ and $\hat{W}$ and hence $r_{2}$ and $r_{3}$ for each phosphate using equations (6) and (8).

### 4.3.2 Derivation of the Co-ordinates of the Base Atoms

The only base atom co-ordinates given by Rodley et al (1976) were those of N3 or N9. It was therefore necessary to derive the coordinates of the other atoms for the present work. A standard base-pair, using the co-ordinates given in Chapter 3, was used in these calculations. No tilt or twist was applied. Each base-pair was fitted into the model with a rotation followed by two translations. The rotation moved the base-pair into the orientation required by the N3 and N9 atoms in a given residue. The first translation moved the newly oriented base-pair parallel to the molecular axis until the $z$ co-ordinate of the base atoms was equal to the average of the published N3 and N9 co-ordinates. The second translation moved the base-pair along the line which is the base diad axis in double-helical DNA until the average calculated radial co-ordinates of N3 and N9 were equal to the average published co-ordinates of these atoms. No attempt was made to produce a random base sequence so this method gives rise to solely purine bases on one strand ( $A$ ) and pyrimidines on the other ( $B$ ).

### 4.3.3 Realignment of the Molecular Diad Axis

In the co-ordinate systen chosen by Rodley et al (1976) the molecular diad axis lies along the line joining the points $(x, y, z)=$ $(0,0,1.70)$ and $(-1.00,-4.20,1.70)$. The Fourier transform calculations to be described are greatly simplified if the diad axis lies along the $x$-axis since then the transform is completely real. Therefore the molecule was rotated about and translated along the $2-a x i s$ by $-77^{\circ}$ and -1.7A respectively,

### 4.3.4 Derivation of SBSO CO-ordinates

The SBS model published by Rodley et al (1976) contains a long-range right-handed twist of approximately $35^{\circ}$ every ten nucleotide pairs. In order to examine the diffraction to be expected from a model with no nett winding of the polynucleotide chains, the $\phi$ co-ordinates within the $q^{\text {th }}$ nucleotide pair were reduced by ${ }^{35} / 10 \times(q-1)^{0}$, $(q=1,2,3,4,5)$. The model so obtained will be referred to henceforth as SBSO and the original model as SBS 36.

### 4.3.5 Examination of the Molecular Stereochemistry of SBS 36

The covalent bond lengths and angles and the van der Waals' contacts between atoms in the molecular asymnetric unit of SBS 36 were calculated using the program BONANG described in Chapter 2. All the backbone atoms, including 02 and 03 in each residue, and the base atoms were included.

Non-bonded interactions between atoms in adjacent helices of SBS 36 were examined using the program IHC described in Chapter 2. Only the crystalline packing scheme was considered. In crystalline specimens the distance between the nearest neighbour helical axes is 19.04 A (Langridge et al, 1960 a). Contacts shorter than 4 A were printed out, therefore only those atoms whose radial co-ordinates were greater than 7.52 A needed to be included in the data. One molecule was displaced along the z-axis by 11.05 A with respect to the other as required by the $X$-ray data (Langridge et al, 1960a). The molecules were rotated about their axes in steps of 10 degrees, maintaining their diad axes parallel to each other.

The backbone torston angles of SBS 36 were calculated by the program PREP.

The stereqchemistry of SBSO was not examined since no attempt had been made in its derivation to preserve the correct bond lengths and angles.

### 4.3.6 Building the CPK Model of SBS 36

A CPK model of SBS 36 similar to the model of Rodley et al
(1976) was built in order to identify the most serious stereochemical difficulties to be expected from such models. Initially a short righthanded segment (corresponding to residues $8,9,10,1$ ) was constructed on a metal rod. The nucleotides were fixed on the rod using a bracket designed by Pigram (1968) for B-DNA and intercalation models. The remainder of the asymmetric unit was then added and adjusted until the stereochemistry was as good as could be achieved. An identical model was then constructed on a second rod, removed carefully from the rod, inverted, and slipped on top of the first model. When the covalent bonds between the two asymmetric units had been made, the model was adjusted again to achieve the best possible stereochemistry taking care to ensure that the diad symmetry element was preserved.

### 4.3.7 Fourier Transform Calculations

The SBS 36 model has a long-range twist of $35^{\circ}$ ten nucleotidepairs. However we will introduce only an insignificant error into the calculation of its transform if we assume it to be a 10 , helix with $h=34.5 A$ and $t=36^{\circ}$. Since the diad axis has been oriented along the $x$-axis the transform will be completely real. The diffraction from such a structure is confined to layer-planes given by $z=\ell / 345 A^{-1}$ where $\ell$ is any integer. The Fourier transform components, $G_{n \ell}(R)$, and the cylindrically averaged intensity, $\sum G_{n \ell}^{2}(R)$, on the layer-planes $\ell=0$ to 125 where calculated by the program Helix 1 described in Chapter 2 assuming the atoms scattered according to the curves given by Langridge et al (1960b) and discussed in Chapter 3.

The SBSO structure is not a helix except in the trivial sense but it is nonetheless convenient to treat it as a $l_{1}$ helix in the transform calculations. In this case, since there is no rotational symmetry, every

Bessel function will contribute to each layer-plane (Vainstein, 1966; Chapter 2). The molecule still retains a diad axis perpendicular to the molecular axis so the asymmetric consists of five nucleotide-pairs. The transform of this molecule is therefore also real and it is confined to layer-planes given by $Z=\ell / 34.5 \mathrm{~A}^{-1}$.

In both calculations Bessel functions were included in the range $|n| \leqslant 16$ and the transform was calculated from $R=0$ to $0.4 A^{-1}$ in steps of $0.1 \mathrm{~A}^{-1}$.

The transforms of the sugar, phosphate and base components of each model were also calculated

### 4.4 Results and Discussion

### 4.4.1 The Co-ordinates of 02 and 03

The values of $\alpha$ and $\beta$ are given in table 1. In order to verify the accuracy of the method, the co-ordinates of 02 and 03 in several published DNA and RNA models were calculated. Table 2 shows a comparison of the observed and calculated co-ordinates of these atoms in B-DNA (Langridge et al, 1960b; Arnott and Hukins, 1972b), A-DNA (Arnott and Hukins, 1972b), RNAlO and RNAll (Arnott et al, 1967b). The largest error in a linear dimension is approximately 0.3 A ( $Z$ of 02 in Langridge's B-DNA) and that in the angular values is approximately $0.5^{\circ}$ ( $\phi$ of 02 in RNAll) In both cases these models were built by hand. The agreement with the more precise computer-derived co-ordinates is much better. When we allow for slight variations in the phosphate stereochemistry between the various models, the agreement is quite satisfactory and confirms that the method will produce accurate values for the SBS co-ordinates. The final co-ordinates of the 02 and 03 atoms are included in table 3.

Table 4.1 : Values of $\alpha$ and $\beta$ (see text)

$$
\alpha=1.20 \mathrm{~A} \quad \beta=0.84 \mathrm{~A}
$$

Table 4.2 : Comparison of Observed and Calculated Co-ordinates of 02 and 03

| Source | Calculated |  |  | Observed |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $R(A)$ | $\phi$ | $Z(A)$ | $R(A)$ | $\phi$ | $Z(A)$ |
| A-DNA ${ }^{1}$ | 9.90 7.73 | 68.1 73.8 | -5.14 -4.59 | 9.96 7.69 | $\begin{aligned} & 67.9 \\ & 73.9 \end{aligned}$ | $\begin{aligned} & -5.10 \\ & -4.53 \end{aligned}$ |
| B-DNA ${ }^{1}$ | $\begin{array}{r} 10.21 \\ 8.89 \end{array}$ | 91.4 103.1 | 1.81 1.26 | 10.20 8.82 | $\begin{array}{r} 91.1 \\ 103.3 \end{array}$ | 1.86 1.29 |
| $B-D N A^{2}$ | $\begin{array}{r} 10.31 \\ 9.11 \end{array}$ | $\begin{aligned} & 54.4 \\ & 65.2 \end{aligned}$ | -1.43 -2.42 | 10.33 9.14 | 54.2 65.3 | $\begin{aligned} & -1.18 \\ & -2.45 \end{aligned}$ |
| RNA10 ${ }^{3}$ | 9.79 7.71 | $\begin{aligned} & 36.7 \\ & 41.6 \end{aligned}$ | $\begin{aligned} & -6.88 \\ & -5.92 \end{aligned}$ | 9.79 7.65 | 36.5 41.4 | $\begin{aligned} & -6.96 \\ & -5.82 \end{aligned}$ |
| RNA11 ${ }^{3}$ | $\begin{aligned} & 9.76 \\ & 7.71 \end{aligned}$ | $\begin{aligned} & 36.5 \\ & 41.7 \end{aligned}$ | -7.49 -6.51 | 9.76 7.64 | $\begin{aligned} & 36.0 \\ & 41.8 \end{aligned}$ | $\begin{aligned} & -7.55 \\ & -6.45 \end{aligned}$ |

1) Arrott and Hukins (1972) Biochem. Biophys. Res. Commun., 47, 1504.
2) Langridge et al (1960) J. Mol. Biol., 2, 38.
3) Arnott et al (1967), J. Mol. Biol., 27, 535.

### 4.4.2 The Co-ordinates of the Base Atoms

The distance between the N3 and N9 atoms in the most recently refined Watson-Crick models is 9.04A not 9.2 A as quoted by Rodley et al (1976). In fact this distance has not been held constant as they claim (table 4). Instead it varies from 9.05A in residue 6 to 9.22A in residue 8. This irregularity makes it difficult to fit the standard Watson-Crick base-pair to the Rodley model, however it is not necessarily a serious flaw since the N3-N9 distance is determined by the purine-pyrimidine hydrogen bonds and the energy required to produce such small distortions could probably be compensated if for example more favourable stereochemistry resulted thereby in the sugar-phosphate chain. It is not clear of course that such compensating effects are present in the SBS model.

It is useful to compare the co-ordinates of N3 and N9 as published by Rodley et al (1976) with those calculated for the present work (table 5). The largest discrepancies ( 0.18 A and 0.17 A ) occur in the residue (number 8 ) where the $\mathrm{N} 3-\mathrm{N} 9$ separation is most distorted from the Watson-Crick value. The other values indicate that the calculated basepair co-ordinates will be satisfactory for the Fourier transform calculations.

### 4.4.3 Final Co-ordinates of the Models

The co-ordinates of the atoms in SBS 36 following the addition of the oxygen and base atoms are given with respect to the system which results from realignment of the diad axis in table 3. The co-ordinates of the SBS 0 model are not given since they may be derived trivially from those of SBS 36.

### 4.4.4 The Stereochemistry of SBS 36

Rodley et al (1976) claim that all the bond lengths within SBS 36 are within 0.15 A and that all the bond angles are within $15^{\circ}$ of the standard values derived by Arnott and Hukins (1972a) from a survey of

Table 4.3 ：Co－ordinates of the Asymmetric Unit of SBS36

RESID：JE AG

|  | $x(4)$ | $Y(A)$ | 7 （A） | $R(4)$ | PHI（DEC） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $05^{\circ}$ | －2．6？ | 5.0 .7 | －2．85 | 6.19 | 115.1 |
| C1． | －2． 43 | 5.12 | －1．69 | 5.58 | 115.8 |
| C2 | －2． 35 | 5.97 | － 6.55 | 6.41 | 111.6 |
| C3： | －1．8．） | 7.13 | －1．25 | 7.35 | 184．？ |
| Cai． | －2．47 | 7.1 ？ | －2．5． | 7.54 | 109.1 |
| C5＇ | －1．83 | 7.69 | －3．7 | 7.74 | 103.4 |
| － | －4．26 | 7.73 | －4．6．1 | 8.8 ？ | 118.9 |
| 01 | －4．09 | 6.27 | －4．2w | 8.11 | 128.5 |
| 02 | －4．4．） | 8.64 | －3．5． | 9.7 1 | 117.9 |
| 03 | －4．87 | 8.27 | －5． $8_{4}$ | 9.67 | 12.9 .5 |
| 04 | －？． 59 | 7.43 | －4．85 | 7．90 | 1.19 .9 |
| PYH11 | 6.127 | 2.44 | －1．7 7 | 2.45 | 83.6 |
| PYCL | －6． 95 | 3.11 | －1．7．1 | 3.16 | 107.6 |
| PYO？ | －2． 7 \％ | 2.35 | －1．7．1 | 3.118 | 13.4 |
| PYVZ | －1．01 | 4.43 | －1．7） | 4.5 ？ | 183.3 |
| PYCA | ¢． 08 | 5.17 | －1．7．1 | 5.17 | 89.2 |
| PYCS | 1.32 | 4.61 | －1．60 | 4.811 | 74.1 |
| THIE | 2.57 | 5.46 | －1．60 | 6.23 | 51.8 |
| PYCG | 1.38 | 3.19 | －1．7．1 | 3.47 | 65.5 |
| Privg | 2.56 | 2.57 | －1．7．1 | 3.65 | 45.2 |
| Prog | 2.50 | 2.57 | －1．7 | 3.63 | 45.2 |

## RESICUE B6

|  | $x(1)$ | $Y(A)$ | Z（A） | R（A） | PHI（DEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $05!$ | 3.26 | －6． 26 | －0．35 | 0.27 | 272.3 |
| C1！ | 3.25 | －5．59 | －1．71 | 5.6 ？ | 272．6 |
| C2 ${ }^{\text {i }}$ | 0.75 | －0．61 | －2．5． | b． 65 | 276.5 |
| C3i | 1.83 | －7．27 | －1．85 | 7.5 － | 254．？ |
| C4i | 1.12 | －7．52 | －6．6：1 | 7.59 | 277.7 |
| $\mathrm{CS}^{\circ}$ | 1.83 | －7．63 | 6.31 | 7.91 | 233.4 |
| P | －3．32 | －3．81 | 1．2． | 8.87 | 267.9 |
| 01 | －1．63 | －7．34 | ？．8．1 | 8.11 | 258.3 |
| 02 | ？． 1.3 | －9．79 | 0.14 | 9.717 | $27{ }^{\prime \prime}$ |
| 03 | －1．6．4 | －9．58 | 2.4 .1 | 9．6？ | 260．？ |
| ${ }^{2} 4$ | 1.74 | －7．151 | 1． 15 | 7．9：4 | 276.4 |
| PiJ！ 1 | 1.15 | －6． 37 | －1．7．1 | 1.35 | 32\％．3 |
| Puc ？ | －1．62 | －1．26 | －1．7 | 1.30 | 247．6． |
| CiJ？ | －1．63 | －1．05 | －1．7 | 1.9 .4 | 197.5 |
| Pu：！ 3 | －$\%$－ 93 | －2．53 | －1． 6 （i） | 2.64 | 25． 4 |
| Puca | 4．67 | －3．．7 7 | －1．7 | 3.15 | 28？． 7 |
| Plic 5 | 1． A$)$ | －2．17 | －1．7 | 3． 2 | 37 ¢． 1 |
| P Jf．\％ | 1.76 | －．1．97 | $-1.7$ | 2．11 | 331.1 |
| 「ばい | 7.72 | －1．1？ | －1．7 | 2．1？ | 357.5 |
| Pנח\％ | 2．7？ | －． 1 12 | －1．7．1 | 2.72 | 3ら7．5 |
| คu゚！ 7 | 3.11 | －3．19 | $-1.7$ | 4.38 | 313.3 |
| Pur．s | ？． 11 | －4． 42 | －1．7．1 | 5.14 | 208．5 |
| Fille | 1.94 | －4．4： | －1．711 | 4.63 | 283．3 |

RESIDUE AT

|  | $x(4)$ | $Y(4)$ | 7．${ }^{\text {a }}$ ） | R（A） | PHI（DEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| OS！ | －4．87 | 3.41 | －6．3：1 | 5.95 | 145.0 |
| C1i | －4．45 | 2．9．1 | －5．35 | 5.31 | 146.9 |
| C2 ${ }^{\text {i }}$ | －4．82 | 3.84 | －4．0） | 0.15 | 141.5 |
| C3！ | －1．65 | 5.15 | －1．85 | 6.37 | 132.6 |
| C4！ | －5． 13 | 4． 17 | $-5.95$ | 7.16 | 139.3 |
| C5 ${ }^{\circ}$ | －5．11 | 5.47 | －7．30 | 7.48 | 133.1 |
| P | －7．45 | 4.55 | －7．7\％ | 8.73 | 148.6 |
| 01 | －7．46 | 2.85 | －7．50 | 7.90 | 159.1 |
| 02 | －7．64 | 5.13 | －6．46 | 9.23 | 146.1 |
| n3 | － 8.57 | 4.95 | －8．6．7 | 9.90 | 154.0 |
| $\mathrm{n}_{4}$ | －6．09 | 4.87 | －8．3！ | 7.8 可 | 141.4 |
| PYN1 | － 9.94 | 1.76 | －5．22 | 1.99 | 118：7 |
| PYC2 | －2．79 | 1.72 | －5．23 | 2.36 | 143.1 |
| PYOZ | －2． 92 | 1.05 | －5．22 | 2.99 | 167.4 |
| Priz | －3．3．7 | 2．9？ | －5．22 | 4.19 | 135.8 |
| FYCa | －2．35 | 4.11 | －5．22 | 4.73 | 119.8 |
| rycs | －．1．79 | 4.18 | －5．22 | 4.29 | 133.3 |
| THMF． | －3． 26 | 5.59 | －5．22 | 5.54 | 92.7 |
| PYCG | －6． 29 | 2.93 | －5．？2 | 2.95 | 95.0 |
| PYido | 1.04 | 2.91 | －5．22 | 3.19 | 714.4 |
| PYOG | 1.64 | 2.91 | －5．22 | 3． 59 | 70.4 |

## RESIDIJE 37

|  | X（A） | $Y(A)$ | 7（A） | R（A） | PHI（DEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $05!$ | 3.11 | －5．94 | －4．45 | 0.71 | 297.6 |
| C1： | 2.53 | －5．46 | －5．25 | 6.42 | 294．9 |
| C2 | 3.35 | －6． 619 | －6．35 | 6.95 | 298.8 |
| C3i | 4.58 | －6． 35 | －5．5！ | 7.50 | 307.1 |
| C．${ }^{1}$ | 4.25 | －0．6． $0^{4}$ | －4．25 | 7.89 | 3.2 .5 |
| C．5 ${ }^{\circ}$ | 5.37 | －6．71 | －3．16 | 8.55 | $3: 8.3$ |
| P | 3.89 | －3．58 | －1．9．1 | 9．4？ | 294． 1 |
| 01 | 2.39 | －8．32 | －2．5．1 | 8.66 | 285.4 |
| 0 | 4.63 | －9．54 | －2． $\mathrm{Ba}_{0}$ | 11．6：1 | 295.9 |
| 13 | 3．Ris | －7．14 | －5．5： | 9．0．） | 29？．6 |
| 11.4 | 4.59 | －7．18 | －1．9．1 | 6． 52 | 3：2．6 |
| PUN1 | 1．6．3 | －1．59 | －5．22 | 4.41 | 31.8 |
| 0．jr2 | 1.03 | －1．3？ | －5．2 | 1.89 | 27：8 |
| Fojup | －1．12 | －1．94 | －5．？？ | 2.14 | 233．9 |
| $07: 3$ | $\because .28$ | －3．19 | －4．22 | 3.21 | 775. |
| DJEA | 1.93 | －2．97 | －5．？？ | 3.51 | 3，3． |
| ロ．J！ | 2.65 | －1． 12 | －5．2．3 | 3.22 | 325.7 |
| nucu | 1.03 | －．．ol | －！． 2 ？ | 2.13 | 342.0 |
| －u゙す | 2．17 | 1．5？ | －5． 23 | 2.47 | 13.8 |
|  | 2.41 | $\therefore \square^{5}$ | －5． 23 | 2.17 | 13.8 |
| 1－J 17 | 4．？ | －2．${ }^{\text {a }}$ | －5．2．3 | 1． 5.3 | 332.5 |
| 「いく』 | 1．117 | $-3.39$ | －5． 2.3 | 5.314 | $321 . ?$ |
| P 149 | 2．014 | －5．3n | －5．？ 2 | $4.9: 1$ | 3：5．n |

RESIDUE AB

|  | X(A) | $Y(A)$ | Z ( ${ }^{\text {a }}$ ) | $R(A)$ | PHI (DEG) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $05!$ | -5.59 | 1.78 | -9.85 | 5.64 | 172.1 |
| C1i | -5.39 | .1.64 | -9.45 | 5.13 | 172.8 |
| c. ${ }^{\text {i }}$ | -5.87 | 1.33 | -7.45 | b.t? | 167.3 |
| C3i | -6. 15 | 2.27 | -8.4d | 6.84 | $16 \% .6$ |
| C4: | -6.53 | 1.49 | -9.6! | 6.84 | 167.4 |
| C5 ${ }^{\text {i }}$ | -6.75 | 2.45 | $-19.411$ | 7.19 | 169.3 |
| P | -7.23 | 4.91 | -11.75 | 8.74 | 145.8 |
| 01 | -6.27 | 5.69 | -18.75 | 3.41 | 138.2 |
| 02 | -6.59 | 4.54 | -13.13) | 8.619 | 145.4 |
| 03 | -8.45 | 5.72 | -12.0.4 | 10.211 | $145 . ?$ |
| 014 | -7.58 | 3.59 | -10.85 | 8.39 | 154.7 |
| PYNI | -1.47 | 1.35 | -8. 65 | 1.89 | 144.5 |
| PYCZ | -2.69 | 6. 45 | -B. 65 | 2.72 | 179.5 |
| PYO2 | -2.81 | -11.73 | - 8.65 | 2.92 | 195.6 |
| Pru3 | -3.83 | 1.25 | -8.05 | $4.6 ?$ | 162.17 |
| PYC4 | -3.73 | 2.64 | -8.65 | 4.54 | 145.1 |
| PYCS | -2.52 | 3.22 | -8.64 | 4.619 | 128.9 |
| THME: | -2.4 | 4.73 | -8.64 | 5.37 | 116.9 |
| Pyco | -1.30 | 2.38 | -8. 8.65 | 2.74 | 119.8 |
| ryidg | -2. 15 | 2.92 | -8.65 | 2.92 | 93.4 |
| PYOg | -1). 15 | 2.92 | -8.65 | 2.92 | 93.1 |


|  | X (A) | $Y(A)$ | 7 ( $\mathrm{A}^{\text {) }}$ | $R(A)$ | PHI (DEG) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 051 | 5.26 | -4.61 | -7.8i | 0.99 | $318 . ?$ |
| C.1: | 4.41 | -4.31 | - 8.85 | 5.98 | 317.5 |
| C. ${ }^{\text {i }}$ | 5.36 | -1. 5.57 | -10.05 | 6.77 | 317.5 |
| C3! | 6.31 | -4.92 | -9.70 | 8.65 | 322.6 |
| C4! | 6.28 | -5.24 | - 3.35 | 3.18 | 32.7 |
| $\mathrm{Cb}^{\circ}$ | 7.67 | -5. 5.11 | -7.65 | 9.16 | 326.8 |
| P | 7.02 | -0.31 | -5.9.1 | 9.78 | 315.9 |
| 91 | 5.10 | -13.36 | -6. 15 | 4.79 | 3 c 8.6 |
| 02 | 7.75 | -7.011 | -f.rin | 11.31 | 314.8 |
| 03 | 7.32 | -7.11 | -4.5:1 | 14.29 | 315.9 |
| 0.1 | 7.45 | -5. 37 | -6.25 | 9.18 | 321.? |
| 『J!! | ${ }^{11} .94$ | -1. 55 | -8.64 | 1.19 | 33.10 |
| Pucz | 3.92 | -1.87 | -R.64 | 2.119 | 296.2 |
| Gus? | -7. 37 | -\%. 52 | -8.65 | 2.55 | 261.7 |
| P1P. 3 | 1.109 | -2.95 | - 4.6 .1 | 3.461 | 299.9 |
| P.IC. 1 | 3.1 .1 | -2.35 | -8.65 | 3.72 | $32 \% .3$ |
| F.Jcu | 3.27 | -.1.71 | -8.65 | 3.37 | 347.8 |
| Puca | 2. 13 | $\because \cdot 19$ | -8.65 | 2.13 | 2.3 |
| Padm | 2.115 | 1.37 | - - .65) | 2.47 | 33.7 |
| runt | P.45 | 1.37 | - 2.65 | 2.47 | 33.7 |
| Pい1\%7 | 1.64 | -1.11 | -3.6's | 4.65 | 355.1 |
| rucis | 5.23 | -1.5ts | -8. 60 | 4.45 | 343.4 |
| P!Jas | 4.30 | -2.01 | - H .65 | 6.15 | $329^{\prime}$ |

RESIDUE A9

|  | x（A） | $Y(A)$ | 2（A） | R（A） | PHI（DEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $05{ }^{\text {i }}$ | －4．15 | 4.72 | －13．11 | 6.29 | 131.4 |
| C1 ${ }^{\text {i }}$ | －3．91 | 3.69 | －12．1i | 5.38 | 136.6 |
| C2 ${ }^{\text {i }}$ | －4．15 | 4.51 | －113．85 | 6.13 | 132.7 |
| C3！ | －4．83 | 5.79 | －11．15 | 7.54 | 129.8 |
| C4： | －4．49 | 5.92 | －12．6．1 | 7.43 | 127.2 |
| C5 ${ }^{\text {i }}$ | －3．59 | 7．19 | －12．90 | 7.99 | 117.5 |
| P | －1．25 | 8． 118 | －12．25 | 8.19 | 98.8 |
| 01 | －1．43 | 8.60 | －13．85 | 8.72 | 99.4 |
| 02 | －1．29 | 9.21 | －11．3．1 | 9．3！ | 98.3 |
| 03 | －n．an | 7.40 | －12．111 | 7．4： | 9月．？ |
| 04 | －2． 51 | 7.11 | －12．84 | 7.54 | 139.5 |
| PYN1 | － 7.58 | 2.19 | －12．11 | 2.18 | 185.5 |
| PYCZ | －1．93 | 2.23 | －12．13 | 2.95 | 130.8 |
| Proz | －2． 69 | 1.26 | －12．111 | 2.97 | 155.7 |
| Prus | －2． 48 | 3.51 | －12．1w | 4.33 | 125.2 |
| PYC4 | －1．68 | 4.61 | －12．11 | 4.91 | 110.9 |
| PYC5 | －0．32 | 4.50 | －12．39 | 4.51 | 94.1 |
| Trime | 0.58 | 5.71 | －12．以 | 5.74 | 84．？ |
| pyeg | 9.22 | 3.17 | －12．11 | 3.18 | 86.1 |
| PYidg | 1.53 | 2.99 | －12．11 | 3.35 | 62.9 |
| PYó | 1.53 | 2.99 | －12．13 | 3.35 | 62.9 |

## RESIOUE R9

|  | $x(4)$ | $Y(A)$ | Z（A） | R（A） | PHI（OEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $00^{1}$ | 2.61 | －-21 | －11．4 ${ }^{1}$ | 6.74 | 292.8 |
| C1： | 1.77 | $-5.28$ | －12．95 | 5.57 | 288.5 |
| C2； | 1.74 | －5．8．3 | －13．35 | 6.46 | 286.7 |
| C3i | 2.75 | －7．12 | －13．55 | 7.52 | 291.7 |
| C4： | 3.35 | －7．23 | －12．20 | 7.35 | 292.8 |
| C5＇ | 4.61 | －7．28 | －12．11 | 8.62 | 312.3 |
| P | 6.53 | －5．64 | －12．15 | 8.6 .3 | 319.2 |
| 01 | 6.7 \％ | －5．91 | －10．511 | 8.93 | 318.5 |
| 02 | 7.15 | －0．52 | －12．34 | 9.90 | 318．8 |
| 03 | 6.89 | －4．26 | －12．4．1 | 8.10 | 328.3 |
| 14 | 5．：76 | －5．94 | －12．611 | 7.859 | $31 \% .4$ |
| P（1is） | A． 62 | －．1． 5.3 | －12．19 | 6． 82 | 310.1 |
| Pus？ | －．）．1． | －1．65 | －17．17 | 1．6＇s | 265.5 |
| GJり？ | －1．51 | －1．51 | －12．11 | 2.16 | 224.1 |
| Pats | －1．9？ | －2．97 | －12．17 | 2.97 | 269.7 |
| Puta | 1.54 | －2．96 | －12．1．1 | 3.39 | 209.1 |
| Pucs | 2．52 | －1．9？ | －1？．11 | 3.17 | 322.7 |
| PJCo | 1.95 | －il． 6 ？ | －12．1．1 | 2．r5 | 342.3 |
|  | 2.53 | 13．59 | －12．11 | 2.02 | 11.1 |
| Puos | 2.513 | 11.51 | －12．1い | 2．62？ | 11.1 |
| PU17 | 3.43 | －2．35 | －12．1．1 | 4．5＇3 | 328.3 |
| Puca | 3.71 | －3．0it | －12．11 | ¢． 21 | 315.1 |
| Fリリ9 | 2.42 | －4．89 | －12．11 | 4.75 | $34: 1.5$ |

RESIDJE A1N

|  | $x(4)$ | $Y(A)$ | 7（ ${ }^{\text {a }}$ | $R(A)$ | PHI（DEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Os： | －6． 52 | 0.71 | －16．7．1 | 6.73 | 94.4 |
| C1！ | －1．17 | 5.78 | －15．8：1 | 5.93 | 121.4 |
| C2 ${ }^{\text {＇}}$ | －1．72 | 6.37 | －14．65 | 6.60 | 105.1 |
| C3． | －1． 93 | 7.59 | －14．65 | 7.05 | 97.3 |
| C4！ | －7． 59 | 7.38 | －16．45 | 7.9 \％ | 94.3 |
| C5＇ | 9．72 | 8.40 | －16．135 | 8.43 | 85.1 |
| P | 3.18 | 7.83 | －15．75 | 8.42 | 67.8 |
| 01 | 3.14 | 7.94 | －17．43 | 8.54 | 6 \％． 4 |
| 02 | 4.04 | 8.82 | －15．24 | 9.70 | 65.4 |
| 03 | 3.73 | 6.51 | －15．4n | 7.50 | 6\％． 2 |
| 04 | 1.70 | 7.96 | －15．17 | 8.14 | 77.9 |
| PYM1 | ． 688 | 2.61 | －15．52 | 2.79 | 75.4 |
| PYCZ | －0． 34 | 3.50 | －15．53 | 3.52 | 95.6 |
| PYu2 | －1．53 | 3.15 | －15．5？ | 3.53 | 115.9 |
| PYM3 | －2．94 | 4.86 | －15．52 | 4.86 | 94．5 |
| PYC4 | 1.25 | 5.29 | －15．5？ | 5.14 | 76.8 |
| PYCS | 2.29 | 4.41 | －15．52 | 4.97 | 62.6 |
| TrMEE | 3.48 | 4． 57 | －15．52 | 5.74 | 52.7 |
| PyCG | 1.95 | 3.32 | －15．52 | 3.69 | 57.2 |
| Prisg | 2.91 | 2.11 | －15．52 | 3.59 | 35.9 |
| PYOG | 2.01 | 2.11 | $-15.52$ | 3.59 | 35.9 |

RESIDUE RIA

|  | X（A） | $Y(A)$ | 7 （A） | R（A） | PHI（DEG） |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $00^{\circ}$ | －1．45 | －6．41 | －14．85 | 6.57 | 257.2 |
| C1： | －1．68 | －5．24 | －15．64 | 5.53 | 252．2 |
| C2i | －2． 75 | －15． 13 | －17．29 | 5.87 | 249.3 |
| C． 3 ！ | －1．4．1 | －0．66 | －17．111 | 6.814 | 258．1 |
| C4！ | －1．53 | －7．49 | －15．819 | 7.56 | 258.3 |
| C．5 ${ }^{\circ}$ | －7． 7.25 | －8． 83 | －15．02 | 8.13 | 268.2 |
| $p$ | 2.29 | － 3.29 | －15．85 | 8.41 | 285.8 |
| 01 | 2.07 | －3．304 | －14．29 | 8．3．］ | 284．4 |
| 02 | 2.4 .1 | －9．59 | －15．1） | 9.89 | 284．2 |
| n3 | 3.46 | －7．32 | －10．3i | 3.10 | 205.3 |
| $\mathrm{n}_{4}$ | $\therefore$ ， 0 | －7．46 | －15．15 | 7.45 | 276.4 |
| Pudi | 1.13 | －1． 23 | －15．5？ | 9.20 | 299．4 |
| Puc？ | －1．11 | －i1． 72 | －15．57 | 1．3？ | 213.4 |
| 「びに | －2．？．i | 11.23 | －15．52 | 2.21 | 171.1 |
| 「いり3 | －1．8．9 | －1．84 | －15．52 | 2.58 | 225.6 |
| いUC4 | －${ }^{-1.45}$ | －2．0．1 | －15．52 | 2.34 | 264.9 |
| יucs | $\because .87$ | －2．46 | $-15.53$ | 2.61 | 239.5 |
| pueg | 1.16 | －1．113 | －15．52 | 1.59 | 317.1 |
| ＂」゙に | 2.32 | － 1.52 | －15．53 | 2.38 | 347.3 |
| Fuge | ？． 37 | －1．5？ | －15．43 | 2.34 | 347.3 |
| いJ゚\％ | 1.68 | －3．58 | －15．53 | 3.35 | 295.1 |
| Puca | $\because \cdot 8.3$ | －1．57 | $-15.53$ | 4.05 | 239.3 |
| ［J！49 | －1．43 | －1．17 | －15．52 | 4．2） | 263.5 |

## Table 4.4 ; N3-N9 Separation in the Published Model

| Residue | Separation (A) |
| :---: | :---: |
| 6 | 9.05 |
| 7 | 9.07 |
| 8 | 9.22 |
| 9 | 9.06 |
| 10 | 9.15 |

Table 4.5 : Comparison of the Published and Calculated N3 and N9 Co-ordinates

| Residue | Calculated |  |  | Published |  |  | Discrepancy <br> $(A)$ |
| :--- | ---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $r$ | $\phi$ | $z$ | $r$ | $\phi$ | $z$ |  |
| A6 | 4.52 | 103.3 | -1.70 | 4.55 | 103.0 | -1.70 | 0.04 |
| B6 | 4.52 | 283.3 | -1.70 | 4.55 | 283.0 | -1.70 | 0.04 |
| A7 | 4.19 | 135.8 | -5.23 | 4.20 | 135.4 | -5.20 | 0.04 |
| B7 | 4.90 | 305.5 | -5.23 | 4.91 | 305.1 | -5.25 | 0.04 |
| A8 | 4.02 | 162.0 | -8.65 | 4.11 | 161.4 | -8.50 | 0.18 |
| B8 | 5.08 | 329.0 | -8.65 | 5.16 | 328.8 | -8.80 | 0.17 |
| A9 | 4.30 | 125.2 | -12.10 | 4.31 | 124.8 | -12.00 | 0.10 |
| B9 | 4.75 | 300.6 | -12.10 | 4.76 | 300.1 | -12.20 | 0.11 |
| A10 | 4.86 | 90.5 | -15.53 | 4.92 | 90.1 | -15.50 | 0.08 |
| B10 | 4.20 | 263.5 | -15.53 | 4.25 | 263.1 | -15.55 | 0.06 |

single crystals of nucleotides and nucleosides. We now discuss this contention and consider whether the tolerances allowed by Rodley et al are acceptable.

The covalent bond lengths and angles of SBS 36 are shown in figure 4 and those imposed in double helical models are shown in figure 5 for comparison. It is immediately clear that there is a wide variation in the covalent parameters of the side-by-side model. Bond lengths differing by about 0.1A from the accepted value are common and in some cases the discrepancy is around 0.3 A (B6C4-B6C5, 0.29A; A9C3-A801, 029A; A10C2-A10C3, 0.33A). Bond angles generally differ from the accepted value by about $8^{\circ}$ and some discrepancies are as high as $18^{\circ}$ (A8N-A8C1-A805 and A1005-A10C4-A10C5). The worst discrepancies are not localised in the bend-regions where one might expect them. Indeed some are in the right-handed helices. Examination of the endocyclic and exocyclic sugar rings suggests that the puckers may differ considerably from the standard ones.

Fixed sugar puckers are incorporated into the refinement of double-helical models in order to reduce the number of variable parameters and therefore to improve convergence. Energy barriers between the different puckers are small so variation of the puckers in the side-by-side model may not be serious. However, the wide variations in the covalent bond lengths and angles are unacceptable. Arnott and Hukins (1972a) have found that the largest standard deviation observed in the bond lengths of furanose rings is 0.023 Ahich is an order of magnitude smaller than the worst discrepancies in SBS 36 . The shortest observed C1-C2 bond length for a C3'-endo sugar is 1.50 A and the mean value is 1.525 A with estimated standard deviation 0.017A, All the C1-C2 bonds in SBS 36 are shorter than 1.48A. One C2-C3 bond length of 1.46 A has been qbserved but it is quite distinct from the lengths of all other C2-C3 bonds the average value of which is 1.528 A with standard deviation 0.019 A . All but one of these
bands in SBS 36 are less than 1.48A, The smallest length of any observed C3-C4 bond is 1.49A and the mean value is 1.529A with standard deviation 0.021A. Most of these bonds in SBS 36 are no greater than 1.47A. The smallest observed C4-05 bond is 1.44 A , the mean value is 1.457 A and the standard deviation is 0.020 A . In contrast, most of these bonds are about 1.42A in the side-by-side model with one as short as 1.23A. The majority of the 05-C1 bond lengths in SBS 36 agree well with the observed values but two are 1.51A and 154A which is significantly greater than the largest observed value of 1.45A. Comparison of figures 4 and 5 shows that the exocyclic bond lengths are in no better agreement with experimental results. The observed bond angles, which tend to cluster within about $5^{\circ}$ of each other, are also at variance with those in the SBS model.

Table 8, showing the backbone torsion angles of SBS 36 , illustrates the irregularity within the molecule. It is rather surprising that the standard double-helical B-DNA conformation was not imposed in the right-handed regions. The choice of the C3'-endo sugar pucker probably encourages the bases to adopt a higher displacement than is observed in B-DNA models but in SBS 36 the bases were constrained to have a low displacement so the backbone contains a number of torsion angles close to the eclipsed conformation. It is possible that C3'-exo sugar puckering would lead to a more satisfactory conformation in the right-handed regions. Indeed Millane and Rodley (1981) have changed the pucker at B10 to C2'-endo. The modelbuilding studies described in the previous chapter indicate that left-handed backbones with acceptable stereochemistry are possible however no refinement has been carried out on the bend regions. Future work on the conformation within bends would be usefur.

Rodley et al (1976) stated that their model contained only four contacts less than 2.7A between non-bonded atoms. Table 6 shows the worst contacts found in the present work with those mentioned by Rodley et al underlined. This table also shows the equilibrium separation for
these contacts using accepted values (see for example Campbell-Smith and Apnott, 1978). It is clear that a large number of bad contacts occur, some of them quite unacceptable. Bates et al (1977) pointed out that the omission of 02 and 03 atomic co-ordinates from their original publication lead them to overlook some of these contacts, however, the table shows that many more exist, It must be admitted that several of these involve base atoms whose positions were derived by the author. The analysis is section 4.2 suggests that the bases are fairly accurately placed but it is conceivable that by varying the base twists, tilts and displacements many of the contacts could be relieved. Nonetheless the stereochemistry of the sugar-phosphate chain is also unsatisfactory. This is in contrast to the claims of Rodley et al since they chose to use the arbitrary value 2,7A (above which contacts were not considered) which is unsatisfactory when considering contacts between phosphorous and carbon atoms for example.

Table 7 shows the worst intermolecular contacts which occur when SBS 36 is packed into the unit cell in the same manner as WatsonCrick B-DNA (Langridge et al, 1960a) i.e. the relative displacement between the molecule at the centre of the cell and those at the corners is 11.02A and the molecular diad is oriented along the b-axis. The $A$ - and B-chains are then arranged as shown in figure 6. Each of the ten residues referred to in columns 1 and 3 of table 7 contains the twenty nucleotides Al-A10 and B1-B10, appropriately oriented and translated, to give one complete pitch of SBS 36. The residue number increases with $z$. If one complete pitch is taken as the repeating unft then PSHIFT $=0$ when a contact is between units in the same unit cell. PSHIFT $=-1$ represents a contact between a unit at the corner of the cell and one in the centre of the cell below. PSHIFT $=1$ represents a contact with a centre unit in the cell above.

In this orientation many short contacts occur, some shorter than 1A. The details of intermolecular stereochemistry in other orientations



Figure 4,4a: Stereochemisttry of SBS36 residtue (6)







Fiqure 4.5 : stereochemistiry of the sugar-phosphate chain in dioubble melical modeds. Untinacketed ficgures corresppond to celt-endo sugian puachering and the bracketed figgures corresppond to

C2"-endo sugars


Ejgurs 4．6：Projection down the $\underline{c}$－axis of the SBS36 unit cell．The molecule in the centre is displaced by（33．8／3）A out of the plane of 销e paper：

Table 4.6 : Intramolecular Contacts in SBS 36

The contacts underlined are those mentioned by Rodley et al (1976).

| Atoms involved | Separation <br> (A) | Equilibrium <br> Separation (A) |  |
| :--- | :--- | :---: | :---: |
| Bl0C3 | Bl004 | $\frac{2.58}{2.30}$ | $\frac{3.00}{3.00}$ |
| A1005 | A10PYC4 | 3.17 | 3.50 |
| A10P | Al0PYC4 | 2.71 | 3.00 |
| Bl0C5 | B1001 | 2.66 | 3.20 |
| B10PUC8 | B9C2 | 2.77 | 3.00 |
| A10PYC4 | A1003 | 2.74 | 3.00 |
| Al0PYC4 | A1004 | 2.96 | 3.20 |
| A10PYC4 | A10C2 | 2.70 | 3.00 |
| A1004 | A10C3 | 2.90 | 3.50 |
| A10C2 | A9P | 3.28 | 3.60 |
| Al0C2 | A9THME | 2.77 | 3.00 |
| B9C3 | B904 | 2.43 | 3.00 |
| A905 | A9PYC4 | 2.54 | 3.00 |
| A802 | A9C4 | 2.69 | 3.00 |
| B904 | B9PUC8 | 3.04 | 3.50 |
| A9C4 | A8P | 2.81 | 3.50 |
| B9P | B8C2 | 2.59 | 3.20 |
| B9PUC8 | B8C2 | 2.64 | 3.00 |
| A9PYC4 | A904 | 2.78 | 3.20 |
| A9PYC4 | A9C2 | 2.79 | 3.00 |
| A904 | A9C3 | 3.23 | 3.50 |
| A8P | A9C2 | 2.94 | 3.20 |
| A9C2 | A8PYC4 | 2.83 | 3.60 |
| A9C2 | A8THME | 2.80 |  |
| A801 | A704 | 2.65 | 2.80 |
| A805 | A8PYC4 | 2.62 | 3.20 |
| A8PYC4 | A8C3 | 2.76 |  |

Cont.

| Atoms involved |  | Separation (A) | Equ†librium Separation (A) |
| :---: | :---: | :---: | :---: |
| A8PYC4 | A8C2 | 2.77 | 3.20 |
| A8PYC4 | A705 | 2,74 | 3.00 |
| A8PYN3 | A8C3 | 2.83 | 3.10 |
| A8THME | A7C5 | 3.11 | 3.60 |
| A8C3 | A704 | $\underline{2.63}$ | 3.00 |
| B8C4 | B8P | 3.00 | 3.50 |
| A7P | A7C4 | 2.68 | 3.50 |
| A8C2 | A705 | 2.58 | 3.00 |
| A7C5 | A702 | 2.71 | 3.00 |
| A702 | A7C4 | 2.30 | 3.00 |
| B7C2 | B7PUC8 | 2.99 | 3.20 |
| B8P | B7C4 | 3.22 | 3.50 |
| B8P | B7C5 | 3.29 | 3.50 |
| B7C3 | B7Puc8 | 2.73 | 3.20 |
| B7C3 | B7PuN9 | 2.72 | 3.10 |
| A7PYC4 | A7C3 | 2.51 | 3.20 |
| A7PYC4 | A7C2 | 2.75 | 3.20 |
| A7PYN3 | A7C3 | 2,73 | 3.10 |
| A7THME | A604 | 3.13 | 3.40 |
| A7THME | A6C5 | 3.09 | 3.60 |
| A6P | A6C4 | 2,82 | 3.30 |
| B803 | B7C5 | 2.49 | 3.00 |
| B7C4 | B7P | 3.07 | 3.50 |
| A602 | A6C4 | 2.66 | 3.00 |
| B7C5 | B702 | 2.27 | 3.00 |
| B704 | B6C3 | 2.75 | 3.00 |
| B7P | B6C5 | 3.14 | 3.50 |
| B6C3 | B6PUC8 | 2.91 | 3.20 |
| A6PYC4 | A6C3 | 2.75 | 3.20 |
| A6PYC4 | A6C2 | 2.80 | 3.20 |
| A6PYN3 | A6C3 | 2.86 | 3.10 |
| B6C4 | B602 | 2,51 | 3.00 |

Cont.

| Atoms involved | Separation <br> (A) | Equilibrium <br> Separation (A) |  |
| :--- | :--- | :---: | :---: |
| B6C4 | B6P | 2.59 | 3.50 |
| B703 | B6C5 | 2.58 | 3.00 |
| B605 | B6P | 3.04 | 3.30 |
| B605 | B604 | 2.49 | 2.80 |
| B602 | B6C5 | 2.73 | 3.00 |

Table 4.7a - The Worst Interhelical Contacts in SBS 36

> Molecular orientation : $\phi_{1}=\phi_{2}=90^{\circ}$
> Molecular displacement : $\Delta z=c / 3$

| Residue | Atom | Residue | Atom | Distance ( A ) | PSHIFT |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | B7P | 1 | A8P | 2.4 | 0 |
| 1 | B7P | 1 | A803 | 1.8 | 0 |
| 1 | B701 | 1 | A803 | 2.2 | 0 |
| 1 | B702 | 1 | A9C5 | 1.3 | 0 |
| 1 | B702 | 1 | A904 | 2.5 | 0 |
| 1 | B703 | 1 | A8P | 1.6 | 0 |
| 1 | B703 | 1 | A801 | 1.5 | 0 |
| 1 | B703 | 1 | A803 | 1.2 | 0 |
| 1 | B9P | 10 | A703 | 1.9 | -1 |
| 1 | B902 | 10 | A703 | 0.7 | -1 |
| 2 | B6P | 2 | A902 | 2.1 | 0 |
| 2 | B601 | 2 | A902 | 2.5 | 0 |
| 2 | B602 | 2 | A9P | 2.3 | 0 |
| 2 | B602 | 2 | A902 | 1.3 | 0 |
| 2 | B2P | 2 | A602 | 12.3 | 0 |
| 2 | B202 | 2 | A6P | 2.2 | 0 |
| 2 | B202 | 2 | A401 | 2.5 | 0 |
| 2 | A902 | 2 | A502 | 1.4 | 0 |
| 3 | B3P | 3 | B4P | 2.4 | 0 |
| 3 | B3P | 3 | A403 | 1.6 | 0 |
| 3 | B301 | 3 | A403 | 1.5 | 0 |
| 3 | B303 | 3 | A4P | 1.8 | 0 |
| 3 | B303 | 3 | A401 | 2.2 | 0 |
| 3 | B303 | 3 | A403 | 1.2 | 0 |
| 3 | B2C5 | 3 | A402 | 1.3 | 0 |
| 3 | B204 | 3 | A402 | 2.5 | 0 |
| 4 | B403 | 3 | A2P | 1.9 | 0 |
| 4 | B403 | 3 | A202 | 0.7 | 0 |
| 5 | A801 | 4 | A202 | 2.3 | 0 |

Cont.

## Table 4.7a (Cont.)

| Res idue | Atom | Residue | Atom | Distance (A) | PSHIFT |
| :---: | :---: | :---: | :--- | :---: | :--- |
| 5 | A803 | 4 | A2P | 1.8 | 0 |
| 5 | A803 | 4 | A201 | 1.8 | 0 |
| 5 | A803 | 4 | A202 | 1.8 | 0 |
| 5 | A803 | 4 | A203 | 2.1 | 0 |
| 6 | A6C4 | 6 | B902 | 2.5 | 0 |
| 6 | A602 | 6 | B902 | 1.7 | 0 |
| 6 | A702 | 6 | Bl002 | 1.9 | 0 |
| 6 | A703 | 6 | B7P | 2.4 | 0 |
| 6 | A703 | 6 | B703 | 1.9 | 0 |

Table 4.7b - Interhelical Contacts in B-DNA

| Residue | Atom | Residue | Atom | Separation (A) | PSHIFT |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | B-C4 ${ }^{\prime}$ | 7 | B-02 | 3.7 | 1 |
| 1 | B-02 | 8 | B-02 | 2.9 | 1 |
| 1 | B-02 | 8 | B-P | 3.9 | 1 |
| 1 | B-C31 | 8 | B-03 | 3.4 | 1 |
| 2 | B-01 | 8 | B-03 | 3.3 | 1 |
| 2 | B-02 | 8 | B-03 | 3.8 | 1 |
| 2 | B-02 | 8 | B-THME | 3.9 | 1 |
| 2 | B-03 | 8 | B-03 | 3.8 | 1 |
| 2 | B-P | 8 | B-03 | 3.8 | 1 |
| 6 | A-03 | 2 | A-01 | 3.3 | 0 |
| 6 | A-03 | 2 | A-02 | 3.8 | 0 |
| 6 | A-03 | 2 | A-03 | 3.8 | 0 |
| 6 | A-03 | 2 | A-P | 3.8 | 0 |
| 6 | A-THME | 2 | A-02 | 3.9 | 0 |
| 6 | A-02 | 3 | A-02 | 2,9 | 0 |
| 6 | A-03 | 3 | A-C3' | 3.4 | 0 |
| 6 | A-P | 3 | A-02 | 3.9 | 0 |
| 7 | A-02 | 3 | A-C41 | 3.6 | 0 |
| 7 | A-01 | 6 | B-C41 | 3.9 | 0 |
| 8 | A-02 | 6 | B-02 | 3.0 | 0 |
| 8 | A-C4: | 7 | B-01 | 3.9 | 0 |

The relattye positions of the molecules at the centre and corner of the cell are the same as for part (a).

Table 4.8 : Backbone torsion angles of SBS36

|  | $\alpha$ | $B$ | $\gamma$ | $\delta$ | $\varepsilon$ | $\zeta$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $A 6$ | - | - | -9.0 | -140.7 | 33.0 | - |
| $A 7$ | -162.7 | -81.3 | -15.4 | 166.7 | 51.4 | 77.2 |
| $A 8$ | 28.1 | 116.5 | -86.2 | 62.0 | -49.1 | 87.9 |
| $A 9$ | -110.9 | -16.3 | -106.8 | 37.0 | -179.0 | 128.0 |
| A10 | -94.0 | -40.1 | - | - | - | 84.8 |
| B6 | - | - | -96.5 | 44.8 | 177.4 | 79.0 |
| B7 | 90.4 | -49.1 | 87.1 | 60.1 | 171.5 | 69.6 |
| B8 | -84.1 | -59.5 | -81.2 | 61.2 | 179.3 | 86.9 |
| B9 | 171.4 | -82.3 | -50.0 | 162.0 | 59.0 | 86.9 |
| B10 | -154.0 | -99.1 | -55.8 | 177.6 | 28.4 | - |

are not given but no arrangement of the molecules gave significantly more satisfactory packing, For comparison the worst contacts between double-helical B-DNA molecules in the same orientation are also shown in table 7, It ts clear that there is little difficulty in accommodating a double-helix in the unit cell whereas the packing of SBS 36 is unacceptable.

The results of the present section have shown that both the covalent and non-bonded stereochemistry of SBS 36 are much worse than in double-helitcal B-DNA and on these grounds alone the model of Rodley et al (1977) could be rejected, However, it would be dangerous to claim that the side-by-side hypothesis is thereby refuted since it may be possible to refine Rodley's co-ordinates using, for example, the method of Levitt (1978) to obtain a model with standard bond lengths and angles and better intermolecular contacts.

### 4.4.5 The CPK Model of SBS 36

CPK components accurately represent the space occupied by atoms in a polynucleotide so if such a model of SBS 36 can be built it suggests that it is not a stereochemically impossible conformation. However, CPK models do not allow us to decide, for example, whether repulsion between adjacent phosphate groups is compensated by more efficient base-stacking.

Plates la-c show three views of the model built by the author. As described previously, the bases were all constrained to stack on each other with the same displacement from the helix axis. As a result the amount of variation possible in the sugar-phosphate chain is very small. Considerable difficulty was experienced in obtaining a satisfactory conformation in the backbone, especłally in the bend regions. A large amount of strain is indicated in many of the bonds by the difficulty in joining the atoms involved, It is true, as suggested by Sasisekharan
et al (1978) that short contacts are present in this type I model between $\mathrm{C}^{\prime}$ ' and purine C 8 or pyrimidine C 6 in the left-handed regions. However these may be removed by adjusting the angle $x$ slightly and there is no need to rotate the sugans to the type II conformation as they suggested. We found no contacts between base atoms and C3' in cqntrast to the claim of Sasisekharan et al. Although not correct in detail, this model is a good representation of Rodley's structure. It is probable that an acceptable conformation with approximately $36^{\circ}$ long-range twist may be defined, particularly if the bases were given more freedom to move, however in many cases this leads to an increase in the radial position of the phosphate groups which suggests that the intermolecular stereochemistry in the unit cell would be worsened. We also twisted the bases to remove the long-range turn of $36^{\circ}$ in order to see whether a model with no nett twist was feasible. Under these circumstances the sugar-phosphate chain was only satisfactory if the bases unstacked by 1-2A which suggests that no model with zero twist may be built.

Several comments about the structure are in order. First the bases are not stacked efficiently as claimed by Rodley et al (1976). They suggested that slightly unsatisfactory stereochemistry within the backbone could be compensated by the bases stacking directly upon each other rather than being rotated by $36^{\circ}$ with respect to each other in B-DNA. In fact this stacking does not generally occur. In both the left- and right-handed regions the angle between successive base-pairs is about $30^{\circ}-40^{\circ}$ but in the bend region (where the molecule is changing direction) the angle is only about $10^{0}$. It is unlikely therefore the base energy in SBS 36 is any more favourable than that in double-helical B-DNA. When the model was distorted so that the angle between successive bases was close to zero, unstacking was once again necessary so such structures are also unlikely to be possible.


It is clear from the photographs that the sugar-phosphate chain can change from left- to right-handed very quickly. As a result phosphate groups in the bend regions are clustered closely together. Since these groups are charged it would be essential for them to be shielded from each other by $\mathrm{Na}^{+}$for example. The conformation of the backbone leads to a fairly narrow groove, corresponding to the narrow groove of B-DNA, along sịde which the phosphates are arranged. By contrast, the other side of a twenty nucleotide-pair section is relatively devoid of backbone : instead the base-pairs are exposed. This may be an attractive feature in proteịn recognition processes for example but it would appear to be undesirable in a general conformation of DNA. This is because the bases contain the genetic message which must be preserved with high fidelity. Under these circumstances it would presumably be safer to maintain the bases within the backbone as in B-DNA, where they are relatively protected from chemical attack.
4.4.6 $\frac{\text { X-Ray Diffraction from the Side-by-Side Model }}{\text { It has been suggested that the side-by-side model is capable }}$ of accounting for the X-ray diffraction patterns of B-DNA (Rodley et al, 1976). In this section we compare the cylindrically averaged intensity of both the original model (SBS 36) and the distorted model (SBSO) with that of B-DNA and we discuss the level of agreement between the diffraction from these models and the observed intensities. First we compare the calculated intensities with the crystalline data from the lithium salt of B-DNA and then criticise the suggestion that comparison with the diffraction patterns from less-well-ordered B-DNA specimens would be more appropriate (Bates et al, 1980).

Figure 7 shows the cylindrically averaged intensity transform of SBS 36. The ptt=h of SBS 36 is $345 A$ whereas that of B-DNA is an order of magnitude smaller (the precise value is 33.7 A but this small discrepancy
does not significantly affect the argument). Since the layer-plane spacing is inversely proportional to the pitch, it immediately follows that the SBS 36 planes will be separated by $1 / 345 A$, approximately ten times smaller than the spacing observed in diffraction patterns from the lithłum salt (Langridge et al, 1960a), The only possibility for removing this major discrepancy would lie in the extremely unlikely event of the SBS 36 model having a Fourfer transform on all layer-planes for which \& $\neq 10 p$ ( $p$ an integer, including zero) which was accidentally zero. It can be seen from figure 7 that such a fortuitous occurrence is not the case for the published model co-ordinates. A particularly serious deficiency of the SBS 36 model is the occurrence of substantial diffraction on layer-planes which are neither observed nor predicted by models of the Watson-Crick type, i.e. on $\ell=8,12,18,22,26,78$ and 88 in figure 7. There is no doubt that the quality of the $X$-ray diffraction patterns from crystalline fibres of LiDNA is sufficient for the SBS 36 model to be eliminated on the basis of the non-observation of these layer-planes alone. But even if it were not possible to detect such layer-line splitting (as a result of layer-line breadth or disorientation, for example) the model could still be rejected since it fails to predict the observed relative intensities. In particular, the diffraction predicted to occur at $R<0.1 A^{-1}$ and near $Z=(10 / 345) A^{-1}$ and $(20 / 345) A^{-1}$ consists of peaks of equal magnitude whereas the observed diffraction in this region (corresponding to $\ell=1$ and $\ell=2$ in $B-$ DNA) shows a strong peak on $\ell=2$ and a very weak one on $\ell=1$. In addition the SBS 36 model predicts little diffraction at $R \simeq 0.1 A^{-1}$ and $Z=(30 / 245) A^{-1}$ where substantial diffraction is observed (corresponding to $\ell=3$ in B-DNA). Nor does it predict diffraction corresponding to that observed on $\ell=5,6$ and 8. The \$BS 36 model also predicts substantial meridional intensity on $\ell=60,70$ and 90 (corresponding to $\ell=6,7$ and 9 in the B-DNA nomendature). This is also at variance with the observed diffraction.

Figure 8 shows the cylindrically averaged intensity transform of SBSO. This model, which has the biologically attractive feature that the sugar-phosphate strands are not intertwined, was contrived to predict intensity only on those layer-planes actually observed. The transform Qf B-DNA (calculated usting the co-ordinates of Arnott and Hukins (1972b)) has been superimposed for comparison. The magnitudes of the relative intensities predicted by this model are seriously in error. These discrepancies are particularly acute on $\ell=2,3$ and 8 where the overall calculated intensity is much less than that observed and on $\ell=4$ for which substantial intensity is predicted where the observed diffraction is essentially zero. In addition the positions of the peaks in the transforms are not in agreement with observation. The diffraction along the meridion in these patterns corresponds to that from a projection of the molecule onto the helix axis, Since the distortion of the SBSO model from SBS 36 involved no change in the atomic $z$ co-ordinates, meridional intensity is still predicted on $\ell=6,7$ and 9.

Figure 9 shows the level of agreement which was obtained between the observed and calculated diffraction in the initial refinement of the Watson-Crick model for LiDNA (Langridge et al, 1960b). Whilst there are significant differences between the transforms of the two Watson and Crick models, the differences are not large and for both models the calculated diffraction is similar in position and intensity to that observed. Even the most superficial examination of figures 7, 8 and 9 shows that the level of agreement achieved by the SBS models is markedly inferior to that achieved by the Watson-Crick models.

Bates et al (1980) and Bates, Rodley and McKinnon (1980) have rejected the analysts given above on the grounds that fibre intensities cannot be placed on an absolute scale. They suggest that the transforms should be normalised such that their peak values on the second layerplane are the same. It is true that determining the scale of fibre





Figure 4.8 : Comppalison of the cylimdricallyy averaged squared Fourrier tramsfform of SBSO (solich curve) with that of the MatsonCrick moattel descrilbed by Arnott and Hukkins (brolken curve)

Figure 4.9 : Comparison of the cylindrically averaged squared Fourier transform of the Langridge et al model of B-DNA with the observed intensities. Observed intensities in order of decreasing reliability are indicated by 0 and $: 2$

patterns is difficult and no attempt was made by Greenall, Pigram and Fuller (1979) to claim otherwise. Instead, the relative sizes of the peaks in the calculated transform were compared with the relative sizes of the peaks observed, Normalising the patterns according to the diffraction on any one layer-plane is open to objections - instead it would be preferable to scale the data such that the sum of the observed intensities is equal to the sum of the calculated - however, even if this procedure is accepted, it does not affect the conclusions reached earlier.

Bates and McKinnon (1978) have published the cylindrically averaged intensity transform of an SBS model. The only co-ordinates presented are those of the phosphorous atoms so it is impossible to repeat their calculations, however comparison of these co-ordinates with those of Rodley et al (1976) suggests that the models are different. In general the positions and magnitudes of the peaks are in quite close accord with those predicted by the double-helical model. The major exception is on $\ell=1$ where their model predicts very high intensity. A remarkable assertion (which is reiterated by Bates et al (1980)) is that the transform changed very little whether or not the scattering factors were corrected for the effect of water. Wilkins and co-workers recognised at an early stage in their studies that failure to take account of the water resulted in the lower layer-planes being too intense relative to the higher ones, and they developed the correction technique (Langridge et al, 1960b) discussed in Chapter 3. Fuller (1961) has presented a comparison of the corrected and uncorrected transforms of A-DNA which
 of Bates and McKinnon's calculation. However, if the agreement between the observed and calculated diffraction of this model is as close as the authars suggest, then it is clearly essential that the full co-ordinate set be published in order to allow closer scrutiny by other workers.

Arnott (1979) claims that the R-factor of the side-by-side model published by Rodley et al (1976) is $48 \%$ whereas that of B-DNA is 28\%. It is not clear that this comparison is meaningful. First, calculation of $R$ requires that the predicted structure factors be known, however no model for the packing of SBS DNA into the crystalline unit cell has been proposed and Arnott gives no details of the relative orientation and translation of the molecules which he assumed. Second he does not state whether the long-range twist was preserved in his calculations. If so then it was necessary to renumber the $\ell$ index in the observed structure factors to be consistent with the longer c-period of the SBS 36 model. However the calculation would still not be valid since it compares only the observed structure factors with the values predicted at those points. Figure 7 shows that the most significant peaks in the molecular transform do not co-incide with the observed lattice points (Langridge et al, (1960a); Arnott and Hukins (1973)) so they would not be included in the R-factor calculation. There is no doubt that the clearest way forward in choosing between the double-helical and SBS models is to compare the full molecular transforms as we have done here.

We now consider whether the crystalline data used above in comparing the observed and calculated diffraction is indeed the best to use. First it is necessary to clarify the nomenclature we will employ. The term "crystalline" has been applied to those specimens which contain small crystallites wherein the DNA molecules are arranged with three-dimensional order. These crystallites have random azimuthal orientation with respect to the fibre axis as discussed in Chapter 2. Bates and co-workers have consistently referred to such specimens as "paracrystalline". This imprecise term is no longer in general use in nucletc acid crystallography and we will retain the normal term. The same workers also refer to "fibre" specimens. It is clear from their papers that this refers to specimens consisting of DNA molecules randomly oriented with respect to the fibre
axis and randomly translated along the fibre axis, i,e, an array of molecules with both rotational and slippage disorder according to the classification in Chapter 2,

Rodley et al (1976) daim that their model is capable of accounting for the $X$-ray data from B-DNA. The B conformation has been observed in crystalline fibres of the lithium salt of DNA, semicrystalline fibres of the sodium salt and, with less detailed characterisation of the conformation, in gels and poorly oriented fibres of NaDNA and in complexes of DNA with protein. The X-ray data from crystalline fibres are particularly extensive for a fibrous structure, with sharp reflections extending to spacings as low as 3.3 A and, in the meridional directional, less well-defined data to spacings of 1.lA (Wilkins, 1961). Therefore the diffraction intensities place rigorous constraints on the detailed molecular conformation. The observed diffraction from semicrystalline fibres of NaDNA has been interpreted in terms of scattering from a hexagonal array of helical molecules exhibiting screw disorder, i.e. packed as if they were completely smooth helices (Langridge et al, 1960a). In particular this accounts for the presence of sharp spots in the centre of the pattern with continuous scattering on layer-planes elsewhere. After allowance has been made for the effects of molecular packing, the X-ray scattering from a molecule of DNA in a semicrystalline fibre of NaDNA is, within the limits of the data from these fibres, essentially identical to that from a DNA molecule in a crystalline fibre of LiDNA, indicating a very similar if not identical nolecular structure. The diffraction from fịbes of DNA which, while exhibiting molecular orientation, show little or no crystallinity can also be accounted for by assuming that the DNA has a conformation of the $B$ type. The data from these specimens is necessartly less extensịve than that from crystalline and semicrystalline fibres and it may be that there are small but nevertheless significant differences from the $B$ model determined by analysis of the crystalline
fibres. However, while it ís perfectly reasonable to postulate that the molecular conformation of the B form of DNA differs slightly in specimens with differing degrees of order, it must also be emphasised that these various types of specimen can all be made from the same sample of DNA. It would therefore be rather implausible to propose that despite their very similar X-ray diffraction patterns there are major differences in the molecular conformation associated with these different types of packing, e.g. a Watson-Crick model accounting for data from crystalline fibres and a side-by-side model with its major differences in helix sense accounting for that from concentrated gels. There is therefore a compelling requirement that if the SBS model is to have any significance for the structure of extracted DNA, it, or a simple variant of $i t$, should be able to account for diffraction from crystalline fibres of DNA.

Bates et al (1980) have argued that (even should we accept that very similar DNA conformations are present in each type of specimen giving the $B$ pattern) adjudication between the double-helical and SBS models should nonetheless be based on "fibre" and not crystalline data since they believe that the former give rise to smaller errors. This argument is based on a false premise. Since their analysis is presented in a source which is not readily accessible, it is reproduced in full in the appendix to this chapter. The assumptions made by Bates et al are open to question and indeed it is doubtful whether their approach is sensible simply because it considers the errors in the amplitudes whereas the quantities observed are the intensities, However a more fundamental refutation is available. Equations (A5) and (A9) compare the error in the amplitude at a point in reciprocal space from a single oriented molecule with that from a cylindrically averaged specimen, Now even the most disordered fibre specimen is unlikely to contain no intermolecular interference effects (for instance on the equator) but, more important, crystalline specimens of DNA consist not of a single oriented molecule but of an ordered array of oriented molecules.

The effect of the three-dimensional ordering is to produce a local amplification of the diffracted signal at the lattice points in reciprocal space. There is no doubt that this enhancement of the signal-to-noise ratio ensures that intensities from crystalline samples may be measured far more accurately than those from fibrous specimens. The analysis of Bates et al is therefore false simply because their calculation of the discrepancy between the true and observed amplitude of diffraction from crystalline flbres is not a correct model of the physics of the process as a result of their neglect of the sampling effect.

The same misconception is also apparent in a number of other statements made by Bates and co-workers. For example, Rodley et al (1976) state that the diffraction pattern from a fịbrous specimen is "roughly proportional to the intensity of the diffraction pattern of a single molecule averaged by rotation". In fact it is very difficult to conceive of any solid state sample of DNA which would not exhibit some degree of partial ordering thus leading to intermolecular interference effects in the diffraction pattern. In contrast to the claims by Rodley et al, these effects are significant. Indeed, even when measuring the Gaussian portion of the scattering profile of particles in solution it is extremely important to ensure that the concentration is low otherwise interparticle effects distort the curve (Guinier and Fournet, 1955). For the reasons given in the previous paragraph, Bates and co-workers (Bates and McKinnon, 1978; Bates, Rodley and McKinnon, 1980; Bates et al, 1980) have chosen to compare the diffraction pattern predicted by their model with data from fibrous specimens obtained by Bram (unpublished), Zimmerman and Pheiffer (1979) and Feughelman et al (1955). The implication by Bates et al that errors of measurement are the sple reason for the differences between the intensity distributions observed in these patterns is unlikely to be correct. Zimmerman and Phetffer have suggested that the molecules within their specimens have random shift and rotation relative to each other giving a
diffraction pattern which consists of continuous transform on all layer-lines except the equator. But Feughelman et al (1955) refer to diffraction patterns containing as many as sixteen reflections indicating a relatively high degree of order within the fibre resulting in sampling of the molecular transform. Before one can compare the transforms of molecules in such specimens it is necessary to correct for this effect but to achieve this one needs to propose a model for the partial order. In view of the low quality of the data and the large number of types and degrees of partial ordering, it is extremely unlikely that this can be done with accuracy. Therefore one is forced once again to the conclusion that crystalline spectmens should be used since their high degree of order enables one to obtain the molecular transform relatively easily.

Rodley et al (1976) and Bates et al (1977) suggest that the only requirement placed on $B-D N A$ models by the diffraction data is that they be capable of predicting the cross-shape which led Watson and Crick to propose their double-helical model. They have stated that diffraction data which was sufficient to establish the double-helix hypothesis in the 1950's should also be a sufficient test of the side-by-side hypothesis. The pattern obtained by Franklin and Gosling (1953a) was obviously important since it apparently suggested to Watson and Crick that DNA was a helical molecule (Watson, 1968). They then devised their model, largely by building wire models once the base-pairing scheme had been discovered, to agree with the gross helical parameters suggested by the diffraction pattern. However, patterns of this kind were not the sole X-ray evidence upon which later checking and refinement of the model was based, Langridge et al (1960a, b) have given a very detailed account of their use of crystalline DNA samples in this process. Nor is it true to suggest, as Bates and co-workers have done, that the helical hypothesis was accepted immediately and no alternatives were considered once cross-shaped patterns were obtained. It was no doubt apparent to the crystallographers of the
time that other structures could give rise to crosss-shaped patterns. Examples are gi̇ven by Harburn, Taylor and Welberry (1975) of the optical transforms of a number of such structures. It was precisely because a wide range of models could possibly account for the two-dimensional fibre patterns that it was essential that three-dimensional data from crystalline samples were used to support the double-helix. It is worth noting that the angle of the cross predtcted by SBSO , is incorrect whilst SBS 36, the model proposed by Rodley et al, does not predict a cross at all (Figures 7 and 8).

Bates et al (1980) have referred to two features on the tenth layer-plane of crystalline DNA samples in terms which suggest they are seriously at variance with the double-helix hypothesis. The first concerns the indexing of reflections on this layer-plane and the second concerns the agreement between the observed and calculated intensities near the meridion. As discussed in Chapter 3, Langridge et al (1960a) and Arno:t and Huk ns (1973) differ in the indices they have assigned to the $\ell=10$ reflections, Arnott and Hukins have assigned the indices ( $1,0,10$ ), $(0,1,10)$ and $(1,1,10)$. These spots, which all have similar $\rho$-values, tend to be arced in even tie best diffraction patterns and therefore it. is difficult to be certain about the indexing. However, they all sample the molecular transform at a point in reciprocal space where it is extremely intense, so no serious discrepancy between the observed and calculated diffraction will occur. Although the indexing may have implications for the refinement of double-helical models, it is of little importance in choosing between the Watson-Crick and side-by-side hypotheses.

The calculated diffraction from double helical B-DNA indicates that the most intense paint on $\ell=10$ occurs at $R \simeq 0.07 A^{-1}$ and not on the meridion (figure 8), Examination of the transforms of the phosphate and base shows that this occurs not simply because the bases are tilted as suggested by Bates et al (1980) but because these groups are scattering
with opposite phases at the meridion and so cancellation reduces the total transform on the axis. As a consequence, the highest amplitude on this plane occurs at $R \simeq Q .07 A^{-1}$ where the phosphate and base are scattering in phase, By contrast, the transforms of the phosphates and bases are in phase at $R=O A^{-1}$ in the $\$ B S$ model (figure 10) and so the highest intensity occurs on the meridion (figure 7). Bates et al (1980) have pointed out that the SBS transform appears to be in better agreement with the observed diffraction. However this is incorrect : the precise shape of the curve is unimportant since it cannot be determined experimentally with any certainty. To measure the profile of such an intense reflection as the tenth layer-plane meridional would in any case present formidable difficulties but more fundamental problems arise from the Lorentz correction. The Ewald construction (Chapter 2) shows that a reflection can only be observed when the corresponding reciprocal lattice point cuts the sphere of reflection. The reciprocal lattice point of a true meridional reflection lies on the axis of rotation in reciprocal space, therefore if this point cuts the sphere of reflection at all (either by specifically tilting the fibre to observe the reflection or by misorientation of crystallites within the fibre) then it must do so for the duration of exposure. Clearly the Lorentz correction, which measures the relative speed with which lattice points cut the sphere, is indeterminate under these circumstances so it is impossible to place the meridional intensity on the same scale as the rest of the pattern. We conclude that the tenth layer-plane is not in disagreement with the double-helix hypothesis. Indeed the difficulties outlined above emphasise that more attention should be paid to the rest of the pattern where the data is in distinctly superior agreement with the double-helix rather than the SBS model,

It is significant for the refinement of any future SBS model to consider the prigin of the meridional intenstty predicted on $\ell=6,7$ and 9 of SBSO. Some highly crystalline B-DNA samples do show weak

Figure 4.10 Meridional amplitudes of the Fourier transforms of the sugar, phosphate and base groups of SBS-DNA

meridịnals on layer-planes other than the tenth (see for example plate IV of Langrtdge et al (1960a)), However, those predicted by the SBS models are considerably stronger. The observed meridional reflections may be explained within the Watson-Crick framework by small distortions in the molecular symmetry, by scattering from water or ions within the unịt cell whose symmetry is different from that of DNA or by disorder of the crystallite orientation giving rise to apparently meridional reflections. The less well-defined patterns of Bram and Zimmerman and Pheiffer mentioned earlier also contain meridional intensity. Bram's pattern has meridional intensity on all layer-lines, whereas that of Zimmerman and Pheiffer has such intensity on $\ell=1,2,3,4,9$ and 10 . The discrepancy between these two patterns is a further indication that the fibres from which they were obtained suffered from different types and degrees of partial order. The most likely explanation for the observed intensity on the meridion is that individual molecules within the fibre are misaligned with respect to the fibre axis giving only apparent rather than true meridional diffraction. This simple explanation needs to be discounted before any suggestion that the double helical model is at variance with the observed diffraction would be plausible.

Figure 10 shows the amplitudes of the Fourier transforms along the meridion of the sugar, base and phosphate groups of the SBS models. These transforms depend only on the $z$ co-ordinates of the atoms so the curves for SBSO and SBS 36 are identical. They were calculated (using Helix 1) from the co-ordinates in table 3 so the positions of the three components relative to the origin was preserved. Under these conditions the three curves may simply be added together without the introduction of an additional phase factor to give the total molecular transform. The base transform shows significant peaks only at the origin and on $\ell=10$ (corresponding to $Z=1 / 3,45 A^{-1}$ ) and $t t$ is essentially zero elsewhere. This behaviour is to be expected since the bases in the SBS models are
held perpendicular to the helix axis and are stacked upon each other with a separation of $3,45 A$ with only small deviations from perfect regularity, Therefore they scatter like a perfect 3.45 A diffraction grating where the $\ell=10$ peak is the first order. In B-DNA both successive sugars and phosphates are also regularly spaced with a period of 3.4 A and so their diffraction follows the same pattern. As a result no diffraction is observed on the meridion except $\ell=10$. The sugar-phosphate chain in the SBS model is less well-ordered hence both the sugar groups and the phosphate groups behave like imperfect diffraction gratings. The theory of such gratings has been outlined by Sparrow (1919) and a treatment of the effect of disorder in helices has been given by Vainstein (1966) and Goodwin (1977). We may define two idealised types of disorder. In disorder of the first kind the diffracting units are displaced from the true lattice points with no correlation between the displacements of nearest neighbours. Such structures, similar to the thermal disorder observed in crystals, give rise to continuous diffuse scattering and the Bragg peaks are increasingly attenuated as the scattering angle increases. In disorder of the second kind the displacements of the diffracting units from the true lattice points are not independent. So the displacement of a given unit is a linear sum of its own displacement from its ideal position and that of all the other units between it and the origin. Such structures, which are similar to liquids in showing short-range order but long-range disorder, give rise to intermolecular interference effects which decrease very rapidly with increasing scattering angle, Figure 11 shows the distribution of phosphorous atoms along the $z$-axis in both the $A$ and $B$ chains of SBS. The irregularity, which is immediately apparent, is more graphically represented in the next line where vertical bars show the phosphorous positions, The corresponding diagram for B-ONA would show a regular 3,4A lattice with two bars per lattice point separated by 0.36A. Since there ts a diad axis at the origin we may restrict our consideration

$$
\because=-\square
$$

$$
\text { : }:: \quad=-
$$

n SBS36


$$
\text { 2• } \quad \text { •吕 }
$$


to the phosphorous atoms in just one chain of the molecule. The A-chain atoms are shown in the next line and the bottom line shows the corresponding B-DiAA lattice with the origin arbitrarily set at AIOP. It is then clear that the major source of irregularity is a dislocation between atoms A9 and A8 : within the group A8-A3 the interphosphorous spacing is relatively regular but all the atoms are displaced about $3 A$ down the helix. The region $A 7-A 9$ is the sharp $q$-bend region. A second discontinuity occurs between $A 3$ and $A 2$ : the atoms $A 2$ and $A 1$ are displaced about $1 A$ up the helix from the regular positions. The region between A2 and A4 corresponds to the relatively open p-bend region. Therefore within the repeating unit of 10 nucleotides the irregularity bears some similarity to disorder of the second kind. The effect of the disorder on the diffraction pattern (figure 10 ) is to produce positive peaks at $\ell=2,3,6$ and 9 and negative peaks at $\ell=4$ and 7. Since the disorder within the sugars is likely to follow the same pattern as that within the phosphates, the sugar scattering curve is similar in profile to that of the phosphates. However, since the scattering power of sugars is lower than that of phosphates, their effect on the total transform is less significant. The ratios of the absolute amplitudes of the peaks $\ell=1$ to 9 of the total transform are approximately 1:2:1:2:0:6:5:0:3 respectively. The corresponding intensities therefore will be 1:4:1:4:0:36:25:0:9 respectively in agreement with the strong meridionals predicted by the cylindrically averaged transform on $\ell=6,7$ and 9 . We nay now consider how to eliminate the predicted meridional intensity. We have observed in Chapter 3 that the total transform may be equal to zero in essentially two ways. First, the base, sugar and phosphate scattering curves may all pass through zero simultaneously. Second, the curyes may be non-zero but with their amplitudes and phases being such that cancellation reduces the total transform to zero. In order to satisfy the first possibility the phosphate groups must be regularly spaced along the helix axis. The work with the CPK model
suggests that this will be very difficult to achieve particularly in the $q$-bends. Any attempt to impose regularity in this region of the backbone resulted in unstacking of the bases which is unlikely to be energetically satisfactory, The second possibility will also be difficult to satisfy since any trregularity in the phosphate spacings is likely to be copied by the sugars which will therefore tend always to scatter roughly in phase with the phosphates thus removing any chance of cancellation. Therefore one would need to arrange the bases such that they negate the phosphate contribution at those points where it is significant without introducing appreciable meridional diffraction on the other layer-planes. Any future model must satisfy these constraints in addition to predicting correctly the diffraction observed elsewhere in the pattern.

We have not yet calculated the $X$-ray diffraction to be expected from the most recent side-by-side model of Millane and Rodley (1981), however, consideration of its helical parameters suggests that it will be no more successful than its predecessors in accounting for the observed diffraction from fibres in the $B$-form. This model contains a long-range right-handed twist of about $48^{\circ}$ every ten nucleotide pairs. It may therefore be described in terms of a 340 A pitch helix with $360 / 48=7.5$ residues per turn. The model contains two diad axes perpendicular to the helix axis every 34A. It is therefore a $15_{2}$ helix with 5 nucleotidepairs in the asymmetric unit. Table 9 shows the order of the Bessel function contributions to the first 16 layer-planes from various $m$-families of such a helix. The layer-plane spacing is $1 / 340^{A^{-1}}$ so the first meridional is predicted to occur at $z=15 / 340=0.044 \mathrm{~A}^{-1}$ whereas the first meridional observed in B-DNA patterns is at $z=10 / 34=0.29 A^{-1}$ (Langridge et al, 1960a), Successive meridional reflections will aiso be predicted by the SBS 48 model at $Z=p \times 0.044 A^{-1}$ where $p$ is any integer. As discussed earlier, it is unlikely that all these meridionals will have low intensity. In addition the observed strong meridional reflection

Table 4.9: The orders of Bessel functions contributing to SBS 48 layer-planes

| m | $\ell$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| -1 |  | 8 |  | 9 |  | 10 |  | 11 |  | 12 |  | 13 |  | 14 |  | 15 |
| 0 | 0 |  | 1 |  | 2 |  | 3 |  | 4 |  | 5 |  | 6 |  | 7 |  |
| 1 |  | -7 |  | -6 |  | -5 |  | -4 |  | -3 |  | -2 |  | -1 |  | 0 |
| 2 | -15 |  | -14 |  | -13 |  | -12 |  | -11 |  | -10 |  | -9 |  | -8 |  |

(on $\ell=10$ in the B-DNA nemenclature) occurs when $p=6.6$ which is not allowed, so SBS 48 does not even predict this reflection and in this sense it is in even worse agreement than SBS 36 with the observed patterns,
4.5 Constraints Imposed by X-ray Diffraction Data on Future Side-by Side Models

The SBS models published so far have attempted to account only for the B-type diffraction patterns. However DNA is possibly the most polymorphic biological macromolecule. As discussed in more detail in Chapter One there are four well-defined conformations : A(Fuller et al, 1965), $B($ Langridge et al, 1960b), C(Marvin et al, 1965) and D(Arnott et al, 1974) which have been explained by the Watson-Crick paradigm. In addition minor modifications of these structures account for the $B^{\prime}$ (Arnott and Selsing, 1974) C' and C" (Leslie et al, 1980) diffraction patterns. Recently a further conformation, E-DNA, has been observed and this too may be explained in terms of a Watson-Crick double-helix (Leslie et al, 1980). Were these isolated conformations then it could be argued that the side-by-side model is a peculiar structure accessible only under conditions which yield B-type diffraction patterns. However it is well established that the same DNA specimen may undergo transitions between conformations. In particular, DNA may make reversible $A \rightarrow B$ transitions as a function of salt content and ambient humidity (Franklin and Gosling, 1953b; Cooper and Hamilton, 1966). It is now known that this transition is only a special case of a family of allowed transitions. For example the $C \rightarrow A \rightarrow B$ transitions have been observed repeatedly with DNA from a wide variety of sources and with various base composition and sequence (Rhodes et al, 1981) and the $D \rightarrow A \rightarrow B$ family of transitions has also been reported (Leslie et al, 1980), The significance of these results for the SBS hypothesis is clear : the transitions may easily be explained
in terms of the Watson-Crick model by a small winding or unwinding of the helix with a concomitant modification of the nucleotide conformation so that the number of nucleotide-pairs per helix pitch is varied; but a change from a B-type SBS cQnformation to an A-type Watson-Crick conformation for molecules constrained wịthịn a fibre is very implausible and therefore if the SBS structure is to be established as an alternative to the Watson-Crick model it would seem essential to develop SBS models which account for diffraction patterns of the $A, B, C$ and $D$ type. However, it is not easy to suggest even the broad outlines of such models. The difficulty arises because the base-pairs in A- and D-DNA appear to be highly tilted. This information comes from the general intensity distribution around the meridional reflections in these patterns. In A-DNA the meridional intensity on $\ell=11$ is very weak. The intensity on $\ell=10$ is essentially zero and that on $\ell=7,8$ and 9 is strong (Fuller et al, 1965). In D-DNA patterns the intensity on $\ell=8$ is very low whereas that on $\ell=7$ is high (Arnott et al, 1974; A. Mahendrasingham, personal communication). As argued in the previous section, base-pairs stacked perpendicular to the axis would give rise to a very intense meridional reflection. The only possibility of diminishing this intensity would be by cancellation or systematic absence. If indeed the bases are highly tilted, as in the double-helical models for $A-$ and $D-D N A$, then steric hindrance is likely to be a major problem in the bend regions of SBS models. This is because highly tilted bases will only pack comfortably into a helix if the sense is such that the angle between the normal to the base-pairs and the tangent to the helix passing through the centre of the base is relatively small, i,e, the bases must tilt roughly 'with the helix' (see Chapter 3), Therefore bases in left-handed regions tend to tilt in the opposite sense to those in right-handed ones. There are no çvious stereochemical difftculties in achieytng this if the molecule is of a single hand, but in a structure with alternating right- and left-handed
regions short contacts in the bend region are likely which will force the bases to reduce their tilt if stacking is to be maintained. If the stacking constraint is removed then the irregularity in the molecule will give rise to diffuse scattering which does not appear to be present in, for example, the best A-DNA patterns (Fuller et al, 1965).

Further constraints on alternatives to the double-helix arise from crystal symmetry and the packing of molecules within the unit cell. DNA in the A conformation crystallises in a unit cell which is compatible with the monoclinic space group C2 (Franklin and Gosling, 1953b; Fuller, 1961; Fuller et al, 1965). If the molecule is two-stranded this space group requires that two identical molecules pass through each unit cell and that each molecule should contain at least one two-fold rotation axis perpendicular to the helix axis per 5iz nucleotide-pairs (in an ll-fold helix) oriented along the unique axis $b$. The unit cell of crystalline B-DNA contains symmetry elements which are consistent with the orthorhombic space group $\mathrm{P}^{2} \boldsymbol{1}^{2} \boldsymbol{1}^{2} \boldsymbol{1}^{\text {. }}$. This space group does not of itself require a twofold axis in the DNA molecule. However, the conditions limiting possible reflections are (h00), (0kO) and (00l) where $h, k$ or $\ell=2 n$ where $n$ is any integer. Each one of these conditions arises from the presence of one, and only one, set of parallel screw diads. For example, in the space group $\mathrm{P}_{1}$, which contains a screw diad along $\underline{b}$, reflections of the type $F(O k O)$ are only observed if $k$ is even. But the diffraction patterns from crystalline B-DNA contain extra systematic absences when $h+k$ is odd and $\ell=3 m$ where $m$ is any integer. These extra absences imply that for each molecule at $(x, y, z)$ there exists an identical one at $\left(x+\frac{1}{2}, y+\frac{1}{2}, z+\frac{1}{3}\right)$. If for example there is one molecule at the origin then the second is at $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{3}\right)$ and:-

$$
F(\text { hkl })=f_{m}^{2}\left[1+\exp 2 \pi i\left(\frac{n}{2}+\frac{k}{2}+\frac{k}{3}\right)\right]
$$

which clearly predicts the absences described above, But the screw axes in $P 21^{2} 1^{2}$ 1 invert the molecule at the centre of the cell with respect to that at the corner (Figures $12 a$ and $b$ ). Therefore if the two molecules are to be identical, as the absences require, it follows that a diad rotation axis must also be present oriented along a or b. Figures $12 c$ and $d$ show the effect on the equivalent positions within the cell if a diad is introduced along bwich is the diad orientation chosen by Langridge et al (1960a). Therefore both A- and B-DNA must contain at least one diad axis perpendicular to the helix axis in each helix pitch. Fortunately both SBSO and SBS 36 contain such a diad. However it is clear from figure $12 b$ that the B-DNA molecule must also contain a diad screw axis along the helix axis. Whilst such an axis is present in SBS 36, there is none in SBSO so it is inconsistent with the requirements derived from the $X$-ray data.

Whilst there is evidence from a statistical analysis of the distribution of diffracted intensities which supports the assignment of the space groups $C 2$ and $P 2,1^{2} 1^{2}$, to $A-$ and $B-D N A$ respectively (Appendix to Fuller et al, 1965; Arnott, 1971) this view has been challenged (Donohue, 1969, 1971) and the suggestion made that the symmetry in these two structures is no higher than that of the triclinic space groups Pl with two (in the case of $A-D N A$ ) or three (in the case of B-DNA) of the unit cell angles accidentally equal to $90^{\circ}$. Nonetheless there is striking evidence from the lattice parameters of DNA and RNA crystals which suggests both that the molecules are regular objects and that they pack together so as to optimise the intermolecular stereochemistry. This is particularly so in the case of B-DNA where $\arctan (b / a)=36.15\left( \pm 0.25^{\circ}\right)$. Dover (1977) has shown how the closeness of this angle to the turn per residue in the Watson-Crick B-DNA pelix maximises the number of equivalent contacts between adjacent molecules. The SBSO model provides no such natural explanation for this remarkable co-incidence but this argument does not

Figure 4.12 : (a) Symmetry elements of $\mathrm{P}^{2} 1^{2} 1^{2}$ ( (with the origin in a different position from that in International Tables) and (b) the corresponding equivalent positions. (c) Twofold rotation axes oriented along $\underline{b}$ added to space group $P 2,2,2$ and (d) the corresponding equivalent positions.
decisively refute the SBS36 model, as suggested by Arnott (1979), since it tqo has a turn per residue of about $36^{\circ}$ and could therefore maximise favourable contacts between successive ten base-pair units in adjacent molecules, It does however provide an important constraint which immediately throws doubt on the acceptability of the SBS 48 structure of Millane and Rodley (1981). Stmilar intuitively expected relationships between polynucleotide and crystal symmetry are realised as pointed out by Arnott (1978). For example D-DNA which has been described as an eight-fold Watson-Crick helix by Arnott et al (1974) crystallises in a tetragonal lattice; 12-fold $A^{\prime}-R N A$ double-helices crystallise with rhombohedral symmetry (Arnott et al, 1973); 1l-fold A-RNA helices pack into molecular triads within which the contacts are identical - the triads are arranged on trigonal lattices (Arnott et al, 1967c). The importance of such arguments is that they depend only on the lattice parameters which can be determined with precision and accuracy from crystalline samples.

The molecular symmetry itself may also be affected by the chemical structure of polynucleotides. That alternating purinepyrimidine sequences can give rise to the 8 -fold D-DNA helices which are apparently not accessible to polynucleotides with random sequences has been known for some time (Arnott et al, 1974). Recent work by Leslie et al (1980) and Arnott et al (1980) has shown that such alternating sequences may also adopt the left-handed S-DNA helix which, unlike D-DNA, has a dinucleotide asymmetric unit. In addition, Leslie et al (1980) have discovered that poly $d(A-G-C) p o l y d(G-C-T)$ and poly $d(G-G-T) p o l y d(A-C-C)$ form helices with 9 , symmetry and paly $d(A-G) p o l y d(C-T)$ forms a helix with ${ }_{2}$ symmetry reflecting the dinucleotide and trinucleotide chemical repeats respectively within these polymers in contrast to the 28 helices adopted under similar circumstances by DNA with a randon base sequence (Marvin et a1, 1961).

In summary, X-ray analysis has provided a wide range of evidence concerning the molecular and crystal symmetry of polynucleotide conformations and the transitions between them. If the SBS model is to be regarded as a serious competitor to the double-helix, as originally suggested by Rodiey et al (1976), then stereociemically acceptable models must be devised which conform to the rigorous constraints imposed by the X -ray intensities at least as well as the Watson-Crick models. The work described in this Chapter shows that no such success has yet been achieved even with the one diffraction pattern which the proponents of the side-by-side hypothesis claimed that their model explained.
4.6 Discussion of Other Experiments Relevant to the SBS Controversy

In this section we will discuss very briefly evidence other than $X$-ray diffraction which has been cited in favour of the SBS model. This includes aspects of the biology of DNA, for example the topological problem of unwinding during replication and transcription, the difficulty of recognition by proteins of specific sites in a highly symmetrical molecule and the packing of DNA in chromatin in addition to more physical considerations such as optical rotation and electron microscopy of nucleic acids. An additional physical parameter, the electrophoretic mobility of covalently-closed circular DNA, has been described in detail by Crick et al (1979) and Wang (1979) in their refutation of the SBS model and so it will not be discussed here.

The SBS model is undoubtedly less aesthetically pleasing than the double-helix. Naturally this is not a decisive argument against it despite the importance of, for example, symmetry in nature. However symmetry is widely exploited by biological macromolecules. Fibrous proteins and viruses, polysaccharides, muscle and flagellae are all believed to consist of regular helices. In addition oligomeric proteins (such as ferritin) and icosahedral viruses consist of repeating units which are arranged in
equivalent or quasi-equivalent positions. This is unsurprising since it follows that if one of the units is in a minimum energy configuration then the whole system or macromolecule will also have a low energy. It is also significant for the efficient assembly of macromolecular systems. It would be strange if DNA did not also exploit symmetry in this way. The SBS model contains changes of handedness at arbitrary points along the axis yet it maintains a fixed pitch of 34A. It is not obvious how this can occur. The proposers of the SBS model have suggested that it might be stabilised by proteins but this is clearly impossible in fibres of pure DNA. It is conceivable that favourable interactions between well stacked bases could compensate for irregularity within the backbone for example but this contention needs to be examined quantitatively. This has not so far been attempted and it would therefore be necessary to perform an energy refinement of the model using a method such as that of Levitt (1978).

It is known that protein molecules such as repressors interact with DNA at specific sites. If the protein is to recognise its site then clearly there must be some distinguishing features along the genome, but the refined structures of DNA in fibres appear to be too symmetrical in this sense. However it should be emphasised that the assumption of perfect regularity in fibre diffraction studies of DNA is to a large extent determined by the paucity of data. Recently more attention has been paid to irregularity in the molecule. In particular the combination of both X-ray diffraction and energy constraints in the refinement procedure has improved the parameter/data ratio so that helical constraints may be relaxed. This has shown that the base sequence modifies the local conformation in a less dramatic sense than the polymorphism observed in fibres. The X-ray studies on a DNA dodecamer have conftrmed that such small variations occur (Wing et al, 1980). Other studtes have shown that the helical parameters of DNA may be a function of base sequence, (Rhodes and Klug, 1981) and that the dynamic structure of poly $d(G-C) \cdot p o l y d(G-C)$
resulting from thermal fluctuations is a function of salt content (Ramstein and Leng, 1980). These subtle features together with the long-range polymorphism observed in fibres and the alternating $B$ and Z structures (Klug, et al, 1979; Wang et al, 1979; Drew et al, 1980) indicate that the problems involved in recognition may not be so severe as suggested by the proponents of the SBS model. Indeed even in a perfectly regular helix it is still possible to discern the base sequence since the distribution of hydrogen bonding acceptor and donor groups within the major groove is different in the two major types of base-pair. This may be a significant feature in the specificity of recognition processes. Two DNA-binding proteins have recently been solved to about 2.9 A resolution. Modelbuilding studies have suggested that the $\lambda$ cro repressor binds to right-handed B-DNA (Anderson et al, 1981) whereas the E.coli catabolite activator protein binds to left-handed B-DNA (McKay and Steitz, 1981). In both cases the protein interacts with the DNA via a pair of $\alpha$-helices which dock in the major groove. This refutes the suggestion by Bates et al (1977) that steric hindrance from the sugarphosphate chain prevents intimate contact between DNA and protein over the several nucleotide-pairs present in binding sites. The cro dimer covers one pitch length of the DNA so it is unlikely that it would fit into SBS-DNA since one of the two $\alpha$-helices would be incorrectly oriented, The importance of the $\alpha$-helical regions in the building of CAP and cro repressor is particularly interesting since the protamine-tRNA complex (the only other nucleic acid-protein complex to have been solved so far) also contains $\alpha$-helix within one of the grooves (Warrant and Kim, 1978). Other modelbuilding studies have suggested that $\beta$-sheet may also be significant in the binding process (Carter and Kraut, 1974; Church et al, 1977; Blake and Oatley, 1977) and in the cro repressor a pair of $\beta$-sheet strands appear to interact with the DNA. Although we still know few
details of the mechanism of interaction of nucleic acids with proteins, it is clear from the above that double-helical DNA can accommodate proteins at least as well as the side-by-side model.

The primary reason for the invention of the SBS model was the topological problem of strand separation during replication and transcription (Rodley et al, 1976). This may be illustrated by an example. The covalently-closed circular chromosome of E.coli contains at least 300,000 helical turns which are removed at a rate of about 4,000 per minute during replication. The problem is twofold. First, the rate of unwinding is extremely high yet the strands separate without tangling, Second, since the duplex is covalently-closed at least one single-stranded break must be present if the two chains are to separate. In eucaryotic cells the problem is yet more pronounced since the chromosomes consist of complexes of DNA and proteins. These apparent difficulties were realised when the double-helix was devised (Crick, 1954). The resolution of the problem is now generally believed to lie with the various DNAbinding proteins which have been discovered comparatively recently. There are two main classes of such proteins - (1) those which bind strongly to single strands of polynucleotides and hence destabilise the duplex structure; and (2) the helicases which unwind DNA in reactions driven by ATP hydrolysis. Short reviews are available which discuss such proteins from procaryotes (Abdel-Monem and Hoffman-Berling, 1980) and eucaryotes (Falaschi et al, 1980). However, Pohl and Roberts (1979) have argued that the topological problem persists despite these proteins since it is necessary for the replication forks of $\theta$-shaped intermediates to be maintained in extremely precise alignment in the case of cccDNA to avoid "knotted" daughter molecules. To achieve this alignment, they suggest, would require that the proteins at the replication forks are able to serse the global DNA conformation. Since this appears impossible they conclude that DNA
is not a double helix at least during replication. In the place of the double helix they propose that the SBS model is capable of explaining replication. The important feature of the molecule in this respect is the asymmetry of the duplex which enables the protein to sense to which side of the molecule it is attached. It should be re-emphasised however that the distinction between the topologies of the double helix and the SBS model is one of degree rather than kind as a result of the long-range twist which modelbuilding suggests will be a feature of any SBS structure. Thus any topological constraints in replication which arise from the helical nature of the duplex would apply to both the double helical and SBS models. The solution of the unwinding problem may be that duplexes supercoil to such an extent that the nett topological winding is zero. Under such circumstances the strands may be separated with no difficulty. It is possible that left-handed regions in the DNA duplex may be incorporated at certain regions (under conditions controlled by the cell) in order to induce compensating supercoils elsewhere. This idea receives some support from the recent observation of Z-DNA regions in polytene chromosomes of Drosophila (Nordheim et al, 1981) which provides the first evidence that left-handed helices are of biological significance. It is clear that we need a detaileddescription of the unwinding process. A first step in this direction has been made by McPherson et al (1979) who have solved to 2.3 A resolution the gene 5 product of bacteriophage fd, a DNA-unwinding protein. The active form of the protein appears to be a dimer which binds to DNA via the aromatic amino residues (within a threestranded antiparallel $\beta$-sheet) which may intercalate into the duplex. It has been proposed that successive dimers along the duplex also bind weakly to each other to produce co-operative unwinding. To some extent this mechanism may satisfy the requirement of Pohl and Roberts that the replication complexes need to have large scale information about the DNA conformation and it will therefore be of great interest to determine whether
it is a common feature of all unwinding proteins. Until such work has been extended it does not appear fruitful to speculate about the conformation of DNA during replication.

DNA in chromatin is arranged into a linear chain of bead-like structures which consist of histones (01ins and Olins, 1974). Each bead or nucleosome is about 100 A in diameter and is attached to about 170 base-pairs of DNA (Noll, 1974). It is clear that the DNA must be highly compressed and its arrangement has been the subject of speculation. Crick and Klug (1975) suggested that the DNA is wrapped around a protein core with the DNA being kinked every 20 nucleotide-pairs by unstacking the bases and modifying the sugar-phosphate conformation in the kink region, The helix axes of adjacent segments of DNA then form an angle of $90^{\circ}$ giving a left-handed superhelix. Both Bates et al (1977) and Sasisekharan et al (1977) have pointed out that such unstacking would be likely to be energetically unfavourable and they suggest, without justification, that the SBS model is less rigid than the double-helix and could be continuously deformed to form a left-handed superhelix of maximum diameter 50 A and pitch 100 A which could wind around nucleosome cores. As a result, they suggest, SBS-DNA may be the favoured conformation in nucleosomes. Sussman and Trifonov (1978) and Levitt (1978) have since shown by energy calculations that B-DNA may be continuously deformed without unstacking or producing sharp irregularities in the structure to give a supercoiled conformation which would accommodate the nucleosome core. Therefore the SBS model offers no better explanation than the double-helix for the packing of DNA in chromatin. That the bases are stacked approximately with a spacing of 3.4 A has been shown by X-ray diffraction from crystals of nucleosome cores plus DNA (Finch et al, 1981) but the resolution is not sufficiently high to yield detailed information about the ONA conformation.

Electron microscopic evidence has also been cited tn favour of the SBS model by Sasisekharan et al (1978) who suggested that the features
observed in a high resolution electron micrograph were more consistent with the rather irregular proftle of the SBS model than with the regular Watson-Crick structure. However the difficulty of interpreting such micrographs is greater than Sasisekharan and co-workers imply. Two major problems concern the uniformity of the distribution of stain along the molecule and damage to the sample from the electron beam. But even if these two points were to present no problem it is necessary to determine that the features observed in the micrograph are true detail and not artefacts or noise. The most satisfactory approach is to examine the optical diffraction pattern of the micrograph. This would be a crucial test of the SBS hypothesis if high quality, high resolution micrographs of the DNA molecule could be obtained since the optical diffraction patterns predicted by the SBS and double-helical models are significantly different. In an electron micrograph one observes the profile of the molecule hence the repeating unit in the dquble-helical model is 17 A along the molecular axis whereas the repeat in the SBS model is 34 A (or 340 A if the long-range twist is taken into account) (see figure 13). Therefore the layer-line spacing in the optical diffraction pattern from a micrograph of SBS DNA would be one half (or one twentieth) that in the pattern from double-helical B-DNA. A further advantage of this method is that an image of the molecule with greatly reduced noise may be obtained by optical filtering. The noise in the micrograph will tend to be spread over the whole optical diffraction pattern whereas diffraction corresponding to repeating detail will be concentrated at points in reciprocal space. A recombined image utilising only the information at these points will therefore contain less noise than the original. Unfortunately it is doubtful whether the micrograph discussed by Sasisekharan et al or any other is of sufficiently high resolution to yield to this type of analysis and so we cannot dectde between the double-helical and SBS models on the basis of electron microscopy.


Figure 4.13 : Comparison of the profiles of the side-by-side and double helical models

The 260 nm branch of the circular dichroism spectrum of DNA and its synthetic analogues has been extensively examined. Rodley et al (1976) and Bates et al (1977) proposed that the SBS model readily explains the low optical activity of DNA since, they suggested, the contributions of the right-handed segments would tend to cancel those from the lefthanded segments. However the c.d. spectrum of synthetic polynucleotides is very intense (Johnson and Tinoco, 1969). These authors have explained the form of the spectrum in terms of base-base interactions involving $\pi \rightarrow \pi^{*}$ electronic transitions. Only the nearest neighbour interactions need to be taken into account in the case of B-DNA since contributions to the rotational strength vary inversely as the square of the distance between the bases. There is a higher number of possible combinations of nearest neighbours in DNA with random base sequence than in synthetic polynucleotides with repeating sequences. Some of the combinations add a positive contribution to the rotational strength and others are negative. The difference in the strength of the signals observed from natural and synthetic polynucleotides may be explained in terms of cancellation between contributions of opposite sign. In DNA with a random base sequence the large number of combinations of neighbours gives a high probability that cancellation will be relatively efficient giving rise to a low signal whereas in synthetic polymers with a smaller number of combinations the cancellation will be less efficient giving an intense signal. When the base tilt is increased as in A-DNA and RNA second and third neighbour interactions become more significant and in addition the signal changes from the conservative spectrum exhibited by B-DNA to a nonconservative form (using the nomenclature of Bush and Brahms, 1967). The calculated spectra of B-DNA and RNA agree well with experimental results (Johnson and Tinoco, 1969). Since the base tilt appears to be a dominant factor determining the shape of the observed spectrum of natural

DNA it is possible that the SBS B-DNA model may explain the signal equally well. However, Johnson and Tinoco have found that circular dichroism is not a relfable indicator of either the molecular symmetry nor indeed the number of strands in a polynucleotide, therefore we cannot predict with confidence the shape of the spectrum of the SBS model on general grounds alone. Instead one must calculate the spectrum analytically. Since this may prove an important test of the SBS hypothesis it will be undertaken by the author at a later date.

As described in Chapter 1, the c.d. spectrum of poly d(G-C).poly d(Gindicates that the molecule undergoes a reversible transition as the salt concentration is increased (Pohl and Jovins, 1972; Pohl, 1976). This was interpreted in terms of a transition between left- and right-handed conformations. The ${ }^{32} \mathrm{P}$ nmr studies of Patel et al (1979) suggest that the left-handed conformation is a member of the Z-family (Wang et al, 1979; Drew et al, 1980). As the salt content was increased the number of resonances changed from one (indicating that all the phosphorous atoms were in identical chemical environments as in B-DNA) to two (indicating two types of environment as in Z-DNA). Rodley et al (1976) and Bates et al (1977) suggested that the transition observed by Pohl and co-workers may be easily explained by the SBS model since a relatively small change in the base orientation could lead to a change in the handedness of the longrange twist. However in this conformation there are more than two phosphate environments therefore one might expect to observe a corresponding number of resonances in the n.m.r. spectrum so the SBS model does not appear to be consistent with the experimental results. In addition no SBS model has been published with a left-handed twist. Indeed the most recent attempt to produce a stereochemically acceptable model (Millane and Rodley, 1981) has required a greater right-handed twist than in SBS 36. Work by the auth: 2 whflst building the CPK model described earlier suggested that there is insuffictent flexibility within the sugar-phosphate backbone to allow a stereochemically acceptable SBS confomation with a left-handed
long-range twist without base unstacking.

### 4.7 Conclusions

In this Chapter we have shown that the SBS model is not a satisfactory alternative to the double helix for the structure of DNA in fibres. However, as Crick et al (1979) have pointed out, the SBS model has been useful in forcing us to re-examine the evidence in favour of the double-helix. The data obtained from $X$-ray fibre diffraction is inferior to that from single crystal analysis since the resolution is insufficient to allow the determination of atomic positions. Instead we must decide between competing models by comparing their predicted diffraction with the observed. On this criterion the SBS model is inferior to the double helix since it predicts the incorrect layer-line spacing, relative intensities and peak positions in the molecular transform. We have rejected the recent argument of Bates et al (1980) that "fibrous" diffraction patterns should be used and instead we agree with their earlier contention : "The only way to use the measured $X$-ray data to make a critical assessment of the Watson-Crick and SBS models is to calculate the structure factor amplitudes for crystalline specimens. . . . We intend to make comparisons with what appear to be the best observational data and the most highly refined version of the Watson-Crick model. (Arnott and Hukins, 1973)," (Bates et al, 1977).

Although the SBS model is not the structure of DNA in fibres it, or certain features of $i t$, may nonetheless have a function in cell biology. The discovery of Z-DNA and the possibility of a $B \rightarrow Z$ transition has shown that bends between left- and right-handed regions are likely to be important. It is possible therefore that small stretches of SBS-like DNA may be present, at least transiently, in the cell.

## Appendix to Chapter IV

The Error Argument of Bates et al (1980)

In the following pages the argument of Bates et al (1980) concerning the errors involved in diffraction patterns from fibrous and crystalline samples is reproduced verbatim for convenience. The nomenclature they use is slightly different from ours. The azimuthal angle in reciprocal space has been denoted by $\psi$ (rather than $\Phi$ ) and the cylindrically averaged squared molecular transform is $\Omega$. Any other synbols used are either the same as in this thesis or defined in the text.

In this appendix the functional dependence of qualities is suppressed for convenience, i,e. the complex amplitude of the diffraction from a model of a polynucleotide structure is written as $E$, rather than $E_{\ell}(R, \psi)$. The true complex amplitude is denoted by $\tau$ :

$$
\begin{equation*}
\tau=E+\varepsilon \tag{Al}
\end{equation*}
$$

where $\varepsilon$ is the error, i.e. it is a measure of the discrepancy between the model structure and the actual structure.

Angular brackets denote an angular average, e.g.
$\langle\tau\rangle=(1 / 2 \pi) \int_{0}^{2 \pi} \tau_{\ell}(R, \psi) d \psi$
Consequently, E can be written as
$E=\langle E\rangle+\tilde{E}$
where $\langle\tilde{E}\rangle=0$ by definition.
The measurable intensity $M$ at a point in reciprocal
space is
$M=\tau \tau^{*}$
where the asterisk denotes the complex conjugate. It follows from equation (Al) and (A3) that

$$
\begin{equation*}
M=I+\langle E\rangle \varepsilon^{*}+\varepsilon\left\langle E^{\star}\right\rangle+\tilde{E} \varepsilon^{*}+\varepsilon \tilde{E}^{*}+\varepsilon \varepsilon^{*} \tag{A5}
\end{equation*}
$$

where I = EE*.
The angular average of the error $\varepsilon$ can be expected to be small, and $\varepsilon$ and $\tilde{E}$ can be expected to be almost angularly independent:

$$
\begin{equation*}
\langle\varepsilon\rangle \approx\left\langle\tilde{E}_{\varepsilon}\right\rangle \tilde{\sim} 0 \tag{A7}
\end{equation*}
$$

The measurable fibre pattern is <M>. The ftbre pattern of the model structure is:

$$
\begin{equation*}
\Omega_{2}=\left\langle E E^{\star}\right\rangle=\langle E\rangle\left\langle E^{\star}\right\rangle+\langle\tilde{E N}\rangle \tag{A8}
\end{equation*}
$$

because <E> $=0$ by definition, It follows from equations (A5) through (A8) that:-
$\langle M\rangle=\Omega+\left\langle\varepsilon \varepsilon^{\star}\right\rangle$

Comparison of equations (A5) and (A9) shows that the discrepancy between $<M>$ and $\Omega$ is of smaller order than is the discrepancy between $M$ and $I$.

## CHAPTER V

## THE APPLICATION OF PATTERSON FUNCTIONS IN THE DETERMINATION

OF NUCLEIC ACID STRUCTURE

### 5.1 Introduction

A criticism of molecular models derived from fibre diffraction analysis is that they are based on a number of assumptions which are not objective. Many assumptions have been implicit in the work of those involved in determining the structures of the nucleic acids. For example, until recently it was natural to impose right-handedness, anti base-sugar orientations and regular mononucleotide repeats as constraints when building models. It is important to ensure that these constraints are not weighted too highly especially since the discovery of Z-DNA and its relatives (Wang et al, 1979; Drew et al, 1980; Arnott et al, 1980) has shown that they are not absolute. This implies that we now face a much wider choice of plausible structures when analysing new diffraction patterns and indeed well-established ones. It is therefore desirable that we extract from each diffraction pattern the maximum amount of information whilst arrogating the minimum amount of supposition. A similar problem in a different guise is present in the crystallography of small molecules. In this case the experimenter frequently obtains a large number of structure factor amplitudes but initially at least no phases. One possible solution is to find the position of one atom in the structure and use it as a 'seed' about which to base a Fourier synthesis. Patterson (1934) discovered that this may be achieved in principle by examining the Fourier transform of the obseryed intensities which gives a map of interatomic vectors. The advantage of his method is that it requires the assignment only of the Miller indices to the observed reflections.

Patterson functions have been widely used in the solution of the crystal structures of small molecules but they present difficulties when the molecule under consideration contains a large number of atoms as in the case of the biological macromolecules. However they have been particularly important for example to protein crystallographers in determining the positions of heavy atoms in the multiple isomorphous replacement method (Blundell and Johnson, 1976). They have also been applied to the nucleic acids and their components. Tollin has devised procedures which enable the position and orientation of planar groups in crystals to be determined by molecular replacement methods (Tollin and Cochran, 1964; Tollin, 1966, 1969). These techniques have been used in studies on nucleic acid components and the planar dye proflavine (Tollịn, Wilson and Young, 1968; Young, Tollin and Wilson, 1969; Munns and Tollin, 1970; Rahman and Wilson, 1972; Young, Tollin and Wilson, 1974; Young and Wilson, 1975; Jones and Neidle, 1975). In addition Jacobsen (1976) has applied the method in a fibre study of poly(I).poly(5-iodo C) in order to determine the base tilt. Franklin and Gosling (1955) used the three-dimensional Patterson function of crystalline sodium DNA in the A-form to obtain experimental support of the Watson-Crick model and to determine the orientation of the molecule in the unit cell.

MacGillavray and Bruins (1948) devised a cylindircally averaged Patterson function (CAPF) which is particularly appropriate in the analysis of fibre data. It has been used in the determination of the lattice parameters of A-DNA and to support the contention that DNA is helical (Franklin and Gosling, 1953c). Sato et al (1966) have performed a similar analysis on the data from rice dwarf virus RNA to find the relative displacement of the sugar-phosphate chains along the helix axis. Comparison of the calculated CAPF of an $\alpha$-helix with maps derived from the observed intensities from collagen and poly-y-methyl-L-glutamate demonstrated that the former contains no $\alpha$-helical structure whereas the
latter is very similar to an $\alpha$-helix (Yakel and Schatz, 1955). The CAPF has also been used in the determination of a $\beta$-helix model for feather keratin (Schor and Krimm, 196la, b) and to find the radius of tobacco mosaic virus and to show that the virion contains a helical groove (Franklin, 1955), The radial autocorrelation function, which is a spectal case of the CAPF, has been applied to the interpretation of the equatorial diffraction from biological membranes giving information about the size of the scatterers, their lateral association and their rotational symmetry element (Kataoka and Uekị, 1980). Finally Namba et al (1980) have applied both a CAPF and a difference CAPF to find the distribution of the myosin heads about the F-actin helix in both the relaxed and rigor states of striated crab-leg muscle.

Since the CAPF has been found to be useful in such a wide range of systems, it is of interest to apply it to the known polynucleotide structures, It is unlikely that these low resolution maps will be sufficient to allow a unique determination of the molecular structure however it may be possible to identify important inter- and intra-molecular vectors. A stimulus to this work has come from the suggestion of Bates et al (1977) that the CAPF of A-DNA (Franklin and Gosling, 1953c) is at variance with the molecular structure of Fuller et al (1965). We have therefore calculated the CAPF of A-DNA using the more extensive data of Fuller (1961). In addition Patterson maps of B-DNA, D-DNA, B-A-RNA and A'-RNA are presented.

Bates, Rodley and McKinnon (1980), Bates et al (1980) and Rodley and Bates (1980) have defined and discussed the Axial Patterson Function of B-DNA. They suggested that their results were in slightly better agreement with the SBS model of Rodley et al (1976) than with the double helix, These claims will be examined in this chapter and the analysis of fibre diffraction data by Bates and co-workers will be crittcised.

### 5.2 The Theory and Interpretation of Patterson Functions

We consider first the three-dimensional Patterson function in a Cartesian co-ordinate system. A vector in Patterson space will be denoted by $\underline{u}$ and its components relative to the co-ordinate axes are ( $u, v, w$ ). The Patterson function is then defined by the equation:-
$P(\underline{u})=\frac{2}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} F^{2}(h k \ell) \cos 2 \pi(h u+k v+\ell w)$

It is therefore the Fourier transform of the diffraction intensities. Since no phases are required it may be calculated once Miller indices have
 also be written in the form:-

$$
\begin{equation*}
P(u v w)=\int_{v} \rho(x y z) \rho(x+u, y+v, z+w) d V \tag{2}
\end{equation*}
$$

where $V$ is the volume of the unit cell and $\rho(x y z)$ is the electron density at a point $(x, y, z)$ within it. The Patterson function is therefore a three-dimensional map and $\mathrm{P}(\mathrm{uvw})$ is large if the points ( $x, y, z$ ) and $(x+u, y+v, z+w)$ are in regions of high electron density so peaks in the map represent interatomic vectors. From equation 2 it can be seen that:-

$$
\begin{equation*}
P(000)=\int \rho^{2}(x y z) d V=\sum_{j=1}^{N} z_{j}^{2} \tag{3}
\end{equation*}
$$

where $z_{j}$ is the atomic number of the $j{ }^{\text {th }}$ atom in a unit cell containing $N$ atoms. This origin peak will be the largest since it represents the contribution of the vectors from each of the atoms to itself. The magnitude of the peak representing a vector connecting atom $i$ (whose atomic number is $Z_{i}$ ) to atom $j$ (whose atomic number is $Z_{j}$ ) will be $Z_{i} Z_{j}$,
therefore contributions from the heavy atoms tend to dominate Patterson maps. Patterson space is very crowded. There are $N(N-1)$ peaks apart from the origin and the peaks are more diffuse than those in an electron density map giving rise to considerable overlapping and hence the difficulty in solving complicated structures using the Patterson function alone. Sharpening procedures may be utilised which tend to reduce the amount of overlap. This may be done by performing the Fourier summation using the unitary structure factors $U(h k l)=F(h k l) / \Sigma f_{\vec{j}}^{2}$ in place of $F(h k l)$ in equation 1. In principle this gives the Patterson function of a structure containing point atoms but in practice it has the deleterious effect of introducing false peaks as a result of series termination errors. A compromise may be achieved by modulating the observed intensities according to an empirically chosen smoothly varying function in reciprocal space. The relatively low resolution of diffraction patterns from biological macromolecules is a further impediment. Rather than representing interatomic vectors, Patterson peaks may then correspond to vectors between different groups of atoms. Finally, the Patterson function is centrosymmetric even if the crystal contains no centre of symmetry so the space group of Patterson space is generally different from that of the crystal.

The function described so far is that used by Franklin and Gosling (1955) in their determination of the molecular orientation of A-DNA within its unit cell. Specialised versions of the function may be defined which are more appropriate to particular experimental conditions. One example is the cylindrically averaged Patterson function of MacGillavray and Bruins (1948) which has been applied to a number of fibrous systems. An ideal specimen is assumed which has strict periodicity along the fibre axis and random azimuthal orientation. The Patterson function is then defined by the equations:-

$$
\begin{equation*}
P\left(r^{\prime}, z^{\prime}\right)=\sum_{\ell} P_{\ell}\left(r^{\prime}\right) \cos \left(\frac{2 \pi \ell z^{\prime}}{c}\right) \tag{4}
\end{equation*}
$$

where

$$
\begin{equation*}
P_{\ell}\left(r^{\prime}\right)=\frac{2 \pi}{N V} \int_{0}^{\infty} I_{\ell}(R) J_{0}\left(2 \pi R r^{\prime}\right) R d R \tag{5}
\end{equation*}
$$

$I_{\ell}(R)$ is the continuous intensity on layer-plane $\ell ; N$ is the number of periods of length $c$ in the fibre direction; $V$ is the irradiated volume and $r^{\prime}$ and $z^{\prime}$ are the radial and axial components respectively of a vector in Patterson space. Since fibre diffraction intensities are not generally placed on an absolute scale, the constants in equation 5 may be ignored. Should crystalline regions in the sample give rise to sampling then equation 5 may be replaced by:-

$$
\begin{equation*}
P_{\ell}\left(r^{\prime}\right)=\sum_{R} I_{\ell}(R) J_{0}\left(2 \pi R r^{\prime}\right) \tag{6}
\end{equation*}
$$

where $I_{\ell}(R)$ is the integrated intensity of a spot with radial co-ordinate $R$ on the $\ell^{\text {th }}$ layer-plane.

The interpretation of the CAPF is less straightforward than for the three-dimensional Patterson function. We consider first the simplest case of identical atoms situated on an N-fold helix. Figure la shows two atoms, $i$ and $j$, on a projection perpendicular to the helix axis and figure 16 is a projection down the helix axis. The position vector $\underline{r}_{i j}$ between these two atoms may be written as:-

$$
\begin{equation*}
\underline{r}_{i j}=\underline{r}_{i j}^{\prime}+z_{i j}^{\prime} \underline{k} \tag{7}
\end{equation*}
$$

where $z_{i j}^{\prime}$ is the difference between the $z^{\prime}$ co-ordinates of the atoms; $k$ is a unit vector along the $z^{\prime}$-axis and $\underline{r}_{i j}^{\prime}$ is the vector between the atoms in projection down the helix axis. Peaks in the CAPF corresponding to interatomic vectors between all the atoms will be distributed along the
(a)

(b)


Figure 5.1 : Components of the vector between two atoms situated in a helix
(a) Projection perpendicular to the helix
(b) Projection along the helix axis
z'-axis according to the rule:-

$$
\begin{equation*}
z^{\prime}=\frac{m P}{N} \tag{8}
\end{equation*}
$$

where $P$ is the pitch of the helix and $m$ is any integer. Application of the cosine rule shows that the radial distribution of the peaks is given by:-

$$
\begin{equation*}
r_{i j}^{\prime}{ }^{2}=2 R_{h}^{2}\left(1-\cos \theta_{i j}\right) \tag{9}
\end{equation*}
$$

where $R_{h}$ is the radius of the helix and $\theta_{i j}$ is the angle between the atoms in projection down the axis. Since $\theta_{i j}=2 \pi z^{\prime}{ }_{j j} / P$ this may be rewritten as:-

$$
\begin{equation*}
r_{i j}^{\prime}=R_{A l}\left[2\left(1-\cos \frac{2 \pi z_{i j}^{\prime}}{P}\right)\right]^{\frac{1}{2}} \tag{10}
\end{equation*}
$$

The locus of peaks in the Patterson map is shown in figure 2. Since the vectors from atom 1 to atom $j(j=2,3, \ldots, N)$ are equivalent to those from atom $\mathbf{i}$ to atom $\mathbf{j}(\mathbf{j}=\mathbf{i}+1, \mathbf{i}+2, \cdots, \ldots, \mathbf{N})$ only one series of peaks is observed. The peaks lie on a curve of period $P$ with a minimum radial co-ordinate at $r^{\prime}=0$ and a maximum at $r^{\prime}=2 R_{h}$. The effect of cylindrical averaging is that information on the direction of $\underline{r i j}_{i j}^{\prime}$ is lost. Patterson maps showing the curves upon which phosphorous-phosphorous intramolecular vectors lie have been published for A-DNA (Franklin and Gosling, 1953c) and rice dwarf virus RNA (Sato et al, 1966).

Further peaks are introduced if the molecule contains atoms on several coaxial helices or is double helical. Consider first a molecule which consists of identical atoms on a single helix, which we will call H , and a second helix, H 2 , which is related to the first by a diad axis perpendicular to the helix axis. It is clear that the intrastrand vectors will be described by equation 10 so we need determine the components only of the interstrand vectors. If the co-ordinates of a given atom on $H 1$ are ( $R_{h}, \phi, z$ ) then all atoms on $H 1$ are given by ( $R_{h}, \phi+n \phi_{r}, z+n z_{t}$ ) and those on $H 2$ are given by $\left(R_{h},-\phi+n^{\prime} \phi_{r},-z+n^{\prime} z_{t}\right)$ where $\phi_{r}$
and $z_{t}$ are the rotation per residue and rise per residue respectively of the helices and $n$ and $n^{\prime}$ are integers. We define $\left(\Delta R_{h}, \Delta \phi, \Delta z\right)$ as the difference between the co-ordinates of atoms on different helices, so in general:-

$$
\begin{equation*}
\left(\Delta R_{h}, \Delta \phi, \Delta z\right)=\left(0,2 \phi+\Delta n \phi_{r}, 2 z+\Delta n z_{t}\right) \tag{11}
\end{equation*}
$$

where $\Delta n=n-n^{\prime}$
Peaks between such atoms are distributed along the $z^{\prime}$-axis according to the rule:-

$$
\begin{equation*}
z^{\prime}=2 z+\Delta n z_{\hat{t}} \tag{13}
\end{equation*}
$$

and with respect to the $r^{\prime}$-axis according to the rule:-

$$
\begin{align*}
r^{\prime} & =\left[2 R_{h}^{2}(1-\cos \Delta \phi)\right]^{\frac{1}{2}}  \tag{14}\\
& =\left[2 R_{h}^{2}\left(1-\cos \left(2 \phi+\Delta n \phi_{r}\right)\right)\right]^{\frac{1}{2}} \tag{15}
\end{align*}
$$

Equations 13 and 15 may be combined to eliminate $\Delta n$ giving:-

$$
\begin{equation*}
r^{\prime}=\left[2 R_{h}^{2}\left(1-\cos \left(2 \phi+\left(\frac{z^{\prime}-2 z}{z_{t}}\right) \phi_{r}\right)\right)\right] \tag{16}
\end{equation*}
$$

The locus of Patterson peaks will cut the $r^{\prime}$-axis when:-

$$
\begin{equation*}
2 \phi+\left(\frac{z^{\prime}-2 z}{z_{t}}\right) \phi_{r}=2 \pi m \tag{17}
\end{equation*}
$$

where $m$ is any integer. Hence:-

$$
\begin{equation*}
z^{\prime}=2 z_{t}\left(\frac{\pi m-\phi}{\phi_{r}}+\frac{z}{z_{t}}\right) \tag{18}
\end{equation*}
$$

It will be useful to consider the position of just one of these nodes and so we set $m=0$ giving the 'origin node':-

$$
\begin{equation*}
z^{\prime}=2 z_{t}\left[\frac{z^{z}}{z_{t}}-\frac{\phi}{\phi_{r}}\right] \tag{19}
\end{equation*}
$$

Once again the period of the curves is $P$ and the maximum values $r^{\prime}=2 R_{h}$ occur mid-way between the nodes. Equation 16 describes the locus of peaks representing vectors from H 1 to H 2 . Those from H 2 to H 1 lie on a similar curve and the pair of curves is symmetrical about $z^{\prime}=c / 2$ as required by centrosymmetric property of Patterson space. The locus of $\mathrm{H} 2 \rightarrow \mathrm{H} 1$ peaks is given by:-

$$
\begin{align*}
& r^{\prime}=\left[2 R_{h}^{2}(1-\cos (-\Delta \phi))\right]^{\frac{1}{2}}  \tag{20}\\
& z^{\prime}=-2 z-\Delta n z_{t} \tag{21}
\end{align*}
$$

or:-

$$
\begin{equation*}
r^{\prime}=\left[2 R_{h}^{2}\left(1-\cos \left(2 \phi-\left(\frac{z^{\prime}+2 z}{z_{t}}\right) \phi_{r}\right)\right)\right]^{\frac{1}{2}} \tag{22}
\end{equation*}
$$

The zeros in this curve are given by:-

$$
\begin{equation*}
z^{\prime}=-2 z_{t}\left(\frac{\pi m^{\prime}-\phi}{\phi_{r}}+\frac{z}{z_{t}}\right) \tag{23}
\end{equation*}
$$

where $m^{\prime}$ is any integer and at the origin node:-

$$
\begin{equation*}
z^{\prime}=-2 z_{t}\left(\frac{z}{z_{t}}-\frac{\phi}{\phi_{r}}\right) \tag{24}
\end{equation*}
$$

It is useful at this point to consider the predicted Patterson maps of some model structures. We will use the $B$-DNA parameters $\phi_{r}=36^{\circ}$ and $z_{t}=3.4 \mathrm{~A}$ in these models. Figure 3 shows the interstrand curve of a helix with one atom situated on the diad axis so $\left(R_{h}, \phi, z\right)=(10,0,0)$. In this special case the $\mathrm{H} 1 \rightarrow \mathrm{H} 2$ and $\mathrm{H} 2 \rightarrow \mathrm{H} 1$ curves co-incide both with each other and with the intrastrand $\mathrm{H} 1 \rightarrow \mathrm{H} 1$ and $\mathrm{H} 2 \rightarrow \mathrm{H} 2$ curves since effectively the structure is a single-stranded helix. As the angular co-ordinate of the atom is increased from $\phi=0^{\circ}$ to $\phi=45^{\circ}$ and $90^{\circ}$ (Figures 4 and 5) the phase relationship between the curves also varies.


Figure 5.2 : Locus of vectors in C.A.P.F. between atoms related by the symmetry of a single helix whose pitch is $P$.


Figure 5.3 : Locus of vectors in C.A.P.F. between atoms situated on the diad axis of the B-DNA double
helix


Figure 5.4 : Locus of vectors between symmetry-related atoms placed on the B-DNA helix. The first atom is situated at $\left(R_{h}, \phi, z\right)=(10,450,0)$.


Figure 5.5 : Locus of vectors between symmetry-related atoms placed on the B-DNA double helix. The first atom is situated at $\left(R_{h}, \phi, z\right)=(10,90,0)$

The curves are again co-incident (as in figure 3) when $\phi=180^{\circ}$ but in this case the separation between the origin nodes is $P$ so the relative displacement of the curves has a period of $180^{\circ}$. Similarly the curves are in identical relative positions each time the $z$ co-ordinate of the atom is changed by $P / 2$ if the angular co-ordinate is held constant. In a real molecule the atoms will occupy more general positions and a particularly important series of peaks will arise from the phosphorous atoms in DNA since these are the most electron dense in the structure (figure 6). Franklin and Gosling (1953c) and Sato et al (1966) have published the Patterson maps of A-DNA and rice dwarf virus RNA with the intrastrand phosphorous loci superimposed. The separation between the nodes of these curves is an indication of the relative displacement of the two sugar-phosphate chains if the phosphorous peaks are assumed to predominate. The origin nodes are separated by:-

$$
\begin{equation*}
\Delta z^{\prime}=4 z_{t}\left(\frac{z}{z_{t}}-\frac{\phi}{\phi_{r}}\right) \tag{25}
\end{equation*}
$$

which may be proved using equations 19 and 24 and the two nodes closest to $z=P / 2$ are separated by:-

$$
\begin{equation*}
\Delta z^{\prime}=p-4 z_{t}\left(\frac{z}{z_{t}}-\frac{\phi}{\phi_{r}}\right) \tag{26}
\end{equation*}
$$

Although the phosphorous atoms may make the most significant contributions in regions where the $r^{\prime}$ co-ordinate is high, their peaks are likely to be distorted in the region of low $r^{\prime}$ where the base atom peaks will also be important. Figure 7 shows the curves for adenine N9 of B-DNA. Since its radial co-ordinate is small (4.63A) it cannot make contributions to the map in the regions where $r^{1} \geqslant 9.2 A$.

We now constder briefly intrastrand vectors arising from a structure consisting of two co-axial helices with one atom in the asymmetric unit of each helix. The general co-ordinates of atoms on the first helix

$$
\text { Figure } 5.6 \text { : Predicted mossitions in ©.A. }
$$



Figure 5.7 : Loci of peaks representing vectors between adenine N9 atoms of B-DNA


Figure 5.8 : Racdiial compenremt of the vecter between two atroms on coaxial hellicces
(HI) are ( $R_{1}, \phi_{1}+n \phi_{r}, z_{1}+n z_{t}$ ) and those on the second helix (HZ) are ( $R_{2}, \phi_{2}+n^{\prime} \phi_{r}, z_{2}+n^{\prime} \phi z_{t}$ ) where $R_{1}$ and $R_{2}$ are the radii of the helices. The differences ( $\Delta \mathrm{R}, \Delta \phi, \Delta z$ ) are given by:-

$$
\begin{equation*}
(\Delta R, \Delta \phi, \Delta z)=\left(R_{1}-R_{2}, \phi_{1}-\phi_{2}+\Delta n \phi_{r}, z_{1}-z_{2}+\Delta n z_{t}\right) \tag{27}
\end{equation*}
$$

The Patterson peaks are distributed along $z^{\prime}$ according to:-

$$
\begin{equation*}
z^{\prime}=z_{1}-z_{2}+\Delta n z_{t} \tag{28}
\end{equation*}
$$

Figure 8 shows a projection down the helix axis from which we may derive the relationship between $r^{\prime}$ and $\Delta \phi:-$

$$
\begin{equation*}
r^{\prime}=\left[R_{1}^{2}+R_{2}^{2}-2 R_{1} R_{2} \cos \Delta \phi\right]^{\frac{1}{2}} \tag{29}
\end{equation*}
$$

or

$$
\begin{equation*}
r^{\prime}=\left[R_{1}^{2}+R_{2}^{2}-2 R_{1} R_{2} \cos \left(\phi_{1}-\phi_{2}+\Delta n \phi_{r}\right)\right] \tag{30}
\end{equation*}
$$

Combining equations (28) and (30) gives the relationship between $z$ ' and $r^{\prime}:-$

$$
\begin{equation*}
r^{\prime}=\left[R_{1}^{2}+R_{2}^{2}-2 R_{1} R_{2} \cos \left(\phi_{1}-\phi_{2}+\left(\frac{z^{\prime}-z_{1}+z_{2}}{z_{t}}\right) \phi_{r}\right)\right] \tag{31}
\end{equation*}
$$

Now the maximum value is given by:-

$$
\begin{equation*}
r^{\prime}=R_{1}+R_{2} \tag{32}
\end{equation*}
$$

and the minimum by:-

$$
\begin{equation*}
r^{\prime}=R_{1}-R_{2} \tag{33}
\end{equation*}
$$

which may be verified by inspection of figure 8. The minima occur when:-

$$
\begin{equation*}
z^{\prime}=\left[M \pi+\phi_{2}-\phi_{1}\right] \frac{z_{t}}{\phi_{r}}+z_{1}-z_{2} \tag{34}
\end{equation*}
$$

Once again these equations apply to vectors from atoms of type 1 to those of type 2. Similar equations may be derived which apply to $2 \rightarrow 1$ vectors. The curves derived fron the co-ordinates of $P(8.91,95.2,2.08)$ and Cl' $5.86,67.4,0.47$ ) in B-DNA are shown in figure 9 as an example. The zero nodes occur at $z^{\prime}= \pm 1.02 \mathrm{~A}$ as predicted by equation 34 and its symmetry-related pair.

The analysis presented above may be adapted to apply to vectors between atoms on coaxial double helices or atoms in different molecules. In the latter case additional terms will appear which relate to the intermolecular orientation and translation with respect to the fibre axis. These equations will not be presented here, however it is clear from figure 10 that the maximum and minimum values of $r^{\prime}$ in the case of intermolecular vectors are given by:-

$$
\begin{align*}
& r_{\max }^{\prime}=d+R_{1}+R_{2}  \tag{35}\\
& r_{\min }^{\prime}=d-R_{1}+R_{2} \tag{36}
\end{align*}
$$

where $d$ is, the intermolecular separation. Therefore peaks corresponding to these vectors may contribute to the Patterson map even at very low values of $r^{\prime}$. For example, in the case of crystalline B-DiNA, $d=19 \mathrm{~A}$ for nearest neighbours and $R_{1}=R_{2}=8.9 \mathrm{~A}$ for the phosphorous atoms so the locus of peaks of vectors between these atoms has a minimum value $r_{\text {min }}^{\prime}=1.2 A$. It should also be noted that several similar values of $d$ may occur in the case of a unit cell containing more than one molecule. For example, in the case of B-DiNA there are three distinct sets of intermolecular vectors between near neighbours corresponding to $d=19 \mathrm{~A}, 22.7 \mathrm{~A}$ and 31.2 A .

It is clear that the various interstrand, intrastrand and intermolecular vectors present in a DNA fibre will give rise to a wide variety of Patterson peak loci which differ in both amplitude and phase. This combined with the low resolution of the data makes unambiguous interpretation of


Figure 5.9 : Locus of vectors between B-DNA phosphorous $(r, \phi, z)=(8.81,95.2,2.08)$ and $\mathrm{Cl}^{\prime}(5.86,67.4,0.47)$


Figure 5.10 : Radial component of the vector between two atoms on adjacent helices.
cylindrical Patterson maps difficult. However we might expect to see the phosphorous atom peaks since these atoms are the most electron-dense in DNA and RNA. In addition, special values of ( $r^{\prime}, z^{\prime}$ ) will occur at which many vectors co-incide giving a significant peak as, for example, in the case of two molecules related by a lattice repeat.

Bates, Rodley and McKinnon (1980), Bates et al (1980) and Rodley and Bates (1980) have suggested the CAPF maps are unreliable except at low values of $r$ '. They have therefore defined the 'on-axis' or axial Patterson function given by:-

$$
\begin{equation*}
P_{a x}\left(z^{\prime}\right)=\sum_{\ell} P_{\ell} \cos \left(\frac{2 \pi \ell z^{\prime}}{c}\right) \tag{37}
\end{equation*}
$$

and

$$
\begin{equation*}
P_{\ell}=\int I_{\ell}(R) R d R \tag{38}
\end{equation*}
$$

where $I_{\ell}(R)$ is the intensity at the point $R$ on layer $\ell$ in the case of specimens giving continuous diffraction along all layer-planes or:-

$$
\begin{equation*}
P_{\ell}=\sum_{R} I_{\ell}(R) \tag{39}
\end{equation*}
$$

in the case of crystalline samples, where $I_{\ell}(R)$ is the integrated intensity of a reflection with radial co-ordinate $R$. Comparison with equations 4, 5 and 6 shows that:-

$$
\begin{equation*}
P_{a x}\left(z^{\prime}\right) \equiv P\left(0, z^{\prime}\right) \tag{40}
\end{equation*}
$$

Therefore peaks in the axial Patterson correspond simply to vectors between atoms with the same radial co-ordinate. Since it is merely a subset of the CAPF the distribution of nodes and peaks along the $z^{\prime}$-axis will be identical to those of the model systems discussed earlier.

### 5.3 Computer Programs to Calculate $P\left(r^{\prime}, z^{\prime}\right)$ and $P a x(z)$

The program CAPF was written in Algol and run on the CDC 7600
at U.M.R.C.C. It is desirable that structure factor amplitudes or intensities or continuous cylindrically averaged squared transforms may be read as data therefore the program was written as a series of modules so as to allow the flexfbility in input and output to be transparent to the main body of the program. A schematic flow chart is shown in figure 11.

The procedure READDATA inputs (i) the diffraction data in the form ( $h, k, \ell, F),(h, k, \ell, I)$ or $\left(I_{\ell}(R), R\right)$; (ii) the lattice parameters if structure factors are being used; (iii) the range of $r^{\prime}$ and $z^{\prime}$ over which the map is to be calculated and the corresponding step-sizes $\Delta r^{\prime}$ and $\Delta z^{\prime}$. If lattice parameters have been read then REALRECIP converts them to reciprocal lattice parameters. PRINTDATA simply outputs the data which has been read. The procedure FILL IR TABLE then reads the diffraction data into an $N \times 2$ array where $N$ is the number of data points. $I(j, 1)$ contains the intensity of the $j^{\text {th }}$ point and $I(j, 2)$ contains its $R$ co-ordinate. CALCULATE 30 stores $J_{0}(x)$ from $x=0$ to 90 in steps of 0.1. The main body of the program is contained within the two loops, FIND PL OF $R$ evaluates and stores $P_{\ell}\left(r^{\prime}\right)$ as a function of $\ell$ for a particular value of $r^{\prime}$. These values are then used by $P$ OF RZ which evaluates $P\left(r^{\prime}, z^{\prime}\right)$ using equation 4 . These operations are repeated cyclically until the complete map has been accumulated whereupon PRINT $P$ OF RZ outputs the array in a form suitable for contouring. The procedure was designed so that both axes are on approximately the same scale if $\Delta r^{\prime}=2 \Delta z^{\prime}$. The origin peak was set equal to 1000 . The maps were contoured by hand.

The program AX PAT was written in BASIC and run on an APPLE microcomputer. The values of $I_{\ell}$, range of $z^{\prime}$ and $\Delta z^{\prime}$ were read as data and the program simply calculates the sum in equation 37 . If the origin peak was included in the range then the values were modified so that $P(0)=1000$. The program TRAP INTEGRAL accepts a file of type CAST as

Fiqure 5.11: Flowecthent of the Program CAPF

input and evaluates $I_{\ell}$ for each layer by trapezoidal integration. The axial Patterson function may of course also be evaluated using CAPF but it was generally more convenient to use AX PAT.

### 5.4 Results of the Cylindrical Patterson Function Calculations

5.4.1 A-DNA

It is important to recollect that a Patterson map contains no more information than the structure factor amplitudes, indeed the cylindrical Patterson function contains less, instead it is simply a different representation. Therefore it is an alternative to the comparison of observed and calculated diffraction patterns. Patterson maps are difficult to interpret due to the large number of possible vectors in a complex structure and the overlap of vectors caused by cylindrical averaging presents a further impediment. This is particularly important in the case of a molecule, such as DNA, in which no atom contains ${ }^{\text {. }}$ significantly more electrons than the rest. Nonetheless it is possible to proceed in the normal manner of fibre diffraction analysis and compare maps derived from the observed and calculated intensities. If the maps agree well it will increase our confidence in the correctness of the model.

The cylindrical Patterson function of A-DNA was calculated by Franklin and Gosling (1953c, d) using 100 observed reflections as data. We repeated that calculation (figure $14 a$ ) to confirm the correctness of the computer program. More extensive data ( 300 reflections) for A-DNA have been collected by Fuller (1961) and a detailed model for this conformation has also been presented (Fuller et al, 1965). The Patterson functions of both the observed and calculated intensities are shown in figures 14 b and c .

It is necessary to attempt to assess the effect on the maps of errors in the intensities. Franklin and Gosling (1953c) did this
by comparing the map calculated using the observed intensities with that calculated after the two strongest reflections had been reduced by $30 \%$ of their observed values. The most significant peaks of these maps were closely similar so they concluded that the effect of errors was small. One would intuitively expect that the effect would be diminished as the number of reflections in the data set increased. Therefore the maps calculated using the data set of Fuller (1961) should be more reliable than that of Franklin and Gosling (1953c). However, whilst the position and magnitude of the major peaks are insensitive to errors, the general background upon which they are superimposed consists of rather low magnitude "ripples" (typically an order of magnitude smaller than the dominant peaks) and only small changes are necessary to change the shape of contours of this height. When comparing maps computed from observed and calculated intensities we will frequently see, for example, that three predominant peaks, $A, B$, and $C$ may be present in both maps, but that in one $A$ and $B$ are connected and $C$ is distinct whereas in the other $B$ and $C$ are connected and $A$ is distinct. This effect, which may be due to errors in the intensities or series termination errors in the Patterson synthesis, can dramatically change the appearance of two maps which are quite similar. In all the maps presented in this chapter the origin peak has been arbitrarily normalised to 1000 . As a result, both positive and negative contours are present.

We now compare figure 14a with figure 3 of Franklin and Gosling (1953c). The major features of both maps are identical. For example, both have strong peaks at $\left(r^{\prime}, z^{\prime}\right)=(0,14),(12,7),(23,14)$, $(23,0),(35,7),(45,14)$ and $(45,0)$. But Franklin and Gosling have a pQsitive, although low magnitude, strip connecting the peak at $(12,7)$ with tts symmetry-related pair at $(12,21)$ and to the peak at $(23,14)$ whereas in figure 14 a the $(23,14)$ peak is distinct from these two and is connected instead to the symmetry-related peaks at $(23,0)$ and $(23,28)$.

This and similar discrepancies in the finer details of the maps may arise from the $\Delta r^{\prime}$ and $\Delta z^{\prime}$ step-sizes used in the calculation, Franklin and Gosling used $\Delta r^{\prime}=1 A$ but the $\Delta z^{\prime}$ value is not recorded. In this thests we have used $\Delta r^{\prime}=1 A$ and $\Delta z^{\prime}=0.5 A$ in all calculations. None of the observed discrepanctes affects the conclusions of Franklin and Gosling (1953c) concerning the indexing of the A-DNA lattice. In addition it should be noted that the curve representing the locus of vectors between phosphorous atoms on the same helix plotted by Franklin and Gosling (1953d) passes through a negative region at $\left(r^{\prime}, z^{\prime}\right) \simeq(18,14)$ and several peaks are predicted to occur here. However, whilst these points are negative according to the convention we have adopted, they are nonetheless local maxima and therefore not at variance with the conclusions of Franklin and Gosling (1953d) concerning molecular symmetry.

In comparing the maps calculated from the $A-D N A$ intensities observed by Franklin and Gositing (1953c) and Fuller (1961) certain difficulties arise. The reflections observed in these two studies are not equivalent so it is not possible to scale the two data sets as described in Chapter 3, As a result there will inevitably be a difference in the position of, for example, the zero contour in the two maps. Nonetheless several facts are clear. The major peaks in the map calculated from the Franklin and Gosling data are also present in the Fuller data map although the latter contains far more detail arising from the more extensive data used in the Patterson synthesis. For example, the large positive area in the Franklin and Gosling map at $\left(r^{\prime}, z^{\prime}\right) \simeq(34,7)$ has more structure in the fuller map, Also the five distinct alternating bands of positive and negative vectors in the Franklin and Gosling map have largely disappeared in the Fuller map, It is interesting to nate that the positive region at $\left(r^{\prime}, z^{\prime}\right) \simeq(17,14)$ in the map published by Franklin and Gosiing (1953c) which is missing from that calculated here using the same data (figure 14a) is present in the Fuller map as a well
defined pasitive area, This region is important since the curve representing vectors between phosphorous atoms on the same helix passes through it, More structure is also present in the fuller map along the $z^{\prime}$-axis which contajns nine peaks in contrast to the three in the Franklin and Gosling map,

The maps calculated from the observed and calculated intensities of Fuller (1961) also contain the same major peaks fndicating that the overall molecular conformation and crystal packing determined by Fuller et al (1965) are not at variance with the observed diffraction. However, there are clearly differences in detail between these maps for example along the $r^{\prime}$-axis. This may arise from errors of measurement in the intensities, small deficiencies in the model or inadequate allowance for the effect of scattering by water or ions in the structure. It should be noted that near the $z^{\prime}$-axis the agreement is good suggesting that the molecular model is satisfactory since this region is largely dominated by intramolecular vectors. Previous cylindrical Patterson functions of nucleic acid data (Franklin and Gosling, 1953c, d; Sato et al, 1966) have employed only observed intensities. This is the first such study in which maps from both the observed intensities and the intensities calculated from a specific model have been compared. A major difficulty in such a comparison is the lack of any objective criterion which may be applied to determine the quality of agreement between the maps.

The Patterṣon map should contain strong peaks corresponding to the numerous identical vectors between equivalent atoms on adjacent molecules, i,e, lattice and pseudo-lattice vectors. Indeed one of the main objectives of the work undertaken by Franklin and Gosling (1953c) was to assign indices to the Bragg reflections. The three shortest such vectors in the A-DNA crystal are shown in figure 12a. The ( $r^{\prime}, z^{\prime}$ )
components of $f_{1}$ are simply ( $\mathrm{b}_{1}, 0$ ) and consideration of figure 13 a shows that the components of $\underline{r}_{2}$ are $\left(\operatorname{acos}\left(\beta-90^{\circ}\right)\right.$, asin $\left.\left(\beta-90^{\circ}\right)\right)$, The components of $\underline{r}_{3}$ may eastly be shown to be:-

$$
\begin{align*}
& r^{\prime}=\left(p^{2}-z^{\prime 2}\right)^{\frac{1}{2}}  \tag{41}\\
& z^{\prime}=\frac{1}{2} \operatorname{ain}\left(90^{\circ}-\beta\right) \tag{42}
\end{align*}
$$

where:-

$$
\begin{equation*}
p=\frac{1}{2}\left(a^{2}+b^{2}\right)^{\frac{1}{2}} \tag{43}
\end{equation*}
$$

(refer to figure 13b). From the lattice parameters of Fuller (1961) $\left(a=22.24 \mathrm{~A}, \mathrm{~b}=40.62 \mathrm{~A}, \mathrm{c}=28.15 \mathrm{~A}, \mathrm{~B}=97^{\circ}\right.$ ) the components of the three vectors are as shown in table 1. The peaks corresponding to these three vectors and an additional one at $2 \underline{r}_{2}$ are present in the maps computed from the observed and calculated intensities of Fuller (1961) (figure 14b and c), the map calculated from the Franklin and Gosling data (figure 14a) and the map published by Franklin and Gosling (who used slightly different lattice parameters) although in the latter two cases the $2 \underset{\sim}{r}$ vector is not in a very large peak. This suggests that the generally accepted lattice parameters of A-DNA are correct, but it is certainly not clear that this indexing is unique. It is interesting to consider whether the discrepancies between the observed and calculated maps from the Fuller data might arise from incorrect indexing. Sasisekharan et al (1982) have examined a precession photograph of A-DNA which suggested to them that the lattice of Fuller et al (1965) is wrong. The magnitudes they propose for the lattice parameters are not dramatically different from the original ones, but the $\underline{a}$ and $\underline{b}$ vectors have been interchanged giving the final values refined against the same 28 reflections used by Fuller (1961) as $a=40.75 \mathrm{~A}, \mathrm{~b}=22.07 \mathrm{~A}, \mathrm{c}=28.16 \mathrm{~A}, \beta=97.5^{\circ}$. The cylindrtcally


Figure 5.12 : Intermolecular vectors in A-DNA (top) and B-DNA (bottom)

(b)


Figure 5.13 : Derivation of the components of intermolecular
vectors in A-DHA.
(a) Projection parallel to $\underline{b}$
(b) Projection perpendicular to the $\underline{a}$ - $\underline{b}$ plane
averaged Patterson function may readily be used to test this assertion, The components of the $\underline{r}_{1}, \underline{r}_{2}$ and $\underline{r}_{3}$ vectors in this indexing scheme are shown in table 1. The difference between the predicted positions of $\underline{r}_{3}$ using the Fuller and Sasisekharan schemes is very small so it cannot be used to distinguish between the two lattices. However, the other vectors are quite different. The vector $\underline{r}_{9}$ predicted by Sasisekharan et al is barely within a positive region and $2 r_{1}$ is at a local minimum. The vector $\underline{r}_{2}$ is also within a negative region (although it is adjacent to a significant peak). Thus the agreement between the map from the observed intensities and the predicted vectors is worse in the case of the lattice proposed by Sasisekharan et al than in the case of the Fuller (1961) lattice suggesting that the former is likely to be incorrect.

The DNA molecule has high symmetry so one would expect to see significant peaks corresponding to vectors between identical atoms within successive molecular asymmetric units. As suggested by Franklin and Gosling (1953d) the phosphorous atoms should tend to give rise to dominant peaks and in section 2 we derived equations relating atomic co-ordinates to the curve upon which $P \rightarrow P$ vectors should lie. Figure $14 c$ shows that vectors between phosphorous atoms on the same chain fall within the positive regions, although they are not always centred on the predicted positions, but the peaks between atoms on different chains often fall within negative regions. Bates et al (1977) have claimed that the Fuller model of A-DNA is at variance with the cylindrical Patterson function of the data. They suggest that strong peaks corresponding to cross-vectors between the phosphate groups should be present at $\left(r^{\prime}, z^{\prime}\right) \simeq(19,12)$ and $(19,15)$. However, the Patterson function of the predicted intensities presented here shows that the model only predicts weak peaks at these points as observed.

Table 5.1 : Components of lattice vectors of Fuller (1961) and Sasisekharan et al (1981) A-DNA lattices

| Vector | Component | Fuller lattice | Sasisekharan lattice |
| :---: | :---: | :---: | :---: |
| $r_{1}$ | $r^{\prime}$ | 40.62 | 22.07 |
|  | $z^{\prime}$ | 0 | 0 |
| $\underline{r}_{2}$ | $r^{\prime}$ | 22.07 | 40.40 |
|  | $z^{\prime}$ | 2.71 | 5.32 |
| $\underline{r}_{3}$ | $r^{\prime}$ | 23.04 | 23.17 |

All distances are in Angstrom units

### 5.4.2 B-DNA

The only cylindrical Patterson functions of nucleic acids hitherto calculated have been members of the A-family: A-DNA (Franklin and Gosling, 1953c, d) and RNA (Sato, et al, 1966). It is therefore of interest to examine the Patterson functions of other conformations. Three sets of data and three models for B-DNA have been presented by Langridge et al (1960a, b), Marvin (1961) and Arnott and Hukins (1972b, 1973). Maps have been calculated for the observed and calculated intensities from each set. The suspect (110) reflection in the Arnott data has been omitted from all calculations.

The maps calculated from both the observed and calculated intensities of Langridge et al (figure 14d and e) and Arnott and Hukins (figure 14h and i) contain both the same major features and roughly the same shaped contours even at high $r^{\prime}$ despite the differences between the data sets discussed in Chapter 3.

The peaks in both the Langridge et al and Arnott and Hukins maps show a well-developed 3.4 A periodicity parallel to the $z^{\prime}$-axis. Below $r^{\prime}=14 \mathrm{~A}$ two rows of parallel peaks occur, one at the $z^{\prime}$-axis and the other at $r^{\prime} \simeq 10 \mathrm{~A}$, the two rows being displaced by about 1.7 A along the $z^{\prime}$ direction with respect to each other. These features are quite distinct from those observed in the maps of A-DNA.

The maps calculated from the Marvin data (figure 14 f and g ) are different from those discussed above. Peaks with 3.4 A periodicity are still present close to the $z^{\prime}$-axis but the second set of peaks at $r^{\prime} \simeq 10 \mathrm{~A}$ has disappeared, At high values of $r^{\prime}$ the maps contain simply a broad section of alternating positive and negative regions. The reason for this difference is not clear. The obvious explanation is that the resolution of the data is poor, but in fact the resolution of Marvin's data does not appear to be significantly inferior to that of the other two sets.

Figure 12b shows the expected intermolecular vectors of B-DNA, The vector $\underline{r}_{1}$ is present in all six maps but although $\underline{r}_{2}$ is present in the Arnott and Marvin maps, it lies in a negative region in both Langridge maps. In addition, $\underline{r}_{4}$ is present in the Langridge and Marvin maps, but in a negative region in both Arnott maps. The vector $\underline{r}_{3}$ is present in all maps although in both Arnott maps this peak is only one of many of roughly the same height separated by 3.4 A along the $z^{\prime}$ direction at $r^{\prime} \simeq 19$ A. Finally $\underline{r}_{5}$ is absent from the Arnott maps but present in the other four. Thus the cylindrical Patterson map of B-DNA is less successful than that of A-DNA in determining the strongest intermolecular vectors. Inis feature requires further investigation.

It is also of interest to study the Patterson map of the lefthanded $B$-stacked B-DNA model discussed in Chapter 3 (figure 14j), This model is similar to the Langridge model in predicting $\underline{r}_{1}$ and $\underline{r}_{3}$ to be at peaks in the map whilst $\underline{r}_{2}$ is in a negative region. The $\underline{r}_{4}$ vector is within a positive region although it is not at a peak, Instead the peak occurs at $(34.0,0)$. The map is similar to that of the Arnott model below $r^{\prime} \simeq 14 \mathrm{~A}$. There are more discrepancies at higher values of $r^{\prime}$ but the agreement with the map from the Arnott data does not appear to be manifestly inferior to that achieved by the Arnott model.

A major difficulty in the interpretation of the Patterson maps We have discussed hitherto is the appearance on each map of both intramolecular and intermolecular vectors. This arlses because the co-efficients employed in the synthests are structure factor intensities. If we were able to separate the molecular and crystalline features within the diffraction pattern then we could compute two Patterson functions: one containing peaks corresponding to intramolecular vectors and the other containing crystalline or intermolecular vectors. The first function may be obtained by using $f_{m}^{2}$ as Patterson co-effictents, where $f_{m}$ is the amplitude
of scattering at a point in reciprocal space due to a single molecule. Two maps (figures $15 a$ and b) were computed using the observed and calculated $f_{m}^{2}$ values for B-DNA presented by Langridge et al (1960a). The structure factor intensities, I(hk\&), are given by:-

$$
\begin{equation*}
I(h k l)=f_{m}^{2} P(h k l) \tag{44}
\end{equation*}
$$

where $P(h k l)$ is the packing factor given by:-

$$
\begin{equation*}
P(h k \ell)=1+\exp \pi i\left[(h+k)+\frac{\ell d}{c}\right] \tag{45}
\end{equation*}
$$

Following a procedure similar to that proposed by Giglio et al (1958) we may attempt to eliminate the intramolecular vectors by using $I(h k \ell) / f_{m}^{2}$ as Patterson co-efficients.

The shapes of the contours and the connectivity between the various regions in the maps calculated using $f_{m}^{2}$ (figure $15 a$ and $b$ ) differ from those in figures 14 d and 14 e which were calculated directly from structure factor intensities, but the major features have not changed. In particular the peaks corresponding to the strongest intermolecular or crystalline vectors are still present. Unfortunately, therefore, it appears that the effect of lattice sampling has not been removed to any significant extent by this procedure so we are unable to use it to determine solely intramolecular vectors. This suggests that the position of peaks within the maps may be dominated by the position of the points at which the molecular transform is sampled. It is posstble that the presence of intramolecular peaks would be more apparent if we were to interpolate into the observed molecular transform at regular points along each layer-plane. (Indeed the Shannon sampling theorem (Shannon and Weaver, 1949) shows that we could obtain the full molecular transform were it possible to observe the
intensities at the half-lattice points but this is precluded by Bragg's Law). Since the lattice points are unevenly spaced and the molecular transform changes rapidly in some areas it would be difficult to ensure that any such interpolation was meaningful and so this was not pursued. The map using $\mathbb{l}(h k l) / f_{m}^{2}$ co-efficients of the Langridge data is more promising (figure 16). The peaks corresponding to $\underline{r}_{1}, \underline{r}_{2}$ and $\underline{r}_{4}$ are clearly present, In addition the vectors $\underline{r}_{3}$ and $\underline{r}_{5}$ are also apparent confirming the correctness of the conclustons drawn from the $I$ (hkl) maps. This is a useful result since the $\underline{r}_{2}$ peak at $(31,0)$ is obscured in the $I$ (hkl) maps. The features below $r^{\prime}=10 \mathrm{~A}$ are also interesting. Since the asymmetric unit of the B-DNA unit cell is one polynucleotide chain, one would expect to see peaks at low $r^{\prime}$ corresponding to vectors between two chains within the same molecule. The strong and broad peak at $z^{\prime}=11 \mathrm{~A}$ represents such a vector. It does not correspond to the separation between the helices upon which the phosphorous atoms are situated (which would give a peak at $z^{\prime} \simeq 13.5 \mathrm{~A}$ ). Instead it represents the average chain separation. Unfortunately $I(h k l) / f_{m}^{2}$ maps will not be helpful in the early stages of structure analysis since it is necessary to know $f_{m}^{2}$ and hence the molecular structure before the co-efficients may be calculated. However, it may be useful, for example, in orienting a proposed model correctly in a unit cell.

The majority of vectors between phosphorous atoms on the same chain fall within positive regions in both the Arnott and Langridge maps, but in both cases the vectors between phosphorous atoms on different chains frequently lie in negative regions, However, these vectors are also predicted to be in negative regions according to the Patterson maps of the calculated intensities.

### 5.4.3 D-DNA

The most recent model of D-DNA is that of Arnott et al (1974). The Patterson maps calculated from the observed and calculated intensities of this conformation are shown in figures $14 k$ and $14 \ell$. There is some
disagreement at low $r^{\prime}$, for example the positive region at $z^{\prime} \simeq 17 \mathrm{~A}$ extends to higher $r^{\prime}$ in the observed map than in the calculated one. In addition there are seven positive peaks along $r^{\prime}=0$ in the map from the observed intensfties whereas the negative regions in the calculated map do not extend to the $z^{\prime}$-axis so there are fewer peaks. Various other discrepancies are also present; for example the peak at $\left(r^{\prime}, z^{\prime}\right)=(5,6.5)$ is present in both maps, but it is isolated in the calculated map whereas it is connected to the peak at $(0,7,5)$ in the observed map. The negative valley between the two peaks in the former is quite deep so this probably does not arise simply from notse in the data. At higher values of $r^{\prime}$ the calculated map contains more discrete peaks than the observed one.

The disagreements between the maps may indicate anomolies in the model. Work in this laboratory has resulted in a model which is distinctly different from that of Arnott and co-workers (Mahendrasingam et al, manuscript in preparation). We have calculated the cylindrical Patterson function of one model in a series which were built during the refinement process. This model has very good intermolecular and intramolecular stereochemistry and agrees well with the observed diffraction pattern, but subsequent models have been devised which are superior in both respects. The model is a left-handed eightfold helix with $\beta$-stacked basepairs. The Patterson map (figure 14 m ) still differs from that calculated from the observed intensities. For example, the negative region originating at $(0,12)$ merges with the region at $(15,9)$ but the agreement is superior to that achieved by the Arnott model, particularly at low values of $r^{\prime}$. It is also interesting to note the behaviour of the maps along the $r^{\prime}$-axis. The curve $P\left(r^{\prime}, 0\right)$ is shown for both models and the observed intensities in figure 17. One can see tmmediately that both models predict the vectors at $r^{\prime}=17 \mathrm{~A}, 24 \mathrm{~A}, 34 \mathrm{~A}$ and 38 A which arise from the intermolecular vectors, shown in figure 12c, As one would expect from the definition of
the cylindrical Patterson function the heights of the first three peaks fall as $r^{\prime}$ increases, but the fourth peak is large because there are eight intermolecular vectors of this length whereas there are only four of each in the other cases. The left-handed model is in better agreement than the Arnott model with the curve from the observed data at $r^{\prime}=34 \mathrm{~A}$. In addition the Arnott model predicts a substantial broad peak at $r^{\prime}=9 \mathrm{~A}$. This is neither predicted by the left-handed model nor present in the curve from the observed intensities. A peak at this position is probably, but not necessarily, due to anomolies in the molecular conformation of the Arnott model. It would be incautious to say decisively that the Arnott model is incorrect on the basis of this evidence, however it is remarkable that the agreement between the curves from the data and the left-handed model is so good when one considers the major conformational novelties which it contains. This investigation will be extended in the future when data is available from the high quality D-DNA patterns which have recently been obtained in this laboratory.

The vectors between phosphorous atoms efther on the same or different chains qf both Arnott's model and the left-handed model are in much worse agreement with the theory developed in section 2 than in the case of either A-DNA or B-DNA with many of the vectors being in negative sections, but once again they agree with the predictions of the model map.

### 5.4.4 RNA

Two well-defined double helical RNA conformations, designated A- and $A^{\prime}-$ RNA, have been determined from fibre diffraction studies (Arnott et al, 1973). The A'-RNA helix has 12 , symmetry and 36 A pitch. The A-RNA helix is now believed to have 11 , symmetry and its pitch is 30 A.
$A^{\prime}-$ RNA crystallises in space group R3, the helix axes being placed at the $3_{1}$ positions. Since the molecules occupy special positions the intermolecular separations may be determined once the lattice parameters
are known. Each of the molecules is one member of a triad which surrounds a threefold rotation axis hence their displacements with respect to the $z$-axis are identical but corresponding diads perpendicular to the helix axis are rotated by $120^{\circ}$ with respect to each other. Therefore a right-handed rotation of any one molecule by $120^{\circ}$ makes its diad parallel to that of one of its neighbours. Such a rotation is equivalent to a translation of c/3 parallel to the helix axis so a strong intermolecular peak is expected at $\left(r^{\prime}, z^{\prime}\right)=(d, c / 3)$ where $d$ is the separation between the molecules. The lattice parameters are $\mathrm{a}=\mathrm{b}=39.4 \mathrm{~A}$ so $d=a / \sqrt{3}=22.8 \mathrm{~A}$. There are six nearest neighbours to any one molecule which contribute to this peak, The six next nearest neighbours give rise to a peak at $\left(r^{\prime}, z^{\prime}\right)=(a, 0)=(39.4,0)$. The maps of the observed and predicted intensities of $A^{\prime}-$ RNA presented by Arnott et al (1973) are shown in figures 140 and $p$. The expected peaks are clearly present in both maps.

Most of the phosphorous-phosphorous vectors lie within positive regions of the maps. Those in negative regions in the map calculated from the observed intensities are predicted by the model map.

The helical symmetry of A-RNA has been the subject of some debate. The early studies of Langridge and Gomatos (1963) on reovirus RNA suggested that the molecule was a tenfold right-handed helix. Subsequent analysis by Arnott et al (1967a, b, c) showed that the tenth layer-line meridional reflection (upon which Langridge and Gomatos based their assumption of tenfold symmetry) should be systematically absent. They showed that both tenfold and elevenfold models could explain the diffraction pattern although in both cases the meridional reflection must be explained in terms of molecular distortion or scattering from ions or water which have the symmetry of the molecule but not of the space group. As in the case of $A:-R N A$, the $A-R N A$ molecules are arranged in a triad within the unit cell. If the molecule has terfold symmetry a $3_{2}$ axis is placed
at the centre of the triad but if the molecule has elevenfold symmetry the triad is centred by a threefold rotation axis, On the basis of intermolecular contacts and Fourier synthesis studies Arnott and coworkers found the elevenfold model to be slightly favoured. Subsequently better quality diffraction patterns from synthetic polyribonucleotides showed the elevenfold model to be decisively superior. It is useful to consider whether the cylindrical Patterson maps may be used to distinguish between the $3_{2}$ and 3 axes which may centre the triad and thus distinguish between the 10 and 11 , molecular helices, Consider the three diagrams in figure 17 which show triads of molecules each of which has a rotation or screw axis at the centre. The symbols beside the molecules denote their relative displacements with respect to the plane of the page. Corresponding molecular diads are shown arbitrarily pointing along the sides of the triangle. The separation between the molecules is $d$. In order to determine the position of the main intermolecular peaks in the Patterson maps we need to determine the vector between equivalent positions within adjacent molecules. The diads of molecules 1 and 2 will be parallel following an anticlockwise rotation of molecule 1 by $120^{\circ}$. This is equivalent to a displacement of $c / 3$ parallel to the $z$-axis out of the page. So the equivalent positions of molecules surrounding a $3_{2}$ axis (figure 18a) are separated by the vector $\left(r^{\prime}, z^{\prime}\right)=(d, c / 3)$. A similar analysis performed on the other two cases gives the results shown in table 2. Unfortunately there is no difference between the vectors predicted by the 3 and $3_{2}$ axes so it is not possible to distinguish between the $10_{1}$ and 11, models on the basis of the Patterson function. Were the molecules left-handed then the rotation of molecule 1 by $120^{\circ}$ is equivalent to a displacement of $c / 3$ into the paper so the predicted vectors do then discriminate between the molecular symmetries (table 2),

The lattice parameter of A-RNA is $\mathrm{a}=39.9 \mathrm{~A}$ and che corresponding vector is present in the maps of both the observed and calculated intensities

Table 5,2 : Components parallel to $z^{\prime}$ of vectors between equivalent posttions in A-RNA triads

| Helix symmetry | Triad symmetry element |  |  |
| :--- | :---: | :---: | :---: |
|  |  | $3_{1}$ | 3 |
| Left-handed | 0 | $c / 3$ | $c / 3$ |
| Right-handed |  |  |  |

Fiqure 5.14 : Cylindrically averaged Patterson functions
using I(hkl) co-efficients

Shaded areas are negative
a) Franklin and Gosling A-DNA
b) Fuller A-DNA, observed intensities
c) Fuller $A-D N A$, calculated intensities
d) Langridge $B-D N A$, observed intensities
e) Langridge $B-D N A$, calculated intensities
f) Marvin B-DNA, observed intensities
g) Marvin B-DNA, calculated intensities
h) Arnott B-DNA, observed intensities
i) Arnott B-DNA, calculated intensities
j) $\beta$-LHE16, calculated intensities
k) Arnott D-DNA, observed intensities

1) Arnott D-DNA, calculated intensities
m) $\beta$ LHD-DNA, calculated intensities
o) $A^{\prime}-R N A$, observed intensities
p) $A^{\prime}-R N A$, calculated intensities
q) A-RNA, obseryed intensities
r) A-RNA, calculated intensities


## $14 b$



$14 d$

$14 e$


$149$


## 14 h h


$14 i$


14j










Figure 5,15 : Cylindrically averaged Patterson function of Langridge B-DNA with $\mathrm{f}_{\mathrm{m}}{ }^{2}$ co-efficients
a) Observed intensities
b) Calculated intensities
$15 a$


## 15 b



Figure 5.16 : Cylindrically averaged Patterson function
of Langridge B-DNA observed intensities
using I(hkl)/fm ${ }^{2}$ co-efficients



Figure 5.17 : $P\left(r^{\prime}, 0\right)$ of D-DNA
Top : Calculated from BLHD intensities
Middle : Calculated from the observed intenstities of Arnott et
al 1974 )
Bottom : Calculated from the intensities predicted by the model
of Arnott et al (1974)

(figures $14 q$ and $r$ ), There are four nearest neighbour molecules at $d=22.0 \mathrm{~A}$ and two next nearest neighbours at $d=25,2 \mathrm{~A}$ so peaks in the maps are expected at $\left(r^{\prime}, z^{\prime}\right)=(22.0,10.0)$ and $(25,2,10.0)$. The second peak is present in both maps, but the first is present in neither. on the contrary, additional peaks are observed in the vicinity at ( $r^{\prime}, z^{\prime}$ ) $=(22,8.5)$ and $(22,12)$.

Most of the peaks corresponding to phosphorous-phosphorous vectors lie in positive regions. Those which do not are predicted to be in negative regions by the map calculated from the model intensities.

### 5.5 Results of the Axial Patterson Function Calculations

### 5.5.1 B-DNA

The axial Patterson functions of B-DNA computed from the observed and calculated intensities of Arnott and Hukins (1973) are shown in figure 19a. The Patterson co-efficients, $P_{\ell}$, defined in equation 39 are recorded in table 3. The (110) reflection has once again been omitted from the calculations. In order to compare the curves we must define a normalisation condition. Bates et al (1980) have chosen to set $I_{2}=100$ and then multiply one of the curves by a constant factor which minimises the discrepancy between the observed and calculated Patterson functions. We do not adopt that procedure here for two reasons: (i) it places undue emphasis on just one layer-line of the diffraction pattern; and (ii) the second layer-line is not strong in all nucleic acid patterns so rather than normalise against a weak layer-line one would need to define a separate normalisation condition for each type of pattern, Instead we have chosen to adopt the condition defined earlier, that is the data set is scaled so $\Sigma I_{0}=\Sigma I_{c}$ Where the summation is taken over all the reflections in the set. In addition we arbitrarily set $P_{a x}(0)=1000$ in all the calculations described here.

The Patterson function of Arnott's B-DNA consists of a series of ripples separated by $\Delta z^{\prime}=3.4 \mathrm{~A}$. The agreement between the positions of the peaks in the observed and predicted curves is good, but the predicted peak amplitudes differ from those demanded by the observed intensities. Patterson functions were also calculated using the observed and predicted intensities of Langridge et al (1960) and Marvin (1961) (figures $19 b$ and 19 c ; table 3). Although the general appearance of the curves is unchanged, the relative peak heights vary somewhat amongst the three data sets. For example, only in the case of the Langridge data set is the $z^{\prime}=10.2 \mathrm{~A}$ peak larger than those at $z^{\prime}=3.4 \mathrm{~A}$ and 6.8 A .

The Patterson curve for the left-handed model BLHB16 is compared with that calculated from the observed intensities of Arnott and Hukins in figure 20, (the $P_{\ell}$ co-efficients are given in table 3). Once again the positions of the predicted peaks agrees with those calculated from the observed diffraction pattern, but there are discrepancies in the peak amplitudes.

It may be argued that the comparison between the Patterson functions of the three data sets given above is unfair since the number of reflections observed is different in each case. Therefore we have also calculated the curves from reduced data sets each of which contains only those reflections common to all three sets. The reduced sets were then scaled as described as above. Comparison of the curves shows little difference between calculations based on full or reduced data sets (figures 21a, b and $c$; the co-efficients for the reduced data sets are given in table 4). This probably occurs because the reflections which are not common to all sets tend to be rather weak and therefore they have little effect on the synthes is.

Bates et al (1980) claimed that the axial Patterson function Was capable of discriminating between the double hellcal and side-by-side models of DNA and that the latter was in superior agreement with the
function calculated from fibrous data, therefore we calculated the Patterson of the SBSO model described in the previous chapter. No model has been proposed for the packing of any SBS model within the B-DNA unit cell so it is not possible to calculate the structure factors and hence the Patterson of predicted Bragg intensities as for the other models, therefore the co-efficients, $P_{\ell}$, were derived from the cylindrically averaged intensity transform. These co-efficients are shown in table 5. Their Patterson function (figure 22) is quite different from that published by Bates et al (figure 30 ) who used the SBS model with long-range twist in their calculations. Since the pitch of this model is approximately 340 A the number of co-efficients in the synthesis must be increased to account for the tenfold increase in the predicted number of observable layer-lines within the resolution limit of the diffraction data. Bates et al found that the difference between the axial Pattersons of the two models were insignificant, We have not tested this, but the differences between the relative amplitudes of the peaks published here and by Bates et al suggest either that it is untrue or that one of the calculations is incorrect. Extensive checks revealed no errors in our calculations. It is immediately clear that the axial Patterson function of S8SO agrees with neither that calculated from the crystalline data of Arnott and Hukins nor with that from the fibrous data published by Bates et al (figure 29 ). However, for reasons we detail later, we feel that even if such comparisons were to yield good agreement it would not be compelling evidence in favour of the SBS model.

It is useful also to determine the effect of the water correction to the atomic scattering factors on the axial Patterson function. Consideratipn of table 3.3 shows that the phosphorous atomic scattering factor is substantially reduced by the water correction, but the axial Patterson function of the B-DNA phosphorous is essentially unaffected by the correction (figure 23 , table 5 ). It is reasonable to suppose therefore

Table 5.3 : Axial Patterson Co-efficients, $P_{\ell}$, of B-DNA using the Complete Published Data Sets

| $\ell$ | Source |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Arnott and Hukins (1973) |  | Langridge et al (1960a) |  | Marvin (1961) |  | BLH16 <br> Calculated |
|  | Observed | Calculated | Observed | Calculated | Observed | Calculated |  |
| $\begin{array}{r}0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ \hline\end{array}$ | 217 49 89 106 13 105 58 11 95 14 244 1001 | 136 73 98 172 16 80 65 15 64 9 273 1001 | 217 24 89 132 4 72 102 10 67 20 264 1001 | 224 30 112 172 4 59 111 11 70 11 195 999 | 158 39 65 88 18 96 60 9 88 15 358 994 | $\begin{array}{r} 154 \\ 44 \\ 100 \\ 132 \\ 21 \\ 98 \\ 56 \\ 10 \\ 103 \\ 11 \\ 268 \\ 997 \end{array}$ | $\begin{array}{r} 199 \\ 77 \\ 144 \\ 133 \\ 17 \\ 88 \\ 66 \\ 11 \\ 88 \\ 11 \\ 166 \\ 1000 \end{array}$ |

Table 5.4 : Axial Patterson Co-efficients of B-DNA using the reflections common to each data set

| $\ell$ | Source |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Arnott and Hukins (1973) |  | Langridge et al (1960a) |  | Marvịn (1961) |  |
|  | Observed | Calculated | Observed | Calculated | Observed | Calculated |
|  |  |  | 230 |  |  |  |
| 1 | 27 95 | 30 114 | 230 24 94 | 238 25 | 178 18 | 187 25 |
| 3 | 95 75 | 114 | 94 | 124 | 70 | 107 |
| 4 | 1 | 125 3 | 90 3 | 124 | 61 | 105 |
| 5 | 115 |  | 73 | 3 66 | 1 | 2 |
| 7 | 50 | 48 | 103 | -66 | 88 53 | 98 |
| 7 | 9 | 15 | 10 | 112 | 53 8 | 54 8 |
| 8 9 | 100 | 65 | 71 | 70 | 84 | 8 |
| 9 10 | $\begin{array}{r}12 \\ 282 \\ \hline\end{array}$ | ${ }^{9}$ | 19 | 10 | 14 | 14 |
| $\Sigma$ | 282 1002 | 336 1001 | 281 | 216 | 425 | 320 |
| L |  |  | 1001 | 1001 | 1000 | 1000 |

Table 5.5 : Axłal Patterson Co-efficients of SBSO

| $\ell$ | $P_{\ell}$ |
| :---: | ---: |
| 0 | 531 |
| 1 | 28 |
| 2 | 30 |
| 3 | 20 |
| 4 | 18 |
| 5 | 15 |
| 6 | 19 |
| 7 | 17 |
| 8 | 8 |
| 9 | 807 |
| 10 | 1001 |
|  |  |

Table 5.6 : Axial Patterson Co-efficients of B-DNA
Phosphorous

|  | Scattering <br> Factor |  |
| :---: | :---: | :---: |
|  |  |  |
| $\ell$ | $f$ | $f 1$ |
| 0 | 228 | 213 |
| 1 | 102 | 100 |
| 2 | 164 | 147 |
| 3 | 193 | 190 |
| 4 | 24 | 19 |
| 5 | 126 | 128 |
| 6 | 76 | 85 |
| 7 | 24 | 28 |
| 8 | 43 | 62 |
| 9 | 5 | 5 |
| 10 | 14 | 24 |
| $\sum$ | 999 | 1001 |

Figure 5.19 : Axial Patterson functions of B-DNA

Source of data :-
(a) Arnott and Hukins (1973)
(b) Langridge et al (1960a)
(c) Marvịn (1961)

All intensities were included in the calculations except in (a) where the (110) reflection was omitted.
__ Calculated from the observed intensities
--------- Calculated from the predicted intensities


Figure 5.19a


Figure 5.19b


Figure 5.19c


Figure 5.20 : Compaanison of the axial Patterssomm functimams of BLHB16 and the observed B-DNA intemssities of Arnottt and Hulkins (1973).

Source of data:-
a) Arnott and Hukins (1973)
b) Langridge et al (1960a)
c) Marvin (1961)

Only those reflections common to all three data sets were included in the calculations.
$\qquad$ Calculated from the observed intensities
calculated from the predicted intensities


Figure 5.21a


Figure 5.21 b


Filggure 5.21c



Figure 5. 5 : 3 : Comppanison of the axial Pattersson functions of the E -DMA Mhosphorouls calculated with wather-corrrected ( -- ) and uncomrected ( $=-$ ) scattering factifs
that the curves for DNA are insensitive to the effect of water. The phosphorqus Pattersons consist of just two peaks: the origin and one at $z^{\prime}=12.5 \mathrm{~A}$, which cQrresponds to vectors crossing the minor groove.

### 5.5.2 A-DNA

The axial Patterson functions of A-DNA, which were calculated from the observed and predicted intensities of Fuller (1961), are shown in figure 24. The Patterson co-efficients are given in table 7. It is noteworthy that the positions of the peaks are not related to the internucleotide spacing in contrast to the case of B-DNA. There is excellent agreement between the two curves (which is superior to that achieved by any other conformation) except in the region close to $z^{\prime}=c / 2=14 \mathrm{~A}$ where the predicted intensities require a plateau but the observed intensities show a large peak. This region is dominated by vectors crossing the major groove which is hollow in A-DNA and is therefore probably filled by water and ions. It is possible that inadequate allowance for this in the fourier transform calculations may give rise to the observed anomoly in the Patterson function. Alternatively a small change in the base tilt may serve to reduce the discrepancy in this region without seriously worsening the agreement elsewhere.

### 5.5.3 D-DNA

The axial Patterson functions of D-DNA (computed from the observed and predicted intensities of Arnott et al (1974)) are shown in figure 25. The Patterson co-efficients are given in table 8. Although the amplitudes of the two curves agree relatively well, the positions of the peals are different. As in the case of A-DNA the position of the peaks is not related to the internucleotide separation.

The curves calculated from two left-handed $\beta$-stacked $D$ forms are compared with that calculated from the observed intenstities in figures

Table 5.7 : The Patterson Co-efficients of A-DNA

| $\ell$ | Observed | Calculated |
| :---: | :---: | :---: |
| 0 | 115 | 124 |
| 1 | 87 | 110 |
| 2 | 194 | 152 |
| 3 | 39 | 46 |
| 4 | 93 | 46 |
| 5 | 25 | 23 |
| 6 | 130 | 132 |
| 7 | 126 | 149 |
| 8 | 159 | 141 |
| 9 | 27 | 43 |
| 10 | 0 | 0 |
| 11 | 8 | 33 |
| $\sum$ | 1003 | 999 |

Table 5.8 : The Pattersqn CQ-efficients of D-DNA

|  | Arnott Model |  |  |
| :---: | :---: | :---: | :---: |
|  | Observed | Calculated | BLHD |
|  |  |  |  |
| 0 | 146 | 206 | 170 |
| 1 | 43 | 71 | 80 |
| 2 | 35 | 22 | 34 |
| 3 | 80 | 102 | 93 |
| 4 | 150 | 114 | 180 |
| 5 | 96 | 39 | 84 |
| 6 | 100 | 163 | 148 |
| 7 | 304 | 204 | 198 |
| 8 | 0 | 0 | 0 |
| 9 | 46 | 79 | 12 |
| $\sum$ | 1000 | 1000 | 999 |
|  |  |  |  |



Figure 5.24 : Axial Patterson functions of $A-D N A$
Calculated from the observed (-) and predicted (---) intensities of Fuller (1961)


Figure 5.25 : Axial Patterson functions of D-DNA
Calculated from the observed ( - ) and predicted ( --- ) intensities of Arnott et al (1974)


Figure 5.26 : Axial Patterson function of left-handed, $\beta$-stacked D-DNA (-) compared with that of observed intensities of Arnott et al (1974)


Figure 5.27 : As for figure 26 except that the predicted intensities were derived from a later, improved model

26 and 27. The model used in figure 26 is the same one whose cylindrical Patterson function was described earlier whereas that used in Figure 27 is a latter model whose cylindrical Patterson has not yet been calculated. The cq-efficients are given in table 8. A detailed description of the model will be presented elsewhere (Mahendrasingam et al, manuscript in preparation). The predictions of both models are in close agreement with the curve calculated from the data. The largest discrepancy between the earlier model and the data occurs at $z^{\prime}=2 A$ whereas in the later model agreement is better at this point, but worse at $z^{\prime}=10.5 \mathrm{~A}$.

### 5.6 Discussion

The ajms of this chapter have been twofold. First, to evaluate the usefulness of the cylindrical and axial Patterson functions in the determination of nucleic acid structure; and second, to examine the claims by Bates and co-workers that the accepted A-DNA model is in disagreement with the cylindrica Patterson function of the data and that the axial Patterson function of B-DNA indicates that the molecule is in a side-by-side rather than double helical conformation.

A major obstacle in theuse of the cylindrical Patterson function is that no objective way may be easily defined with which to assess the measure of agreement between maps based on calculated and observed intensities. In addition there is considerable detail in most of the maps, but it is not clear if it all represents real information or is an artefact of the analysis. Therefore we chose to concentrate on the major features: (i) peaks representing vectors between phosphorous atoms and (ii) peaks arising from lattice or pseudolattice vectors.

The prediction of the positions of peaks corresponding to vectors between phosphorous atoms on the same chain according to the theory derived in section 2 was successful in the case of A-DNA, A-RNA and A'-RNA but less so in the case of B-DNA and D-ONA. In all the maps vectors between phosphorous atoms on different chains frequently fell in regions of low
amplitude but there was always good agreement in such cases between the maps from the observed and predicted intensities. As a result, given only the observed intensities of an unknown structure, it would be unwise to attempt to determine such parameters as the helical symmetry and radius on the basis of a cylindrical Patterson map alone. Prediction of the interchain separation parallel to the molecular axis from the position of intense peaks long the $z^{\prime}$-axis is also prone to error: only in the case of $A^{\prime}-$ RNA does such a peak accurately predict the chain separation of the accepted model; the D-DNA and B-DNA predictions are approximately correct (but note that although the two D-DNA models discussed here have different chain separations, they both agree equally well with the map from the data) and A-RNA is quite wrong. It is clear therefore that vectors between phosphorous atoms are not always dominant in cylindrical Patterson maps.

It is equally difficult to draw definite conclusions concerning crystal symmetry. Only in the case of D-DNA, A-RNA and A'-RNA are all the lattice vectors unambiguously present. It is generally unclear whether the indexing which has been assigned is unique but the A-DNA map does suggest that the accepted lattice is correct whereas that proposed by Sasisekharan and co-workers is less successful in predicting peak positions.

We now consider the relevance of the results in this chapter to the side-by-side model of DNA. In the previous chapter we argued that SBS-like models of other allomorphs of DNA would need to be determined if the SBS rather than the doublehelical model were to describe the general structure of DNA. Although the model of Radley et al (1976) was proposed as the structure of $B-D N A$, Bates et al (1978) suggested that the cylindrical Patterson map of A-DNA contained features which were also at variance with the predictions based on the double helix, One of these features concerned the absence of vectors which might be expected at $\left\langle r^{\prime}, z^{\prime}\right\rangle=(19,12)$ if the phosphorous-phosphorous vectors were the most
significant in the map. But we have shown that this assumption is unfounded and that the model of Fuller et al(1965) predicts their absence, The second feature occurs in the axial Patterson region where the agreement between the observed and predicted vectors is excellent except at $z^{\prime} \simeq c / 2$. It is unlikely that this agreement is fortuitous and so we suggest that the discrepancy may arise from errors in the intensities or be remedied by small modifications in the molecular structure or by accounting for the effect of ions and water in the Fourier transform calculations. Neither of the points raised by Bates et al refutes the A-DNA model of Fuller et al (1965).

The major emphasis of the work of the New Zealand group has concerned the axial Patterson function of B-DNA. They have ignored the high quality $X$-ray data available from crystalline fibres and instead they have concentrated on the fibrous patterns obtained by Bram (unpublished), Feughelman et al (1955) and Zimmerman and Pheiffer (1979), They have calculated the $P_{\ell}$ co-efficients from microdensitometer traces of the patterns. In addition they have calculated the axial Patterson functions of B-DNA using the structure factors published by Arnott and Hukins (1973). We reproduce the results of their calculations in figures 29 and 30 to facilitate comparison with the work described here. Figure 29 shows a comparison of their Patterson curve of the observed data of crystalline B-DNA with the average of the curves calculated from the observed fibrous data. Figure 30 shows their predicted Patterson functions of both double helical B-DNA and the SBS mqdel compared with the average Patterson of the fibrous data. They claim that this figure indicates that the double helix is incapable of accounting for the average Patterson.

This work may be criticised on several points. We argued in the previous chapter that there were strong indications that the molecular assemblies which gave the three fibrous patterns studied by Bates and co-workers exhibited different types of molecular disorder. Since this
disorder manifests itself in the diffraction pattern it is to be expected that it will also affect the Patterson function. It is unlikely therefore that the average Patterson function of these pattersn is a suitable curve against which to compare those predicted by models.

Bates et al do not describe how they determined the co-efficients of double helical and SBS DNA used to calculate the curves in figure 30. Since it is not pQssịble to calculate structure factors for the SBS model the co-efficients were probably derived from the cylindrically averaged intensity transform of the model. If this is so then the same procedure should have been adopted for the double helical model which may explain the large discrepancy between the Watson-Crick model curves in figures 19a and 30. However, we adopted this procedure when calculating the axial Patterson co-efficient of SBSO (figure 22 ) but the result is quite different from the curve of Bates et al (figure 29) since we found all the peaks to have the same amplitude. A further discrepancy exists between the axial Patterson of the crystalline data of Arnott and Hukins (1973) presented here (figure 19a) and by Bates et al (figure 29).

Figure 31 shows two typịcal densịtometer traces (provided by Dr. G.A. Rodley) of the Bram pattern from which Bates et al determined the Patterson co-efficients of fibrous DNA. The first layer-plane shows an intense peak at the meridion. This diffraction is almost certain to be due largely to low-angle background scatter, (If it is a true feature of the molecular transform then both the double helical and SBS models must be incorrect since neither predicts such a peak). In addition the second layer-plane contains a peak whose shape is quite unlike that normally seen in such patterns, Finally, in all the traces provided the baseline has been assumed to be constant which is unlikely to be correct, One would generally expect the background to be at fts highest near the main leam and to fall off as the scattering angle is increased. It is worthwhtle to examine the effect of baseline errors on the axial

Patterson function in order to determine whether they may affect the conclusions of Bates et al.

In the Appendix expressions are derived which describe the effect of the background intensity on the Patterson functions. Figure 28 shows the predicted discrepancy for B-DNA calculated from equation A20. Three curves are shown : (i) with identical background on each layer-line, i.e. $f_{\ell}=1, \ell=0-10$; (ii) with $f_{\ell}=1 /(\ell+1)$; and (iti) with $f_{\ell}=1 /(\ell+1)^{2}$. The second two cases are included to show the effect of a background intensity which decreases as both $R$ and $\ell$ increase. These examples are intended to be representative rather than exact physical models of the background. The discrepancy functions shown in the figure would be added to the true axial Patterson to give the axial Patterson of the uncorrected intensities. Each of the curves shows a tendency to enhance the size of the measured axial Patterson at low values of $z^{\prime}$. This is precisely the effect observed at $z^{!}=3.4 \mathrm{~A}$ in the curves for fibrous DNA calculated by Bates et al. The curves in figure 28 simply show the form of the effect of background intensity but the size of the constant $K$ in equation $A 20$ determines the relative amplitudes of the true and background axial Patterson curves: if the background is low then $K$ will be small and the correction will be negligible. It is not possible to assess the value of $K$ without access to the original diffraction patterns but this analysis does indicate the danger inherent in uncritical use of continuous intensity data. If one concentrates on crystalline data as we have done in this chapter the baseline error, whilst still probably the greatest source of error in the intensity measurements, is less significant. Our calculations show that different crystalline data sets for B-DNA produce similar, but çertafnly not identical, axial Patterson curves (the differences being indicative of the effect of errors in the intensities) and that the agreement between the observed and predicted Pattersons is


Figure 5.28 : The effect of baseline errors in intensity measurements on axial Patterson functions. The curves are calculated using B-DNA parameters from equation $\mathbf{A} 20$.

$$
\begin{array}{ll}
\text { Top } & : f_{\ell}=1 \quad \text { for all values of } \ell \\
\text { Middle }: f_{\ell}=1 \\
\text { Bottom }: & f_{\ell}=1 /(\ell+1) \\
& =1 \ell+1)^{2}
\end{array}
$$



Figure 5.29 : Comparison of the "average" axial Patterson function of B-DNA from various "fibrous" patterns (-) with the axial Patterson function of the crystalline intensities observed by Arnott and Hukins (1973) (---). (From Bates et al, 1980).


Figure 5.30 : Axial Patterson function of SBS36 (--) and double helical B-DHA (.......) compared with the average curve from figure 29 (-).
(From Bates et al, 1980).


Figure 5.31 : Densitometer traces (provided by Dr, G.A. Rodley) of two layer-lines of the B-DNA pattern obtained by Bram (unpublished). Note that the baseline ( --- ) has been assumed to be constant.
good for each of the Arnott, Langridge and Marvin models. This indicates that the limited information which may be extracted from the axial Patterson function does not disagree with the predictions of the double helical model.

### 5.7 Conclustons

The Patterson functions described here have been shown to be of limited utility in the analysis of nucleic acid structure, We find that such analyses are likely to be unreliable sịnce even such gross parameters as interchaìn separation and lattice constants may not be determined with confidence. There would be little support for the contention that the SBS model is more successful than the double helix in accounting for the Patterson function of the observed B-DNA diffraction pattern even if accurately determịned crystalline intensities were used, The contrary argument by Bates and co-workers is not only based on poor quality diffraction data, but it also relies on the assumption that disorder within the fibre may be ignored. Therefore their conclusions are unreliable.

## APPENDIX TO CHAPTER V

The Effect of Errors in the Measured Intensities on the Patterson Functions

In this appendix we derive expressions which describe the effect of errors in the intensity measurements on the Patterson functions. Our principal aim is to examine the error introduced into the axial Patterson function by incorrect allowance for the background intensity. We will concentrate on measurements of continuous intensity rather than Bragg reflections so, whilst we shall derive equations applicable to the cylindrical Patterson function, we shall not explore their implications since all such maps presented in this chapter have utilised crystalline intensity data.

Suppose that there are randomly distributed errors of magnitude $\Delta I_{\ell}(R)$ in the measured intensities so that:

$$
\begin{equation*}
M_{I_{\ell}}(R)=T_{I_{\ell}}(R)+\Delta I_{\ell}(R) \tag{AI}
\end{equation*}
$$

Where the letters $M$ and $T$ refer to the measured and true values respectively of the intensity I at point $R$ on layer-1ine $\ell$. Equation 5.38 shows that the axial Patterson co-efficients are given by:-

$$
\begin{equation*}
P_{\ell}=\int_{0}^{\infty}\left[T^{T} I_{\ell}(R)+\Delta I_{\ell}(R)\right] R d R \tag{A2}
\end{equation*}
$$

But since the errors are random it follows that:-

$$
\begin{equation*}
\int_{0}^{\infty} \Delta I_{\ell}(R) R d R=0 \tag{A3}
\end{equation*}
$$

so

$$
\begin{equation*}
P_{\ell}=\int_{0}^{\infty} T_{I_{\ell}}(R) R d R \tag{A4}
\end{equation*}
$$

which shows that the axial Patterson function will not be affected by randomly distributed errors. Such errors would affect the cylindrical Patterson since the integrand in equation $A 2$ would contain $J_{0}\left(2 \pi R r^{\prime}\right)$ and so the integral would not necesśarily be zero.

The major error in the intensities is likely to be a systematic one arising from uncertainty in the position of the baseline. Suppose the background intensity is described by ${ }^{8} I_{l}(R)$ so that the measured intensity is given by:-

$$
\begin{equation*}
M_{I_{\ell}(R)}=T_{I_{\ell}(R)}+{ }^{B} I_{\ell}(R) \tag{A5}
\end{equation*}
$$

Substitution of this into equation 5.5 shows that the measured Patterson co-efficient ${ }^{M_{P}}{ }_{\ell}\left(r^{\prime}\right)$ contains a term related to the background intensity:

$$
\begin{align*}
M_{P_{\ell}}\left(r^{\prime}\right) & =\int_{0}^{\infty} M_{I_{l}}(R) J_{Q}\left(2 \pi R r^{\prime}\right) R d R \\
& =\int_{0}^{\infty}\left[T_{I_{\ell}}(R)+B_{I_{l}}(R)\right] J_{0}\left(2 \pi R r^{\prime}\right) R d R \tag{A6}
\end{align*}
$$

$$
=T_{P_{\ell}}\left(r^{\prime}\right)+{ }^{B} P_{P_{\ell}}\left(r^{\prime}\right)
$$

He may determine the effect of the background on the Patterson function by assuming the form of ${ }^{B}{I_{f}}(R)$ and substituting into $A 6$. We consider two simple cases.

In the first case we let the background be a constant $\mathrm{K}_{1}$ along each layer-line, so:-

$$
B_{F_{\ell}}(R)=K_{1}
$$

Then

$$
\begin{align*}
B_{P_{\ell}}\left(r^{\prime}\right) & =K_{1} \int_{0}^{\infty} J_{0}\left(2 \pi R r^{\prime}\right) R d R  \tag{AB}\\
& =K_{2} \frac{\delta\left(r^{\prime}\right)}{r^{\prime}} \tag{AQ}
\end{align*}
$$

where $K_{2}$ is another constant and $\delta$ is the Dirac $\delta$-function.
The first case is unlikely to be physically realistic so instead we may select a function which decreases smoothly as the angle of diffraction increases. For the sake of mathematical simplicity we choose an exponential decrease:-

$$
B_{I_{\ell}(R)}=K_{3} \exp (-a R)
$$

where $\mathrm{K}_{3}$ and a are further constants. Then:-

$$
\begin{align*}
{ }^{B} P_{\ell}\left(r^{\prime}\right) & =K_{3} \int_{0}^{\infty} \exp (-a R) J_{0}\left(2 \pi R r^{\prime}\right) R d R  \tag{AlD}\\
& =\frac{K_{4}}{\left(a^{2}+r^{\prime 2}\right)^{3 / 2}} \tag{All}
\end{align*}
$$

Note that in both these cases we have assumed the background to be identical on each layer-line. It would be more realistic to allow for a drop in the background as $\ell$ increases but this complicates the analysis since we must then perform an integration in cylindrical co-ordinates of a function which varies in a spherical or pseudo-spherical manner. We
prefer to circumvent this by effecting an empirical correction at a later stage.

We now consider the effect of the background on the measured Patterson which is given by:-

$$
\begin{equation*}
M_{P}\left(r^{\prime}, z^{\prime}\right)=\sum_{\ell}\left[{ }^{T} P_{\ell}\left(r^{\prime}\right)+{ }^{B} P_{\ell}\left(r^{\prime}\right)\right] \cos \left(\frac{2 \pi \ell z^{\prime}}{c}\right) \tag{Al2}
\end{equation*}
$$

In the first case, with a constant background along the layer-line, this may be replaced by:-

$$
\begin{align*}
M_{P}\left(r^{\prime}, z^{\prime}\right) & =\sum_{\ell}\left[T_{\ell}\left(r^{\prime}\right)+\frac{K_{2} \delta\left(r^{\prime}\right)}{r^{\prime}}\right] \cos \left[\frac{2 \pi \ell z^{\prime}}{c}\right]  \tag{A13}\\
& =T_{P\left(r^{\prime}, z^{\prime}\right)+\frac{K_{2} \delta\left(r^{\prime}\right)}{r^{\prime}} \sum_{\ell} \cos \left(\frac{2 \pi \ell z^{\prime}}{c}\right)} .\left\{\begin{array}{l}
\end{array}\right) \tag{A14}
\end{align*}
$$

So the background has no effect except at $r^{\prime}=0$, i.e. the axial Patterson region. But in the case of the exponentially varying background the correction ${ }^{B} P_{\ell}\left(r^{\prime}\right)$ is finite at all values of $r^{\prime}$ and so $i t$ has an effect at all points in the cylindrical Patterson map:-

$$
\begin{equation*}
M_{P}\left(r^{\prime}, z^{\prime}\right)=\sum_{\ell}\left[{ }^{T} P_{\ell}\left(r^{\prime}\right)+\frac{K_{4}}{\left(a^{2}+r^{\prime 2}\right)^{3 / 2}}\right] \cos \left(\frac{2 \pi \ell z^{\prime}}{c}\right) \tag{A15}
\end{equation*}
$$

We shall not consider the cylindrical Patterson further here but instead we concentrate on the axial Patterson region:-

$$
\begin{equation*}
{ }^{{ }^{M} P\left(0, z^{\prime}\right)}=\sum_{\ell}\left[{ }^{T} P_{\ell}(0)+K_{5}\right] \cos \left[\frac{2 \pi \ell z^{\prime}}{c}\right] \tag{A16}
\end{equation*}
$$

where $r^{\prime}$ has been set to zero in equation $A 15$ and $K_{5}=K_{4} / a^{3}$, Comparison of equations A14 and Al6 shows that the effects of the two types of
baseline on the axial Patterson functions are formally identical. So the error $\Delta P$ fntroduced into the axial Patterson at $z^{\prime \prime}$ by the background intensity is given by:-

$$
\begin{align*}
\Delta P\left(z^{\prime}\right) & =\left|T^{T} P\left(0, z^{\prime}\right)-M_{P}\left(0, z^{\prime}\right)\right|  \tag{A17}\\
& \left.=\left|\begin{array}{l}
K \\
\ell
\end{array}\right| \frac{2 \pi \ell z^{\prime}}{c}\right] \mid \tag{A18}
\end{align*}
$$

where we have dropped the subscript in the constant which appears in equations A14 and A16. In a typical case the summation runs from $\ell=0$ to $\ell=L$ (where $L$ is the highest observed layer-line number) but it is easier to evaluate if the limits are symmetrical. Using:-

$$
\sum_{\ell=0}^{L} \cos x=\frac{1}{2}\left[1+\sum_{\ell=-L}^{L} \cos x\right]
$$

it is straightforward to show that:-

$$
\begin{equation*}
\Delta P\left(z^{\prime}\right)=\frac{k}{2}\left[1+\frac{\left.\sin \frac{(2 L+1) \pi z^{\prime}}{c}\right)}{\sin \frac{\pi z^{\prime}}{c}}\right] \tag{A19}
\end{equation*}
$$

So the effect of the baseline error is to superimpose upon the true Patterson function a series of ripples similar in form to the diffracted intensity from a finite lattice of point particles. Two points remain to be made. First, equation A19 describes the form of the effect of the background but this correction may be insignificant if $K$ is small. Second, we need to include the effect of the reduction in background intensity as \& increases. We may do this empirically by inserting a factor $f_{\ell}$ in Al9 giving:-

$$
\begin{equation*}
\Delta P\left(z^{\prime}\right)=\frac{1}{2} K f_{\&} \phi\left(z^{\prime}\right) \tag{A2O}
\end{equation*}
$$

where $\phi\left(z^{\prime}\right)$ is the term in square brackets in $A 19$ and $f_{\hat{x}}$ decreases in some well-defined way with increasing layer-line number. The form of A2O is shown in figure 5.23 with the B-DNA parameters $c=34 \mathrm{~A}$ and $L=10$ for a number of different $f_{\&}^{\prime} s$.

## CHAPTER VI

## THE MOLECULAR AND CRYSTAL STRUCTURE OF

DNA FROM BACTERIOPHAGE $\phi W-14$

### 6.1 Introduction

The lytic bacteriophage $\phi w-14$ has the bacterium Pseudomonas acidovorans as host. The detailed phage structure is unknown but it has a regular icosahedral head which is about 850 A in diameter and a contractile tail which is 1400A long and about 200A in diameter. There is a baseplate at the end of the tail which appears to carry pin-like structures (Kropinski and Warren, 1970).

The DNA from $\phi \mathbf{W}-14$ is unusual. Estimates of the $G+C$ content from buoyant density and melting temperature measurements give widely divergent values : 4.5 per cent and 72.9 per cent respectively. This anomoly was resolved by Kropinski et al (1973) who showed by chemical analysis that the $G+C$ content was in fact 56.2 per cent. They explained the discrepancy between the values obtained by physical techniques in terms of the presence of a hypermodified pyrimidine : 5-(4-aminobutylaminomethyl uracil in which a putrescine molecule ( 1,4 diamino-n-butane) is covalently bonded to the uracil at C5 (fig. 1).

Putrescine is an aliphatic diamine which, together with the other polyamines spermine, spermidine and cadaverine (fig. 2), is found in all procaryotic and eucaryotic cells. It is a metabolic precursor of spermine and spermidine. The concentration of polyamines in a cell appears to be related to both the type of cell and the stage to which the organism has developed. Polyamines have been found in association with membranes, organelles, ribosomes and nucleic acid. Although their exact function is


Figure 6.1 : Structure of the hypermodified pyrimidine 5-(4-aminobutylaminomethyl) uracil

Spermine:

$$
\mathrm{H}_{3} \mathrm{~N}^{+}\left(\mathrm{CH}_{2}\right)_{3} \mathrm{~N}^{+} \mathrm{H}_{2}\left(\mathrm{CH}_{2}\right)_{4} \mathrm{~N}^{+} \mathrm{H}_{2}\left(\mathrm{CH}_{2}\right)_{3} \mathrm{~N}^{+} \mathrm{H}_{3}
$$

Spermidine:

$$
\mathrm{H}_{3} \mathrm{~N}^{+}\left(\mathrm{CH}_{2}\right)_{3} \mathrm{~N}^{+} \mathrm{H}_{2}\left(\mathrm{CH}_{2}\right)_{4} \mathrm{~N}^{+} \mathrm{H}_{3}
$$

Putrescine:

$$
\mathrm{H}_{3} \mathrm{~N}^{+}\left(\mathrm{CH}_{2}\right)_{4} \mathrm{~N}^{+} \mathrm{H}_{3}
$$

Cadaverine:

$$
\mathrm{H}_{3} \mathrm{~N}^{+}\left(\mathrm{CH}_{2}\right)_{5} \mathrm{~N}^{+} \mathrm{H}_{3}
$$

Figure 6.2 : Structures of some polyamines
unknown, their abundance in proliferating and embryonic tissues suggests that they may be important in the regulation of template dependent syntheses (Sakai and Cohen, 1976). Polyamines are known to stimulate RNA synthesis (Krakow, 1963) but Shalitin and Sarid (1967) have found that bacterial ONA synthesis after phage infection is stimulated by spermidine and is inhibited by high concentrations of putrescine. Addition of spermine and spermidine to DNA solutions raises the melting temperature ( $T_{m}$ ) of the ONA. Putrescine also raises $T_{m}$ but to a lesser extent (Stevens, 1967). Minyat et al (1978) have used circular dichroism spectra to follow the $B$ to A transition of DNA in water/ethanol solutions in the presence of polyamines. They found that spermine and spermidine tended to stabilise the DNA in the A-family of conformations whereas putrescine, cadaverine and hexamethyl diamine tended to stabilise the B-family. Zhurkin et al (1980) have suggested that these results may be explained by considering the energy of interaction of spermine molecules with DNA in the $A$ and $B$ forms. The spermine, which was placed in the minor groove, was hydrogen bonded to two phosphate groups. The energy of the complex was calculated by the atom-atom potentials method as a sum of the van der Waals interaction within the spermine molecule and that between DNA and spermine, the torsional strain Within the spermine and the hydrogen bond energy. Both the separation and the orientation of the phosphates in A-DNA give rise to an interaction between the two molecules which is energetically more favourable than when the DNA is in the B form.

Tsuboi (1964) was the first to notice that the distance between the amino groups in the trimethylene segment of the spermine chain is approximately equal to the distance between the phosphate groups in successive nucleotides of B-DNA. He proposed that each trimethylene group hydrogen bonded to two phosphates and that the butyl segments of the spermine bridged the wide groove. Subsequent workers (Liquori et al, 1967;

Suwalsky et al, 1969) pointed out that the butyl groups will not bridge the wide groove and therefore all recent models of the complex have placed the spermine in the narrow groove.

Suwalsky et al (1969) carried out an X-ray diffraction study of fibres formed from the sodium salt of calf thymus DNA and spermine. They found the DNA to be in a hexagonal semi-crystalline B form at and above 92\% relative humidity. At and below $76 \%$ relative humidity the DNA gave C-like patterns. No A forms were observed. However, salmon sperm DNA behaved differently (Huse et al, 1978). At a relative humidity of $92 \%$ or above it was in the B form whereas below $75 \%$ it was in the $A$ form. The $A$ to $B$ transition within the fibre was reversible.

The crystal structure of putrescine diphosphate has recently been solved by Woo, Seeman and Rich (1979). The putrescine molecules contain two protonated amino groups both of which form three hydrogen bonds with phosphate groups arranged tetrahedrally about the nitrogen atoms. The crystal, which belongs to space group $\mathrm{P} 2 \mathrm{f} / \mathrm{a}$, consists of hydrophobic layers formed by the butane segment of putrescine, surrounded by hydrophilic layers composed of the amino and phosphate groups. The phosphate groups are each hydrogen bonded to three putrescine molecules and three additional phosphates. Putrescine molecules in the observed conformation have been fitted into double helices of B-DNA (Arnott and Hukins, 1972 b), A-RNA (Arnott and Hukins, 1972 c) and ApU-RNA (Rosenberg et al, 1976) to see whether they can bridge the grooves and form hydrogen bonds with phosphate groups on opposite strands. Putrescine can fit across the grooves of the two RNA helices forming one hydrogen bond with a phosphate group on each strand. In the ApU helix, the putrescine can also bridge the groove and make two hydrogen bonds with one strand and one with the other giving rise to an arrangement similar to that proposed by Liquori et al. (1967) for DNAspermidine complexes. The putrescine molecule in the observed conformation
could not bridge the groove in B-DNA in a satisfactory manner but if all the putrescine torsion angles were trans (as, for example, is observed in the putrescinyl segment of the spermine molecules in the spermine phosphate hexahydrate crystal (Itaka and Huse, 1965)) then a bridge could be formed with two hydrogen bonds.

Many bacteriophages contain DNA with chemically modified bases (Warren, 1980). These modifications give rise to unusual chemical and physical properties and it is possible that they are also of biological significance. For example, they may change the susceptibility of the DNA to nuclease action, alter behaviour during transcription or replication, or confer advantages in terms of telestability, packaging within the phage head or injection of the DNA into the host. It is therefore of great interest to determine the conformation of DNA from such phages in order to see if their structure suggests the functional role of the modifications. The DNA from bacteriophage $\phi \mathrm{W}-14$ is worthy of study not only because it is the first to be discovered which contains a covalently bonded group from the polyamines described above but also because the modified base is electrically charged.

The studies described in this chapter suggest that the chemical modification does not affect the DNA conformation but it does have an effect on the $A$ to $B$ transition. These results are discussed in terms both of the likely effect of the charged group on the role of ions in the fibre and the possibility that a hydrogen bonding bridge might stabilise the structure. Molecular models have been built and their calculated diffraction patterns are compared with the observed data.

The work described in this chapter has been performed in conjunction with Drs. D.C. Goodwin and C. Nave in this laboratory and Prof. R.A.J. Warren of the University of British Columbia, Vancouver. The work at Keele is currently being extended by Mr. A, Mahendrasingam.

### 6.2 Experimental Studies

6.2.1 Methods

### 6.2.1.1 Phage Extraction and DNA Purification

Bacteriophage $\phi W-14$ DNA was prepared and partially purified as described by Kropinski et al (1970). Calf thymus DNA, for use as a control, was from Miles Laboratory Incorporated. Both types of DNA were further purified by phenol extraction as described in Chapter 2.

DNA gels were obtained by ultracentrifugation of a solution of the DNA ( 1 mg in 10 mls of buffer). The buffer solutions contained either 0.01 M or 0.02 M NaCl and 0.002 M Tris-HCl, pH 7.6

To obtain the lithium salt, excess lithium chloride solution was added to DNA solutions and the DNA was precipitated by adding propanol. The LiDNA was then redissolved in a tris-lithium chloride buffer and a gel was made by ultracentrifugation.

### 6.2.1.2 Preparation of Acetylated $\phi W-14$ DNA

\$W-14 DNA was dialysed exhaustively against 0.1 M triethanolamide hydrochloride, pH 8.0 , and it was then treated at room temperature with a thousand fold molar excess (based on the molarity of the DNA putrescinyl groups) of acetic anhydride. The acetic anhydride was added in small portions to a magnetically stirred DNA solution and the pH was monitored with a Pye 79 pH meter. The pH of the solution was maintained at 8.0 by addition of 4 M NaOH from a syringe. The acetylated DNA was dialysed against $0.01 \mathrm{M} \mathrm{NaCl}, 0.002 \mathrm{M}$ Tris $-\mathrm{HCl}_{\ell}, \mathrm{pH} 7.6$. The extent of acetylation was determined by estimation of the unreacted amino groups with trinitrobenzene sulphuric acid (Fields, 1971).

### 6.2.1.3 X-Ray Methods

The preparation of fibres, recording of diffraction patterns and refinement of lattice parameters have been described in Chapter 2.

### 6.2.1.4 Measurement of the Sodium to Phosphate Ratio

The measurement of the ratio of sodium ion to phosphate ion content in a fibre has been described in detail by Blakeley (1976). In the present work, the gels were redissolved in 3 mls of distilled water. The sodium concentration was determined with a Unicam SP1900 flame emission spectrophotometer using the radiation emitted at 589 nm . The corresponding phosphate concentration was determined using a Cary 118 spectrophotometer assuming a value of $6600 \mathrm{M}^{-1} \mathrm{~cm}^{-1}$ for the extinction co-efficient of a nucleotide (average molecular weight $=330 \mathrm{da}$ ) at 260 nm .

### 6.2.1.5 Melting Temperature Measurement

The melting temperatures of normal and acetylated $\phi w-14$ DNA were measured simultaneously (Mandel and Marmur, 1968) using a Gilford 2400 automatic recording spectrophotometer fitted with a Model 2417 thermosensor.

### 6.2.2 Results

X-ray diffraction patterns of fibres of $\phi \mathbf{W}-14$ DNA are shown in plates 1-5. Plate 1 shows a crystalline $A$ pattern and plate 2 a semicrystalline B pattern from the sodium salt. Plate 3 shows a crystalline B pattern and plate 4 a $C$ pattern from the lithiun salt. Plate 5 shows a C pattern from the sodium salt. Similar patterns were obtained from calf thymus DNA.

The $C$ confomation of DNA can exist in a number of related foms distinguished by the number of nucleotide-pairs per helix pitch and the nature of the molecular packing (Marvin et al, 1961). The $C$ patterns obtained from $\phi W-14$ DNA and calf thymus DNA were found to lie within the range previously observed from DNA in this family of conformations. In addition, $C$ patterns were obtained from the sodium salt which were similar to those now commonly observed with other DNAs (Arnott and Selsing, 1975; Leslie et al, 1980; Rhodes and Mahendrasingam, unpublished results).

The diffraction patterns observed for the $A$ and $B$ conformations of OWA are much better defined than those for the $C$ form. The unit cell


Plate 6.1 : A-form Diffraction Pattern of NaDNA from Bacteriophage $\phi w-14$


Plate 6.2: Semi-crystalline B-form Diffraction Pattern of NaDNA from Bacteriophage $\phi \mathrm{w}-14$


Plate 6.3 : Crystalline B-form Diffraction Pattern of
LIDNA from Bacteriophage $\phi W-14$


Plate 6.4 : C-form Diffraction Pattern of LiDNA
from Bacteriophage $\phi w-14$


Plate 6.5 : C-form Diffraction Pattern of NaDNA
parameters were determined and refined from measurements on the crystalline reflections in these patterns. They are compared in table 1 with the values obtained in the detailed analyses of these conformations. Line 1 of the table shows the refined values obtained by Fuller (1961) for the monoclinic A-DNA lattice. In order to test the lattice refinement program, this calculation was repeated with the results shown in line 2 . The refined values obtained for the $\phi w-14$ lattice are shown in line 3 . It is clear from comparison of lines 2 and 3 that the calf thymus and $\phi w-14$ A-DNA lattices are very nearly isomorphous. The slight discrepancies may be explained by the quality of the patterns used in determining the lattice. The pattern used by Fuller (1961) was extremely well oriented and highly crystalline whereas those obtained from $\phi W-14$ DNA are less well defined. In addition, fewer reflections were used in the $\phi \mathbf{w}-14$ lattice refinement. The lattice parameters of orthorhombic B-DNA obtained by Langridge et al (1960a) and those determined for $\phi W-14$ DNA in the B-form (compared in lines 4 and 5 of table 1) are also approximately isomorphous. The correctness of the orthorhombic refinement program was confirmed accidentally. The initial parameters for the refinement were taken from table $1^{\prime}$ of Langridge et al (1960a). However, the $a$ and $b$ values in this table have been mistakenly transposed. Despite starting from incorrect values, the program refined quickly to the accepted values. The semicrystalline patterns obtained from B-DNA are usually of poorer quality than those from crystalline specimens and this is reflected in the larger errors evident in the lattice parameters of the former. Within the limits of experimental error the helix pitch and molecular packing in the crystalline $A$ and $B$ forms and the semi-crystalline $B$ form of $\phi W-14$ are identical to the corresponding parameters for calf thymus DNA. The lattices of the $C$ form were not refined.

Since the lattices of the two types of DNA are so similar, it is

TABLE 6.1 : Refinement of lattice parameters
(i) A-DNA

|  | $a(A)$ | $b(A)$ | $c(A)$ | $B(\operatorname{deg})$ |
| :--- | :---: | :---: | :---: | :---: |
| (1) Fuller (1961) | 22.24 <br> $( \pm 0.06)$ | 40.62 <br> $( \pm 0.10)$ | 28.15 <br> $( \pm 0.16)$ | 97.0 <br> $( \pm 0.4)$ |
| (2) Refined by <br> present author <br> using Fuller <br> (1961) data | 22.24 <br> $( \pm 0.07)$ | 40.61 <br> $( \pm 0.12)$ | 28.16 <br> $( \pm 0.12)$ | 97.0 <br> $( \pm 0.3)$ |
| (3) фw-14 | 22.78 <br> $( \pm 0.07)$ | 40.31 <br> $( \pm 0.12)$ | 28.33 <br> $( \pm 0.09)$ | 97.3 <br> $( \pm 0.3)$ |

(ii) Crystalline B-DNA

|  | $a(A)$ | $b(A)$ | $c(A)$ |
| :---: | :---: | :---: | :---: |
| (4) Calf thymus <br> (Langridge et al <br> 1960a) | 31.22 <br> $( \pm 0.1)$ | 22.72 <br> $( \pm 0.1)$ | 33.70 <br> $( \pm 0.1)$ |
| $(5) \quad \phi W-14$ | 31.34 <br> $( \pm 0.16)$ | 23.50 <br> $( \pm 0.12)$ | 34.00 <br> $( \pm 0.18)$ |

(iii) Semi-crystalline B-DNA

|  | $a(A)$ | $c(A)$ |
| :--- | :---: | :---: |
| (6) Calf thymus <br> (Langridge et al, 1960a) | 46.0 | 34.6 |
| (7) $\phi w-14$ | 47.9 <br> $( \pm 1.0)$ | 32.3 <br> $( \pm 0.7)$ |

reasonably simple to compare the intensities visually. The general overall similarity of the intensity distributions suggests that the presence of the modified base does not induce changes in the conformation of the \$W-14 DNA in fibres.

A difference between calf thymus and $\phi W-14$ DNA was apparent in the induction of the $A$ to $B$ transition. Table 2 shows that the A form tends to persist to much higher humidities for $\phi w-14$ DNA than it does for calf thymus DNA. However, the transition depends not only on the relative humidity of the fibre but also on the ionic content (Cooper and Hamilton, 1966). Therefore, in making this comparison, great care was taken to ensure that in preparing gels from which fibres were to be arawn both DNA's were centrifuged fron solutions of identical ionic strength. The ionic strength of the initial solution affects the ionic content of the fibres produced from it and a number of experiments were performed in which this initial ionic strength was systematically varied. The results summarised in table 2 are for an initial concentration which gave fiores with an ionic content close to that which had previously been found to be optimum for observing the $A$ to $B$ transition when the relative hunidity of the fibre environment was increased from $75 \%$ to 92\%. For ionic contents significantly less than this, the A conformation for calf thymus DNA will persist to relative humidities of $92 \%$ or even higher. The $B$ conformation is favoured in fibres which contain excess salt.

The modification of the $A$ to $B$ transition in $\phi W-14$ DNA could have resulted, for example, either from a direct hydrogen bonding interaction between putrescinyl and phosphate groups tending to stailise the A form, or from a decreased sodium ion content in the fibres. In the original studies on the $A$ to $B$ transition as a function of salt content, the excess salt was estimated by the chloride content (Cooper and Hamilton, 1966). In

TABLE 6.2 : Variation of $\phi W-14$ and Calf thymus DNA conformation with relative humidity

| Relative <br> humidity (\%) | фW-14 DNA | Calf thymus DNA |
| :---: | :---: | :---: |
| 66 | A | A |
| 75 | A | A (occasionally B) |
| 92 | A | B (occasionally A) |
| 95 | B or A/B | B |
| 98 | B | B |

TABLE $6.3:\left[\mathrm{Na}^{+}\right] /\left[\mathrm{PO}_{4}^{-}\right]$of $\phi \mathrm{w}-14$ and Calf thymus DNA in gels from two different salt concentrations

|  | Initial salt concentration <br> 0.01 M |  |
| :--- | :---: | :---: |
| $\phi W-14$ | 0.42 | 0.52 |
| Calf thymus | 0.88 | 1.24 |

All values are $\pm 0.09$
From Goodwin (1977)

West Yorkshire, LS23 7BQ

www.bl.uk

# $3^{\text {RD }}$ PARTY COPYRIGHT MATERIAL EXCLUDED FROM DIGITISED THESIS 

$$
\begin{gathered}
\text { PLEASE REFER TO THE } \\
\text { ORIGINAL TEXT TO SEE } \\
\text { THIS MATERIAL }
\end{gathered}
$$

