STUDIES ON DNA POLYMORPHISM

by

David Charles Goodwin

Being a thesis

Submitted to the University of Keele

for the Degree of Doctor of Philosophy

Department of Physics,
University of Keele,
Keele, Staffordshire,
England.

April 1977

ABSTRACT

Possible variations in the secondary structure and the A $\rightleftharpoons$ B transition in DNA's of varying primary base sequence and composition have been studied by the techniques of X-ray diffraction and molecular model building. The DNA's studied are from both bacteria and from eukaryotic cells. In addition a DNA of viral origin, from the bacteriophage ØW-14, has been investigated.

A computerised model building study of the changes induced in DNA secondary structure by the binding of intercalating drugs has also been carried out.

A linked atom least squares routine has been extended and used to refine the models presented. The routine enables standard values for the parameters defining the covalent stereochemistry of the structure to be retained. Methods of calculating the Fourier transforms of the models produced are discussed, and this enables some comparisons to be made between the observed diffraction data and those predicted by the models.

Structures studied include the intercalation complexes involving ethidium or daunomycin, general intercalation models for DNA and models for the conformation of the putrescene groups in ØW-14 DNA.

ACKNOWLEDGEMENTS

I would like to thank the Science Research Council for a Scholarship for training in Research Methods. I am deeply grateful to Professor W. Fuller for the provision of excellent research facilities, for allowing me the opportunity to participate in the research described here, and for his supervision, advice and criticism throughout the duration of this project. I would particularly like to record my gratitude to Dr. W.J. Pigram for his unfailing patience and for his continual readiness to give advice, criticism and encouragement in times of stress. I should like to thank Dr. C. Nave for his cheerful and enthusiastic help in all aspects of the work of this project. I gladly record my appreciation of the advice and discussion given by many colleagues, particularly Professor R.A.J. Warren, Drs. P.J. Blakeley, E.F. Slade, H. Porumb and T. Porumb. My sincere gratitude is given to all members of the Computer Centre and of the Department of Computer Science at the University of Keele for their unfailing assistance and advice. I should also like to thank all members of the Workshop of the Department of Physics for their cheerful assistance whenever required. Invaluable advice on the drawing of diagrams was given by Mr. N. Banks. All photographic work in this thesis was by Mr. M. Daniels. I should like to thank Mr. David Emley of the Department of Geology, University of Keele for his help with use of the Flame Emission Spectrometer. I sincerely appreciate the work of Mrs. S. Jackson who carefully typed this manuscript. I would finally like to thank my parents for their foresight and their unfailing support and encouragement which have made this thesis possible.

CONTENTS

CHAPTER I

INTRODUCTION

1.1 DNA Structure and Function

Since Watson and Crick (1953a, 1953b) proposed their model for the double helical structure of DNA and a mechanism for its replication, a number of conformations have been identified and their structures elucidated in detail. Their theory of DNA replication involved the proposal of a specific pairing scheme between bases in each stand of the duplex, in which adenine or guanine in one chain is hydrogen bonded to a thymine or cytosine respectively in the other chain. All DNA conformations whose structures have been elucidated incorporate base pairs having Watson and Crick geometry. The importance of the base pairing scheme with regard to DNA replication is that it allows the DNA to act as its own template. Previous theories of DNA replication (Pauling and Delbruck, 1940; Fridrich-Freska, 1940; Muller, 1947) had postulated the existence of complementary sequences between proteins and nucleic acids. Under these schemes, protein synthesis and DNA replication were considered as occurring concomitantly in a series of processes involving the alternate synthesis of protein and nucleic acid segments.

The scheme proposed by Watson and Crick is semiconservative in that one of the parental strands is conserved in each of the progeny. Meselson and Stahl (1958) showed that DNA replication was in fact semiconservative and the scheme proposed by Watson and Crick (1953b) for specific base pairing and for a replication mechanism involving an internal template is now widely accepted.

The Watson Crick base pairing scheme also solves, in a very elegant manner, an apparent conflict in the properties required of a genetic molecule. On the one hand it must allow great variation in some aspect of its structure in order to code for all the genetic information necessary to the development and function of a particular organism; while, unless the process by which the information is read is extremely complex and variable, the molecule must in some ways be extremely regular. Both the adenine: thymine (A-T) and the guanine cytosine (G-C) base pairs proposed by Watson and Crick have very similar overall geometry, which evokes the possibility that the secondary structure adopted by the DNA duplex might be identical whatever base sequence is present. This regularity of secondary structure could allow a standard mechanism for reading the genetic information to be used; while the genetic information can be coded for in terms of the base sequence in each polynucleotide chain. The genetic code must consist of at least one distinct "word" for each of the naturally occurring amino acids which constitute proteins. Since there are only four different kinds of base in DNA, it is not possible for the genetic "words" to be composed of one base. A greater number of words may be produced by taking sequences of more than one base to code for an amino acid. The minimum number of bases which can produce a sufficient variety of words is three, and Crick et al. (1961) have shown that such base triplets (codons) are the units in which genetic information is expressed.

Although the above scheme shows how DNA combines great regularity with flexibility to allow its function as a carrier of genetic information, it does not provide insight into methods by

which the expression of genetic information may be controlled. Such control is clearly exerted in differentiated cells of higher organisms in which only a certain portion of the genome is expressed in any particular type of cell. In higher organisms, DNA is present as chromatin, a structure composed of DNA complexed with specific proteins, particularly histones. It is probable that control of gene expression is exerted by means of this bound protein. An example of a protein attaching to DNA to control the expression of a series of genes in a bacterial cell is provided by the repressor protein of the lac operon system in E. Coli. At present it is unclear how the inhibition produced by such a protein is exerted at the molecular level, or how the specific DNA sequence to which it attaches is recognised.

It is of interest with regard to these problems to determine the way in which small molecules attach to DNA and to elucidate the changes, if any, induced in the DNA conformation at the site of attachment. The validity of the dogma that DNA secondary structure is invariant with base composition should also be reassessed, since small variations in secondary structure along the length of a DNA duplex corresponding to regions of differing base sequence might well provide a structural basis for recognition sites for molecules such as the lac repressor which control gene expression.

1.1.1 Diffraction Patterns from Fibres and their Analysis

It is difficult to prepare macroscopic crystals of DNA, and the most ordered state easily obtainable is that of a fibre in which the DNA is packed regularly in small domains (crystallites) which have a random azimuthal orientation about the fibre axis. The crystallites are packed with their c axes (the axis of the DNA

helix) aligned approximately parallel to the fibre axis. Since the alignment is not perfect, the Bragg reflections from the crystallites are drawn out into arcs and the spots tend to merge at higher diffraction angles. Intensity data from diffraction patterns can be measured most accurately when they consist of distinct Bragg diffraction spots which do not overlap. Accurate diffraction intensity data tend to be limited, therefore, even in the most ordered case, in comparison with equivalent data from single crystals.

A number of techniques have been developed which maximise the amount of information which can be derived from the limited data available. The diffraction theory for helical molecules formulated by Cochran et al. (1952), also Stokes (1952, unpublished), shows how the helical parameters may be derived from a relatively superficial examination of the diffraction data. By use of this information in conjunction with knowledge of the stereochemical parameters (bond lengths and bond angles) it is often possible to build a molecular model which can then be refined. Such stereochemical information is normally available from related small molecules whose structure has been solved to atomic resolution by means of their single crystal diffraction data. The inclusion of known stereochemical information in this way compensates to some extent for the information deficiency in the diffraction data. A number of computer routines have been developed, in particular Arnott and Wonacott (1966), which allow the derivation of atomic coordinates and the subsequent refinement of the model to be performed analytically. This is analogous to the least squares routines used in single crystal structure refinements. A computerised model building routine has been used in a number of studies described in this thesis. It will be discussed in chapter 3.

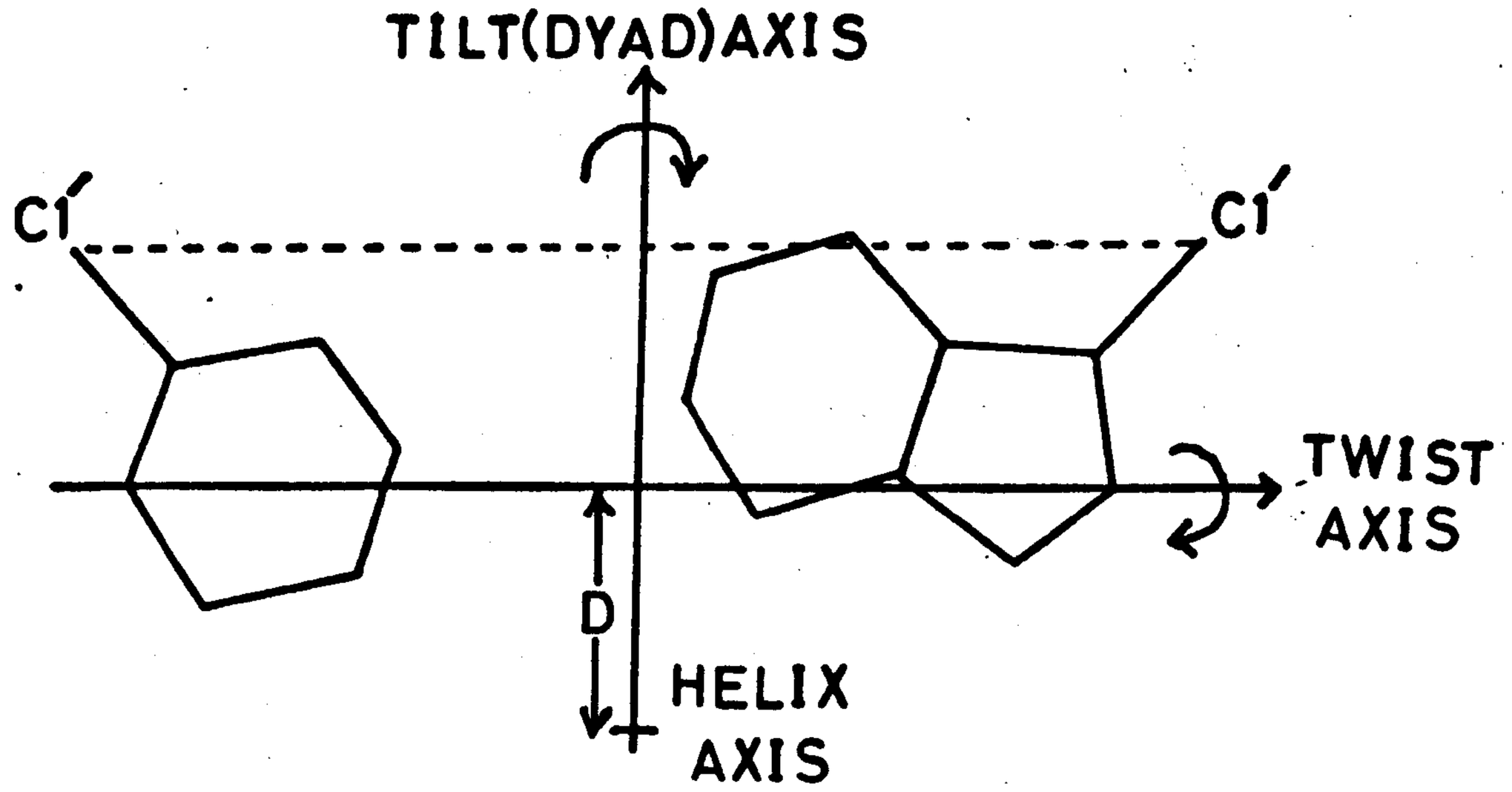


Fig. 1·1 Parameters Defining Base Disposition To The Helix Axis. Positive Values of The Parameters Are Indicated.

1.1.2 DNA Conformation in Fibres

The structures of a number of DNA conformations are now known in detail; while some others have been suggested without their detailed structures having been elucidated. A list of the published conformations for DNA are given in Table 1.1. All work on structure determination was originally performed using natural DNA, though later work also included investigation of the conformations adopted by a number of synthetic DNA's of known base sequence. Early diffraction studies established the fact that there were three distinct conformations, the A, B and C forms, for natural DNA (Franklin and Gosling, 1953; Marvin et al. 1958). The structure of the A form was described in detail by Fuller et al. (1965): that of the B form by Langridge et al. (1960); and that of the C form by Marvin et al. (1961). Recently, more refined structures for the A and B forms have been presented (Arnott and Hukins, 1972).

The transition between the three forms appears to be a function of the type and concentration of countercation present, and of the relative humidity. At low ionic strengths (<5% excess NaCl) the A form was always observed at 92% relative humidity and occasionally at 98% (Cooper and Hamilton, 1966); while fibres containing >9% excess NaCl gave B patterns at 75% relative humidity. Fibres containing salt contents intermediate between these two gave the A form at 75% relative humidity and the B form at 92% and above. The A form is crystalline, giving Bragg diffraction spots over much of the diffraction pattern; whereas the B form observed in fibres of Na DNA is non-crystalline and gives diffuse diffraction intensity over most of reciprocal space.

A crystalline form of B DNA may be observed in fibres of the lithium salt of DNA: the A form is not observed. At high salt

Table 1.1 - Published Conformations for DNA

DNA Type	Pattern	Genus	Pitch	Residues per turn	Rise per residue	Distance of base pair from helix axis (D) Å	Tilt Angle (Deg.)	Twist Angle (Deg.)	Comment
Eukaryotic	A	A	28.16	11	2.56	4.72	20	-1.2	Only with Na DNA low salt and humidity.
DNA	B	B	33.80	10	3.40	-0.16	-5.9	-2.1	High NaCl conc. high humidity obtained with
	C	B	30.90	9.33	3.22	-2.13	-6.0	5.0	Li salt No X.S. LiCl Li salt some X.S. LiCl
Poly(dA).Poly(dT)	B'	B	32.90	10	3.29	0.2	-7.9	-1.0	No A-form
Poly(dA).Poly(dT)^2	A'	A	39.10	12	3.26	4.0*	8.0*		No B-form
Poly(dA-dT) Poly(dA-dT) and Poly(dG-dC) Poly(dG-dC)	 D A	 B A	 24.30 28.16	 8 11	 3.03 2.56	 -1.8 4.72	 -16 20	 -1.2	 Low humidity form not stable
Cl Perfringens	A	A	28.16	11	2.56	4.72	20	-1.2	<2% and >1% X.S. NaCl
	P	B+	30.0	9^S	3.4	-	-	-	<1% X.S. NaCl 66% R.H.
	T	B+	28.0	8^S	3.4	-	-	-	>2% X.S. NaCl 66% R.H.
C. Johnson	J_1	B+	33.0	10^S	3.4	-	-	-	0.8% X.S. NaCl 66% R.H.
	J_2	B+	31.0	9^S	3.4	-	-	-	3.0% X.S. NaCl 66% R.H.

DNA Type	Pattern	Genus	Pitch	Residues per turn	Rise per residue	Distance of base pair from helix axis (D) Å	Tilt Angle (Deg.)	Twist Angle (Deg.)	Comment
S. Aureous	s	B+	33.0	9^s	3.4	-	-	-	2% NaCl 66% R.H.

* Approximate values

\+ Assigned by the author because of intense 3.4Å meridional

s Approximate value assigned by author on the nearest integer to the value of the ratio, pitch/rise per residue.

content (>3% excess LiCl) the B form is observed, whereas fibres containing little or no excess salt give one of the semi-crystalline forms of C DNA (Marvin et al. 1961).

It has become apparent that the A and B forms are members of two different types of DNA conformation (hereafter called A and B genus structures) which are characterised by the type of furanose ring puckering present, and the disposition of the base pairs relative to the helix axis. The base disposition can be fixed by the parameters defined in figure 1.1. In B DNA the distance, D, of the base pairs from the helix axis is small and negative (c.f. Fig. 1.1) while the furanose ring pucker is C3' exo. The A form in contrast, has C3' endo sugar pucker and the base pairs are moved nearly 5Å forward (D positive) from the helix axis and have a large positive value for the tilt angle.

Although the C form has a number of characteristics which make it distinct from the B form, it is similar in having C-3 exo furanose ring puckering and base pairs which have a negative value for the distance, D, from the helix axis and for the tilt angle. On this basis the C form is defined as being of the B genus of DNA structures. It has become apparent that all double helical nucleic acid conformations can be classified as belonging to the A or B genus on the basis of the values of these three parameters. In the case of the distance and tilt parameters for the base pairs, it is the sign of their values which are the important criteria. The magnitudes of these quantities often vary considerably between different conformations of the same genus. All structures reported for natural and synthetic double helical RNA are of the A genus. The A, B and C forms remain the

only conformations for natural DNA which have been obtained reproducibly and for which detailed molecular structures have been published.

Hamilton et al. (1959) showed that the above conformations could be obtained from natural DNA's from a wide variety of sources and the idea became established that the secondary structure of DNA was a function only of the environment. However, a number of different conformations have been obtained for synthetic DNA's having a simple repeating base sequence. Arnott et al. (1974a) have proposed a structure for the D form, first discovered by Davies and Baldwin (1963) in fibres of poly (dA-dT). poly (dA-dT), and have shown that it may also be obtained from poly (dG-dC). poly (dG-dC) so that the criteria for its formation appears to be an alternating sequence of purine and pyrimidine bases. The structure is an eight fold right handed helix and is of the B genus by the criteria outlined above. Mitsui et al. (1970) on the basis of the unusual optical rotatory dispersion data given by the D form concluded that it was a left handed helical structure and built a model to explain the X-ray data. In order to accommodate the requirements of a left handed helix, however, it was necessary to postulate an unusual pucker for the furanose ring. The A form, a right handed helix, has been observed in poly (dA-dT). poly (dA-dT) fibres Davies and Baldwin (1963), but was found to revert spontaneously to the D form. It appears unlikely for this reason that the D form could be a left handed helix, and the alternative structure of Arnott et al. (1974a) which is a right handed helix having standard stereochemistry is more likely to be correct.

Another structure of the B genus, the B' form, which is very similar to the lithium B form of natural DNA, has been reported by Arnott et al. (1974b) for the sodium salt of the synthetic copolymer poly (dA). poly (dT). It occurs in two crystalline forms: the α-B' form found at relative humidities above 77% and indexing on a hexagonal unit cell; and the highly crystalline orthorhombic β-B' form which is obtained at lower relative humidities. The model for the B' conformation has been refined on the basis of data obtained from the more highly crystalline orthorhombic form. The conformation of the α form is thought to be identical on the basis of a general comparison of the intensity data. The above authors have also given the coordinate of a model of the triple stranded DNA, poly (dT). poly (dA). poly (dT). This structure is of the A genus, a fact which is surprising in view of the failure to obtain the A conformation from the double helical form. This point will be discussed in more detail in chapter 6.

In view of the different conformations adopted by these DNA's of varying base sequence, it is interesting to speculate whether, under certain conditions, different segments of natural DNA might adopt different secondary structures. If this were the case it might prove to be the molecular basis of at least some of the control mechanisms in the processes involved in DNA replication and transcription. Evidence for variations in the secondary structure of DNA as a function of base composition and sequence has been provided by the studies of Bram and his co-workers, of the diffraction patterns obtained from fibres of natural DNA from different organisms. The results of this work, which will be discussed more fully in chapter 6, seem to suggest that there may

a family of related structures of the B type for the sodium salt of natural DNA, and that different members of the set are adopted by DNA's of differing base composition and sequence. Other workers have predicted that the secondary structure of DNA is dependent on its base composition (e.g. Pilet and Brahms, 1972; Pohl, 1976), so that the conclusion reached by Bram, though different from those of other workers in ascribing new conformations to natural DNA's in fibres, are not inconsistent with results obtained in other studies. However, it should be pointed out that the diffraction patterns obtained by Bram give diffuse intensity data often with very poorly defined layer line structure so that it is difficult to measure the helix parameters accurately or to analyse the conformations in detail.

1.2 Drug Binding to DNA

Since the basic structure of DNA has been elucidated, it is of considerable importance to determine ways in which conformational changes, particularly those resulting from the interaction with other molecules, can affect DNA replication and transcription. A number of drug molecules, some having therapeutic uses, interact with DNA in a variety of ways. Potentially the most important of the methods of binding is by intercalation (Lerman, 1961). In this mode of binding the drug molecule is inserted between two adjacent base pairs which thereby have to move apart in order to accommodate the drug chromophore. The separation of the base pairs is probably accompanied by a change in the turn angle per residue at the intercalation site. Fibre diffraction patterns of DNA/drug intercalation complexes exhibit a decrease in the layer line separation, indicating an increase in the helical pitch. A strong meridional reflection at 3.4Å is observed in the patterns which indicates that the DNA is in a B like conformation in which the

bases are not significantly tilted, and that the base separation at the intercalation site is 6.8Å. It is now generally accepted that the figure for the base separation is correct: the point of controversy is the sign, magnitude and extent (in terms of number of base pairs effected) of the change in the turn per residue around the intercalation site, and most studies on intercalation complexes are concerned with attempts to determine the value of this parameter. Estimates of the unwinding angle produced by various intercalating drugs are listed in Table 1.2. Individual estimates of the unwinding angle for particular drugs and their possible biological significance are discussed in detail in chapter 5: only the methods by which such determinations are made will be discussed here.

The most important technique for measuring the unwinding angle is that of binding the drug to closed circular DNA. Any topologically closed DNA system will exhibit supercoiling of the DNA duplex as a means of relieving the strain imposed upon the system. Such strain arises because the number of turns (α) made by the DNA around the duplex axis when it is in a closed system and constrained to lie in a plane is different from the number of turns (β) made when the DNA is "free" (i.e. in an open ended state). The number of turns of the superhelix (τ) is given by (Bauer and Vinograd, 1968):

$$\tau = \alpha - \beta$$

If the drug ethidium is bound to a closed DNA system (e.g. ØX-174 replicating phase) at different phosphate to drug, P/D, ratios and the sedimentation coefficient of the complex determined in each case, a minimum is found in the plot of sedimentation coefficient against P/D (Bauer and Vinograd, 1968; Waring, 1970; Crawford and Waring, 1968).

The initial fall in sedimentation coefficient is interpretted as being due to the removal of the super coiling due to the change in winding angle caused by ethidium. At the minimum points all supercoiling is believed to have been removed and the value of S rises again as the P/D is lowered since the drug now introduces supercoiling of the opposite sense.

In order to calculate the magnitude of the unwinding angle the following equation was used (Bauer and Vinograd, 1968).

$$\tau = \tau_o + \frac{N\emptyset\nu}{2\pi}$$

where τ = number of super helical turns

when ν drug molecules are bound per nucleotide

τ_o = number of super helical turns when no drug is bound

N = number of nucleotides in the DNA circle

$\emptyset$ = angle of untwist per bound drug

If measurements are made at the equivalence points (i.e. when S = a minimum), then $\tau = 0$ so,

$$\emptyset = \frac{-2\pi\tau_o}{N\nu}$$

τ_o cannot be eliminated directly and so it is impossible to determine the sign or the absolute magnitude of $\emptyset$ by this method. In order for progress to be made it is necessary to assume a value for the unwinding produced by one intercalating drug and to calculate the equivalent values for the others relative to it. Normally, the value of 12^o unwinding for ethidium intercalation (Fuller and Waring, 1964) is taken as the standard (Waring, 1970).

Table 1.2

Estimates of Unwinding Angles for Intercalating Drugs

Drug	Unwinding*	Ref.	Comment
Ethidium	12^{o} (i)	1	Molecular model building of X-ray fibre diffraction
Ethidium Bromide	29^{o} (1)	2	Single crystal X-ray diffraction solution of cocrystal of 5-iodouridylyc (3'-5') adenosine and ethidium bromide
Ethidium	24^{o}-36^{o}	3	Binding of ethidium to closed circular DNA with known sense to the supercoiling
Actinomycin	44^{o} (4)	4	Molecular model building
Daunomycin	12^{o} (1)	5	Molecular model building/X-ray fibre diffraction
Daunomycin	5.2^{o} +	6	Binding to ϕX-174 RF closed circular DNA
Proflavine	8.4^{o} +	6	" " " " " "
Mycanthone	6.8^{o} +	6	" " " " " "
Nogalamycin	8.1^{o} +	6	" " " " " "
Propidinium	12.0^{o} +	6	" " " " " "
Actinomycin	11.4^{o} +	6	" " " " " "
Dimidium	11.5^{o} +	7	" " " " " "
M & B 2421	8.3^{o} +	7	" " " " " "
Phenidium	8.6^{o} +	7	" " " " " "
M & B 3492	11.8^{o} +	7	" " " " " "
M & B 4594	8.1^{o} +	7	" " " " " "
RD1601	5.1^{o} +	7	" " " " " "
M & B 3016	10.6^{o} +	7	" " " " " "
M & B 3427	7.7^{o} +	7	" " " " " "
M & B 1765	4.6^{o} (max)+	7	" " " " " "
Echinomycin	21.9^{o} +s	8	" " " " " "
-	18.0 (4)	9	Computer molecular model building: DNA conformation only - no particular drug.

* Figure in parenthesis indicates number of base pairs over which the unwinding is spread.

+ These figures are obtained assuming an unwinding angle of 12 degrees for ethidium.

S A bifunctional intercalating agent: value given corresponds to the total unwinding/drug molecule bound.

References

1. Fuller and Waring (1964).

2. Tsai et al. (1975).

3. Pulleybank and Morgans (1975).

4. Sobell and Jain (1972).

5. Pigram, Fuller and Hamilton (1972).

6. Waring (1970).

7. Wakelin and Waring (1974).

8. Wakelin and Waring (1976).

9. Alden and Arnott (1975).

In an attempt to overcome this difficulty, Pulleybank and Morgans (1975) devised a modified form of the above technique in which a closed circular DNA system was produced which had a known sign (-ve) for its supercoiling and in which the number of supercoils could be estimated. Ethidium produced a minimum in the titration curve (S against P/D) showing that it unwinds the helix. On the basis of their estimate of the number of superhelical turns, the above authors derived a value of between 24° and 36° for the unwinding due to ethidium: a value considerably larger than that first predicted by Fuller and Waring (1964) on the basis of molecular model building.

A number of other specific models have been proposed by other workers for intercalation complexes involving various intercalating drugs including ethidium (e.g. Jain and Sobell, 1972; Sobell and Jain, 1972; Pigram et al. 1972; Tsai et al. 1975; Alden and Arnott, 1975). The study of Alden and Arnott involves a computer model building study of an intercalation site involving a base separation of 6.8Å in the absence of any particular drug. It is the only published work in which a computerised linked atom routine has been applied to a study of intercalation complexes.

1.3 The Need for Further Work

The detailed structures for a number of conformations observed in fibres of natural and synthetic DNA have now been elucidated. The biological significance of these different forms is still unclear, however. In particular, the relevance to the secondary structure of natural DNA of conformations observed only in synthetic DNA's with a simple repeating base sequence has not been decided. The work of Bram, suggesting that the nature of the B

conformation observed in fibres of the sodium salt of natural DNA's might be a function of the base composition and sequence, leads to the possibility that short sections of the duplex of natural DNA might adopt a conformation similar or identical to those only associated so far with synthetic forms. Variation in DNA secondary structure as a function of base composition and sequence has implications for control of gene expression in cells. Since the work of Bram has not been repeated by any other workers, while the analysis given is not sufficiently rigorous to permit a detailed evaluation of the significance of the results, there is a need for Bram's work to be repeated and any resultant new conformations analysed in detail.

In addition to information about DNA conformation in the native state, it is of interest to determine ways in which DNA secondary structure can be modified by interaction with other organic molecules. One mechanism by which certain molecules may bind to DNA and significantly alter its conformation is that of intercalation: a mechanism that is now widely recognised as being the mode of attachment of a number of drugs which bind to DNA. Although the general nature of the modifications brought about by an intercalating drug are known, few well defined models have been produced for the DNA conformation at the intercalation site in a full DNA duplex. The elucidation of the conformational changes induced by intercalating drugs is of more than academic interest since many of them have a therapeutic use which may be related to their ability to bind to DNA. Moreover, knowledge of the way in which these drugs modify DNA secondary structure may throw light upon the manner in which proteins attach to DNA and exert their biological effect. It is difficult to obtain sufficient information from fibre diffraction data from intercalation complexes to build a model since the pattern

obtained are of a relatively poor quality and it is in any case difficult to calculate the diffraction pattern for the disordered helical structure.

There is a need, therefore, to examine methods of calculating the diffraction pattern for such complexes and to obtain additional information relating to the possible conformation at the intercalation site. Such information is provided by the results of Sobell and his co-workers in solving the single crystal structure of actinomycin D bound to deoxyguanosine (Jain and Sobell, 1972) and of ethidium to a double helical dinucleotide fragment (Tsai et al. 1975). This has allowed the direct visualisation of an intercalative process in these cases. It seems reasonable to use information derived from the results of these studies as a basis for the construction of models for drug intercalation into a full double helical DNA structure.

The work relating to intercalation complexes involves a model building study using a computerised linked atom least squares refinement routine described in chapter 4; combined with Fourier transform calculations for certain of the models built.

CHAPTER II

MATERIALS AND METHODS

2.1 DNA Purification

All DNA, except that from the bacteriophage ØW-14 was obtained commercially from either the Sigma Chemical Company, B.D.H. Biochemicals or Miles Laboratories Incorporated. ØW-14 DNA was obtained from Dr. R.A.J. Warren of the Department of Microbiology, University of British Columbia, Vancouver, Canada.

Since the DNA's had been obtained from a variety of sources and included natural as well as synthetic forms, it is desirable to subject all samples to a standard purification procedure to extract protein and excess inorganic salts. Massie and Zimm (1965) have reported that phenol is an effective reagent to remove proteins from samples of DNA, and a phenol extraction stage was included in the purification procedure. Analar grade phenol was used in the extraction and was distilled just before use to remove oxidation products. The distillate is allowed to drip into 0·1M NaCl solution and the resultant brine/phenol mixture shaken thoroughly and allowed to stand until the phenolic (lower) layer separate out from the mixture.

A solution of DNA is prepared (concentration approximately 1mg/ml) in 0·002M NaCl. It is useful to use a low salt concentration since the DNA dissolves more slowly if the ionic strength is high; whereas if the salt concentration is too low the DNA is denatured. The salt concentration specified above was found to be an effective compromise. Before adding to the phenol, the salt concentration in the DNA solution was raised to 0·1M by addition of the appropriate volume of 2·0M NaCl solution.

The DNA solution is added to the same volume of the phenolic layer removed from the phenol/brine mixture and shaken gently for 20 minutes. The mixture is then centrifuged at 3,000 r.p.m. for 15 minutes after which the upper (aqueous) layer is removed and centrifuged a second time.

The solutions from the centrifuge tubes are pooled and the DNA precipitated by the addition of two and a half times the original solution volume of 95% ethanol. If the salt concentration of the original DNA solution is increased to 0·1M as described, the ethanol precipitation gives a good (approx. 85%) yield. It is important to raise the salt concentration before the phenol extraction stage: raising the salt concentration after this stage results in a poor (less than 5%) yield.

The precipitated DNA is wound on to a glass rod and washed in 80% ethanol for 30 minutes and then in acetone for 15 minutes. It is dried over phosphorus pentoxide for 12 hours and is then ready to be dissolved in the appropriate buffer solution for use. The final value of the ratio $O.D._{260}/O.D._{280}$ obtained after the purification procedure was around 1·9 for all samples.

Leng et al. (1974) have indicated that phenol may interact with DNA and denature it, particularly if the DNA has a high A-T content. No evidence of denaturation of the DNA under the conditions used was apparent in the diffraction photographs. Some DNA samples gave spectra with a value of around 1·9 for the $O.D._{260}/O.D._{280}$ ratio before purification, and solutions were prepared from these when the phenol extraction stage was omitted. The behaviour and quality of the fibres from such solutions was not measurably different from those of fibres from material which had undergone the phenol extraction.

2.2 Preparation of DNA For Fibre Pulling

Purified DNA was dissolved in a tris-sodium chloride buffer, pH 7.6. Since it was desired that the major ionic component of the solution, other than the DNA, should be the sodium chloride, the concentration of tris was kept low (0.002M) and the sodium chloride concentration varied from 0.005M to 0.1M. Although the tris concentration was low it was found to be adequate to maintain the pH within ±0.2 and so to retain it near to physiological levels.

In special cases, unbuffered solutions were used and sometimes had more extreme pH values. The characteristics of particular solutions are discussed in the appropriate chapters.

Solutions of the DNA were prepared in the appropriate buffer at a DNA concentration of approximately 1mg/ml. The DNA may take up to seven days to fully dissolve depending upon the salt concentration in the solution. Gels were prepared by sedimentation of the DNA in an M.S.E. 50 preparative ultracentrifuge. The solutions were normally centrifuged at 40,000 r.p.m. for 12 hours, in a 10 x 10ml angle rotor giving an average sedimentary force of 105,000g. These conditions result in the sedimentation of over 95% of the DNA as measured by comparison of the ultraviolet absorption spectrum of a supernatant with that of the original solution. Gels could also be obtained by centrifugation in the same rotor for 5-6 hours at 50,000 r.p.m. The author has obtained gels which are more workable using the former conditions and hence they are preferable except in special circumstances where it is essential to produce gels quickly.

The supernatant is removed from the gels by gentle pouring. It is important that all the supernatant is drained from the gel and the tube walls. If this is not done the gel eventually redissolves

and is no longer of a suitable consistency for fibre pulling. Gels which are well dried can be used even for several weeks of storage provided that the tubes are sealed and are kept at $4^{o}C$.

2.3 Preparation of Fibres

Fibres were made in a frame similar to that used by Fuller et al. (1967) and shown in figure 2:1. Two glass rods of diameter 150-200μ are mounted in plasticine inside the frame as shown, and their end separation can be adjusted by rotating the knurled wheel outside the frame. A drop of gel, containing approximately 0.1mg of DNA, is placed between the ends of the glass rods and allowed to dry to form a fibre.

The precise method adopted for the drying process depended partially upon the DNA used but more particularly on the consistency of the gels. There is an optimum gel consistency which, if achieved, normally gave good crystalline fibres when allowed to dry uncovered at room temperature. To slow the rate of drying, the frame could be covered with a glass plate and occasionally a small pot containing a saturated solution of sodium tartrate was placed in the cell. Alternatively a flow of air of controlled humidity could be passed through the cell. This facility was used in certain cases as described in later chapters since it has been reported (Fuller et al. 1967; Bond et al. 1976) that such a procedure leads to fibres of greater crystallinity if the relative humidity is greater than 80%.

The distance between the ends of the glass rods is increased as the fibre dries. If this is not done the molecules will not be well oriented if the fibre is thick; while a thin fibre may buckle. Since, however, the fibre may be damaged by overstretching, the

strategy adopted was to allow the gel to dry with only the minimum of stretching.

If the gel is too gelatinous and is not allowed to dry slowly, then it dries more quickly on its outer surface forming a skin which is too dry to be stretched. Any attempt to elongate the fibre in this state results in rupturing of the surface skin and eventually, of the central part of the fibre which may not be sufficiently dry to withstand the strain of stretching.

2.4 The Taking of X-Ray Fibre Diffraction Photographs

All diffraction photographs were taken using nickel filtered copper K α radiation (λ = 1·5418Å) generated either by a Hilger and Watts micro-focus generator or by an Elliott G x 6 rotating anode generator. Sets were operated at 35KV and at a tube current which gave near maximum power allowable for the particular generator. Typical tube currents were 3·0mA and 10·0mA for the Hilger and Watts generators using low and high power tubes respectively, and 60mA on the Elliott rotating anode machine using the 200μ line focus.

Cameras used were either Searle X-ray diffraction cameras employing Elliott toroidal optics (Elliott, 1965) or Frank's optics (Frank, 1958); or one of a set of pinhole cameras made in the workshop of the Department of Physics, University of Keele, which were similar in design to those described by Langridge et al. (1960a).

The Searle cameras were operated using toroidal optics for the majority of fibres. Frank's optics were used, however, for particularly thin fibres (60μ), but gave longer exposure times than the Elliott optics for fibres of diameter around 100μ.

Alignment of the fibres in the pinhole cameras was effected by passing light though the pinhole system and viewing it by means of a low power microscope the axis of which is collinear with that of the pinhole system. With this arrangement it is possible to position the fibre on the axis of the pinhole system, and hence in the path of the eventual X-ray beam without significant parallax error.

The fully assembled camera is aligned on the X-ray generator by placing the camera in front of the X-ray window. When the camera is aligned, the X-ray beam will pass through the pinhole system and through a 200μ hole in the back of the camera over which is placed the detector tube of a Geiger counter. The position of the camera is adjusted until a maximal reading is obtained with the Geiger counter. A brass stop is inserted over the exit hole during use to prevent the emergence of the X-ray beam from the camera.

The Searle cameras were aligned in the manner described in the relevant manuals.

In order to eliminate air scatter the cameras were flushed out with helium gas which gives a very low intensity of scattered radiation due to the small number of electrons in the helium atom. The relative humidity inside the camera must be controlled since nucleic acid secondary structure is a sensitive function of the relative humidity. This is effected by passing the helium through a saturated solution of an appropriate salt. The following salts were used giving the indicated relative humidity values at 20°C (O'Brien, 1947):-

SALT	R.H. (%)
Calcium chloride	33
Potassium carbonate	44
Sodium bromide	57
Sodium nitrite	66

SALT	R.H. (%)
Sodium chlorate	75
Potassium chloride	86
Sodium tartrate	92
Potassium chlorate	98

2.5 Measurement of Diffraction Patterns

In order to obtain the reciprocal space coordinates of specified points on the diffraction pattern, it is necessary to obtain an accurate value for the specimen to film distance and to correct for distortions due to the film geometry. To determine the specimen to film distance which result from radiation diffracted through a known scattering angle. The calibration was performed by spraying vaterite powder (a crystalline form of calcium carbonate) onto the fibre and using one of the resultant diffraction rings as the locus of reference points.

The X and Y coordinates of specified points on the diffraction pattern, and of a number (typically eight) of points on the vaterite calibration ring are measured using a two dimensional travelling microscope. A programme has been written by Dr. W.J. Pigram to find the best value (in the least squares sense) for the specimen to film distance and for the coordinates of the film centre. The programme uses these values to derive the reciprocal space coordinates, corrected for the film geometry, of the measured points on the diffraction pattern.

The intensity data was obtained from the pattern by means of a Joyce-Loebl recording microdensitometer. Wilkins (1961) has described a method, given in Marvin et al. (1961), of obtaining the integrated intensity of a Bragg diffraction spot, and the intensity per A^{-1} along a diffraction streak. This method has been used in all measurements of intensity data reported in this thesis.

2.6 Computer Programmes

All large computer programmes were written in Algol 60 language and were usually run initially on the Elliott 4130 machine at the Computer Centre of the University of Keele, but were later transcribed to run on the more powerful C.D.C. 7600 machine at the University of Manchester Regional Computer Centre.

Although programmes written in the Fortran language are more rapid in compilation and running, Algol was preferred because of its facility of dyanmic array bounds. For most of the programmes, storage space and not time is at a premium and hence it is useful to have a language which allows the amount of storage space used to be related to the size of the data set in any particular run.

Individual programmes written are described in the appropriate sections of this thesis.

2.7 Model Building

Since it is not possible to solve the phase problem directly for diffraction patterns from nucleic acid fibres, the method of analysis used is to build a good trial model from general considerations and then to refine it by a detailed comparison of the observed and calculated intensity data. Moreover, even if the phase problem could be solved, the quality and quantity of the intensity data obtained from fibres, unlike those from a good single crystal, are inadequate to describe the structure accurately to atomic resolution. The model building technique allows information about the known stereochemical features of the structure to be incorporated into the analysis and thus excludes from the set of possible solutions any coordinate set involving impossible geometrical relationships between bonded atoms. Moreover, it is possible to check distances of separation between pairs of non-bonded atoms so that energetically unfavourable short

contacts may be avoided. This inclusion of stereochemical information allows better use to be made of the limited diffraction data obtainable. Standard values for bond lengths and bond angles are derived from studies on related small molecules whose structures have been solved to atomic resolutions using single crystal X-ray diffraction techniques.

Preliminary structures were always built using wire models, in which bond lengths and bond angles are represented accurately (scale 4cm to 1A), prior to refinement. Coordinates are obtained from the wire model from which a set of torsion (dihedral) angles is calculated and used to define a starting model for a linked atom least squares computer refinement routine.

The model building programme is that described by Pigram (1968) with modifications made by the author. It will be described in chapter 4 of this thesis.

CHAPTER III

Fourier Transform Calculations

3.1 Introduction

It is not feasible to solve the phase problem for fibre diffraction data, and, hence, the method of analysis used is to build a model of the structure and then to refine it so as to minimise the discrepancy between the experimental diffraction intensity data and that predicted by the model. This method of approach can only be used effectively if it is possible to obtain sufficient information about the structure to build a starting model which is good enough to be refined. In most problems subjected to X-ray crystallographic analysis, such information is not available until the structure has been solved, so that molecular model building is of limited importance in these cases. However, for helical polymers, the work of Cochran, Crick and Vand (1952) (discussed in section 3.2) has shown that some structural information may be obtained from a fairly superficial analysis of the diffraction data. Hence, molecular model building has been of paramount importance in the analysis of nucleic acid secondary structure. In the case of diffraction patterns from crystalline fibres, computerised refinement routines have been developed (Arnott and Wonacott, 1966) which perform the refinement analytically, and it is now possible to obtain a starting model and to refine it in most cases provided that it gives well defined crystalline data. Computerised refinement routines will be discussed in more detail in Chapter 4.

It is, therefore, necessary to be able to calculate the predicted scattering intensity from a model, and to be able to measure the experimentally obtained intensity data as accurately as possible. For the case of crystalline fibres discussed above, the problems involved have been overcome to a sufficient degree to allow the diffraction data to be used effectively in structural analyses in most cases.

The problems are more acute for fibres in which the molecules are not regularly packed in crystallites. Such fibres give diffraction patterns consisting of continuous intensity streaks along each layer line, which essentially represent the diffraction from a single molecule. In principle such data allows more information to be obtained from the diffraction pattern along each layer line since, unlike the case of crystalline fibres, the intensity is not merely sampled at discrete points. However, any fibre containing a given number of the same molecular unit must give the same intensity of scattering integrated over the whole of reciprocal space. Hence, the sampling of the molecular transform observed in crystalline fibres implies a concentration of intensity into the sampled region, and, hence, the diffraction intensities from Bragg diffraction spots are greater in comparison with the background than are data from diffuse diffraction streaks, and it is difficult to measure such data accurately.

Moreover, although more information is theoretically obtainable along lines parallel to the equator, in non crystalline fibres, there is some information loss in that only two dimensional information is available. The individual molecules will have a random azimuthal orientation about the fibre axis in such specimens

and hence, the resultant diffraction intensity is cylindrically averaged about the meridian and the maximum amount of information obtainable lies in a plane representing a central section through the cylinder. Hence, unlike crystalline fibres from which genuine three dimensional information may be obtained, in non-crystalline fibre specimens it is not possible to derive any information about the variation in intensity of the molecular scattering as a function of the azimuthal angle about the meridian in reciprocal space.

Fraser et al. (1975) have developed a method of processing the data from non-crystalline fibre specimens so as to obtain maximum accuracy in the measurement of scattering intensity. The method involves scanning the diffraction pattern so as to produce a two dimensional quasi-continuous map of a central section of the specimen transform in the space represented by the surface of the recording film. Methods of correcting for distortions due to film shape for a number of camera geometries in common use are discussed; so that after applying a number of classical (e.g. Lorentz) corrections to the intensity data, an undistorted map of the specimen transform, I_s would be identical to the cylindrically averaged intensity transform of a single molecule, I_m. In the above work, methods of deriving I_m from I_s have been derived for a number of different cases of non-crystalline diffraction patterns. The problem of allowing for effects due to molecular misalignment has also been discussed for certain cases by Holmes and Barrington-Leigh (1974).

Provided that the structure of each molecular unit is regular, then it is possible to calculate the predicted values of I_m for any given molecular model, so that it may be compared with the experimental data. In many of the specimens studied in this laboratory, the disorder may be present in the individual molecules

themselves and not merely in their packing. The diffraction theory for helical molecules as derived by Cochran et al. (1952) involves the concept of helix pitch and, hence, presumes the existence of long range order in the structure. It is not possible, in general, to ascribe a meaningful value to the pitch of disordered helices and, hence, the diffraction theory formulation of Cochran et al. (1952) is not suitable, in its present form, for calculating the Fourier transform of such structures.

Hence, if use is to be made of the diffraction data from such fibre specimens in the structural analysis of their molecular components, it is necessary to derive a method of calculating the molecular transform of disordered structures. The work of Fraser et al. (1975) has made such a method even more desirable since it opens up the possibility of measuring more accurately diffuse intensity data obtained from such structures.

Intercalation complexes between DNA and various drugs provide examples of specimens in which molecular as well as packing disorder is present. The molecular disorder arises in this case because changes in the DNA conformation are induced at the site of attachment of the drug. Since the sites at which drug is bound will not, in general, be regularly spaced along the helix, the DNA duplex is perturbed at random points along its length and its regularity is destroyed.

Such molecular disorder may occur in B genus structures of the sodium salt of DNA. This point will be discussed more fully in Chapter 6.

3.2 Fourier Transform Calculations for Regular Structures

It is possible to calculate the Fourier transform of any group of atoms using a general formula of the type:-

$$F(\underline{s}) = \sum_{j=1}^{N} f_j \exp 2\pi i(\underline{r}_j.\underline{s}) \qquad 3.1$$

where: f_j = the scattering factor for the jth atom

$\underline{r}_j$ = a real space vector at the end of which is situated the jth atom

$\underline{s}$ = a reciprocal space vector

There are problems attached to the use of this type of formula, however, and in practice it is rarely used explicitly in any type of X-ray diffraction analysis.

For studies of diffraction data given by single crystals the classical structure factor formula is used.

$$F(h,k,l) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + ky_j + lz_j) \qquad 3.2$$

In equation 3.2 the summation is taken over the atoms in the unit cell only; and not over all atoms in the object as in equation 3.1. Moreover, the calculation for F is only made for points in reciprocal space at which it will be non-zero. These reductions in computation are achieved by utilising the symmetry properties of an infinite three dimensional lattice. Equation 3.2 is the most efficient method of performing the calculation of the Fourier transform for crystals in which the contents of the unit cell do not, themselves, have symmetry properties which may be exploited. If the unit cell contents possess some element of symmetry, then further reductions in the computation may be achieved and methods of calculation formulated in which the general features

of the diffraction data may be interpreted more meaningfully. The helical conformation adopted by many long chain polymer molecules are examples of molecular structures possessing symmetry elements, and Cochran, Crick and Vand (1952) (also Stokes, unpublished) have produced a formulation of diffraction theory applicable to helical structures in which their special symmetry is exploited.

It was shown that the Fourier transform of a right handed, infinitely long, helical molecule is given by:-

$$F(\xi,\psi,\ell/c) = \sum_n \sum_j f_j J_n(2\pi\xi r_j) \exp i\left[n(\psi-\Phi_j + \pi/2) + 2\pi \frac{\ell z_j}{c}\right] \qquad 3.3$$

where J_n = a Bessel function of the first kind of order n.

r_j,ϕ_j,z_j = the real space cylindrical polar coordinate of the jth atom.

$\xi,\psi,\ell/c$ = reciprocal space cylindrical polar coordinates.

ℓ = an integer +ve, -ve or zero

c = the helix pitch (c axis repeat)

For a crystalline fibre with M molecules in the unit cell, the structure factors can be obtained from:-

$$F'(h,k,\ell) = \sum_n \sum_{j=1}^{N} \sum_{m=1}^{M} f_j J_n(2\pi\xi r_j) \exp i\left[n(\psi-\phi_j+\pi/2) + 2\pi \frac{\ell z_j}{c}\right] \exp 2\pi i(hx_m + ky_m + \ell z_m) \exp i(-n\phi_{om}) \qquad - 3.4$$

where x_m, y_m, z_m represent the fractional unit cell coordinates of the mth molecule in the unit cell; and ϕ_{om} is the azimuthal orientation of the mth molecule.

It has been previously stated that for non-crystalline fibres, the specimen transform is representative of the cylindrically averaged intensity transform for the molecular unit. This can be calculated from:-

$$I_{<\psi>}(\xi,\ell) = \frac{\int_0^{2\pi} \sum_n F_m(\xi,\psi,\ell/c)\, F_m^*(\xi,\psi,\ell/c)\, d\psi}{\int_0^{2\pi} d\psi} \qquad 3.5$$

where F_m represents the nth Bessel function component of equation 3.3.

This gives, disregarding constants:-

$$I_{<\psi>}(\xi,\ell) = \sum_n G_n(\xi,n,\tfrac{\ell}{c})\, G_n^*(\xi,n,\tfrac{\ell}{c}) \qquad 3.6$$

where

$$G_n(\xi,n,\tfrac{\ell}{c}) = \sum_{j=1}^{N} f_j\, J_n(2\pi\xi r_j) \exp i(2\pi \tfrac{\ell z_j}{c} - n\phi_j) \qquad 3.7$$

The function G_n of equation 3.7 is often calculated since it is often useful to examine the phase of each Bessel function component. However, only I of equation 3.6 has any physical meaning in relation to actual experimental data. It should be noted that G_n can be used to obtain the structure factors of equation 3.4:

$$F'(h,k,\ell) = \sum_{j=1}^{M} \sum_n \left[G_n(\xi,n,\ell/n) \exp -in(\psi-\Phi_{oj} + \tfrac{\pi}{2}) \right] x$$
$$\exp 2\pi i(hx_j + ky_j + lz_j) \qquad 3.8$$

An important difference between equation 3.4 and 3.7 is that, because of the cylindrical averaging it is only meaningful to add the intensities of each Bessel function component to G_n to obtain

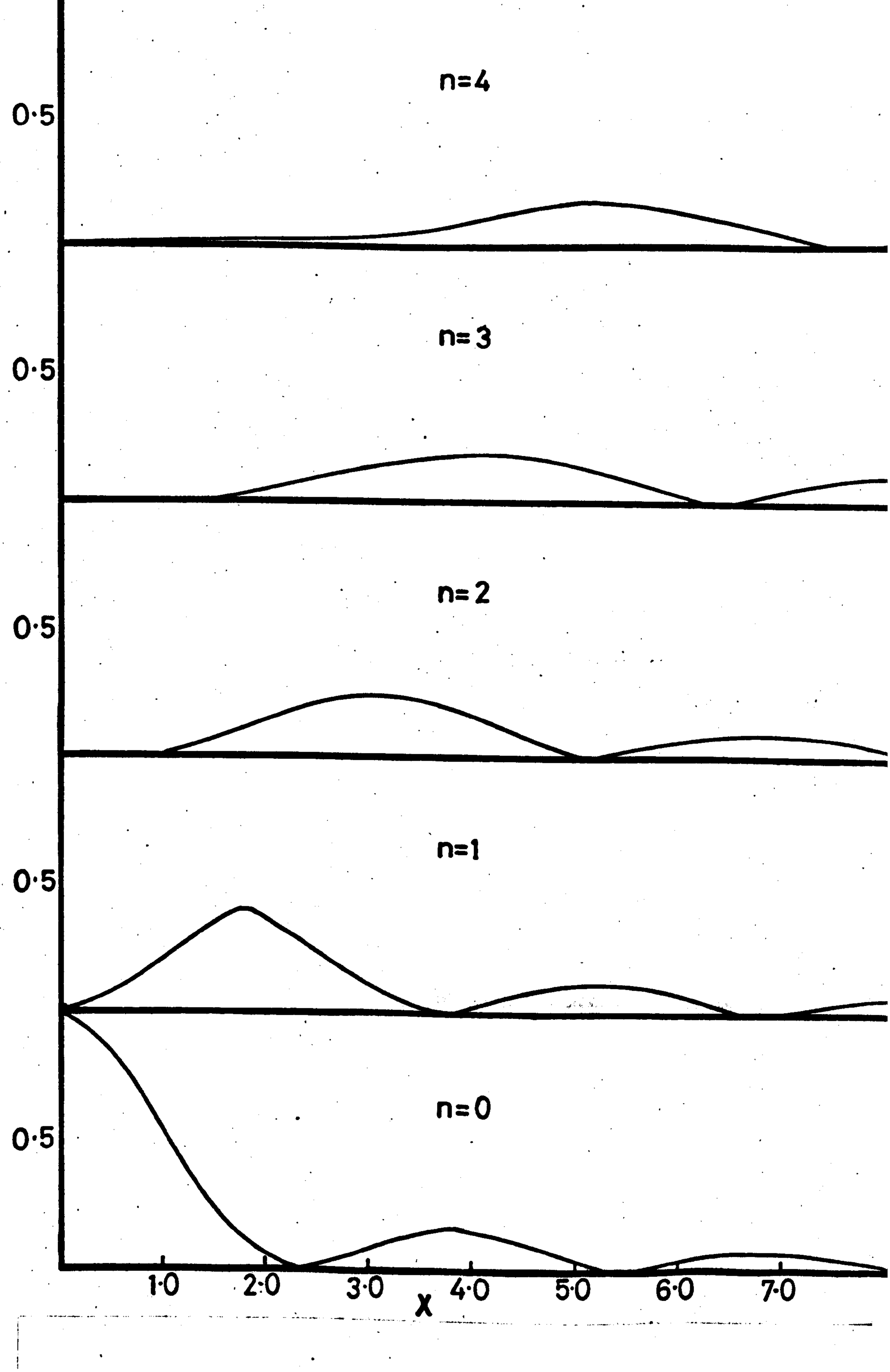


Fig. 3.1 Plot of $J_n^2(x)$ For n = 0 → 4

the total intensity on each layer line. Hence interference between different Bessel function components is not allowed. In crystalline fibres the cylindrical averaging arises because each crystallite has a random azimuthal orientation. Individual molecules within each crystallite, however, have a fixed relative orientation and hence, Bessel function components of G_n are summed in amplitude and phase in equation 3.5 allowing interference effects between different orders to become apparent. This difference can be significant if the intensity transform of equation 3.7 is compared to the intensity data from Bragg diffraction spots in regions of reciprocal space where the contribution of more than one Bessel function component is significant.

For all the above equations, the orders of Bessel functions which occur on each layer line can be derived by the following selection rule.

$$n = \frac{\ell - mN}{K} \qquad 3.9$$

where,

n = Bessel function order

N = the number of units in a repeat

K = the number of turns in a repeat

m = an integer

Hence, a knowledge of the helical parameters, N and K, enables one to predict the Bessel function composition of each layer line. It should be noted that if the helix is integral (K = 1), then when m = 0; n = ℓ so that a ℓth order Bessel function occurs on the ℓth layer line. A plot of Bessel functions of order 0 to 4 is

given in figure 3.1. The important points to note are that the first maxima of the functions occur at increasing values of the argument as n increases; while only a Bessel function of zero order has a non-zero value when the argument is zero. It is possible to draw a straight line which passes approximately through the position of successive Bessel function peaks, and it is this arrangement which when repeated in the four quadrants of a diffraction pattern, produces the "helix cross" which is characteristic of diffraction from a helix. This arrangement is particularly well seen in sodium BDNA (Plate 3.1), though because of disorder in the packing only the first and second layer lines are easily visible. In general, for a structure consisting of a single primitive helix, the intensity of the first maxima on the lower layer lines decreases as ℓ increases. This is due to the decreasing value of the maxima of Bessel functions of higher order. If the structure consists of a double helix in which one helix is related to the other by a translation along their common axis of half a pitch length, then the diffraction pattern consists of the product of the transform of a single helix and the transform of two points of spacing c/2, where c is the helix pitch. The transform of the point function will be zero along lines in reciprocal space represented by $\zeta = \ell/2c$, where ℓ is an integer. Hence, only the even numbered layer lines will be non-zero in this case. BDNA is a double helical structure, but in this case the helices are not separated by exactly half a pitch length and so the first layer line is non-zero but is, nevertheless, visibly less intense than the second.

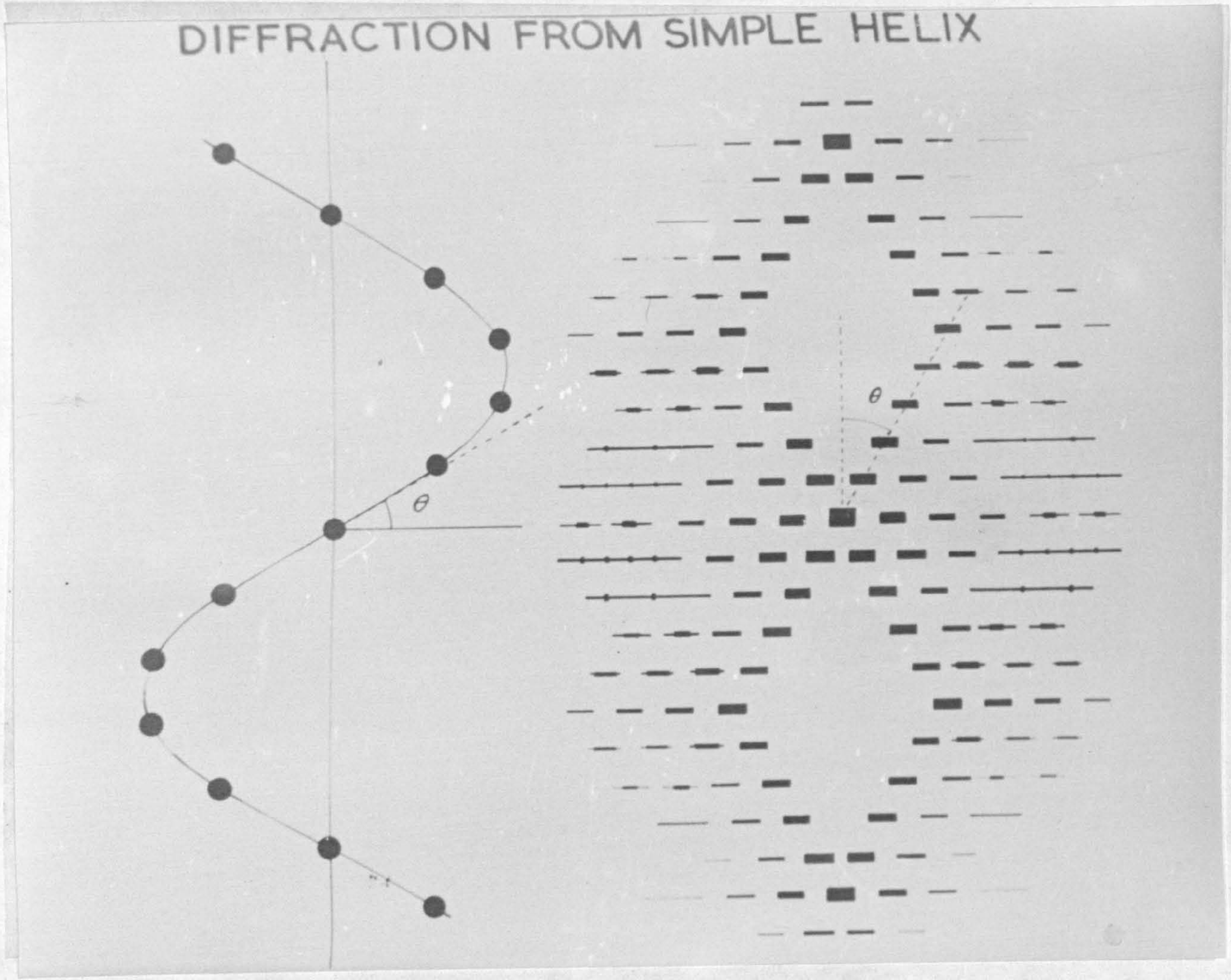


Fig. 3·2

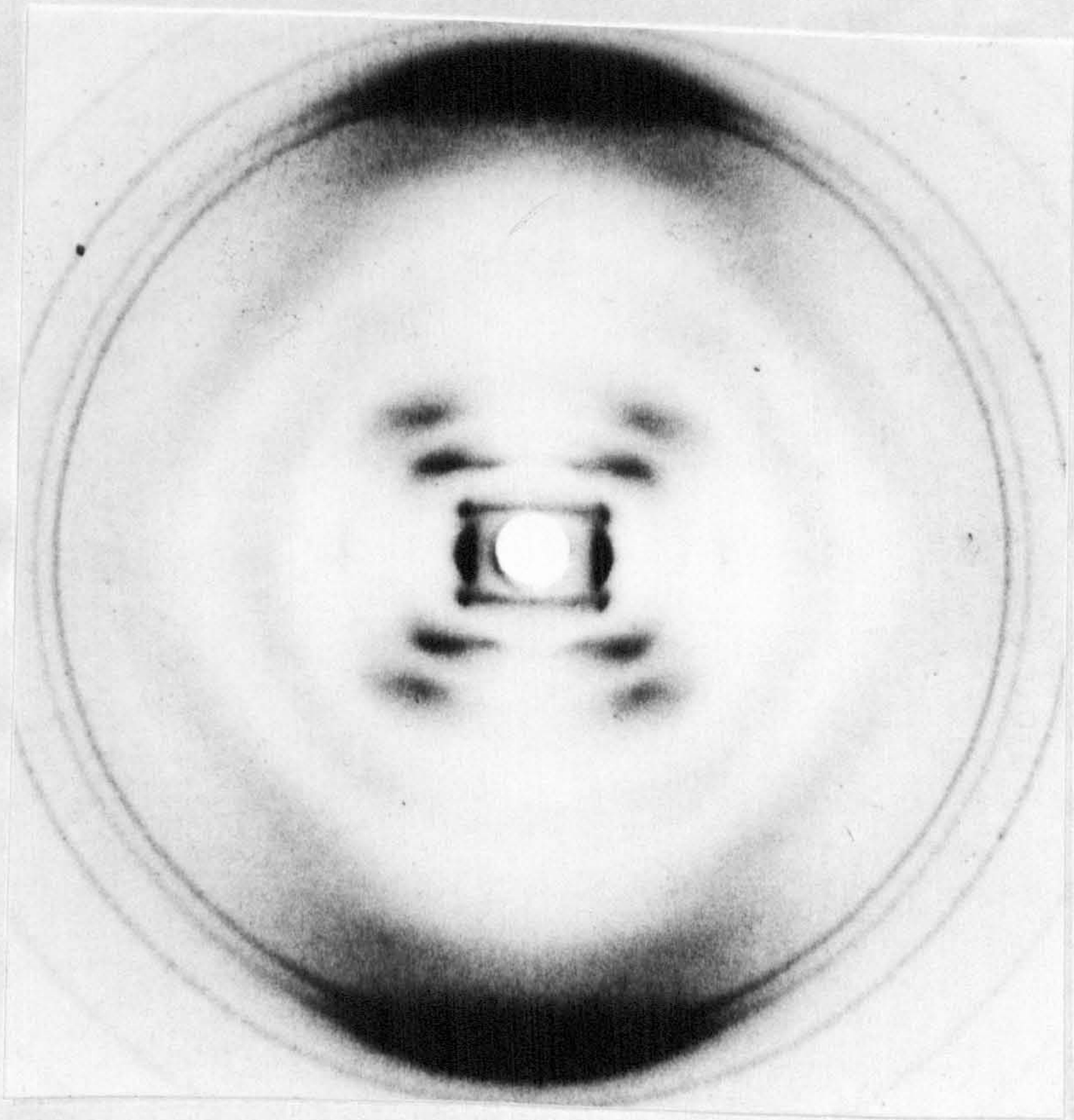


Plate 3.1 Sodium BDNA

For BDNA (N = 10; K = 1), a zero order Bessel function will occur on the 10th layer line when m = 1. Hence, meridional intensity would be expected to be observed on the 10th layer line; while consideration of layer lines around the 10th in terms of the selection rule with m = 1 shows that the distribution of Bessel functions around the centre of the diffraction pattern will be repeated around the 10th layer line (and around ℓ = -10 due to the diffraction pattern symmetry). The meridional intensity on the 10th layer line is visible in plate 3.1, though all detail of surrounding layer lines has been lost due to packing disorders.

An important point to note is that the diffraction from a helical polymer can be considered as the product the Fourier transform of an array of points spaced along a helix, and the transform of the group of atoms representing the repeat unit. Hence, the features of the diffraction pattern predicted from consideration of the selection rule may be masked in certain cases by the nature of the transform of the repeat unit. This is the case in ADNA pattern in which the helix cross around the central area is much less pronounced than in BDNA patterns; while the expected meridional intensity on the 11th layer line is scarcely visible.

The Bessel function argument in Fourier-Bessel transforms for helical structure is given by $2\pi\xi r$ so that a given value of the argument will correspond to a larger value of ξ if r is decreased. Hence, there is a reciprocal relationship between the position of Bessel function peaks and the radius of the helix; while the layer line separation is determined by the helix pitch. For this reason, the angle, θ illustrated in figure 3.2, between the helix and its axis is given approximately by the angle between the central intensity peaks and the equator of the diffraction pattern.

3.3 Types of Disorder in Helical Structures

Two different types of disorder can be distinguished in otherwise regular helical structures: substitution disorder in which the lattice is regular but in which all repeat units are not identical; and displacement disorder in which the lattice sites, themselves, are displaced from their regular position in some way. Methods of calculating the Fourier transform of structures exhibiting these types of disorder will now be discussed.

3.3.1 Substitution Disorder

Methods of treatment of this problem have already been formulated and have now been reproduced in many texts (e.g. Arnott, 1973). Since the resultant equations will be of significance in this chapter, their derivation will be presented here.

The Fourier transform of any structure which is composed of identical units distributed in a known way in space can be obtained by an equation of the form:

$$f(\underline{s}) = \frac{F(\underline{s})}{N} \sum_{j=1}^{N} \exp 2\pi i \, \underline{r}_j . \underline{s} \qquad 3.10$$

where $\underline{r}_j$ is a vector representing the displacement from the origin of the jth unit, and $F(\underline{s})$ is the Fourier transform of the repeat unit.

In order to calculate $f(\underline{s})$, knowing $F(\underline{s})$, it is necessary to perform the summation represented in 3.10 over all N positions. The maximum value of the summation, obtained when all components are in phase, is N; so that division by N as in 3.10 reduces the value of $f(\underline{s})$ to the scattering power of one repeat unit. If N is infinite, or effectively so, it is necessary to find a function representing the infinite sum if the calculation for $f(\underline{s})$ is to be performed. In the case of a crystal, $F(\underline{s})$ represents the transform of the unit cell contents; while the infinite sum of 3.10 for the regular array

produces a Dirac function.

The scattered intensity corresponding to $f(\underline{s})$ can be obtained from the relationship:

$$I(\underline{s}) = f(\underline{s}).f^*(\underline{s})$$

$$= \frac{F(\underline{s}).F^*(\underline{s})}{N^2} \sum_{k=1}^{N} \sum_{j=1}^{N} \exp 2\pi i(\underline{r}_k - \underline{r}_j).\underline{s} \qquad 3.11$$

In the case of an array exhibiting substitution disorder, there will be more than one possibility for the nature of the repeat unit at each lattice site and, hence, for the $F(\underline{s})$ of 3.10. Therefore, we have:-

$$f(\underline{s}) = \frac{1}{N} \sum_{j=1}^{N} F_j(\underline{s}) \exp 2\pi i\, \underline{r}_j.\underline{s} \qquad 3.12$$

where $F_j(s)$ represents the transform of the jth repeat unit.

For the intensity we have:-

$$I(\underline{s}) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{j=1}^{N} F_k(\underline{s}).F_j^*(\underline{s}) \exp 2\pi i(\underline{r}_k - \underline{r}_j) \qquad 3.13$$

Let us define a mean unit cell as being one whose transform fulfils the following conditions (3.14 and 3.15):

$$F_{av}(\underline{s}) = F_j(\underline{s}) - \Delta_j(\underline{s}) \qquad 3.14$$

and

$$\sum_{j=1}^{N} \Delta j = 0 \qquad 3.15$$

also,

$$F^*_{av}(\underline{s}) = F^*_j(\underline{s}) - \Delta^*_j(\underline{s}) \qquad 3.16$$

Substituting into 3.13 and multiplying out gives:-

$$I(\underline{s}) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{j=1}^{N} \left[F_{av}(\underline{s}).F^*_{av}(\underline{s}) + F^*_{av}(\underline{s})\Delta(\underline{s}) + F_{av}(\underline{s})\Delta^*_j(\underline{s}) + \Delta_k(\underline{s}).\Delta^*_j(\underline{s}) \right] \exp 2\pi i(\underline{r}_k - \underline{r}_j)\underline{s} \qquad 3.17$$

If the summation is taken over pairs of repeat units having identical values for $\underline{r}_k - \underline{r}_j$ then the second and third terms in brackets in equation 3.17 vanish because of the condition specified in 3.15. Unless k = j the fourth term also vanishes and so 3.17 reduces to:-

$$I(\underline{s}) = \frac{F_{av}(\underline{s}).F^*_{av}(\underline{s})}{N^2} \sum_{k=1}^{N} \sum_{j=1}^{N} \exp 2\pi i(\underline{r}_k - \underline{r}_j)\underline{s} + \frac{1}{N} \sum_{j=1}^{N} \Delta_j(\underline{s}).\Delta^*_j(\underline{s}) \qquad 3.18$$

Comparison with 3.11 shows that the first term in equation 3.18 is the transform of a regular array having the mean unit cell situated at every lattice point. If the array is spatially regular and infinite in extent, the exponential part of this term sums to give a Dirac function which is non-zero only at special points. Hence, for such a lattice this term consists of a series of sharp Bragg reflections. The second term contains no exponential term and is generaly non-zero giving rise to a diffuse background of intensity over the whole of reciprocal space. In many cases, the contribution of this second term is small and the calculation is then performed using only the first term of equation 3.18

F_{av}, the transform of the mean unit cell is formed as:-

$$F_{av}(\underline{s}) = \sum_{j=1}^{M} W_j \, F_j(\underline{s}) \qquad 3.19$$

where W_j is a weighting factor corresponding to the proportion of the jth repeat unit in the structure, and M is the number of possible repeat units. The summation is performed in amplitude and phase and the weighting factors are normalised so that their sum is unity.

The diffuse scattering term of 3.18 is then given by:-

$$\frac{1}{N} \sum_{j=1}^{N} \Delta_j \cdot \Delta^*_j = \frac{1}{N} \sum_{j=1}^{N} \left\{ \left[F_j(\underline{s}) - \sum_{i=1}^{M} W_i \, F_i(\underline{s}) \right] x \left[F^*_j(\underline{s}) - \sum_{i=1}^{M} W_i \, F_i^*(\underline{s}) \right] \right\} \qquad 3.20$$

The summation of 3.20 is not, in its present form, amenable to computation since N will normally be infinite. However, it can be written in the form:-

$$\frac{1}{N} \sum_{i=1}^{N} \Delta_j \cdot \Delta_j^* = \sum_{j=1}^{M} W_j \cdot \Delta_j \cdot \Delta_j^* = \sum_{j=1}^{M} \left[W_j \left[F_j(\underline{s}) - \sum_{i=1}^{M} W_i \, F_i(\underline{s}) \right] x \left[F_j(\underline{s}) - \sum_{i=1}^{M} W_i \, F_i(\underline{s}) \right] \right] \qquad 3.21$$

Here the summation is taken over the M canonical forms contributing to F_{av}.

It should be noted that the terms which vanish from equation 3.18 only do so if it is assumed that there is no statistical correlation between unit cells; and that there is also no correlation between the nature of the vector separating two lattice sites and the nature of the repeat unit occurring at either of them. If any such correlation does exist, then further

terms from equation 3.17 will be non-vanishing. The final form of the expression for the diffuse scattering depends upon the nature of the correlation effects, and will be discussed for one particular case in section 3.4.

3.3.2 Displacement Disorder

Two types of displacement disorder may be distinguished, and these will be discussed separately.

(i) Disorder of the First Kind

In displacement disorder of the first kind, the lattice points are displaced from their ideal position in such a way that the position of one lattice site does not affect the position of its nearest neighbour. Hence, the mean displacement from the "true" position does not increase with increasing distance from the origin.

An example of this type of disorder is that produced by thermal motion of atoms and which is corrected for by using the Debye temperature factor. In fact thermal motion probably does not represent true disorder of the first kind, since the position of any atom will not be independent of the position of its neighbours, however, in the Debye theory it is assumed that second order effects are negligible.

Some idea of the general method of treatment of this type of disorder may be obtained if it is noted that, essentially, it is only a special case of substitution disorder in which each of the possible unit cells differ only in the phase, and not in the amplitude of their scattering.

(ii) Disorder of The Second Kind

In disorder of the second kind the lattice sites are displaced as before but the displacement is partially or totally carried over to lattice sites further out from the origin. Hence, the displacement of any lattice site is not independent of the positions occupied by its neighbours, but is a linear sum of its own displacement from an ideal position and that of all other lattice sites which lie between it and the origin.

Interference effects due to intermolecular separation decrease, usually very rapidly, with increasing diffraction angle in diffraction patterns from structures exhibiting disorder of the second kind. This can be seen most clearly in diffraction patterns from liquids, all of which possess this disorder to a high degree. At low angles intermolecular interference effects are observable in the form of a diffuse but distinct diffraction ring, but at higher angles interference effects disappear completely and the liquid scattering at these angles is indistinguishable from that of an ideal gas.

The "molecular disorders" discussed in section 3.2 relating to intercalative drug binding and the variation of B genus DNA structures, provide examples of this type of disorder in nucleic acid fibre specimens.

3.4 Calculation of the Fourier Transform of Helical Polymers Exhibiting Disorder of The Second Kind

To treat this problem we begin by considering an infinite series of points lying on the surface of a cylinder of radius R. The nth Bessel function component of the transform of such a series

of points is given by (see appendix),

$$G(\xi,n,\zeta) = J_n(2\pi\xi r) \sum_{j=1}^{N} f_i \exp i(2\pi\zeta z_j - n\phi_j) \qquad 3.22$$

where f_j represents the scattering from one point and will be of uniform value in reciprocal space. The Bessel function term has been removed from the summation since all points are at constant radius, r.

Consider an array of points on the cylinder possessing disorder of the second kind such that each point has a nearest neighbour in one of s possible positions: the probability of finding a nearest neighbour at the jth position being P_j. In a lattice of N points, the contribution to the scattered intensity of each point and its nearest neighbour can be obtained by taking each point in turn as origin; calculating the contribution to the transform of the origin point and its right and left hand nearest neighbours in each case; and, finally, summing the contributions from all origin points to get the total result.

We have, therefore:-

$$I_{0+1}(\zeta,n) = Nf + f \sum_{j=1}^{s} N_j \exp i(2\pi\zeta z_j - n\phi_j)$$

$$+ f \sum_{j=1}^{s} N_j \exp -i(2\pi\zeta z_j - n\phi_j) \qquad 3.23$$

where N_j is the number of times a first neighbour is found at the jth position. Since 3.23 was derived by considering all points in turn as origin, and including nearest neighbour vectors in each case, it gives the intensity of the scattering from each point and its nearest

neighbours. In accord with this it may be noted that the right-hand side of equation 3.23 is real since the first summation, representing the contribution from right-hand nearest neighbours, is the complex conjugate of the transform of left-hand nearest neighbours given by the second summation of 3.23. The two exponential terms could be combined, therefore, but are kept separate for the present. The function of 3.23 can in fact be considered as the inverse of a Patterson function and is analogous to the packing factor of Fuller et al. (1967). As in equation 3.22, the right-hand side of equation 3.23 should be multiplied by a Bessel function term. However, since for the moment we are interested only in the variation of I as a function of ζ and n, this term has been ignored and will be reintroduced later.

If we consider all points as having unit scattering power, and divide by N, the number of units in the lattice, we have:-

$$I_{0+1}(\zeta,n) = 1 + \sum_{j=1}^{s} P_j \exp i(2\pi\zeta z_j - n\phi_j) + \sum_{j=1}^{s} P_j \exp i(2\pi\zeta z_j - n\phi_j) \qquad 3.24$$

The above equation can be used even when N become infinite.

In order to obtain the contribution of second neighbours we proceed in an analogous way to before. Let us consider the right hand neighbours only, and assume that all first neighbours are at position 1; then for the contribution of second nearest neighbours we have:-

$$I_z(\zeta,n) = \exp i(2\pi\zeta z_1 - n\phi_1)\, E(\chi) \qquad 3.25$$

where $E(\chi) = \sum_{j=1}^{s} P_j \exp i(2\pi\zeta z_1 - n\phi_1)$

The first neighbour could be at any one of the s possible positions, however, and for the total contribution from right hand second neighbours we have:-

$$I_2(\zeta,n) = E(\chi) \sum_{j=1}^{s} P_j \exp i(2\pi\zeta z_1 - n\phi_j) \qquad 3.26$$

$$= E^2(\chi) \qquad 3.27$$

It can be shown by a similar analysis that the contribution from the distribution of jth neighbours can be given by $E^j(\chi)$. Therefore, the intensity transform of the whole array is given by:-

$$I_\infty(\zeta,n) = 1 + \sum_{j=1}^{\infty} \left[E^j(\chi) + E^j(-\chi) \right] \qquad 3.28$$

This sums to:-

$$I_\infty(\zeta,n) = 1 + \frac{E(\chi)}{1-E(\chi)} + \frac{E(-\chi)}{1-E(-\chi)} \qquad 3.29$$

Combining complex conjugate terms from left and right hand neighbours gives:-

$$I_\infty(\zeta,n) = 1 + \frac{2P\cos(\chi) - 2P^2}{1 - 2P\cos(\chi) + P^2} \qquad 3.30$$

where $P = |E(\chi)|$ and $P\cos(\chi) = \mathrm{Re}\ E(\chi)$

The above theory is similar to that derived by Zernicke and Prins (1927) for the scattering of X-rays by liquids which also exhibit disorder of the second kind, and a similar equation is also derived by Vainstein (1966) in a general treatment disorder of the second kind.

We can now use this result for a disordered set of points to calculate the transform for a disordered molecular array. The

point function whose transform we have just calculated could be derived by taking each lattice point in turn to be the origin; adding together the different lattice so obtained and normalising to the value of one lattice. This process (neglecting a normalisation factor) is the P self-convolution of the inifinite lattice. According to the convolution theorem, a P self-convolution in real space corresponds to the multiplication of the transform of the function by its complex conjugate in reciprical space. This in turn corresponds to the intensity transform of the function. Hence, the function $I(\ell,n)$ of 3.30 can only be used to obtain the intensity transform of a molecular array.

This is obtained as:-

$$I(\xi,\zeta) = \sum_{n} G(\xi,n,\zeta)\ G^{*}(\xi,n,\zeta)\ I_{\infty}(\zeta,n) \qquad 3.31$$

where $G(\xi,n,\zeta)$ is obtained from 3.5.

For a regular helix: i.e. one in which there is only one possibility for the nearest neighbour position, an analogous procedure to the one outlined above would yield:-

$$I_{\infty}(\zeta,n) = 1 + \frac{\cos(2\pi z\zeta - n\phi) - 1}{1 - \cos(2\pi z\zeta - n\phi)} \qquad 3.32$$

The above expression is non-zero only when the cos term is equal to unity. At such points the expression becomes indeterminate, and the value which should be given to the function will be discussed later in this section. For the moment it is sufficient to accept the fact that it has a positive value when the cos term is unity. If the rotation between units, ϕ, is some rational fraction of 2π we can write:-

$$\phi = \frac{2\pi k}{N}\ , \text{ where } k \text{ and } N \text{ are integers.}$$

For the cos term to equal unity its argument must equal some multiple, m, of 2π.

Therefore,

$$2\pi z\zeta - \frac{2\pi kn}{N} = 2\pi m$$

$$\therefore \quad z\zeta - \frac{kn}{N} = m$$

$$\therefore \quad \zeta = \frac{mN + km}{zN}$$

This equation is satisfied when $\zeta = \frac{\ell}{zN}$, where ℓ is any integer.

Hence, $\ell = mN + kn$, the selection rule of Cochran et al., (1952). A regular infinite lattice is identical with its self-convolution and hence 3.32 can be considered as the transform of the un-convoluted lattice and can be used to obtain the transform of an array in amplitude and phase.

Since no concept of helical pitch has been introduced, the present method of analysis can be used to treat the case of the irrational helix. The analysis gives the following selection rule:-

$$\ell = \frac{m + \mu n}{z}$$

where μ is an irrational number equal to $\phi/2\pi$. A different method of treating the irrational case has been presented by *RAMACHANDRAN* (1960).

When there is more than one possibility for nearest neighbour positions the situation is much less simple and is not reducable to an exact selection rule of any form. However, it is possible to make certain predictions about the general nature of the intensity data from such arrays. In particular it is possible to predict the positions at which maximal intensity will be observed. Equation 3.30 will have a maximum when the term in parenthesis is at

a maximum, This condition will occur when the real part of the transform of the first neighbour distribution given by,

$$\sum_{j=1}^{N} P_j \cos(2\pi z_j \zeta - n\phi_j) = P \cos(\chi)$$

is maximal.

Differentiating and equating to zero give:-

$$- P \sin(\chi) = 0 \qquad 3.33(a)$$

For a maximum the above equation must be satisfied while the second derivative given by

$$- P \cos(\chi)$$

must be negative.

These conditions are approximately satisfied when the following condition is attained.

$$\sum_{j=1}^{N} P_j(2\pi z_j \zeta - n\phi_j) = 2\pi m \qquad 3.33(b)$$

The assumption is made that the value of all the N arguments are near in value to a mutliple of 2π when condition 3.28 is satisfied. At this point only arguments associated with small values of P_j (and hence make a correspondingly small contribution to the transform) can be significantly different from a multiple of 2π.

Rearranging 3.33(b) gives:-

$$2\pi\zeta \sum_{j=1}^{N} P_j Z_j + n \sum_{j=1}^{N} P_j \phi_j = 2\pi m \qquad 3.34$$

But $\sum_{j=1}^{N} P_j Z_j$ and $\sum_{j=1}^{N} P_j \phi_j$ are the weighted averages of the z and ϕ coordinates respectively.

Hence we can write,

$$2\pi\zeta z_{av} - n\phi_{av} = 2\pi m$$

From here one proceeds as before to obtain:-

$$\zeta = mN_{av} + K_{av}N \qquad 3.35$$

where N_{av} and K_{av} are analogous to the N and K of equation 3.9 but refer to a helix having a rotation and translation zero residue of ϕ_{av} and Z_{av} respectively. The helix pitch will similarly be averaged and this will be the value measured from diffraction photographs. This type of averaging has been assumed by a number of authors (e.g. Neville and Davies, 1973; Pigram, 1972; Fuller, 1966), when using pitch values obtained from diffraction patterns of intercalation complexes.

It should be emphasised that, unlike quation 3.9, equation 3.39 only specifies the approximate condition for a maximum and that equation 3.30 will not generally be zero at other points in reciprocal space. The condition specified in 3.33(b) becomes a worse solution to as either ζ, n or m increases and the maxima in 3.30 become smaller and tend to merge with the background for larger values of the scattering angle. A limitation on this is that the function I_∞ of 3.30 may be periodic in simple cases.

Consider a case in which there are only two possibilities for nearest neighbour positions, then for the real part of the nearest neighbour transform we have:-

$$P_1 \cos(2\pi z_1\zeta - n\phi_1) + P_2 \cos(2\pi z_2\zeta - n\phi_2) \qquad 3.36$$

$$= P_1 \cos(2\pi z_1\zeta - n\phi_1) + P_2 \cos(2\pi(z_1+\Delta z) - n\phi_2) \qquad 3.37$$

where $\Delta z = z_2 - z_1$

Suppose that $\frac{z_1}{\Delta_z} = m$, where m is an integer, then if $\zeta = \frac{\ell\, m}{z_1}$ where ℓ is any integer +ve, -ve or zero, we have:

$$P_1 \cos(2\pi m\ell - n\phi_1) + P_2 \cos(2\,(m+1)\ell - n\phi_2) \qquad 3.38$$

which gives

$$P_1 \cos(-n\phi_1) + P_2 \cos(-n\phi_2) \qquad 3.39$$

Thus for any Bessel function order, n, the function 3.30 will be periodic in the ℓ direction over a reciprocal distance of $\frac{m}{z_1}$. If n = 0 then the argument of 3.30 will be zero and the condition specified in 3.33(a) will be satisfied regardless of the values of P_1 and P_2, and will result in a Bragg reflection in which all units scatter in phase. This is significant for drug/DNA intercalation complexes since the values of P_1 and P_2 are determined by the P/D ratio. The position of peaks in the intensity transform which result from the condition specified above will be invariant with P/D and, hence, may be used to determine the value of z (in this case Δz = the increase in nucleotide separation brought about by the intercalating drug). An example of such an intensity peak is the 3.4A meridional reflection observed in all intercalation complexes. The position of the peak does not vary with the limits of experimental error, and shows that the value of Δz is 3.4Å.

Equation 3.33 may also be periodic with regard to Bessel function order and in principle it should be possible to identify a peak on the equator the position of which is not a function of the P/D, from which the degree of untwisting at the intercalation site could be determined. Unfortunately, such a peak would correspond to a Bessel function of very high order (n = 30 in the case of a 12 degree untwist angle with respect to BDNA). Such peaks occur far out in reciprocal space and are not easily recorded.

The physical explanation of the periodicity is illustrated in figures 3.3, in which the transform of the first neighbour distribution is illustrated in vector form for the case where s, the number of possible first neighbour positions equals two. The amplitude of the vectors are given by the probabilities, P_1 and P_2, of the occurence of a first neighbour at each of the possible positions. The sum of the vectors gives another vector representing the amplitude and phase of the first neighbour transform. In figure 3.3(a) the situation where the reciprocal space parameters, ζ and n, equal zero is illustrated. The phase of both vectors is clearly zero and the resultant vector is of unit modulus. At certain values of the reciprocal space parameters, denoted by ζ' and n', the situation illustrated in 3.3(b) will pertain and a maximum will occur in the value of I_∞. Let us suppose that at some other value of these parameters, denoted by ζ^* and n*, both vectors are equal to some multiple of a 2π, then the vectors will be identical to those observed when ζ and n are zero and the position illustrated in 3.3(a) will be regained. As the reciprocal space parameters increase still further, the vectors will behave in an identical manner to that observed in the region around the origin, and when $\zeta = \zeta^* + \zeta'$ and $n = n^* + n'$ the situation illustrated in figure 3(b) will be regained. If there are more than two possibilities for nearest neighbours, then an analogous argument can be used to show that the same conditions might apply. However, as s increases, the regions of reciprocal space at which all the vectors will equal some multiple of 2π will be more widely spaced; and in the limit, if the first neighbour distribution were a continuous function as it is in liquids, then the condition would never be attained for finite values of the reciprocal space parameters. In general, the ζ

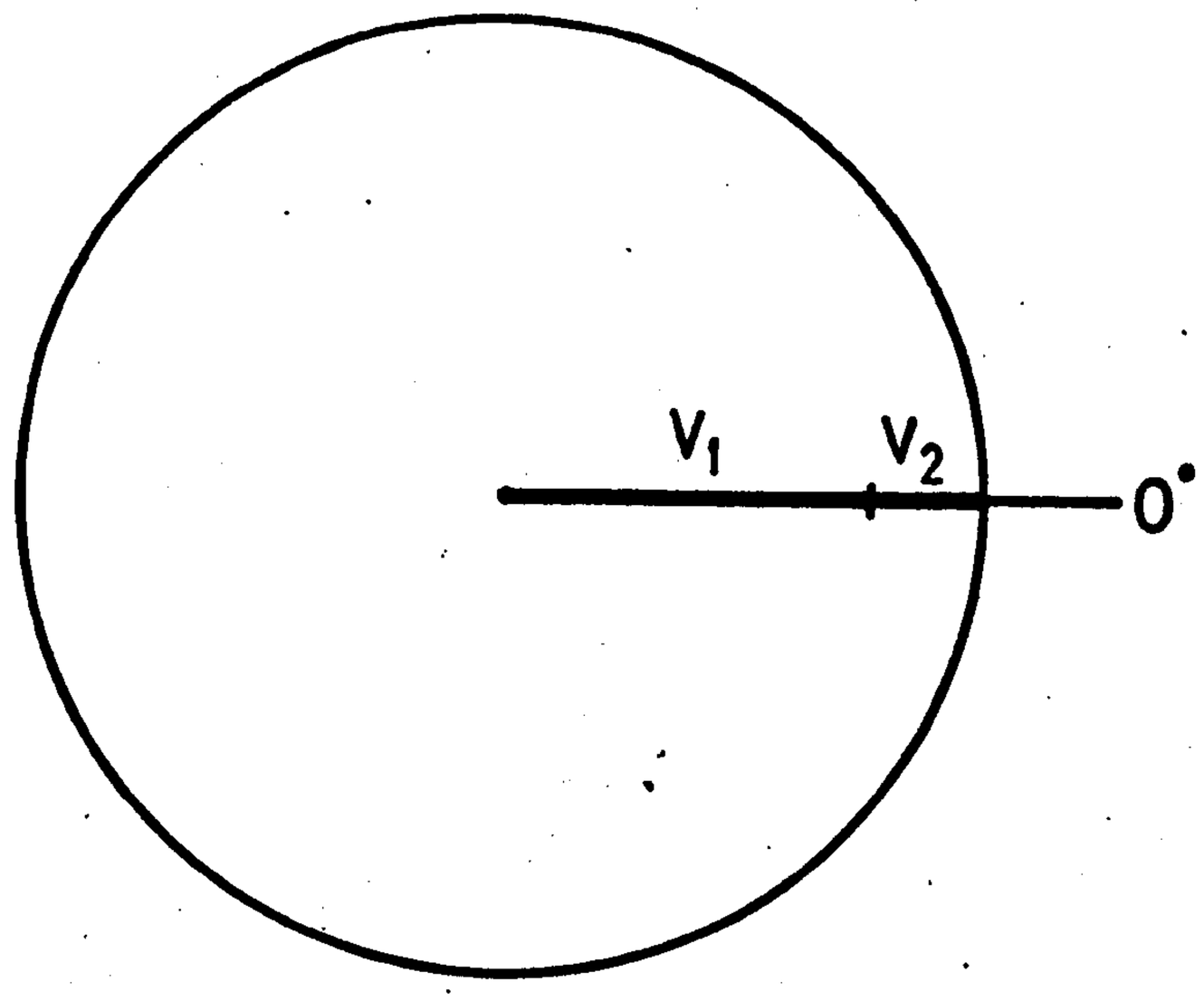


Fig. 3·3(a)

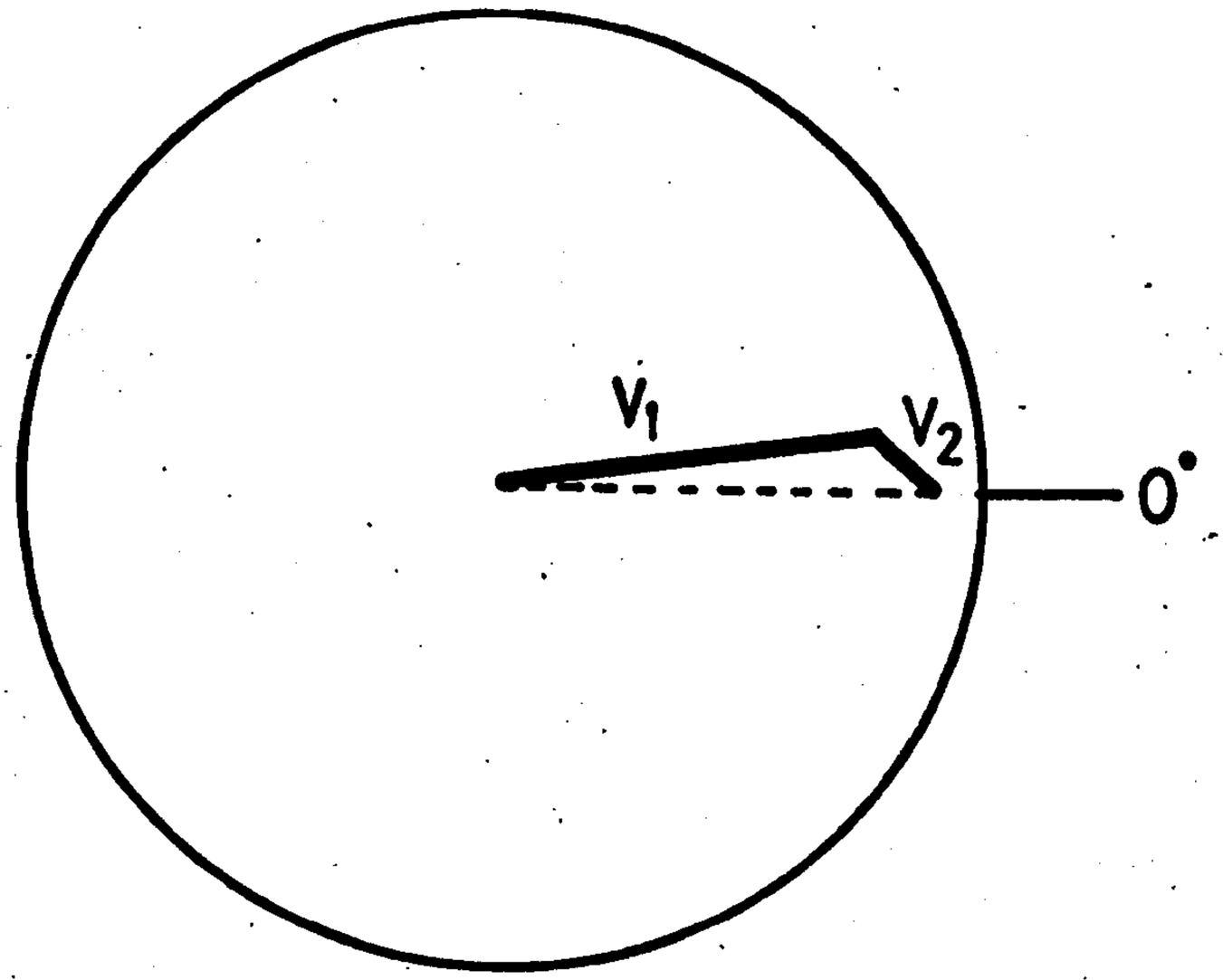


Fig. 3·3(b)

distance over which I_∞ will be periodic is given by $\frac{M}{z_1}$, where $M = \frac{z'}{z_1}$: z' being the lowest common multiple of all possible z values for nearest neighbours.

In diffraction patterns from intercalation complexes, therefore, the layer line structure observed around the centre of the pattern and which would be lost due to the intra-molecular disorder even if the molecules were regularly packed should, in theory, be observed around the meridionals. However, this will not be observed in practice since any layer line structure in the pattern other than that around the centre, will be destroyed by disorder in the molecular packing.

Values for the function $I_\infty(\zeta,n)$ have been calculated for a specified case of disorder of the second kind and are presented in figure 3.4. The periodicity of the function over a distance $\frac{M}{z_1}$ and the infinitely high peaks at $\zeta = \frac{m}{z_1}$ for the n = 0 component, are observable.

If it is desired to calculate the value of $I_\infty(\zeta,n)$ at all points in reciprocal space, then the infinitely high, sharp peaks must be normalised and scaled (in terms of intensity per unit length) relative to the continuous diffraction streaks. This can be done if we note that the integral of the term in parenthesis in equation 3.30 over the repeat period, $\frac{M}{z_1}$, in the ζ direction is zero. Integration confirms this result, though it is intuitively obvious if it is remembered that this term consists of a sum of cosine functions whose integrals over the period $\frac{M}{z_1}$ are zero:

$$\rho_{av} \int_{\zeta=0}^{\frac{M}{z_1}} I_\infty(\zeta,n)\, d\zeta = \frac{\rho_{av}\, M}{z_1} \qquad 3.40$$

where ρ_{av} is the average density of points.

Equation 3.40 gives the value of the average diffracted intensity per unit length to which the calculated data must be normalised. Hence, the "infinite" peaks in G_∞ take the value of $\frac{\rho_{av}\, M}{z_1}$, while the diffuse diffraction takes the value:-

$$I'_\infty(\zeta,n) = I_\infty(\zeta,n)\, \rho_{av}\, \Delta\ell \qquad 3.41$$

where $\Delta\zeta$ is the interval along the ζ direction at which values of $I_\infty(\zeta,n)$ are calculated. The procedure adopted here appears somewhat artificial at first but must be used if any progress is to be made since it converts the calculated data to values of intensity per unit length, and only in this way can data from Bragg type reflections and diffuse scattering streaks be compared. The values of $I^1(\zeta,n)$ calculated from 3.41 are dependent upon the interval, $\Delta\zeta$, at which they are calculated. Although decreasing the value of $\Delta\zeta$ decreases the values of $I'_\infty(\zeta,n)$, the function is sampled at a correspondingly larger number of points and the intensity calculated in a given region of reciprocal space will be the same whatever value is chosen for $\Delta\zeta$, except for changes due to the improved sampling of the function profile for smaller $\Delta\zeta$.

Since ρ_{av}, the average point density, appears in both equation 3.40 and 3.41, it constitutes a common scaling factor and may be omitted in practice.

It is not necessarily vital to perform this scaling at all since, provided it can be measured accurately, the diffuse scattering provides the most information, particularly in the present case in which the Bragg reflections are far apart in reciprocal space. In this case it is not necessary to perform the scaling procedure since it does not alter the position, shape or relative heights of peaks in the diffuse scattering.

Examination of the graphs presented in figure 3.4 show that I_∞ is only fluctuating significantly as a function of ζ for Bessel functions of low order. These occur near the centre of the pattern and hence the affect of I_∞ on the total transform is dominant only in this region. Hence, information about the nearest neighbour positions can only be obtained from studies of the layer lines. At higher angles, the transform of the repeat unit will be fluctuating more rapidly than I_∞ and, hence, variation in the

diffracted intensity in this region will be dominated by the form of the repeat unit transform.

This is analogous to the case of liquid scattering in which information about the intermolecular separation function can only be obtained from the low angle scattering intensity.

It is possible that a different unit cell is associated with each of the nearest neighbour positions. In this case the first neighbour transform is given by:-

$$G_1(\zeta,n) = \sum_{j=1}^{s} I_j P_j \exp i(2\pi\zeta z_j - n\phi_j) \qquad 3.42$$

where I_j = the intensity transform of the jth unit cell.

Consider a case where s = 2; then the first neighbour distribution is given by:-

$$I_1P_1 \exp i(2\pi\zeta z_1 - n\phi_1) + I_2P_2 \exp i(2\pi\zeta z_2 - n\phi_2) \qquad 3.43$$

while for the transform of the second neighbour distribution we have:-

$$I_1P_1^2 \exp i(2\pi\zeta z_1 - n\phi_1) + (I_1 + I_2)\, P_1P_2 \exp i(2\pi\zeta z_1 - n\phi_1)$$
$$\exp i(2\pi\zeta z_2 - n\phi_2) + I_2P_2^2 \exp i(2\pi\zeta z_2 - n\phi_2) \qquad 3.44$$

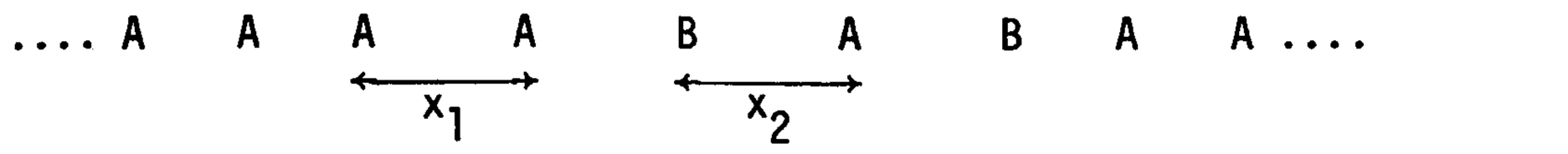
Hence, in this case:-

$$G_2(\zeta,n) \neq G_1^2(\zeta,n) \qquad 3.45$$

This is an example of substitution disorder and can be treated using the method of approach outlined earlier. The function $I_\infty(\zeta,n)$ of equation 3.30 is equivalent to the double summation $\left(\sum_{k=1}^{N}\sum_{j=1}^{N} \exp 2\pi i(\underline{r}_k - \underline{r}_j)\underline{s}\right)$ of equation 3.18 and hence the first term of this equation can be calculated. Some care is required in calculating the contribution from the diffuse scattering, however.

It has been mentioned in the derivation of equation 3.18 that terms involving Δ will only vanish if they are unaffected by any correlation between unit cells or their vector separation. In the case under discussion at present, such correlation does exist and, hence, a number of terms from equation 3.18 need to be included in the calculation of the diffuse component.

The nature of the correlation in this case is best illustrated with reference to an example. Consider the one dimensional infinite lattice illustrated below, which consists of two types of repeat unit denoted by A and B.



The presence of unit B results in larger distance of separation between itself and the next repeat unit on the right hand side, and, hence, the lattice possesses both substitution disorder and displacement disorder of the second kind. We define the transform of the mean unit cell, F_{av}, to be $W_A F_A + W_B F_B$, where F_A and F_B are the transforms of unit A and unit B respectively. W_A and W_B are weighting factors describing the respective probabilities of the occurrence of unit A and unit B at each lattice site. If the function I_∞ is derived, the value of the non-diffuse scattering can be calculated. If we consider any pair of lattice points of the array, then the right hand lattice point of the pair will have the distribution $W_A A + W_B B$ for the probability of finding each repeat unit at the site. The left hand member of the pair will determine the vector separation of the two units, however, so that the nature of the separation vector and of the type of repeat unit at the left hand lattice site are directly correlated. Consideration of equation 3.18 in the light of this correlation shows that the final

equation for the scattering intensity should be:-

$$I_L(\underline{s}) = I_{av}(\underline{s}) \cdot I_\infty(\underline{s}) + 2Re\left[\sum_{j=1}^{M} W_j \Delta_j(\underline{s}) \sum_{n=1}^{\infty} \exp(n x_j \underline{s})\right]$$

$$+ \sum_{j=1}^{n} W_j \Delta_j(\underline{s})\ \Delta_j^*(\underline{s}) \qquad 3.46$$

where $I_{av} = F_{av}\ F_{av}^*$

and I_L = the intensity transform of the lattice.

In order to illustrate the working of the method in practice, it will be applied to two different examples of disorder of the second kind in nucleic acid structures: the case of a DNA duplex in which different nucleotides have a different conformation; and an example of an intercalation complex.

(a) <u>A Disordered Helix</u>

This case is not merely of academic interest, since the possibility exists that such disorder may be present in certain B type diffraction patterns obtained from Na DNA, discussed in chapter 6. Suppose that there are 5 possible conformations for the nucleotide repeated in the disordered helix, not necessarily occurring with the same frequency. Let us further suppose, for the moment, that there is no statistical correlation between repeat units; even between nearest neighbours.

The first stage in the calculation is to obtain values for the function I_∞ of equation 3.30 for a given set of values for ζ and for a given range of Bessel function orders. This can then be multiplied by the transform of the average unit cell which must be calculated for the same Bessel function orders and for the same values of ζ. The transform of the average unit cell is obtained as a sum over all atoms in the s repeat units: the scattering factors for

atoms in the jth unit being multiplied by W_j, the probability of finding the jth repeat unit at a lattice site. The product $F_{av}\ F_{av}^*$ I_∞ is then formed separately for each Bessel function component to give the value of the non-diffuse term.

Since each different nucleotide conformation will result in different characteristic values for the translation and rotation between units, then a correlation exists between the separation vectors and the repeat units, of the type discussed in the previous section. Hence, the total transform should be calculated in accord with equation 3.45.

If, in addition, there is a correlation between the appearance of one particular repeat unit at a lattice site and the probability of the appearance of any repeat unit at adjacent lattice sites, the procedure adopted needs to be modified. Consider one particular case in which each different repeat unit does not occur individually but in groups of N nucleotides along the length of the helix. It is possible to treat his case by further modification of the nature of the terms included in the equation for I_L. However, a better method of approach is to modify the nature of F_{av} by considering each group of N nucleotides as one repeat unit. If this procedure is carried for all s different nucleotide conformations, then a new set of different possible repeat units are generated. If it is supposed that there is no statistical correlation between these new repeat units (N nucleotides in length) then the transform can be calculated as before, using new rotational and translational parameters appropriate to the enlarged repeat units. The transform for each of the repeat units, consisting of N identical nucleotides can be calculated as described in section 3.4

(b) An Intercalation Complex

Consider the intercalation complex illustrated diagrammatically in figure 5.1, for drug intercalation into BDNA. We assume for the moment that the unwinding is restricted to the intercalation site, and that the separation between base pairs is 6.8Å. It is further assumed that the only conformational change induced in the nucleotide repeat is in the phosphate group. There are two different possible repeat units at each lattice point, therefore. One of these is the normal B-DNA nucleotide, which induces a translation and rotation per residue of 3.4Å and 36^{o} respectively; while the other consists of the drug and the modified nucleotide (modified phosphate) inducing a translation and rotation per residue of 6.8Å and 24^{o}. The transform can now be calculated in a manner entirely analogous to that described above, using equation 3.46.

The assumptions made above are probably unrealistic for any actual intercalation complex since changes in the sugar orientation about the glyocosidic bond will amost certainly occur; while other changes might well be introduced. These changes might lead to "excluded site" effects so that the drug distribution would be modified from that of the present example.

These changes can be accounted for by enlarging the size of the repeat unit which includes the drug. If the drug modifies a number of nucleotide pairs adjacent to the site of intercalation, then the whole of the modified region should be included, along with the drug, as being one repeat unit. The values of the translation and rotation parameters associated with this unit should not be those quoted above, but should be new values appropriate to the enlarged unit.

3.5 A Particular method of Calculating the Molecular Transform for Intercalation Complexes

The method of calculation presented in 3.3, is applicable to the treatment of intercalation complexes, but another method of approach can be used by making the following approximation.

Consider an intercalation complex with P/D = 10, giving an average of one drug per five base pairs. The approximation to be made is that it is possible to choose an origin such that every concurrent sequence of five base pairs contains one (no more and no less) bound drug molecule, although the site of attachment within the sequence of five nucleotides is random. The helix would then be made up of a repeating sequence of units consisting of five nucleotide pairs and one drug molecule. These repeat units would have the same translation length and turn angle but would be non-identical and hence cause substitution disorder in the structure. Little discussion of the general method of treatment is presented since it is identical in principle to other cases of substitution disorder already discussed. Some points concerned with the mechanics of the calculation will, however, be presented here.

It is probably adequate, certainly for the preliminary examination of diffraction data, to use only the first term of equation 3.18. Since this represents a regular helix having the average repeat unit at each lattice point, it can be calculated using equation 3.7 derived from the diffraction theory of Cochran, Crick and Vand (1952). For a more detailed analysis of the data, the diffuse scattering should also be calculated and added to that from the non-diffuse component to obtain the final value of the calculated intensity. Since the helix possesses pure substitution disorder, and it can be assumed that there is no statistical correlation between

repeat units, the diffuse component may be calculated by equation 3.21.

The average repeat unit generally contains a large number of atoms; a fact which makes both the data preparation and the computation extremely tedious if all atoms are explicitly included in the summation. It is possible to reduce considerably the work involved by the following method.

Consider a sequence of five bases situated regularly on a smooth helix. The sequence can be considered as the convolution of one base and a point function describing the spatial distribution of the base unit. By the convolution theorem, it is possible to obtain the transform of the sequence as the product of the transform of the base and that of the point function. If the structure is 'statistical' in that more than one position may be occupied by each base, then the function describing the spatial distribution is modified to include all possible points at which each base could be situated: each point being given a weighting factor equal to the probability of binding a base at that particular site.

In practice it may be that different groups of atoms in the intercalation unit have different distribution functions. The transform for each group, multiplied by their appropriate distribution function transform, must then be calculated separately and their sum formed in amplitude and phase to obtain the final result.

Transform calculations for particular models of intercalation complexes have been performed and are presented in chapter 5. This particular method of calculation is preferable to the more general method in the cases discussed in this thesis since comparison between the observed and calculated data is only

made for the peak intensity on the equator and the first two layer lines. When a preliminary comparison of this type is being made, the method presented here, which is less tedious in terms of data preparation and faster in computation, is preferable.

As a test of the theory presented in this chapter, the transform of a series of points of unit scattering power on a cylinder of radius r were calculated by the method given in this section. The points were such as to be related to their nearest neighbours by either a translation of 3.4Å and a rotation of 36^{o}; or of 6.8Å and 24^{o} respectively. Since the points are of uniform, unit scattering power, the function calculated is similar to I_{∞} of equation 3.30. The predicted properties of I_{∞} that it should have intensity peaks at points representative of the average helix; and that it should be periodic over a distance corresponding to 3.4Å are observed.

3.6 Computer Programmes

(a) Helical Transform Programmes

Programmes to calculate the cylindrically averaged molecular transform of equation 3.7, and the structure factors from equation 3.4 have been written by Professor W. Fuller. These programmes have been transcribed by the author, with some modifications, to run on the Elliott 4130 machine at the Computer Centre, University of Keele, and later on the more powerful CDC 7600 machine at the University of Manchester Regional Computer Centre. Since these programmes have been described previously, they will not be discussed in detail here.

The molecular transform programme has been extended to calculate, also the intensity transform of equation 3.6. In addition, a number of modifications have been made to facilitate

the calculations of the transform of intercalation complexes by the method described in section 3.5. In particular, the facility to calculate the distribution function and to multiply this by the transform of an atomic group has been incorporated into the programme. The transform of the atomic group is formed in the normal way, by calculating the contribution from each atom in the group on selected layer lines and Bessel function orders: both being determined by the selection rule of equation 3.9. Contributions from all atoms are added in amplitude and phase to give the total transform for the group. At this stage, the transform of the disposition function is calculated for the same combinations of layer line number and Bessel function order and multiplied by the transform of the atomic group. The procedure is repeated until all atomic groups have been included: the contribution from each group being summed in amplitude and phase to give the total transform.

If a dyad axis exists in the structure (i.e. it is possible to choose a coordinate reference frame such that for every atom with coordinates R, ϕ, z, there is another with coordinate R, $-\phi$, $-z$) then for the molecular transform we have,

$$G(\xi,n,\ell/c) = \sum_{j=1}^{N/2} f_j J_n(2\pi\xi r_j) \left[\exp i\left(\frac{2\pi\ell z}{c} - n\phi_j\right) + \exp i\left(\frac{-2\pi\ell z_j}{c} + n\phi_j\right)\right]$$

$$= 2\sum_{j=1}^{N/2} f_j J_n(2\pi\xi r_j) \cos\left(\frac{2\pi\ell z_j}{c} - n\phi_j\right) \qquad 3.47$$

Hence, the transform of such a structure is real and can be obtained by including only one of the two sets of dyad related atoms.
As can be seen from the above equation it is necessary to double the value of the transform so obtained if the result is required on

an absolute scale. The programme written by Professor W. Fuller has the facility to exploit dyad axis symmetry to reduce the computation by calculating only the real part of the transform for structures possessing a dyad axis. In models of intercalation complexes it is often the case that some groups of atoms have dyad related neighbours, while others do not. The programme has been modified, therefore, to allow calculations of the imaginary part of the transform for some groups in the structure but not for others, thus permitting the use of the dyad simplification in cases where the whole structure does not possess a dyad axis.

Other less important facilities were included in the programme. The ϕ and z coordinates of selected atoms can be modified if desired. This takes the form of values, $\delta\phi$ and δz, which are read by the programme and added to the ϕ and z data for each atom to form the new coordinates. This is useful in special cases where it is necessary to modify atomic coordinates so that the dyad axis of the repeat unit lies along the line (R, 0, 0), which it must do if the dyad symmetry is to be utilised to reduce the computation in the manner discussed above. It is also possible to modify the weighting factors for selected atoms by reading in a value by which these factors are multiplied. This is important when changing the P/D ratio in the calculation of drug/nucleic acid complexes.

(b) <u>Diffuse Scattering Programme</u>

A programme was written to calculate the diffuse scattering component of a structure possessing substitution disorder. In order to perform the calculation it is necessary, in principle to calculate the transform of the average unit cell using equation 3.21 for

a chosen range of Bessel function orders at specified points in reciprocal space. The transforms for each of the possible cannonical forms must then be calculated at the same points in reciprocal space and for the same Bessel function orders. Δ of equation 3.17 for the jth cannonical form is given by:-

$$\Delta_{n,j}(\xi,n,\zeta) = G_{n,j}(\xi,n,\zeta) - G_{n\ av}(\xi,n,\zeta) \qquad 3.48$$

The subtraction in 3.48 is performed in amplitude and phase to give Δ_j. Intensity data are then calculated from:-

$$I_j(\xi,\zeta) = \sum_{N\ min}^{N\ max} W_j \Delta_{n,j}(\xi,\ell,n)\ \Delta^*_{n,j}(\xi,\ell,n) \qquad 3.49$$

and the total intensity by a summation over all the M cannonical forms.

$$I_{tot}(\xi,\zeta) = \sum_{j=1}^{M} I_j(\xi,\zeta) \qquad 3.50$$

In the particular case of intercalation complexes, the R coordinate remains unchanged for equivalent atoms in different unit cells. It is possible, therefore, to reduce the computation in an analogous way to that used for the non-diffuse component, by multiplying the transform of an atomic group by the transform of its disposition function. The transform of each atomic group having a different disposition function is calculated and stored in the machine along with the transform of the average unit cell obtained by the procedure used in the molecular transform programme. The transform for each cannonical unit cell is obtained in an analogous manner by reading in the appropriate disposition function for each atomic group in the particular unit cell under consideration: use being made of the stored transform values for each molecular unit.

Values for each Δ_j are obtained from 3.49 and the diffuse intensity calculated as described.

This method of calculation uses a large amount of computer store because of the requirement for the separate storage of the transform for each atomic group. Occasionally problems were encountered because of the large size of the programme and it was necessary to restrict the number of points at which the calculation was performed and/or the number of Bessel function components included. The problem posed by the size of the programme when this method of approach is used was considered to be outweighed by the reduction in computation achieved and by the very large reduction in the work involved in data preparation. The programme has been written so as to calculate the value of the additional term of equation 3.46 where necessary. The nature of this term, $I'(\xi,n,\zeta)$, in cylindrical polar coordinates is:-

$$I'(\xi,n,\zeta) = 2Re\left[\sum_{j=1}^{m} W_j\Delta_j(\xi,n,\zeta)\sum_{k=1}^{\infty} \exp i(2\pi\zeta z_j - n\phi_j)\right] \quad 3.51$$

The infinite summation of exponential function in equation 3.50 will give a delta function at points in reciprocal space and Bessel function orders specified by the selection rule of 3.9. Hence, the value of I' can be calculated from:-

$$I'_{n,j}(\xi,n,\ell) = W_j\ Re\left[G_{n,j}(\xi,n,\ell) - G_{n_{av}}(\xi,n,\ell)\right] \quad 3.52$$

and

$$I'(\xi,n,\ell) = \sum_{j=1}^{M}\ \sum_{NMIN}^{NMAX} I_{n,j}(\xi,n,\ell) \quad 3.53$$

where the transforms for $G_{n,j}$ and G_{av} have been calculated only for the ℓ values and Bessel function orders specified by the selection rule. The helical parameters associated with each of the M cannonical

forms are different so that the values of the reciprocal space parameters for which equation 3.46 has been calculated will vary for each of the M cannonical forms.

To perform the calculation, the programme first derives the values for each $I_{n,j}$ term using stored values of $G_{n,j}$ and $G_{n_{av}}$, and then performs the summation of 3.52 to obtain the final value.

(c) Programme to Calculate the Transform of Structures Possessing Disorder of the Second Kind

A programme has been written to perform the calculation of the transform for helical structures possessing disorder of the second kind. The programme is very similar to the helical transform programme for regular structures and routines from this programme were used. An important difference between the two programmes, however, is that the programme for regular helices calculate the transform only at points in the ζ direction specified by ℓ/c, where c is the helix pitch, and for Bessel function orders determined from 3.4; while the transform for disordered structures must, in theory, be calculated at all points along ζ and for all Bessel function orders. In practice a range of Bessel function orders and ζ values for which the transform is to be calculated are specified.

A routine is included in the programme to calculate the value of $I_{\infty}(\zeta,n)$, and this is multiplied by the intensity transform of the repeat unit calculated from equation 3.22 for each point in reciprocal space and for each Bessel function order. Bessel function component intensities are then added to obtain the final value of the intensity along the ξ direction at each ζ value. The values are scaled to give intensities per unit length in the manner

described.

The programme will be used to calculate the transforms of disordered helices presented in chapter 6.

3.7 Programme Testing

The helical transform programmes were tested using sets of test data from which results were available from other similar programmes. Other programmes were tested using simple data sets from which the transform values could be checked by hand calculation or by inspection.

APPENDIX 3:1

Equation 3.22 requires some proof. It is analogous to equation 3.7 except that since the structures involved do not possess helical symmetry, in equation 3.22 we have no selection rule for n and the transform in the ζ direction is not confined to a set of planes. Since equation 3.3, from which 3.7 was derived, is itself derived (Cochran et. al., 1952) by utilizing the symmetry properties of a helix, it is not obvious that equation 3.22 is valid.

The transform of an infinitely long, non radially symmetrical, hollow cylinder is given by:-

$$F(\xi,\Psi,\zeta) = \int_{-\infty}^{\infty}\int_{0}^{2\pi}\left[\exp 2\pi i\cos(\psi-\Psi)\right] \times \left[\rho(\psi,z)\exp 2\pi i(\zeta z)\right] \quad 3.54$$

Integration with respect to ψ can be performed using the normal formula for integration by parts:-

$$\int u.v.\,dx = u\int v dx - \int \frac{du}{dx}\left[\int v dx\right]dx \quad 3.55$$

where, $u = \rho(\psi,z)\ \exp 2\pi i(\zeta z)$

$v = \exp 2\pi i\ \cos(\psi-\Psi)$

$x = \psi$

Using the identity for Bessel functions of the first kind:-

$$\int_{0}^{2\pi}\exp\left[iu\cos\phi + i\,n\phi\right]d\phi = 2\pi i^{n}J_{n}(u) \quad 3.56$$

with $2\pi\xi r = u$ and $-(\pi+\phi) = (\psi-\Psi)$

Vainstein (1966) has shown that the integral $\int v dx$ of equation 3.55 becomes

$$\sum_{n=-\infty}^{\infty} J_n(2\pi\xi r) \exp i\, n\, (\Psi + \frac{\pi}{2}) \qquad 3.57$$

This is not a function of ψ and hence, the second term of 3.55 vanishes so that the final form of the equation is:-

$$F(\xi,\Psi,\zeta) = \sum_{n=-\infty}^{\infty} J_n(2\pi\xi r) \exp i\, n\, (\Psi + \frac{\pi}{2}) \quad x$$

$$\int_{-\infty}^{\infty} \rho(\psi,z) \exp 2\pi i(\zeta z)dz \qquad 3.58$$

In the general case the integration of equation 3.58 is difficult to perform. However, if $\rho(\psi,z)$ consists of an infinite series of points we can replace the integral by a summation and accommodate the dependence of ρ upon ψ by the factor exp $i(-n\phi)$.

$$F(\xi,\Psi,z) = \sum_{j=1}^{N} \sum_{n=-\infty}^{\infty} P_j\, J_n\, (2\pi\xi r) \exp i\, n\, (\Psi - \phi_j + \frac{\pi}{2})$$

$$\exp 2\pi i(\zeta,z) \qquad 3.59$$

Equation 3.22 can be derived from this by the same method as was used to pass from equation 3.3 to 3.7.

CHAPTER IV

MOLECULAR MODEL BUILDING

4.1 Introduction

The quantity and quality of diffraction intensity data from fibres are limited in comparison with those obtained from good single crystals. Even in the most ordered case in fibres (e.g. A-DNA) in which the molecules are regularly packed in small domains (crystallites), distinct Bragg reflections are only obtained in a region of reciprocal space corresponding to approximately 3Å or less since there is some misalignment of the individual crystallites which results in the spots being drawn out into arcs which merge at higher diffraction angles. Many fibres possess very little order so that Bragg reflections are almost entirely absent from their diffraction patterns which consist, therefore, largely of diffuse diffraction streaks whose intensity is difficult to measure accurately. Moreover, in fibres of this latter type, the azimuthal orientation of the individual molecules about the fibre axis is random and the diffraction intensity obtained from such specimens do not yield three dimensional information in their diffraction patterns: the maximum information available being contained in a plane representing a central section through the cylindrically averaged intensity transform.

Molecular model building allows best use to be made of the limited diffraction data available by incorporating known stereochemical features of the molecular structure into the analysis. The validity of the method depends on the correctness of the assumption that the values of the stereochemical parameters (bond lengths and bond angles) may be obtained from values derived from

single crystal X-ray diffraction determinations of the structure for related small molecules.

Even knowledge of stereochemical parameters is normally insufficient to build a starting model which is good enough to refine, and some hint about the molecular conformations must be supplied For helical molecules it is normally adequate to obtain values of the translation and rotation parameters relating repeat units in the helix. The diffraction theory for helical molecules of Cochran et al. (1952) shows that such information may be obtained from a relatively superficial analysis of the diffraction data. It is this fact which has enabled the combination of the techniques of molecular model building and of X-ray fibre diffraction to be so powerful for elucidating the structure of helical polymers.

Once a reasonable starting model has been chosen, it is possible to refine it by minimising the discrepancy between the experimentally observed diffraction intensity and that calculated for the model. In the case of highly crystalline fibres, a computer method for performing this refinement, using a least squares function minimisation routine has been developed (Arnott and Wonacott, 1966). Many of the fibre specimens studied in this laboratory give diffuse intensity data which is not amendable to processing in a refinement routine such as the one mentioned above. In this case it is better to perform the refinement so as to improve the stereochemistry (i.e. to minimise the energy of the structure in terms of its non-bonded interactions etc.), and to calculate the diffraction intensity predicted by the refined model. A computer programme for minimising the conformational energy of molecular structures will be described in this chapter.

4.2 Model Building Techniques

Molecular model building can be carried out using some physical structure to represent the molecule, or by a computerised routine in which the model consists of a set of atomic coordinates derived by calculation from the stereochemical data. Computerised model building and refinement procedures have been used to obtain coordinates for all models presented in this thesis and so will be discussed in some detail here. Hand model building has been used but normally only to provide a starting model for the computer refinement.

4.2.1. Hand Model Building

If physical models are to be good representations of molecular structures, it is necessary to use some type of building unit in which bond lengths and bond angles are accurately represented. Space filling models (e.g. Corey, Pauling, Koulton (C.P.K.) models) use units in which the van der Waal's radii of the atoms are also represented. Such models give a better visual representation of a proposed structure than do skeletal models in which only atomic centres are represented along with bond angles and bond lengths. The main disadvantages of C.P.K. models are that it is not easy to measure atomic coordinates accurately, although a method for deriving them has been described (Haen et al. 1976), and that the bond angles and bond lengths are fixed so that slight variations in these parameters cannot be introduced. It is also difficult to check a C.P.K. model to ensure that it corresponds to the intended structure. All hand model building described in this thesis, unless otherwise stated, is performed using skeletal models.

The advantage of hand model building is that it produces a physical model of the structure from which it is often possible to determine by inspection, changes which could be made to the model

to improve its stereochemistry. This is particularly relevant in the early stages of analysis when it may be necessary to make gross changes in the model.

A disadvantage of the method is that it is not easy to assess accurately the conformational energy of a given model in terms of the van der Waals interactions between non-bonded atoms and hydrogen bond energies. It is also difficult to predict accurately changes in the model which would improve its stereochemistry or the agreement between the diffraction intensity calculated from the model and that observed in experiment, particularly in the later stages of analysis when the changes to be made are relatively small.

4.2.2. Computerised Model Building

The primary requirement of any computerised model building routine is a method of calculating the atomic coordinate given the values of bond angles, bond lengths and a set of dihedral (torsion) angles defining the molecular conformation. A method to accomplish this has been suggested by Eyring (1932) for a polymer chain and similar procedures have been used in a number of computer model building routines (e.g. Ramachandran et al. 1966; Arnott and Wonacott, 1966). A procedure for deriving atomic coordinate employing the Eyring transformation matrix will be described in relation to the hypothetical molecule in figure 4.1.

Coordinates are generated for atom 1 in a cartesian coordinate system (x,y,z) having origin on atom 2; the x axis along the bond between atom 2 and atom 1; the y axis perpendicular to the x axis and in the planes of atoms 1, 2 and 3; and the z axis positioned relative to them so as to form a right handed orthogonal set. The coordinates of atom 1 in this coordinate system are

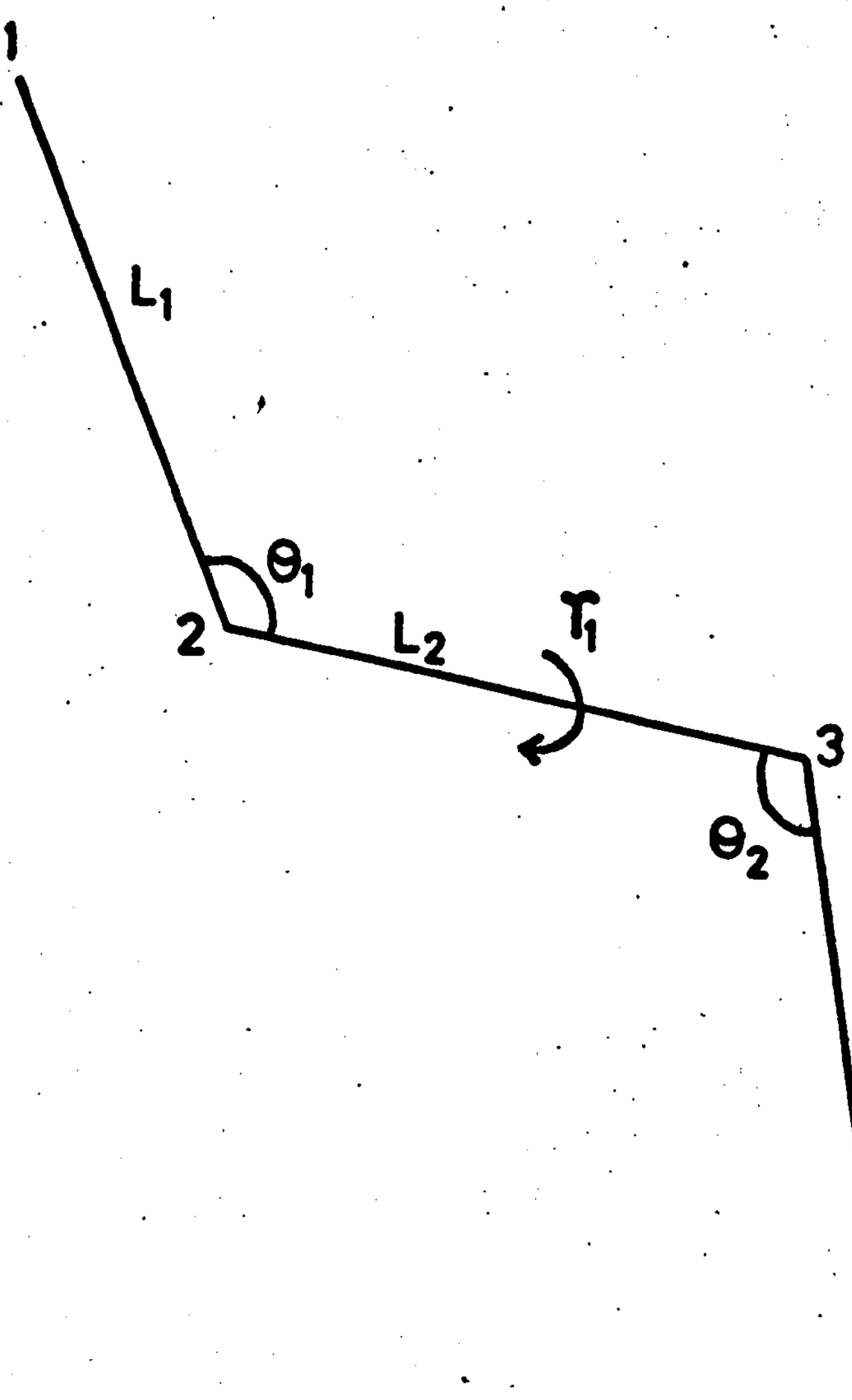


Fig. 4·1 AN IDEALISED CHAIN MOLECULE

evidently $(L_1, 0, 0)$ where L_1 is as defined in figure 4.1.

The coordinate system may now be shifted to have origin on atom 3 and to be oriented relative to atoms 2,3 and 4 in an analogous way to that described above. Atom 2 has coordinates $(L_2, 0, 0)$ in this system; while the coordinates of atom 1 in the system (x,y,z) can be transformed into equivalent coordinates in the set (x',y',z') by the following matrix transformation.

$$\underline{X}'_1 = \underline{A}_{2,3} \cdot \underline{X}_1 + \underline{V}_{2,3} \qquad 4.1$$

where:-

$$\underline{X}'_1 \equiv \begin{pmatrix} x'_1 \\ y'_1 \\ z'_1 \end{pmatrix}, \quad \underline{X}_1 \equiv \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \quad V_{2,3} \equiv \begin{pmatrix} L_2 \\ 0 \\ 0 \end{pmatrix}$$

$$\underline{A}_{2,3} \equiv \begin{pmatrix} -\cos\theta_1 & -\sin\theta_1 & 0 \\ \sin\theta_1 \sin\tau_1 & -\cos\theta_1 \cos\tau_1 & \sin\tau_1 \\ -\sin\theta_1 \sin\tau_1 & \cos\theta_1 \sin\tau_1 & \cos\tau_1 \end{pmatrix}$$

The above process is continued until the coordinates of all the atoms have been derived in a coordinate frame analogous to (x,y,z) but having origin on the final chain atom. Hence, the final coordinates of atoms 1 and 2 in the frame situated on atom 4 in figure 4.1 are given by:-

$$X''_j = \underline{A}_{3,4} \underline{X}'_j + \underline{V}_{3,4} \qquad 4.2$$

$$j = 1,2$$

In a linked atom modelling system, such as the one described here, structure refinement is carried out in terms of changes in the dihedral

angles and in certain refinement procedures the partial derivatives of the coordinates of every atom with respect to each dihedral angle are required. For atom 1 in figure 4.1, the partial derivatives with respect to τ_1 are given by:-

$$\frac{\partial \underline{X'}_1}{\partial \tau_1} = \frac{\partial \underline{A}_{2,1}}{\partial \tau_1} \cdot \underline{X}_1 \qquad 4.3$$

The derivatives of the coordinates of all other atoms with respect to τ_1 are clearly zero. As the coordinate system is shifted along the polymer chain, it is necessary to transform the derivatives to be compatible with the new coordinate frame by multiplying by the transformation matrix $\underline{A}$ so that in moving from coordinate frame (x', y', z') to frame (x", y", z") located on atom 4 in figure 4.1, we have

$$\frac{\partial \underline{X"}_1}{\partial \tau_1} = \underline{A}_{3,4} \frac{\partial \underline{X'}_1}{\partial \tau_1} \qquad 4.4$$

The present method cannot derive the coordinate of atoms which do not form the backbone of a single molecular chain. Coordinate of atoms pendant to an atom of the chain may be calculated when the coordinate system has its origin situated on the chain atom to which they are pendant, by the following transformations.

$$X_j = x_j.L_j, \quad Y_j = y_j.L_j, \quad Z_j = z_j.L_j$$

where X_j, Y_j, Z_j are the cartesian coordinates of the jth atom; L_j is the bond length between the jth pendant atom and the chain atom; and x_j, y_j, z_j are direction cosines defined in figure 4.2.

Branched structures can also be accommodated into the modelling system. A hypothetical structure is illustrated in figure 4.3, which has a branch pendant to the second atom (atom 5 in figure 4.3) of its main chain. The branch chain is built first: its atomic coordinates being derived in a coordinate frame with X axis along the bond between atom 4 and atom 5, and origin situated on atom 5. It is necessary to transform the coordinates so that they are in the coordinate frame adopted by the modelling system while building the main chain and when the coordinate system has origin situated on atom number 5. This can be done by the following matrix transformation provided that the direction cosines x_4, y_4, z_4, analogous to those of figure 4.2 for pendant atoms, are known for atom 4:-

$$\underline{X'}_j = \underline{T} . \underline{X}_j$$

$$j = 2 \rightarrow 4$$

$$\text{where } \underline{T} \equiv \begin{pmatrix} \cos a & \sin a & 0 \\ -\sin a . \cos b & \cos a . \cos b & \sin b \\ \sin a . \sin b & -\cos a . \sin b & \cos b \end{pmatrix}$$

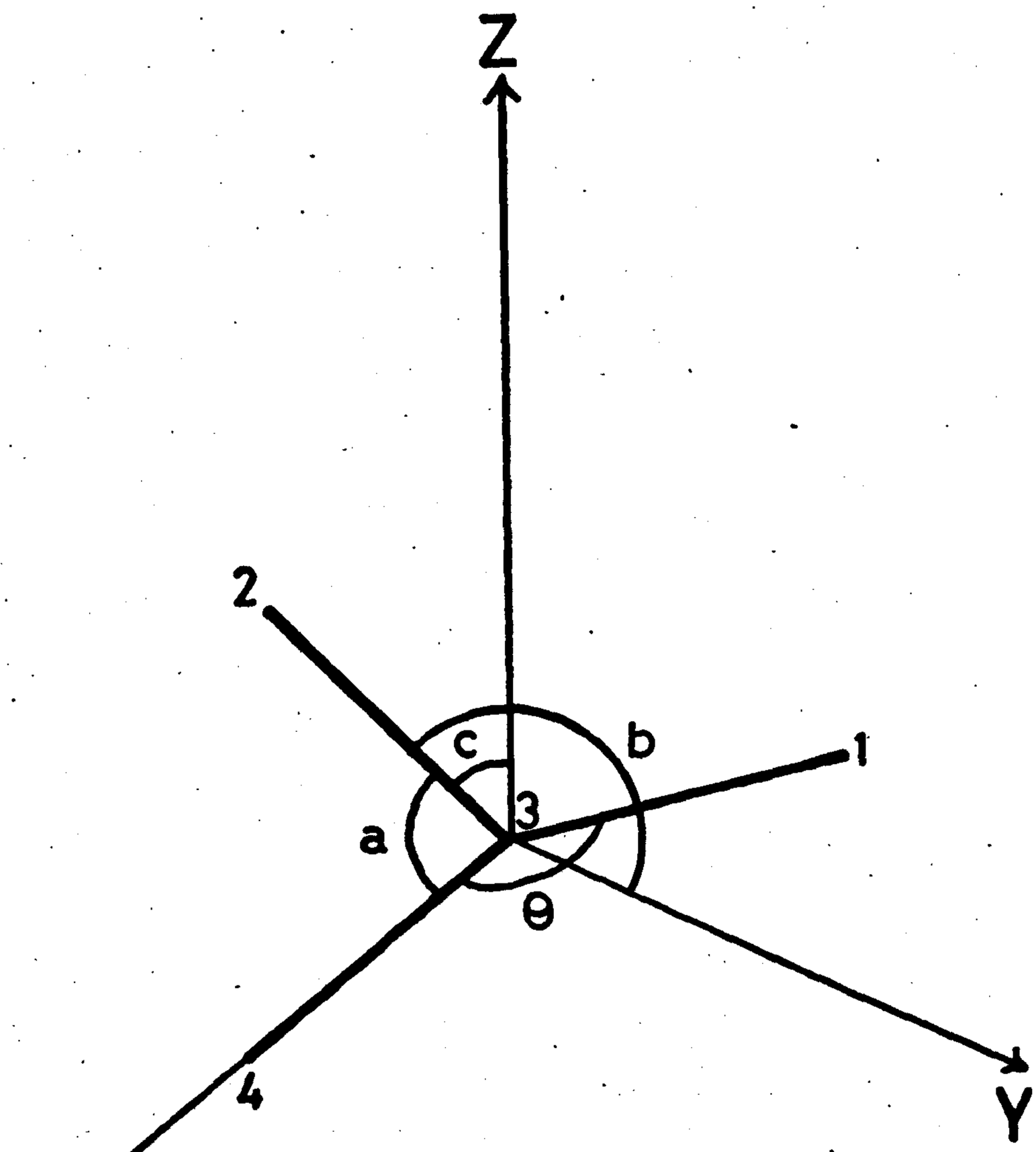
$$a = \cos^{-1} (x_4)$$

$$b = \tan^{-1} \left(- \frac{z_4}{y_4} \right) \text{ if } y_4 \text{ is -ve.}$$

$$\text{or } b = \tan^{-1} \left(- \frac{z_4}{y_4} \right) + 180^{o} \text{ if } y_4 \text{ is +ve.}$$

As described above, the coordinate of pendant or branch atoms are in a coordinate frame situated on atom 5. The coordinate of such atoms, and their derivatives, may be transformed in the same way as the chain atoms, as the coordinate system moves along the main chain.

PENDANT ATOM DIRECTION COSINES



$x_2 = \cos(a)$

$y_2 = \cos(b)$

$z_2 = \cos(c)$

Fig. 4·2

In addition to deriving and refining the atomic coordinates for the structure as described above, it may be desirable to be able to position one group of atoms relative to another by translations along and rotations about the coordinate axes. Atomic coordinates (x,y,z) can be modifed in this way to the new coordinate set (x',y',z') by the following transformation:-

$$\underline{X}' = \underline{R}.\underline{X} + \underline{T} \qquad 4.5$$

where,

$$\underline{X}' \equiv \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}, \quad \underline{X} \equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \underline{T} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix} \begin{pmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{pmatrix} \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\theta_{x,y,z}$ = Eulerian angles describing rotations about the X, Y and Z axes respectively.

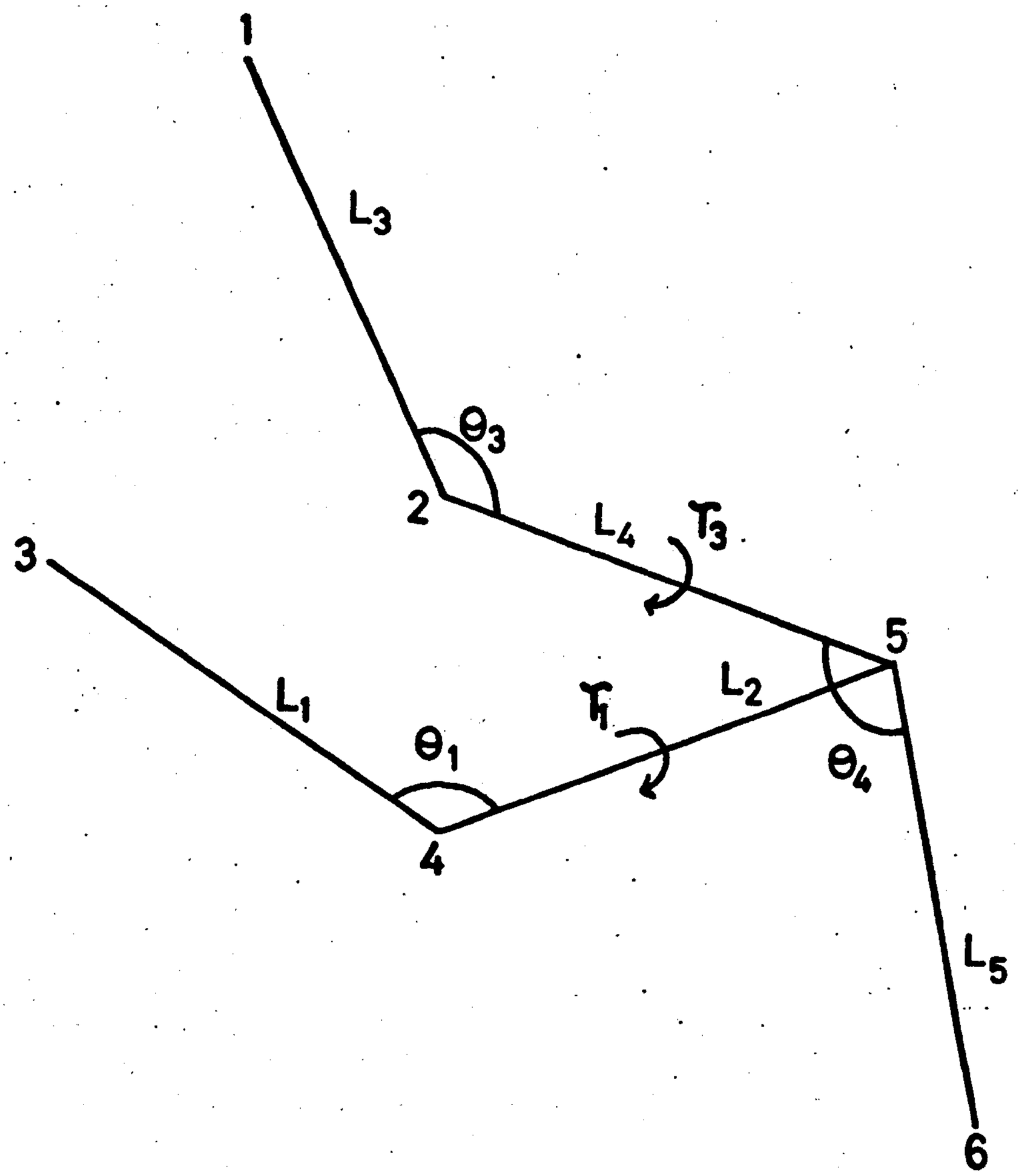
$t_{x,y,z}$ = Translations along the x,y, and z axes respectively.

The derivatives with respect to the Eulerian angles are given by:-

$$\frac{\partial \underline{X}'}{\partial \theta} = \frac{\partial R}{\partial \theta} \cdot \underline{X} \qquad 4.6$$

Also,

$$\frac{\partial x}{\partial t_x} = \frac{\partial y}{\partial t_y} = \frac{\partial z}{\partial t_z} = 1$$



AN IDEALISED BRANCHED STRUCTURE

Fig. 4·3

and

$$\frac{\partial y}{\partial t_x} = \frac{\partial z}{\partial t_x} = \frac{\partial x}{\partial t_y} = \frac{\partial z}{\partial t_y} = \frac{\partial x}{\partial t_z} = \frac{\partial y}{\partial t_z} = 0$$

It is also necessary to modify the derivatives of the torsion angles to be compatible with the new coordinate frame. This is accomplished by multiplying the derivatives by $\underline{R}$. Hence,

$$\frac{\partial \underline{X}'}{\partial \tau} = \underline{R} \frac{\partial \underline{X}}{\partial \tau} \qquad 4.7$$

For computational purposes, the Eulerian angles and translational parameters are treated in an analogous manner to the torsion angles, and any facilities in the programme applying to the latter may also be applied to the positional parameters.

4.3 Computer Programmes

The computer programme used in the refinement of molecular models in this thesis is essentially that written by Dr. W.J. Pigram and has been previously described (Pigram, 1968). The author has transcribed the programme to run on the Elliott 4130 machine at the University of Keele Computer Centre and subsequently to run on the more powerful C.D.C. 7600 machine at the University of Manchester Regional Computer Centre. Several modifications have been made to the programme and will be discussed in later sections of this chapter. A flow diagram for the current version of the programme is presented in figure 4.4.

The molecular conformation is refined primarily so as to minimise the energy of the van der Waals contacts between pairs of non-bonded atoms. In addition, the refinement routine may be constrained to attain or preserve certain desired stereochemical

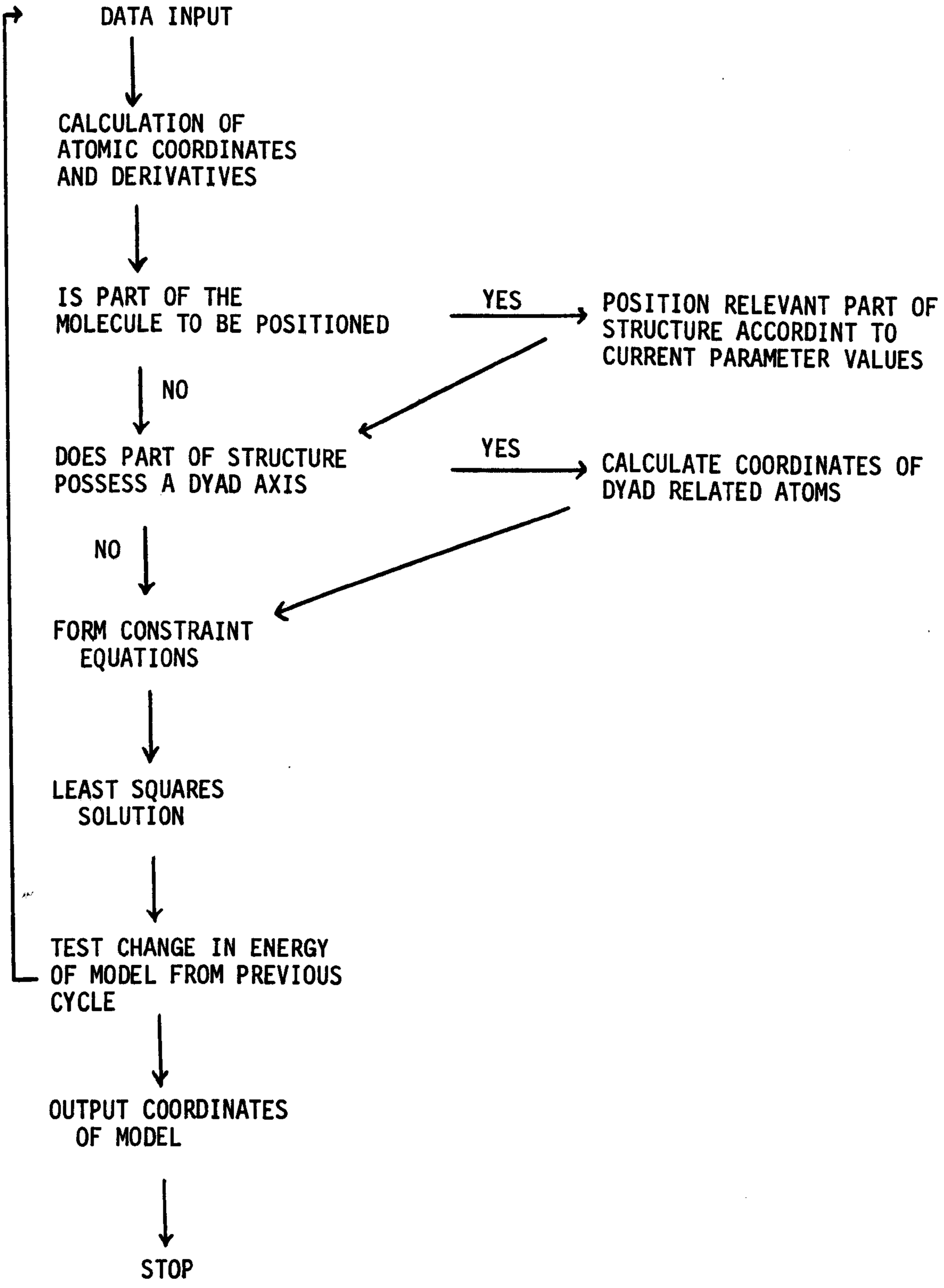


Fig. 4.4 Flow Diagram of The Model Building Programme

features of the molecular structure. Atomic coordinates for the starting model are derived by the linked atom procedure described in the previous section. Interatomic distances between non-bonded atom pairs are then checked and if any is found to be smaller than a given multiple of the sum of the van der Waals radii for the constituent atoms, then details of the interaction are stored in the machine. In fact a maximum number, N, of non-bonded contacts may be included. The value of N is specified at run time: a sort routine being used to select the N worst contacts in the structure. At a later stage the energy of each of the contacts is calculated by means of a Lennard-Jones function:-

$$E = \frac{B}{r^{12}} - \frac{A}{r^{6}} \qquad 4.8$$

where A and B are parameters for any given pair of elements; while r is the distance of separation. This type of energy function has been used in previous computerised refinement routines (Scott and Scheraga, 1966).

Since the constants A and B are related, it is possible to eliminate one of them. If r_o is the separation at which the energy is a minimum, then differentiating 4.8 and substituting for r gives:-

$$\frac{-12B}{r_o^{13}} + \frac{6A}{r_o^{7}} = 0 \qquad 4.9$$

from which we have

$$B = \frac{Ar_o^{6}}{2}$$

Substituting for B in 4.6 gives:

$$E = \frac{-A}{r^6}\left(1 - \frac{r_o^{6}}{2r^6}\right) \qquad 4.10$$

Values of the constant A for different atomic pairs as used by Scott and Scheraga (1966) are given in table 4.1. All values for interactions not involving hydrogen are very similar. Contacts involving a hydrogen atom have values of A reduced approximately to one third of that for contacts not involving hydrogen. The value of A for a H-H contact is reduced to approximately one ninth of the value for a contact not involving hydrogen. It was considered to be a sufficient approximation to take A to be 366 (to give E in K cal/mole) and to multiply this value by one third for each hydrogen atom involved in the interaction. The minimum energy when $r = r_o$ is given by:-

$$E_{min} = \frac{-A}{2r_o^6} \qquad 4.11$$

Hence, the value of the discrepancy between the desired and calculated energy (the constraint value) is given by:-

$$C = E - E_{min} \qquad 4.12$$

The refinement routine requires also the partial derivatives of the constraint with respect to each variable dihedral angle. These derivatives may be obtained by differentiation of 4.12 in terms of the equation for E ($\frac{\partial E_{min}}{\partial \tau} = 0$, for all dihedral angles) and using the partial derivatives of the atomic coordinates with respect to the dihedral angles calculated as described previously. The van der Waals constraint equations are set up automatically during any run. In addition a number of other constraints may be applied to the modelling system to retain desirable or essential features of the molecular structure.

TABLE 4.1

Values for the constant, A, of the Lennard-Jones potential as used by Scott and Scheraga (1966) to give energy values in K cal/mole

Atom Pair	A	r_o (Å)
C-C	370	3.2
C-N	366	3.1
C-O	367	3.0
C-H	128	2.8
N-N	363	3.0
N-O	365	2.9
N-H	125	2.7
O-O	367	2.8
O-H	124	2.6
H-H	46.7	2.4

Moreover, some control over the manner in which the dihedral angles are manipulated, and which interatomic distances are included in the list of van der Waals interactions, is possible. These facilities have been described previously (Pigram, 1968) and so will not be discussed in detail here, though mention will be made of them since they are used in the refinement of some of the models discussed later in this thesis.

It is possible to specify that any particular dihedral angles shall not be varied during the refinement. In this case, derivatives for the constraints with respect to these "fixed" dihedral angles are not calculated and no information relating to them is included in the refinement routine. The facility also exists to specify that a certain dihedral angle, τ_i, shall take the value of another specified dihedral angle, τ_j, at the end of each refinement cycle.

Each atom in the structure is associated with a list of numbers representing atoms with which it is not to be considered in the search for interatomic short contacts. This facility may be used for special reasons in certain cases, but in particular is necessary to ensure that: (a) contacts between covalently bonded atom pairs are not considered; and (b) that atom pairs which are covalently bonded to a common atom are also excluded, since the distance between them will not be a function of any of the dihedral angles. It is also possible to exclude a specified atom entirely from consideration in the search for short contacts. This is accomplished by inserting a negative index into the list of excluded atoms.

A number of geometrical constraints may be applied to the modelling system. It is possible to constrain two atoms to have a desired difference, δx, δy and δz in their x,y and z coordinates

respectively. This routine is often used to ensure that the chain joins up correctly to certain parts of the structure. In this case $\delta x = \delta y = \delta z = 0$, and one of the atoms of the pair is excluded from consideration in the search for short contacts by means of the facility described above. Any combination of the three constraints involved may be excluded so that the pair of atoms may be constrained to lie in the same plane or along the same line.

It may be desirable in certain cases to constrain two atoms to be separated by a given distance without any other specific relationship between their coordinates. A specific use for this facility, which exists in the programme, is to ensure the correct distance of separation for atom pairs involved in a hydrogen bonding interaction.

Since many of the structures built by the programme will be helical polymers, it is occasionally useful to constrain one atom to be the helix repeat of another. Hence, a routine has been included in the programme to constrain two specified atoms to have the same radial polar coordinate (R) and specified differences, $\delta\phi$ and δz, in their ϕ and z coordinates respectively.

It is also possible to place constraints on the value of any specified dihedral angle (τ_i). Its value is constrained to have a desired value, Φ, which may take one of three forms.

(i) It is possible to constrain τ_j to be equal to another specified dihedral angle, τ_i, in the model. In this case ϕ takes the value of τ_i.

The facility to ensure that one dihedral angle takes the value of another torsion angle in the structure already exists by use of the facility already outlined. However, the present routine is useful in that it allows two dihedral angles to be elastically bound to the same value.

The routine has also been written to allow τ_j to be constrained to be equal to $\tau_i + \partial\tau$; where $\partial\tau$ is some specified difference between the two torsion angles.

(ii) It may be desired to constrain τ_j to adopt one of three preferred values θ_1, θ_2, or θ_3. ϕ then takes the value of either θ_1, θ_2 or θ_3 according to which of these three angles is closest in value to τ_j. If there are only two (or one) preferred orientation, then two of the values (or all three) for ϕ_1, ϕ_2 or ϕ_3 can be made the same.

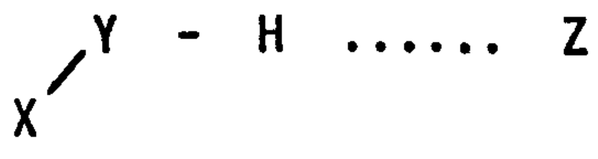
(iii) It is also possible to constrain τ_j to lie in a specified range whose limits are defined by θ_1 and θ_2.

If τ_j is outside these limits then ϕ takes the value of θ_1 or θ_2 which is closer to τ_j; otherwise ϕ is allocated the value of τ_j.

The distance constraint allows the correct distance to be maintained between the atoms involved in a hydrogen bonding interaction. A further two routines relevant to the maintenance of correct hydrogen bond geometry were incorporated into the programme.

(a) <u>H-Bond 1</u>

In the hydrogen bonding interaction:-



where Z is the acceptor atom, Y the donor atom and X an atom covalently bonded to Y, the angle between atoms Y, H and Z, as well as the distance H ..... Z, is important for the strength of the interaction. Generally, the interaction is strongest when the angle Y, H, Z is near to zero. Hence, it is desirable to be able to constrain the angle between three atoms to be a desired value,

and a routine to permit this has been written.

The constraint equation is:-

$$C = \cos^{-1}\left\{ \frac{(x_k-x_i)(x_k-x_j) + (y_k-y_i)(y_k-y_j) + (z_k-z_i)(z_k-z_i)}{d_{i,k} \cdot d_{j,k}} \right\} - \Phi \qquad 4.13$$

where $d_{i,k}$ and $d_{j,k}$ are the interatomic distances and Φ is the desired value of the angle.

The derivative of the constraint with respect to each variable parameter may be obtained from an equation of the form:-

$$\frac{\partial C}{\partial \tau} = - \frac{\partial \alpha}{\partial \tau} \frac{1}{\sqrt{1-\alpha^2}} \qquad 4.14$$

where α represents the term in parentheses in equation 4,13. The differential of α may be obtained by using the derivatives of the coordinates calculated as described in section 4.2.2.

Provided that the appropriate value is given for Φ, the constraint may be applied between atoms Y, H and Z or between X, Y and Z. The combination of this constraint with that relating the distance between atom pairs provide the best method of maintaining hydrogen bond geometry in many cases, since the constraint values are a direct measure of the geometrical parameters involved. However, the disadvantage of the method is that it is not possible to measure the energy of the interaction on the same scale as the van der Waal's interactions. Moreover, the potentials represented in the distance and H-bond 1 constraints are probably not good representations of the energy potential involved in hydrogen bond interactions, and it is desirable to choose a function specifically designed to parallel the hydrogen bond interaction.

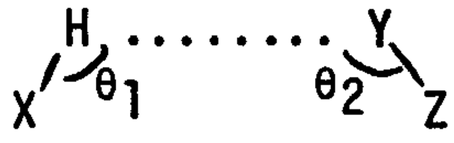
(b) H-Bond 2

De Santis et al. (1965) have used a Stockmayer (1941) type function to represent the hydrogen bonding interaction in a study of the conformations of protein and polypeptides. Scott and Scheraga (1966) in a similar study, have used a function due to Lipincott and Schroeder (1955). A function of the latter type is reported to be more accurate (Ramachandran and Sasisekharan, 1966) than the Stockmayer relation, though relevant parameter values are available for a smaller number of H-bonding systems.

One form of the Lipincott and Schroeder function is:-

$$E(d) = D\left\{1 - \exp\left[-n(r - r_o)^2/2r\right]\right\} - D^*\left\{1 - \exp\left[-n^*(d-r-r_o^*)^2/2(d-r)\right]\right\} + B\exp(-\mu d) - {}^{A}/_{d^m} \qquad 4.15$$

where for the hydrogen bonding interaction



D is the energy of the X-H bond; D* of the H .... Y bond; r the length of the X-H bond, and r_o its optimum value: d and r_o* represent the actual and optimum distance of separation between H and Y; while n and n* are parameters related to the ionisation potential.

Equation 4.15 does not give E as a function of the angles θ_1 and θ_2. Moulton and Kromhout (1956) have introduced this angular dependence by multiplying the first exponential term of 4.15 by $\cos^2\theta_1$ and the second by $\cos^2\theta_2$.

The final two terms of equation 4.15 represent the van der Waals interaction between H and Y in terms of a Buckingham energy function. Since a Lennard-Jones function has been used in the present programme to calculate the energy of such interactions, the Buckingham exponential term, B exp(-μd), representing the repulsive component of the van der Waals interaction has been replaced by the equivalent Lennard-Jones term. Moreover, since bond lengths are not varied during the refinement, the length of the X-H bond is considered as being equal to its optimum value. The value of m, the parameter determining the power dependence of the attractive term of the van der Waals interaction is taken to be 6. These latter two conditions have been imposed by Scott and Scheraga (1966). Hence, the final form of the equation is:-

$$E(d,\theta_1,\theta_2) = D\sin^2\theta_1 - D^*\left\{1-\exp\left[-n^*(d-r-r_o^*)^2/2(d-r)\right]\right\}\cos^2\theta_2 - \frac{A}{d^6}\left[1-\frac{r_o^{*6}}{2d^6}\right] \qquad 4.16$$

This function may be employed in two different ways. It may be used as part of a genuine constraint equation whose value provides data for the refinement routine; or, alternatively, the hydrogen bond geometry may be maintained using the distance and H-bond 1 constraints, with equation 4.16 being used to calculate the energy of the hydrogen bonding interaction at the end of each cycle of refinement. In the latter case, 4.16 is not being used as part of a constraint equation at all since its value is not used in the refinement, and no derivatives of the equation are required.

If the first mode of use is to be employed, then the constraint equations must be set up, which are:-

$$C_1 = E_d - E_{min}$$

$$C_2 = E_{\theta_1} - E_{min}$$

$$C_3 = E_{\theta_2} - E_{min}$$

where E_{min} is the minimum value of $E(d,\theta_1,\theta_2)$ when $d = r_o^*$, $\theta_1 = 180^o$ and $\theta_2 = 90^o$. E_d, E_{θ_1} and E_{θ_2} give the values of E obtained when d, θ_1 and θ_2 respectively take their actual values observed in the structure, while the other two parameters take the value for E_{min}. The constraint equations are, in full, therefore:-

$$C_1 = \frac{A}{d^6}\left[\frac{r_o^{*6}}{2d^6} - 1\right] + \frac{A}{2r_o^{*6}} \qquad 4.17$$

$$C_2 = D \sin^2\theta_1 \qquad 4.18$$

$$C_3 = -D^*\left\{2 - \exp\left[-n^*r^2/2(r_o^* - r)\right]\right\}\cos^2\theta_2 \qquad 4.19$$

The derivatives are given by:-

$$\frac{\partial C_1}{\partial \tau} = \frac{\partial E_d}{\partial \tau} = \frac{\partial E_d}{\partial d}\frac{\partial d}{\partial \tau} = -6A\left[\frac{r_o^{*6}}{d^{12}} - \frac{1}{d^6}\right]\frac{1}{d}\frac{\partial d}{\partial \tau} \qquad 4.20$$

Since we have:-

$$d = \sqrt{(X_H - X_Y)^2 + (Y_H - Y_Y)^2 + (Z_H - Z_Y)^2} \qquad 4.21$$

then we can obtain $\frac{\partial d}{\partial \tau}$ by differentiating 4.21 and making use of the coordinate derivative with respect to the variable parameters. It should be noted that C_1 and its derivative are merely the equivalent equations from the van der Waals constraint routine.

For the derivative of C_2 we have:-

$$\frac{\partial C_2}{\partial \tau} = 2D \sin\theta_1 \cos\theta_1 \qquad 4.22$$

while

$$\frac{\partial C_3}{\partial \tau} = -2D^* \exp\left[-n^* r^2/2(r_o^* - r)\right] \cos\theta_2 \sin\theta_2 \qquad 4.23$$

The facility to use equation 4.16 in the above way, as the basis of a genuine set of constraint equations has been written into the programme and is currently being tested on sets of test data. It has not been used in the building of any models described in this thesis. In addition, the facility to use 4.16 directly as a measure of the energy of the hydrogen bonding interaction without including it in the constraint equations has been incorporated. The atom pair H,Y should be excluded from consideration in the original search for non-bonded short contacts since the energy of the van der Waals interaction between the atoms is accounted for in equation 4.16.

If the energy of a hydrogen bonding interaction is to be on the same scale as those of the non-bonded contacts, then the weighting factors for both of these constraints should be equal.

4.3.1 The Refinement Procedure

The programme employs a least squares refinement procedure which minimises a function of the form:-

$$\Phi = \sum_{j=1}^{M} W_j({}_oP_j - P_j)^2 = \sum_{j=1}^{M} W_j \Delta P_j^2 \qquad 4.24$$

where $_{o}P_j$ = the value of some quantity pertaining to the actual structure

P_j = the desired value of the parameter

W_j = a weighting factor

The function of 4.24 is minimised by making suitable changes (shifts) in the values of the dihedral angles (and/or the positional parameters where appropriate): the sign and magnitude of the shifts being determined by an interactive method involving a Taylor series expansion in M space about the current approximation to the required minimum. A similar routine has previously employed in a programme for the refinement of polymer moleculs (Arnott and Wonacott, 1966).

The expansion of Φ in terms of a particular variable parameter, τ_m, is given by:-

$$\Phi = \sum_{j=1}^{M} W_j \left\{ \Delta P_j - \sum_{i=1}^{N} \Delta\tau_m \frac{\partial \Delta P_i}{\partial \tau_m} \right\}^2 \qquad 4.25$$

where the series has been terminated after the term involving first derivatives.

A minimum value of Φ is given when the N values $\frac{\partial \Phi}{\partial \Delta \tau_m} = 0$, where N is the number of variable parameters. Hence, the condition for a minimum is obtained by differentiating the N equations of the form 4.25 and equating to zero to give N further equations:-

$$-2\sum_{j=1}^{M} W_j \left\{ \Delta P_j - \sum_{i=1}^{N} \Delta\tau_i \frac{\partial \Delta P_i}{\partial \tau_j} \right\} \frac{\partial \Delta P_j}{\partial \tau_q} = 0 \qquad 4.26$$

These can be simply transcribed into matrix form:-

$$D \,.\, D^T \,.\, S - C \,.\, D = 0$$

giving

$$S = (D.D^T)^{-1} D.C$$

where,

$$D = \begin{pmatrix} \sqrt{W_1}\,\frac{\partial \Delta P_1}{\partial \tau_1} & \cdots\cdots\cdots & \sqrt{W_1}\,\frac{\partial \Delta P_M}{\partial \tau_1} \\ \\ \sqrt{W_1}\,\frac{\partial \Delta P_1}{\partial \tau_N} & \cdots\cdots\cdots & \sqrt{W_M}\,\frac{\partial \Delta P_M}{\partial \tau_N} \end{pmatrix}$$

$$C = \begin{pmatrix} \sqrt{W_1}\Delta P_1 \\ \\ \sqrt{W_M}\Delta P_M \end{pmatrix}, \qquad S = \begin{pmatrix} \Delta \tau_1 \\ \\ \Delta \tau_N \end{pmatrix}$$

The procedure outlined above does not necessarily find a set of values for the variable parameters which make all values of ΔP zero. Instead the routine searches for a parameter set giving values of the constraints (ΔP) which define a minimum in Φ: account being taken of the relative importance of each constraint in terms of the weighting factors. Hence, each of the M quantities, P_j of equation 4.24, can be considered as being elastically bound to its desired value. Certain constraints specify geometrical features of the molecular structure which must be satisfied exactly if the resultant model is to be at all valid. It is possible to ensure this by ascribing very large values to the appropriate weighting factors.

A method of ensuring that the required condition is always satisfied, has been used in the routine of Arnott and Wonacott (1966).

Equation 4.24is expanded to become:-

$$\Phi = \sum_{i=1}^{M} W_i \Delta P_i^2 + \sum_{j=1}^{H} \lambda_j \Delta P_j \qquad 4.27$$

where λ_j = a Lagrange multiplier of undetermined value.

The constraints have been split up into two groups, the first of which is treated as before, while the second is processed by the Lagrange multiplier technique.

Expansion by Taylor's series as before shows that the minimum in Φ is given by N equations:-

$$\frac{\partial \Phi}{\partial \tau_m} = \Psi_q + \sum_{j=1}^{H} \lambda_j \frac{\partial G_j}{\partial \tau_q} = 0 \qquad 4.28$$

and by H equations:-

$$\frac{\partial \Phi}{\partial \lambda} = G_r + \sum_{j=1}^{N} \Delta \tau_j \frac{\partial G_r}{\partial \tau_j} = 0 \qquad 4.29$$

where Ψ_q represents the expression on the right hand side of equation 4.24.

The matrix form of these equations is:-

$$D \,.\, D^T \,.\, S + \tfrac{1}{2} N \,.\, L = D \,.\, C \qquad 4.30$$

$$\tfrac{1}{2} N^T \,.\, S + 0 = -\tfrac{1}{2} G \qquad 4.31$$

where D, S and C are as before and,

$$N = \begin{pmatrix} \frac{\partial G_1}{\partial \tau_1} & \cdots\cdots\cdots\cdots & \frac{\partial G_H}{\partial \tau_1} \\ \\ \frac{\partial G_1}{\partial \tau_n} & \cdots\cdots\cdots\cdots & \frac{\partial G_H}{\partial \tau_n} \end{pmatrix}$$

$$L = \begin{pmatrix} \lambda_1 \\ \\ \\ \lambda_H \end{pmatrix}, \qquad G = \begin{pmatrix} -G_1 \\ \\ \\ G_H \end{pmatrix}$$

Equations 4.30 and 4.31 form two simultaneous equations which may themselves, be solved by a matrix method to yield values for the shifts and for the Lagrange multipliers:-

$$\left(\begin{array}{c|c} D.D^T & \frac{1}{2}N \\ \hline \frac{1}{2}N^T & 0 \end{array}\right) \begin{pmatrix} S \\ L \end{pmatrix} = \begin{pmatrix} D.C \\ -\frac{1}{2}G \end{pmatrix} \qquad 4.32$$

Once a set of shifts have been calculated, they are added to the current values of the parameters to give a new set which is used to derive a set of coordinates for the subsequent cycle of the refinement. The total energy of the conformation in terms of the constraints is calculated at the end of each cycle of the refinement, which is continued until either the energy values after two successive cycles of the iteration differ by less than a specified amount, or until a specified maximum number of cycles have been completed.

The refinement procedure described above is of the Newton-Raphson type. Generally, once a starting conformation sufficiently close to the minimum has been found, convergence is rapid. However, the routine does suffer from the usual disadvantages of Newton-Raphson methods in that:-

(a) Since derivatives of higher order than one have been ignored in the Taylor expansion for Φ, the procedure generally, only converges when the first approximation to the exact solution is sufficiently close to the minimum that the convergence of ϕ towards the minimum be described by a quadratic function. Hence, in the present programme, if the starting model is not sufficiently close to the minimum energy conformation, the iteration diverges: the conformational energy increasing after each cycle. In certain cases in which divergence occurs, the routine may calculate parameter shifts which are of the correct sign but which in magnitude are too large. It is possible to attain convergence in such a case by adding a given fraction of the calculated shifts to each variable parameter at the end of each cycle. In the present programme, the facility exists to allow only a fraction of the shifts to be applied in the first few cycles of the iteration: the fraction increasing in each cycle in the order 0.25, 0.33, 0.5, 1.0. Even this procedure does not ensure convergence in many cases, however.

(b) A second difficulty, again associated with a poor first approximation is that the routine may fail in the matrix inversion step if the matrix $D.D^T$ is singular. This problem was found to be particularly serious when using the Lagrange multiplier technique.

(c) A problem common to virtually all minimisation procedures, including the present one, is that the routine only finds the local minimum in the function Φ, so that the calculated minimum at the end of an iteration may not be the global one. In practice, this is not a serious problem in the present programme since the position of the minimum is usually dominated by the need to satisfy a small number of the geometrical constraints which have a unique minimum in the constraint function.

The first two problems constitute the major limitations of the present routine in the author's experience, and in future implementations of the programme it would be desirable to seek ways of modifying the refinement procedure in order to overcome these difficulties if possible.

The Lagrange multiplier technique is designed to be used for constraints, such as those used in certain cases (see chapter 5) to ensure that different parts of the structure join up properly, which must be satisfied at the end of a refinement. In fact it is possible to incorporate all constraints into the first term of equation 4.27 and avoid using the second term at all. The importance of satisfying the constraints which would have been included in the Lagrange multiplier term is specified by including large values of the weighting factors (W_i of equation 4.27).

The programme was written so that any constraint other than those representing the van der Waal's interaction could be included in the refinement routine in either of the above ways. In general, it will not be possible to satisfy all the van der Waal's constraints exactly since this would involve all distances between non-bonded atom pairs being equal to the optimum value. Hence, it is necessary that non-bonded contacts are bound elastically to

their optimum values.

In fact the technique of ascribing large values to the weighting factors of selected constraints was found to work well in practice and the Lagrange multiplier technique was used very rarely in the refinement of models described in this thesis.

Some care is required regarding the relative weighting factors for the different constraints if the refinement is to be meaningful. In routines which refine a molecular conformation so as to minimise the discrepancy between the observed and calculated diffraction pattern, the weighting factors are set to be proportional to the estimated reliability in each experimentally determined datum. No such general scheme is possible in the present refinement routine. The relative values of weighting factors used in individual cases will be discussed in the appropriate chapter, but wherever possible they are set so as to reflect the value of appropriate constants relating to the particular type of interaction or stereochemical feature which the constraint represents.

CHAPTER V

INTERCALATION MODELS AND THEIR TRANSFORMS

5.1 Introduction

The intercalative mode of binding was first proposed by Lerman (1961) to explain the physico-chemical properties and the diffraction data given by complexes of DNA with acridine dyes, and it is now fully recognised as the strong mode of binding in many DNA/drug complexes (Waring, 1974).

Diffraction patterns from fibres of such complexes exhibit a strong reflection on the meridian at 3.4Å. Meridional reflections in diffraction patterns from helical structures occur at distances in reciprocal space indicative of the distance of separation parallel to the helix axis between repeat units of the helix. Such reflections are particularly strong if a number of atoms in the repeat unit have similar values for their z coordinate: i.e. if part of the structure lies in a plane approximately perpendicular to the helix axis. In nucleic acids the repeat unit (one nucleotide pair) contains a planar structure in the base pair. Hence, the intensity of a meridional reflection from a nucleic acid fibre is largely dependent upon the orientation of the base plane with respect to the helix axis. A strong meridional is obtained for BDNA where the bases are nearly perpendicular to the helix axis, while for ADNA where the bases are significantly tilted ($\approx 20^{o}$) from the perpendicular, the meridional reflections are very weak.

The presence of the strong meridional intensity in intercalation complexes suggests, therefore, that the bases are approximately perpendicular to the helix axis and have a similar separation parallel to it to that found in BDNA. It is generally assumed, therefore, that

the form of DNA into which intercalation takes place, and that found between intercalation sites, is that of BDNA or a similar conformation. Hence, values for the perturbation of helical parameters produced by intercalating drugs are normally quoted with reference to the equivalent standard BDNA values.

It is widely accepted, because of the position of the meridional reflection and its invariance as a function of the P/D ratio, that the axial separation of base pairs adjacent to the drug molecule is 6.8Å, and that in order to accommodate this increase the duplex is unwound at the intercalation site and possibly at sites adjacent to the intercalation gap.

The magnitude of the unwinding angle has been a source of some controversy, particularly in the case of the drug ethidium, since the value assumed for the unwinding it induces has been used to determine the degree of supercoiling of closed circular DNA, and, hence, the unwinding angle produced by other intercalating drugs (Waring, 1970).

5.1.1 Published Intercalation Models

Before considering specific models in detail, the definitions of a number of terms used in this chapter will be presented. If the native conformation of DNA is defined as the conformation into which intercalation is assumed to take place, then it is possible to define an intercalation unit as being that section of the DNA which is modified with respect to the native conformation. Implicit in this definition is the requirement that adjacent cohelical regions of the native conformation should be able to join up to the intercalation unit in a stereochemically reasonable manner. The term intercalation site will be restricted to the nucleotide pairs adjacent to the drug molecule and the two phosphate groups between them. All values for the unwinding angle (change in rotation per residue) induced by an

intercalating drug are quoted with reference to BDNA (36^{o} rotation/ residue) unless otherwise stated. A rotation of 24^{o} between two adjacent residues will be described as an unwinding of $+12^{o}$: a negative value for the unwinding therefore indicates that the helix is wound up in comparison with BDNA. This sign convention is used since it seems the most logical and is, hence, less prone to ambiguities. Nevertheless, it is the opposite of the convention adopted by many authors.

To illustrate the use of these terms, consider the sequence of four nucleotides illustrated diagrammatically in figure 5.1 which represents one chain of a section of a DNA duplex. If a drug intercalates between unit 2 and unit 3, then the intercalation site is defined by the base and sugar of residues 2 and 3 plus the phosphate of residue 2 and the equivalent atoms of the complementary chain. The segment consitutes an intercalation unit under the definition, provided that the bases of residues 1 and 4 have the same orientation with respect to the helix axis as is observed in the native conformation, and provided that the angle (χ) defining the rotation of the sugar rings about the glycosidic bonds is that of the native form. No other stipulations concerning the nature of the DNA conformation are required, though the possible number of models will be restricted if the requirement for standard Watson-Crick base pairing is imposed.

Relatively few of the numerous studies of the intercalative binding of drugs to DNA have involved the proposal of a definite molecular model for the complex. Lerman (1961) reports having built a model for the DNA conformation at the intercalation site, but few details are given other than the unwinding angle (45^{o}). In a later paper, Lerman (1963), this value was modifed to 36^{o}. Fuller and Waring (1964) have published a model for the binding of ethidium to DNA; while a model for daunomycin intercalation has been proposed by Pigram et al

(1972). Details of the DNA conformation at the intercalation site in the daunomycin model have been presented by Pigram (1968). In both the ethidium and the daunomycin model, the unwinding at the intercalation site was predicted to be 12^{o}.

Jain and Sobell (1972) have analysed the crystal structure of a complex between actinomycin D and deoxyguanosine. The structure has been solved to atomic resolution and coordinates have been given so that the model is well defined. Some care should be exercised in applying results obtained from studies of complexes such as that one described here, to complexes between the drug and DNA since the stereochemical constraints involved are different. However, the guanine bases stack on either side of the actinomycin chromophore in a way which could easily be incorporated into a DNA duplex involving Watson-Crick base pair geometry. Moreover, the orientation of the bases is such as to allow a strong hydrogen bonding interaction between the guanine 2-amino group and the carbonyl oxygen of the actinomycin L-threonine residue: a second, weaker hydrogen bond being observed between the N-H group of the L-threonine residue and the N3 ring nitrogen. The importance of the first interaction is that it might explain the observed requirement for deoxyguanosine residues in the binding of actinomycin to DNA (Goldberg et al. 1962; Wells and Larson, 1970).

Sobell and Jain (1972) have used the structural information obtained in the above study as a basis for a model of actinomycin intercalation to a hexanucleotide double stranded DNA segment. The model has very good stereochemistry and, since atomic coordinates have been presented, is very well defined. It is not, however, based on any experimental evidence other than that of Jain and Sobell (1972).

The hexanucleotide used consists of two chains ($A_pT_pG_pC_pA_pT$) having complementary base sequence, with the bases arranged C(5'-3')G about the intercalation site. Both deoxygaunosine residues at the intercalation site have sugar rings exhibiting C3' endo pucker; while all other residues have sugar rings with C3' exo pucker. The total degree of unwinding in comparison with an equivalent segment of BDNA is 34^o of which 18^o is at the intercalation site and 8^o at each of the adjacent base pairs. The disposition of the base pairs to the helix axis is not identical with that found in BDNA and hence the model is not, strictly speaking, compatible with the existence of cohelical BDNA regions between the intercalation units.

Tsai, Jain and Sobell (1975) have co-crystallized the drug ethidium bromide with the double helical synthetic RNA fragment 5-iodo-uridylyl (3'-5') adenosine and have solved its structure to atomic resolution. The visualized structure for the complex shows ethidium intercalated between the two base pairs which have a reported unwinding angle of 29^o in comparison with BDNA (36^o), and exhibits the same pattern of furanose ring puckering about the intercalation site (i.e. C3' exo (5'-3') C3' endo), which was included in the model of Sobell and Jain (1972) for the binding of actinomycin.

Computer drawings of the model are given and, hence, the general features of the model are well defined. Neither coordinates nor torsion angle data are presented, however, so that the model is not so well defined as is that of Jain and Sobell (1972) for actinomycin. The ethyl and phenyl groups of the ethidium project into the narrow groove, and the complex is not stabilized by any hydrogen bonding interactions. This is in contrast to the Fuller and Waring (1964) model in which the ethidium ethyl and phenyl groups project into the large groove, and in which the interaction is postulated to involve two hydrogen bonds between the amino groups on the ethidium ring system

and an oxygen atom of the phosphate group.

The lack of a hydrogen bond between these groups in the Tsai et al. (1975) model is a consequence of the backbone conformation induced by the large degree of unwinding. Intercalation models having an unwinding of 12 degrees at the intercalation site, have one of the phosphate oxygens of each chain projecting into the intercalation gap, and hence these are well positioned to form hydrogen bonds with the ethidium amino groups. At larger values of the unwinding angle, the phosphate oxygens gradually move round to project out of the intercalation gap and, hence, away from the ethidium chromphore, making hydrogen bonding impossible.

It should be noted that there is no evidence for the intercalative binding of ethidium to double helical RNA. The fact that intercalation has been visualised in an RNA fragment in this case may illustrate the limitation of studying the intercalative binding of drugs to nucleic acid duplexes by extrapolation from data obtained in binding studies with oligonucleotides. It is also possible, however, that the nature of the sugar moety is insignificant in a dinucleotide. A major difference between RNA and DNA duplexes is the fact that no conformations of the B type (viz. chapter 6) have been observed in RNA. This has been explained in terms of the difficulty of accommodating the extra hydroxyl group of the ribose sugar into a B type helix. In a dinucleotide, however, such steric hindrance may not be present so that the presence of the extra hydroxyl group may be insignificant. Evidence in favour of this argument is the appearance in the complex of a ribose sugar having C3' exo pucker. This is the ring shape associated with B type helices and is not observed in natural or synthetic RNA duplexes.

All the above models have involved the intercalation of a specific drug into DNA. It is also of interest to build intercalation models which do not involve any particular drug, since such a study helps to define the conformational possibilities of the native DNA, and to define any stereochemical features which must of necessity be incorporated into a starting model for a later refinement with a specific drug.

Dr. W.J. Pigram has built a number of models for intercalation sites using an early version of the computer programme for molecular refinement described in chapter 4. The models were built so as to have a dyad axis: the rotations about the glycosidic bonds of the sugars were allowed to vary, but were constrained to be equal for both sugars; and the rotation per residue was constrined to be a given value in any one refinement. Since there is no requirement in these models that the rotation about the glycosidic bonds should be identical to that of the native (B) DNA, the models do not constitute an intercalation unit. Models having C2' endo and C3' endo sugar puckers were built; no cases of mixed sugar pucker being considered. The models had unwinding angles in the range 19.6^{o} to 31^{o} for the models having C2' endo sugar pucker; and values from 21^{o} to 36^{o} for models having C3' endo sugar pucker.

Recently, a general intercalation unit having the Arnott and Hukins (1972) model of BDNA as the native conformation has been published (Alden and Arnott, 1975). A dyad axis is assumed to exist in the structure, so that only one chain of the unit is considered. The unit consists of four base pairs and will be discussed in relation to figure 5.1. The base at the 3' terminus is fixed so as to have the disposition to the helix axis observed in BDNA; while the angle (χ) defining the orientation of the sugar with respect to the base to

which it is attached, is kept at the BDNA value for the sugars of residues 1 and 4. A number of other constraints are applied to the system and are given in table 5.1.

Equivalent atoms in the sugars of residue 1 and 4 are constained to be cohelical, and the bases are all constrained to be in a position to form Watson-Crick pairs with the appropriate bases of the complementary strand. In conjunction with the imposed conditions discussed above, these constraints ensure that the nucleotide pairs at either end of the model are compatible with adjacent cohelical BDNA regions, and the structure therefore constitutes an intercalation unit. Coordinates for the model are not given but values for the torsion angles are quoted. This data has been used to derive the atomic coordinates for the model for comparison with similar results obtained by the author. The model which will be discussed in section 5.4.2, was found to be unsatisfactory in a number of ways and did not appear to correspond to the intended structure.

In this chapter a number of "general" intercalation models built in the absence of any particular drug are described. These have been used as starting models to build intercalation units with specific drugs.

5.2 Intercalation Models Involving DNA Alone

The DNA duplex has a large number of degrees of freedom in changing from the native conformation to the modified form around the intercalation site. A complete multidimensional conformational analysis in which the DNA duplex is allowed a maximum number of degrees of freedom is neither possible nor justified in view of the uncertainties involved in the choice of starting model. Hence, if progress is to be made, a number of simplyfying assumptions must be

TABLE 5.1

Constraint	Where Applied	Desired Value	Function
Distance	GN1 ... CN3) of all GN2 ... CO2) residues GO6 ... CN4)	2.92A 2.85A 2.87A	Maintains correct hydrogen bond geometry between bases of a pair
$\gamma^{(+)}$	Non-root bases	4.0^{o}	Maintain base stacking and disposition to helix axis.
Angle		90^{o}	Maintain correct twist axis
Radius	1C1' ... 4C1' 1C2' ... 4C2' 1O5' ... 4O5'	0.0	Ensures that the ends of the chain are cohelical
Axial separation	1C1' ... 4C1' 1C2' ... 4C2' 1O5' ... 4O5'	13.52 13.52 13.52	Ensures that the ends of the chain are cohelical

(+) γ is the angle made to the helix axis by the normal to the plane of the bases.

formulated and rejected or modified later, if necessary, on the basis of results obtained in a preliminary survey. The assumptions made in the model building survey described in this chapter are indicated below.

(i) It is assumed that the conformation of DNA into which the drug intercalates is that of the B form, and that coaxial BDNA regions exist between each intercalation unit: all intercalation units being assumed to be identical and of finite length.

(ii) The furanose ring pucker at each residue, though changed for different refinement runs, is not allowed to vary during any one refinement.

(iii) The parameters defining the disposition of the bases with respect to the helix axis (i.e. the tilt and twist of the base pairs and their distance from the helix axis; Arnott (1969)) are fixed at standard BDNA values (see chapter 1).

(iv) A dyad axis is assumed to exist in the structure so that only one chain of the intercalation unit need be built.

The constraints on the modelling system implied by the above simplifications reduce the number of degrees of freedom of the system (see sections 5.2.1) and allow more efficient convergence of the refinement routine.

Assumption (iii), concerning the disposition of the bases to the helix axis, is probably a reasonable first approximation. If large changes in the base disposition were introduced by intercalating drugs, the helical structure of the DNA would be lost and the evidence from fibre diffraction data is that approximate helical symmetry is retained. Nevertheless, it is possible that some small changes in the base disposition may be introduced during the unwinding process so that assumption (iii) would need to be relaxed in a detailed refinement

of a particular intercalation complex.

Since any loss of symmetry between the two chains of an intercalation unit is likely to be a reflection of assymmetric features of the drug molecule, the decision to retrain a dyad axis in intercalation units built in the absence of any particular drug is justified. It is probably a good first approximation to make for most cases in which a drug is included, since the chromophores of most intercalating drugs have at least a pseudo-dyad-axis of symmetry.

The retention of a given furanose ring pucker during a refinement run is probably not a serious limitation on the validity of the model building scheme, since the choice of sugar pucker may be varied in different refinement runs. This method of approach has been used in computerised refinement of nucleic acid structures by Arnott and his co-workers (e.g. Arnott and Hukins, 1972).

The assumption has been made that all intercalation units are identical, with coaxial BDNA regions between them. Although it is probable that the DNA conformation at every intercalation site is identical, it is possible that the conformation of the DNA regions between them is not that of the B form, and that the conformation of a region between two particular intercalation sites is a function of the distance between them. Since very little evidence which bears upon this point is available, it is difficult to assess the validity of the assumption contained in (i) concerning the nature of the region between intercalation units.

Models built in accord with the above constraints may be considered as the simplest possible (least modified) intercalation units which can be derived starting from a BDNA conformation. The model of Alden and Arnott (1975), discussed in section 5.1, is of this type.

5.2.1 Methods of Building the Intercalation Unit

Method (i)

In the first method of approach illustrated in figure 5.1, the main chain runs from the C3' atom of the upper residue to the C4' atom of the lowest residue. Co-ordinates of all atoms of the non-terminal sugars which do not form part of the main chain (i.e. all atoms other than C5', C4' and C3'), including hydrogens and the N9 of the appropriate base, are derived by considering the atoms as being pendant to the C3' atom. The O2 and O3 atoms of the phosphate group and the hydrogens attached to C5' are considered as being pendant to their appropriate chain atoms.

Atoms for the lower sugar (except C4' and C5'); all atoms for the upper sugar (including C3' and O1); and the base atoms indicated in figure 5.1, (including the N9 and C1 of all residues), are included in the calculation as fixed atoms. Hence, coordinates for these are read in directly by the programme and are not modified during the refinement.

The coordinate reference frame for the fixed atoms has its origin situated on the helix axis and its x-axis along the local dyad axis for the lowest base pair along the line R, 0, 0. It is clearly necessary that the coordinate of the phosphate backbone derived by the programme are in the same coordinate frame as those of the fixed atoms. This is ensured by extending the chain to terminate on a "dummy" atom situated at the origin, via a similar atom situated on the x-axis. These "dummy" atoms are excluded from consideration in the search for van der Waal's contacts by including a negative index in their atom neighbour lists (viz. chapter 4).

Atom D1 may be placed anywhere on the x-axis, but it is normally given coordinates (X, 0, 0), where X is the x coordinate of atom C4' of the lowest residue. The bond angle between atoms C4', D1

and D2 (at the origin) is then 90^{o}. If the value of τ_{18} of fig. 5.1, is taken as the angle between the X-Y plane of the coordinate system and the plane defined by atoms C4', D1 and D2; and provided that the correct values are given at the bond lengths C4'-D1 and D1-D2, the C4' and C5' atoms of residue 1 are built with the same coordinates as in BDNA. Hence, the coordinate system of the main chain and that of the fixed atoms are identical.

If the final model is to be valid, constraints must be applied to the system to ensure that the chain joins up in the correct way to those parts of the structure defined by the fixed atom system. In order to achieve this, position constraints are applied between the following atoms in the chain and their counterparts in the fixed atom list: the C3' and O1 of the upper residue (residue 4 in fig. 5.1), and the C1' and N9 atoms of the non-terminal residues. The desired position is defined by the coordinates of the appropriate atom in the fixed atom list and, hence, the desired distances of separation in the X, Y and Z directions are set to zero, constraining the atoms of the chain to occupy the same positions as their counterparts in the fixed atom list. In order that van der Waals interactions between atom pairs subjected to the positional constraints are not included in the refinement routine, the fixed atoms concerned are excluded from consideration in the search for short non-bonded contacts by including a negative index as the first element in their atom neighbour list (viz. chapter 4). Hence the rotation and rise per residue in the intercalation unit are determined by the fixed atom positions. The retention of a constant value for the radial coordinate, R, of equivalent atoms in different residues ensures that the disposition of the bases in relation to the helix axis is that of BDNA.

For an intercalation unit of N nucleotide pairs, there are 5N-6 variable torsion angles and 2N-2 positional constraints, giving 3N-4 degrees of freedom to the system. For the case shown in figure 5.1, where N = 4, there are 14 variable torsion angles and 6 positional constraints giving 8 degrees of freedom.

The modelling system described above is efficient in that each cycle of the iteration is rapid since the coordinates of only a relatively small number of atoms are derived: the coordinates for most atoms being given in the fixed atom list. Moreover, since the main chain of the model building system corresponds to the phosphate backbone, the required data for a starting model are easily obtained from a wire model of the structure.

A disadvantage of the method, however, is that the rotations about the N9-C1' glycosidic bonds of the non-terminal base pairs are not explicitly defined in terms of variable torsion angles. Moreover, the length of the main chain increases with an increased number of residues in the intercalation unit. Given changes in torsion angles at the end of the chain induce correspondingly larger changes in the coordinates of atoms at the start of the chain as the chain length increases. For very long chains it is difficult to obtain a starting model from which the refinement will converge since small errors in the starting values for the torsion angles, or in the shifts calculated at the end of any cycle of the iteration, will induce large errors in the coordinates of certain atoms. A third, and perhaps most important, limitation of the system is that although the rotation per residue can be changed at will in different refinement runs, the parameter cannot be varied during any one refinement.

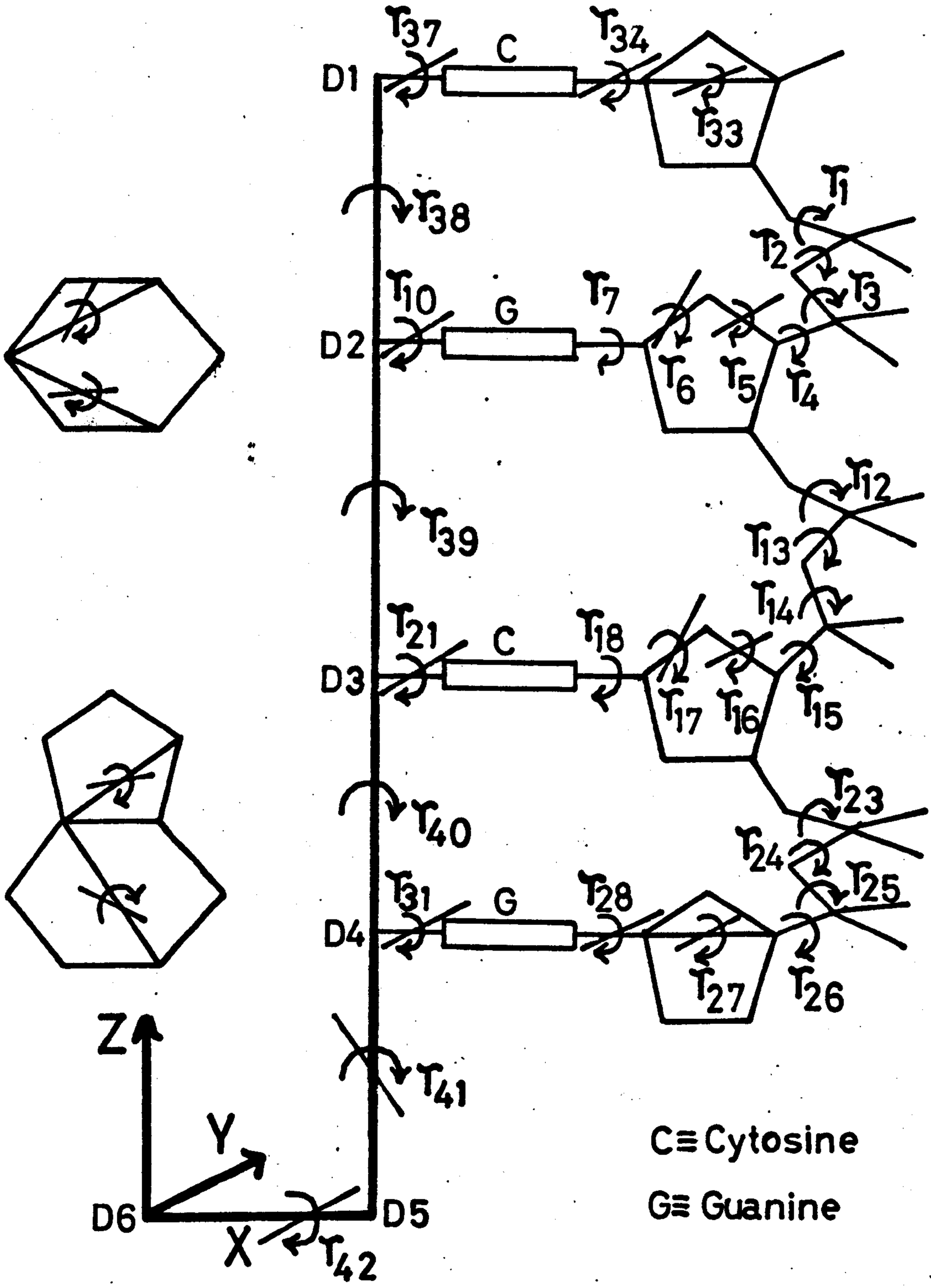


Fig. 5·2

Method (ii)

A second modelling system is illustrated in figure 5.2 for an intercalation unit containing four residues. The main chain runs from the C5' of the upper residue (residue four in fig. 5.2); across the sugar ring via atoms C4' and C1'; then across the base via atoms N9, C5, and N3 (purine base) or atoms N3, C5 and N1 (pyrimidine base) and then via a series of dummy atoms on the helix axis, to the origin.

The sugar atoms not included as part of the main chain are built into the structure as pendant atoms to the C1' atom. Atoms of a pyrimidine base are considered pendant to the base C5 atoms; while for a purine base some atoms are considered pendant to atom C5 and others to N1. Other residues in the structure are built as branches joining the main chain at the appropriate dummy atom on the helix axis as shown in figure 5.2.

Since atoms D_1, D_2, D_3 and D_4 in figure 5.2 all lie on the helix axis, the bond angles between them are equal to 180^0. Hence, the coordinates of the above atoms will not be affected by the values of the dihedral angles τ_{38}, τ_{39} and τ_{40}. The angles cannot be given arbitrary values, however, since they affect the rotation about the helix axis of the branches. It is the required rotation per residue, therefore, which determine the values allocated to the above torsion angles.

The dihedral angles τ_{38}, τ_{39} and τ_{40} which define the rotation per residue in the intercalation unit may be allowed to vary during a refinement, thus allowing the unwinding parameter to be refined in any run. Moreover, an increase in the number of residues included in the intercalation unit is accommodated by an increase in the number of branches without an increase in chain length for each branch. It is necessary to have a dummy atom on the main chain for each branch so that an increase in the number of residues is accompanied

by an increase in the length of the main chain, but not in a part of the chain which defines the conformation of the phosphate backbone.

Since there are no fixed atoms included in this model building system it is not necessary for the purposes of refinement, to adopt any procedure to obtain the atomic coordinates in any particular reference frame. However, the coordinates are much more meaningful, and the atomic positions of the dyad related chain much more easily derived, if the final coordinate system chosen is that of a BDNA base pair in its "dyad" position (i.e. the dyad of the base pair situated along, R, 0, 0). To achieve this a procedure analogous to that described for the previous model building system is adopted, and illustrated in figure 5.2.

The modelling system described here overcomes all the disadvantages of method (i). It converges better for intercalation units containing a large number of residues, though it is probably little better in this respect for intercalation units in which $N \lesssim 5$. The system is much more flexible and allows the rotation per residue to vary during the refinement. This flexibility is paid for by an increase in computation, since the coordinates of all atoms are derived by the programme in each cycle. Nevertheless, on the whole it is a superior method of model building to method (i) for the final refinement of the best model.

In a number of computer model building studies of protein conformation, the effect of hydrogen atoms in the structure is simulated by constraining the torsion angles to be elastically bound to standard values. This type of constraint has been employed in the model building study of Alden and Arnott (1975) on the conformation of DNA at an intercalation site. In the present study, hydrogen atoms were included in the structure and, hence, the dihedral angles are not generally constrained to adopt particular values.

5.2.2 Intercalation Models Including Specific Drugs

Having built the general models for intercalation, it is of interest to use these to build models of intercalation complexes for specific drugs. A number of such models have been built for the drugs ethidium and daunomycin. Work is currently in progress to build a model of an intercalation complex between the drug echinomycin and DNA.

The models built include cases in which the DNA conformation was not varied during the refinement; the drug molecule being built into one of the general intercalation models refined in the absence of a drug. In other cases one of the intercalation models was chosen as a starting model and used in a refinement in which the DNA conformation and drug position were varied simultaneously.

The general method by which the drug atomic coordinates are derived is illustrated in figure 5.3 for a general triple ring system. Suitable atoms in the ring system and in the rest of the drug molecule are chosen to form the chain. Atoms of the drug fused ring system are considered as pendant to one of the chain atoms. Stereochemical data concerning the drug are normally obtained from a single crystal determination of its structure.

It is necessary that the drug coordinates be in the same coordinate frame as that of the DNA. This is ensured by a procedure entirely analogous to that adopted in the building of DNA, in which the chain is extended to terminate on the helix axis. The values of the dihedral angles, τ_1, τ_2 and τ_3 of figure 5.3 are obtained using coordinates measured from a starting model of the complex built by hand model building. During a refinement, these dihedral angles are fixed and the drugs position is refined by means of changes in the values of the eulerian angles defining rotations about the cartesian coordinate axes, and by changes in the parameter values defining translations along these axes.

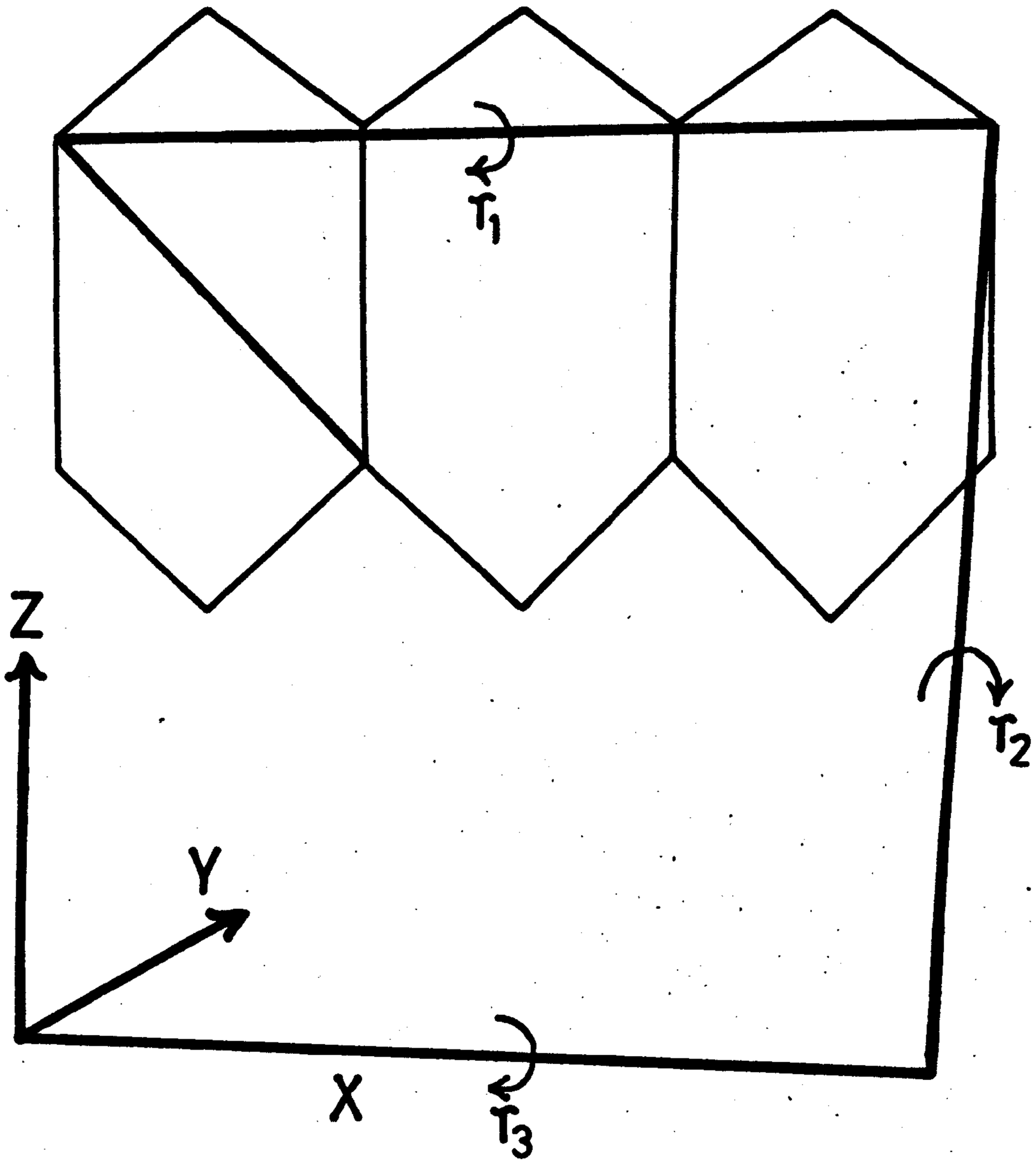


Fig. 5·3

In cases where the DNA conformation is not to be refined, the DNA atomic coordinates are read in as part of the fixed atom list, and the drug only is built by the programme and positioned in relation to the fixed atoms. If the intercalation unit is to be co-refined with the drug molecule, the drug is built as before and included in the model of the total structure as a branch to the main chain of the intercalation unit which is built by the method (ii) above. The branch joins the main chain at the dummy atom situated at the origin (atom D5 of fig. 5.2).

The specific method of building each of the drugs studied will now be discussed.

(a) Ethidium

In studies of drug binding to closed circular DNA, the unwinding angle due to ethidium has been taken as 12^{o}, the value postulated by Fuller and Waring (1964). However, estimates of this parameter have varied from 36^{o} to -13^{o} (Fuller and Waring, 1964; Paoletti and Le Pecq, 1971; Tsai et. al. 1975; Pulleybank and Morgans, 1975; Pigram, Fuller and Davies, 1973; Waring, 1974). The value of -13^{o} (i.e. in which the helix is wound up at the intercalation site) was proposed by Paoletti and Le Pecq (1971) on the basis of fluorescence depolarisation measurements from ethidium/DNA complexes. However, the suggestion that ethidium intercalation winds up the helix has been refuted by a number of authors, in particular Pigram, Fuller and Davies (1973) and Pulleybank and Morgans (1975). It is now widely accepted that the helix is unwound at the intercalation site in ethidium/DNA complexes, but a number of workers (Waring, 1974; Pulleybank and Morgans, 1975, Tsai et. al., 1975) have now provided evidence suggesting that the magnitude of the unwinding angle produced by ethidium may be at least twice the value of 12^{o} proposed by Fuller

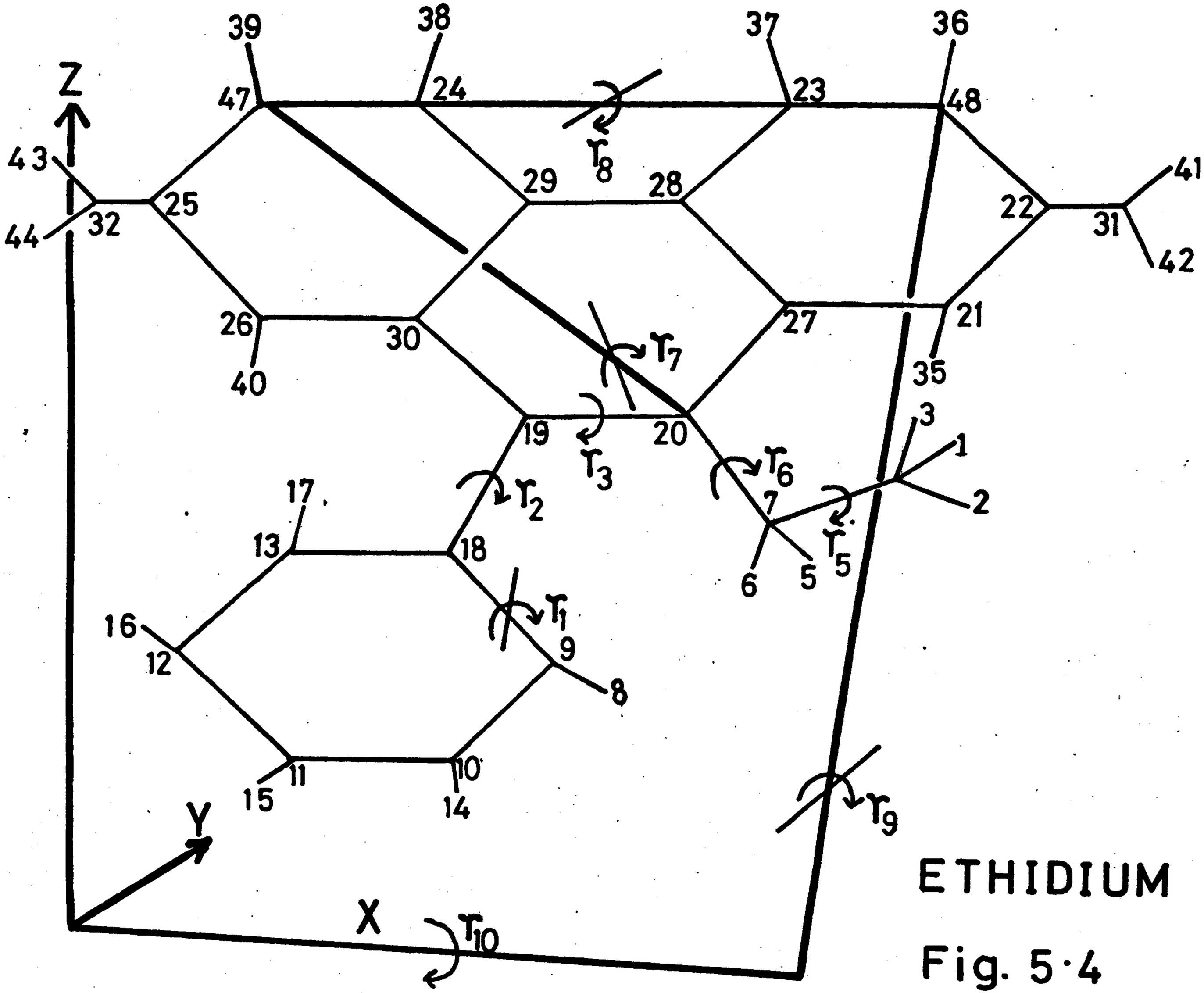
and Waring (1964).

Since the unwinding angles proposed for ethidium either involve values of around 12^{o} or one of $25\text{-}30^{o}$, models with these values for the unwinding parameter were chosen as starting models for the ethidium intercalation complex. In a number of cases where the coordinates of the DNA were not varied during the refinement, the DNA coordinates for one chain were given as fixed atoms and the coordinates of the dyad related chain derived by the programme using the routine described in chapter 4.

The method by which the ethidium molecule is built is illustrated in fig. 5.4. Coordinates for ethidium have been derived by Subramanian et. al. (1971) from single crystal studies and they have been used to determine the relevant data necessary to build the drug into the intercalation complex.

The main chain runs along the ethyl group to atom N10 on the ethidium triple ring system (atom 20 in figure 5.4), to atom C6 (47 in fig. 5.4), then to atom C3 (48). The chain is continued to a dummy atom on the x-axis, and finally to the origin.

A branch runs from atom 10 on the main chain across atoms C9(19), C15(18), C16(9) and H33(8). The coordinates of atoms in the ethyl group not included as chain atoms of the branch are built in as pendant atoms to C15; while atoms of the drug chomophore not forming part of the main chain are considered as being pendant to atom C6. All pendant atom data for the model as well as values of torsion angles defined by atoms internal to the drug structure can be derived directly from the molecular coordinates. Once this data has been obtained the starting model can built. Provided that there are no excessively short contacts, the refinement was begun from the starting model thus obtained: the starting values for the eulerian angles and for the translational parameters which define the position of the molecule are,



ETHIDIUM

Fig. 5·4

hence, set at zero. Torsion angle, τ_2, defining the orientation of the phenyl group and dihedral angles 5 and 6 defining the conformation of the ethyl moety may be allowed to vary during a refinement.

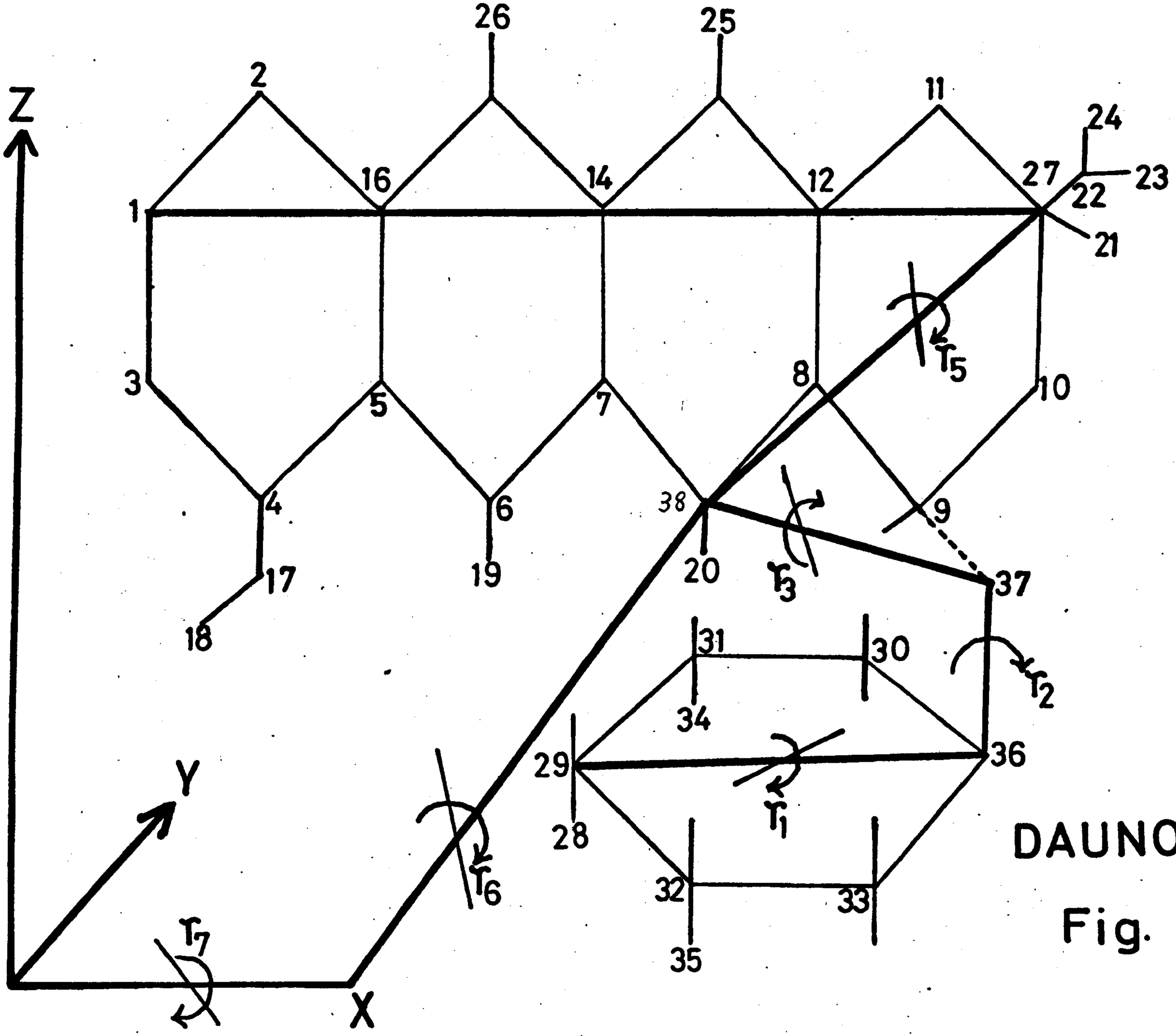
In cases where the DNA conformation was allowed to vary during the refinement, the method by which the drug was built and positioned were identical to that described above. The DNA portion of the complex is built by method (ii) described in section 5.2.1 and coordinates for the dyad related chain were derived by the programme after each cycle of the refinement. In this case the ethidium is built as a branch joining the main chain at atom D5 of figure 5.2, and then positioned according to the eulerian angles and translational parameters.

(b) Daunomycin

The general principles of the method used to build the drug daunomycin into intercalation units are identical to those used to build ethidium models. Only the manner in which the daunomycin molecule is built is significantly different from the equivalent stage in the building of ethidium models. The method by which this is done is illustrated in figure 5.5.

The main chain runs from atom C2 (atom number 1 of figure 5.5) to atom Cl2(27), to atom C8(38) and then via a dummy atom to the origin. A branch chain runs across the ring of the side group and joins the main chain at atom C8. Atoms of the drug chromophore which are not part of the main chain are included as pendant atoms to Cl2, while equivalent atoms in the side group ring are considered as being pendant to C33(29).

Coordinates for the daunomycin molecule are taken from Angiuli et. al. (1971).



DAUNOMYCIN

Fig. 5·5

5.3 Calculation of the Fourier Transform for Models

The cylindrically averaged molecular transform was calculated for a number of the models built which included specific drugs. In all cases the method used was that described in section 3.5. The data set used is similar in all cases and three examples of the method of approach will be given and will serve as models for all Fourier transform calculations mentioned in this chapter.

The size and composition of the repeat unit is dependent upon the P/D at which the calculation is performed. In the case of P/D = 20 the repeat unit chosen consists of 10 mucleotide pairs and one drug molecule. If the drug intercalates at a particular site, it induces a change in the rotation between the base pairs, but no change in the disposition of the bases to the helix axis. In addition it induces characteristic changes in the conformations of the phosphate group at the intercalation site and of the nearest neighbours. The rotation (χ) about the glycosidic bonds of nucleotide pairs at either side of the intercalation site is also modified. Hence, the actual composition of the repeat unit is one drug molecule; ten base pairs, eight BDNA sugar units; seven phosphate groups in the BDNA conformation; three phosphate groups each having one of the modified conformations found in the intercalation unit; and two sugar rings each having one of the modified orientations with respect to its base found in the intercalation unit. In deriving the average repeat unit, we assume that there is an equal possibility of intercalation at any one of the ten possible positions.

Since different groups of atoms have a different distribution in any given repeat unit, it is not possible to derive one unique disposition function which is used for all atoms. Instead the atoms must be divided into groups having a common distribution in the repeat unit, and hence a common disposition function in the average repeat unit.

The different groups of atoms are given, along with their respective disposition function, in tables 5.2 and 5.3 for the cases P/D = 10 and P/D = 20.

The disposition function for the drug atoms and for the modified sugar and phosphate coordinate of the intercalation unit are of a simple form, since they all appear once, and once only, in each of the ten possible canonical repeat units. Hence, their disposition function consists of ϕ and z coordinates specifying 10 different positions and having equal weighting factors. The average unit cell is normalised to the scattering power of one canonical form and, hence, the weighting factors in this case are set at 0.1 for each possible position.

In the case of the base, the BDNA sugar and BDNA phosphate group, the disposition is more complex. The function positions and weighting factors are derived in an analogous manner to that adopted for the disposition functions just described. The function is formed in each case by considering the spatial distribution of the particular group when the drug is intercalated in each one of its possible positions in turn. Relative values of the weighting factors for each position of the atomic group can be obtained as the number of times the group is situated at a given position. In order to obtain the correct absolute values of the weighting factors, the values obtained above need to be divided by ten to normalise the function to the scattering power of one unit as before.

In order to obtain the transform for the model, the programme calculates the transform of each atomic group in turn; multiplies this by the disposition function and stores the result. As the contribution from each atomic group is thus calculated it is added to a running total which gives the total transform when the contribution from the final atomic group has been included. The summation is performed in amplitude and phase, and for each Bessel function component separately

TABLE 5.2

Distribution Functions For Case P/D = 10 - Untwist = 12 degrees

Lower Phosphate and Middle Sugar + Phosphate

Position	Phi (°)	Z (Å)	Weight
1	-36.0	-3.38	0.2
2	0.0	0.0	0.2
3	36.0	3.38	0.2
4	72.0	6.76	0.2
5	108.0	10.14	0.2

Upper Sugar + Phosphate

Position	Phi (°)	Z (Å)	Weight
1	-60.0	-10.14	0.2*
2	-36.0	-3.38	0.2
3	0.0	0.0	0.2
4	36.0	3.38	0.2
5	72.0	6.76	0.2

BDNA Phosphate

Position	Phi (°)	Z (Å)	Weight
1	0.0	0.0	0.4
2	36.0	3.38	0.4
3	72.0	6.76	0.2
4	60.0	10.14	0.2*
5	96.0	13.52	0.4*
6	132.0	16.90	0.4*

BDNA Sugar

Position	Phi (°)	Z (Å)	Weight
1	0.0	0.0	0.6
2	36.0	3.38	0.6
3	72.0	6.76	0.4
4	60.0	10.14	0.2*
5	108.0	10.14	0.2
6	96.0	13.52	0.4*
7	132.0	16.90	0.6*

BDNA Base

Position	Phi (°)	Z (Å)	Weight
1	0.0	0.0	1.0
2	36.0	3.38	0.8
3	24.0	6.76	0.2*
4	72.0	6.76	0.6
5	60.0	10.14	0.4*
6	108.0	10.14	0.4
7	96.0	13.52	0.6*
8	144.0	13.52	0.2
9	132.0	16.90	0.8*

Drug Molecule

Position	Phi (°)	Z (Å)	Weight
1	-36.0	-3.38	0.2
2	0.0	0.0	0.2
3	36.0	3.38	0.2
4	72.0	6.76	0.2
5	108.0	10.14	0.2

* Indicates those values which result from modification due to the presence of a drug.

TABLE 5.3

Distribution Function

P/D = 20 - Untwist Angle = 12 degrees

Tower Phosphate and Middle Sugar + Phosphate

Position	Phi (°)	Z (Å)	Weight
1	-36.0	-3.38	0.1
2	0.0	0.0	0.1
3	36.0	3.38	0.1
4	72.0	6.76	0.1
5	108.0	10.14	0.1
6	144.0	13.52	0.1
7	180.0	16.90	0.1
8	216.0	20.28	0.1
9	252.0	23.56	0.1
10	288.0	26.94	0.1

Upper Sugar + Phosphate

Position	Phi (°)	Z (Å)	Weight
1	-60.0	-10.14	0.1*
2	-36.0	-3.38	0.1
3	0.0	0.0	0.1
4	36.0	3.38	0.1
5	72.0	6.76	0.1
6	108.0	10.14	0.1
7	144.0	13.52	0.1
8	180.0	16.90	0.1
9	216.0	20.28	0.1
10	254.0	23.56	0.1

BDNA Phosphate

Position	Phi(°)	Z(Å)	Weight
1	0.0	0.0	0.7
2	36.0	3.38	0.7
3	72.0	6.76	0.6
4	60.0	10.14	0.1*
5	108.0	10.14	0.5
6	96.0	13.52	0.2*
7	144.0	13.52	0.4*
8	132.0	16.90	0.3*
9	180.0	16.90	0.3
10	168.0	20.28	0.4*
11	216.0	20.28	0.2*
12	204.0	23.66	0.5*
13	252.0	23.66	0.1
14	240.0	27.04	0.6*
15	276.0	30.42	0.7*
16	312.0	33.86	0.7*

BDNA Sugar

Position	Phi (°)	Z(Å)	Weight
1	0.0	0.0	0.8
2	36.0	3.38	0.8
3	72.0	6.76	0.7
4	60.0	10.14	0.1*
5	108.0	10.14	0.6
6	96.0	13.52	0.2*
7	144.0	13.52	0.5
8	132.0	16.90	0.3

Position	Phi (°)	Z (Å)	Weight
9	180.0	16.90	0.4
10	168.0	20.28	0.4*
11	216.0	20.28	0.3
12	204.0	23.66	0.5*
13	252.0	23.66	0.2
14	240.0	27.04	0.6*
15	288.0	27.04	0.1
16	276.0	30.42	0.7*
17	312.0	33.86	0.8*

BDNA Base

Position	Phi (°)	Z (Å)	Weight
1	0.0	0.0	1.0
2	36.0	3.38	0.9
3	24.0	6.76	0.1
4	72.0	6.76	0.8
5	60.0	10.14	0.2*
6	108.0	10.14	0.7
7	96.0	13.52	0.3*
8	144.0	13.52	0.6
9	132.0	16.90	0.4*
10	180.0	16.90	0.5
11	168.0	20.28	0.5*
12	216.0	20.28	0.4
13	204.0	23.66	0.6
14	252.0	23.66	0.3
15	240.0	27.04	0.7*

Position	Phi (°)	Z (Å)	Weight
16	288.0	27.04	0.2
17	276.0	30.42	0.8*
18	324.0	30.48	0.1
19	312.0	33.86	0.9*

Drug

Position	Phi (°)	Z (Å)	Weight
1	-36.0	-3.38	0.1
2	0.0	0.0	0.1
3	36.0	3.38	0.1
4	72.0	6.76	0.1
5	108.0	10.14	0.1
6	144.0	13.52	0.1
7	180.0	16.90	0.1
8	216.0	20.28	0.1
9	252.0	23.56	0.1
10	288.0	26.94	0.1

*Indicates those values which result from modifications due to the presence of a drug.

at each point in reciprocal space at which the calculation is performed.

Since the intercalation unit has a dyad axis, it follows that the DNA portion of the repeat unit has a dyad axis relating the two chains. The drug molecule is not normally symmetrical and hence the total repeat unit does not possess a dyad axis. In order to take advantage of the dyad symmetry to reduce the computation and data preparation, the facility available in the programme to calculate only the real part of the structure for selected atoms is used. If the dyad simplification is to be used it is necessary to ensure that the coordinates of the atoms are in a cylindrical polar coordinate frame in which the dyad lies along the line R, 0, 0.

Weighting factors for the drug atoms are set equal to 0.5 to compensate for the fact that use of the dyad simplification reduces by half the scattering power of the DNA chain to which it is applied. Equivalent data sets for other intercalation models are derived in an analogous manner to that described above.

5.4 Results of Model Building

5.4.1 "Alternate Binding" Models

There is now much evidence (Waring, 1972) to suggest that the binding of an intercalating drug between two base pairs in a DNA duplex prevents intercalation by another drug molecule at adjacent sites. Waring (1965, 1966) has shown that the strong mode of ethidium binding (interpreted as intercalative binding) saturates at P/D between 4 and 6. This behaviour has subsequently been observed for other drugs (e.g. actinomycin: Muller and Crothers, 1968). Drs. T. Sundaresan and W.J. Pigram in this laboratory have shown that pitch values obtained from diffraction patterns of daunomycin/DNA complexes reach a maximum value of 50Å at P/D = 4. These results have

lead to the idea of an "alternate binding" model in which the adjacent sites only are excluded by the presence of an intercalating drug, giving one drug molecule at every other site.

Attempts were made, therefore, to build an alternate binding model by the methods outlined in section 5.2, in accord with the assumptions given in section 5.1: the alternate binding model merely corresponding to the case where N = 3. Models were built based upon the C2' endo BDNA conformation, and on the C3' exo B form of Arnott and Hukins (1972). In the latter case refinements were attempted in which the sugar pucker is mixed: being C3' (5'-3') C3' endo about the intercalation site.

It was not found to be possible to build such a model in the present study, although it is not possible to conclude definitively, that such a structure cannot be built. At least one of the assumptions made, the requirement that the intercalation unit connects up in a stereochemically reasonable way to adjacent cohelical BDNA regions, is probably invalid since in an alternate binding scheme there are no DNA regions between intercalation units. Further attempts to build an alternate binding model in which this requirement was relaxed have not been made since there are some objections to the proposal of an alternate binding scheme for most, if not all, DNA/drug complexes in fibres. The main objection is that even assuming an unwinding angle of 12 degrees at the intercalation site, the minumum value found possible, the pitch value obtained would be about 60Å. No pitch values larger than approximately 50Å have been measured from fibres of DNA/drug intercalation complexes.

The degree of disorder apparent in many intercalation complexes at P/D = 4 does not appear to be in accord with an alternative binding scheme which predicts that the helix is regular at this drug concentration.

Bond et. al. (1975) have published results of an X-ray fibre diffraction/model building study of DNA complexed with 2 hydroxyethanethiolato (2,2',2" - terpyridine) platinum (II) (PtTs) in which an alternate binding scheme is proposed. The smallest value of the unwinding angle in their models is 16 degrees (model 1), corresponding to a pitch value of approximately 65Å.

The element platinum has more electrons (i.e. greater scattering power for X-rays) than does a base pair, and hence tends to dominate the scattering in certain regions. In particular, it may be used as a probe for determining the drug distribution. Since in an alternate binding scheme, the drug should be distributed at a regular repeat distance of 10.2Å (3 x 3.4), there should be meridional reflections at points in reciprocal space corresponding to 10.2Å and 5.1Å. Such meridional intensity should be enhanced by the presence of platinum in the drug structure, and have been observed in the diffraction patterns presented by Bond et. al (1975).

Unfortunately no estimate of the pitch values have been made from the diffraction data. This is a serious omission since knowledge of this parameter would provide a good test of the validity of their molecular models. Moreover, if a pitch value around 65Å was indicated, it would be the largest ever observed in fibres of intercalation complexes. If the data presented by Bond et. al (1975) has been correctly interpreted by them, then PtTS appears to be unusual amongst intercalating drugs in the concentration of drug bound in the fibre state. However, although the absence of a 6.8Å meridional reflection shows that there is an excluded site effect, the relative intensities of the meridional reflections suggests that the drug is not bound to saturation. At saturation (P/D = 4) in an alternate binding model scheme, the 3.4Å meridional intensity would be produced by the DNA base

and from the platinum. Only the platinum would contribute to the other meridional streaks at 5.1Å and 10.2Å. Nevertheless, since the platinum contains almost twice as many electrons as a base pair, these latter two reflections might have been expected to be of comparable intensity to that of the 3.4Å meridional. In fact they are considerably less intense than the 3.4Å meridional, particulary the 5.1Å reflection which appears to have less than 1/100th of the intensity of the 3.4Å meridional. This suggests that the degree of intercalative binding has not reached saturation level.

5.4.2 Intercalation Units in which N=4

Since it proved impossible to construct a satisfactory intercalation unit with N=3, it was decided to attempt to build a unit in which N=4. A wire model was constructed and used to provide a starting model having an untwist of 12 degrees at the intercalation site. The torsion angle data so obtained were used to refine the model which was based upon the Langridge et. al. (1960) (Arnott et. al., 1969 refinement) BDNA structure. It was subsequently found to be possible to build a model having an untwist of 10 degrees at the intercalation site. A series of models were built having values of the unwinding angle ranging from 10 degrees to 36^{o} in intervals of 1 degree. A graph has been plotted (fig. 5.6) of the van der Waal's energy against the unwinding angle for these models, and it can be seen that the energy of the structure in terms of its non-bonded contacts is less at smaller values of the unwinding angle. It proved impossible, with the imposed constraints to build models having an unwinding of 20^{o} or 21^{o} due to steric hinderence in the phosphate backbone. Although the energy of non-bonded contacts decreases with a decreasing degree of unwinding, the geometrical constraints are better satisfied in the models having larger values for this parameter (see table 5.6). It would be of interest to determine whether the energy barrier at 20^{o} unwinding is a

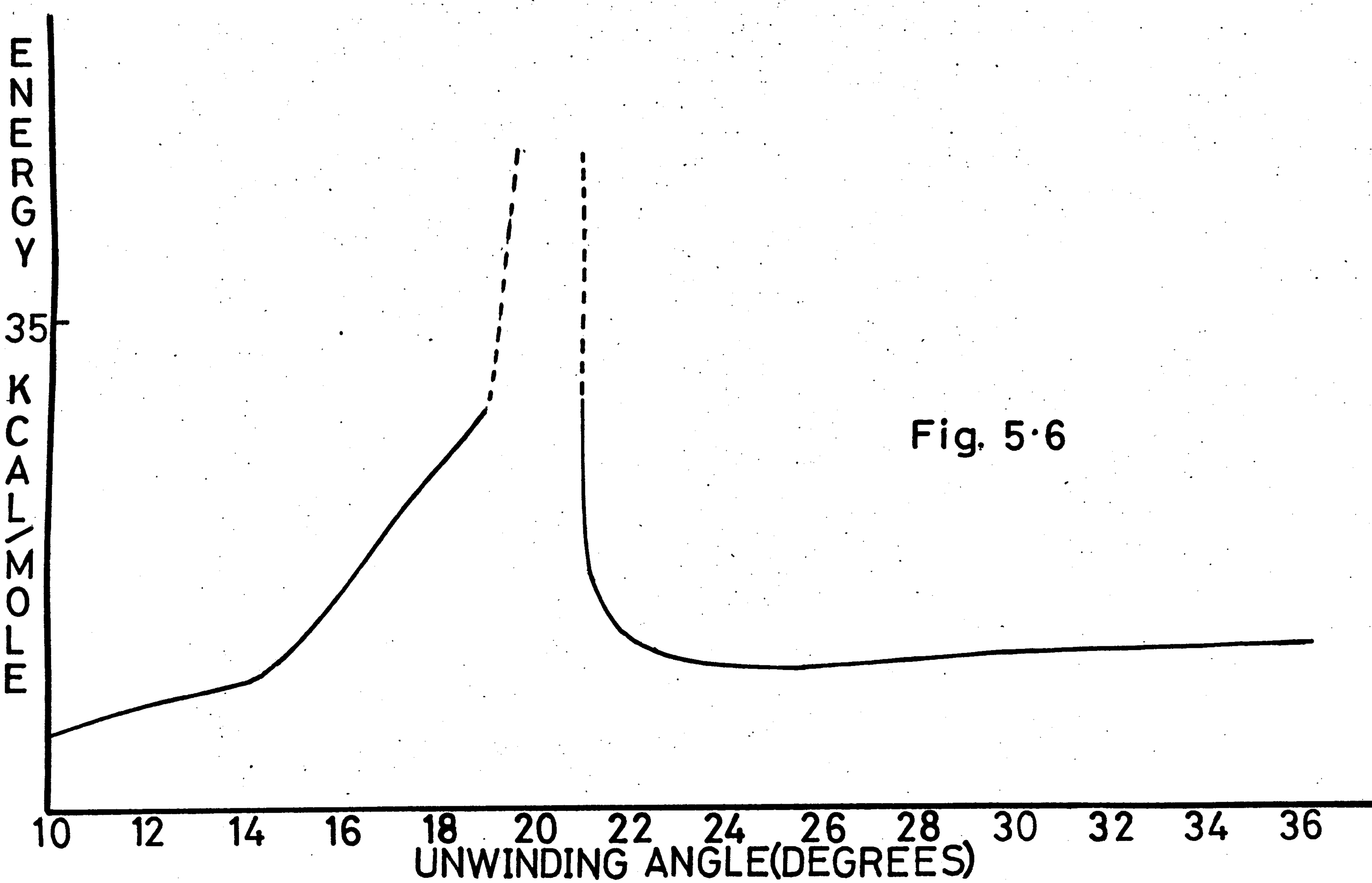

ENERGY
KCAL/MOLE
35
10
12
14
16
18
20
22
24
26
28
30
32
34
36
UNWINDING ANGLE(DEGREES)

Fig. 5·6

real effect or an artefact of the somewhat unnatural constraints applied. Unfortunately, in view of the uncertainty of the experimental determinations of unwinding angle values, and of the small range of "forbidden" values, it is not possible to determine whether values within this region are observed in practice.

The models having an unwinding of 12^{o} and 28^{o} will be used in later studies with particular drugs, and their atomic coordinates are given in tables 5.7 and 5.8. Values for the dihedral angles for these two models and for the one having an untwist of 19^{o} are given in table 5.4; while the final values of the geometrical constraints are given in table 5.5. The model incorporating a 28^{o} untwist can be seen to be superior to the others in the degree to which these constraints are satisfied. The difficult in building models having na untwist of $20^{o} \rightarrow 21^{o}$ arises because of steric hinderence in the phosphate groups, particularly the phosphate of residue 1. As the unwinding angle is increased, the torsion angles attain values which allow more favourable distances between non-bonded atom pairs.

In building the structures, a maximum of 200 non-bonded contacts were included in the refinement routine and any atom pair whose separation was less than three times the sum of their van der Waals radii was considered as a non-bonded contact. None of the models built was entirely satisfactory in this respect. The five worst contacts (those involving interatomic distances which were the smallest percentages of the van der Waals radii sum) for the models having untwisting angles of 12^{o}, 19^{o} and 28^{o} are shown in table 5.6, along with the sum of the energy of the 200 contacts found. The best of the three models is the one having 12^{o} untwisting at the intercalation site, although the worst contact distance even in this case is very short, being only 0.63 times the sum of the van der Waal's radii of the atoms involved. The total energy of all non-bonded contacts in the structure should ideally be negative, although too much

TABLE 5.4

Dihedral Angles For C2' endo Intercalation Models

In the convention of the IUPAC-IUB Commission on Biochemical Nomenclature (1970)

		12° Unwinding	19° Unwinding	28° Unwinding
ϕ	1	-84.27	-75.25	-80.35
ψ	2	-164.50	-157.74	-160.74
θ	3	-223.22	-241.09	-231.31
ζ	4	-217.60	-212.18	-215.45
σ	5	-217.00	-217.00	-217.00
ω	6	-214.32	-187.06	-162.01
ϕ	7	23.91	-23.20	-65.46
ψ	8	-255.40	-256.31	-279.40
θ	9	-268.63	-254.88	-246.64
ζ	10	-184.43	-158.07	-120.39
σ	11	-217.00	-217.00	-217.00
ω	12	-215.31	-232.73	-263.87
ϕ	13	-91.97	-88.52	-75.86
ψ	14	-81.38	-81.68	-83.01
θ	15	-147.61	-143.19	-137.58
ζ	16	-42.99	-34.85	-24.01

TABLE 5.5

Values of Geometrical Constraints For C2' endo Intercalation Models

Atom Pair	Å	12° Untwist	19° Untwist	28° Untwist
	ΔX	0.003	0.005	0.007
3O1-3DO1	ΔY	0.021	0.023	0.022
	ΔZ	0.004	0.002	0.003
	ΔX	0.003	0.003	0.009
3C'-3DC'3	ΔY	0.023	0.020	0.019
	ΔZ	0.003	0.008	0.002
	ΔX	0.003	0.014	0.002
	ΔY	0.006	0.003	<0.001
3N9-3DN9	ΔZ	0.021	0.025	0.008
	ΔX	0.011	0.031	0.002
3C'1-3DC'1	ΔY	0.002	0.005	<0.001
	ΔZ	0.029	0.047	0.011
	ΔX	0.032	0.032	0.018
2N9-2DN9	ΔY	0.006	0.028	0.003
	ΔZ	0.044	0.037	0.023
	ΔX	0.055	0.004	0.028
2C1'-2DC'1	ΔY	0.009	0.024	0.013
	ΔZ	0.070	0.034	0.038

TABLE 5.6

The Worst Non-Bonded Atomic Contacts in the given Models

28^{o} Untwist

Atoms	Distance (Å)	Energy (K cal/mole)
3H22 - 2H52	1.49	5.49
2H22 - 203	1.61	2.78
3H22 - 2C5	1.94	0.81
2H1 - 2O4	1.94	0.40
2H1 - 2P	2.43	0.07

Total Energy = -23.46 K cal/mole

19^{o} Untwist

Atoms	Distance (Å)	Energy (K cal/mole)
2H22 - 203	1.45	20.61
2H22 - 304	1.57	7.87
2H22 - 2P	2.28	0.18
2C2' - 2N9	2.46	0.15
2C2' - 2P	2.81	0.07

Total Energy = 14.9 K cal/mol

12^{o} Untwist

Atoms	Distance (Å)	Energy (K cal/mole)
3H22 - 304	1.63	4.83
2H22 - 203	1.73	2.23
2H22 - 2P	2.31	0.15
3H22 - 3P	2.32	0.15
3C2' - 3N9	2.43	0.22

Total Energy = -28.44 K cal/mole

significance should not be attached to the absolute values presented since they do not include the energy of the base pair hydrogen bonds. Moreover other factors such as the influence of surrounding water and countercations would need to be taken into account if the absolute value of the conformational energy were to have any significance. The figures presented in table 5.6 are useful only to indicate the comparative energy between the different conformations. The worst structure in this respect is that having a 19^{o} untwisting. The oxygen atoms pendant to the lower phosphorous atom tend to interfere with the hydrogen atoms of the sugar rings giving rise to two very short contacts which account for most of the unfavourably high conformational energy.

An estimate of the contribution made by the base pair hydrogen bonds can be made if it is supposed that each one has a bond energy of 4 kcal/mole. Since G:C base pairs include three hydrogen bonds and A:T base pairs two, we make the assumption that the average four base pair intercalation unit contains ten such hydrogen bonds since no assumption has been made regarding the nature of the base pairs around the intercalation site. In this case, the base pair hydrogen bonds would contribute a negative energy term equal to approximately 40 kcal/mole so that incorporating this contribution into the total energy values of table 5.6, gives final values for the conformational energy which are negative in the models having unwinding angles of 12^{o} and 28^{o}. Nevertheless, since the energy contribution from the base pairing is constant in all cases, it is the conformational energy in terms of the van der Waals interactions which is important in distinguishing between the stereochemical acceptabilities of related structures.

For comparison with the above models, a study made of the conformational possibilities of the BDNA model Arnott and Hukins (1972). A general intercalation model based upon this model for BDNA has been built by Alden and Arnott (1975). Their model intercalation unit

TABLE 5.7

Atomic Coordinates for the backbone of the C2' endo intercalation model with 12^{o} unwinding

		X (Å)	Y (Å)	Z (Å)	R (Å)	φ (°)
4DC3	1	-7.952	2.120	13.627	8.230	165.075
401	2	-7.816	3.524	13.787	8.574	155.732
402	3	-8.505	5.867	13.245	10.333	145.400
403	4	-9.449	3.943	11.939	10.239	157.350
4P	5	-8.282	4.534	12.637	9.442	151.304
404	6	-7.000	4.554	11.680	8.351	146.954
3H51	7	-7.487	6.403	10.897	9.851	139.462
3H52	8	-7.070	5.186	9.714	8.768	143.740
3C5	9	-6.833	5.584	10.679	8.825	140.743
3C4	10	-5.404	6.071	10.681	8.127	131.763
3H4	11	-5.373	6.915	11.341	8.757	127.845
305	12	-4.579	5.000	11.218	6.780	132.479
3DN9	13	-3.352	3.204	10.242	4.638	136.293
3C1	14	-3.559	4.659	10.281	5.863	127.375
3H1	15	-2.634	5.134	10.535	5.770	117.161
3C2	16	-4.029	5.146	8.924	6.536	128.057
3H21	17	-4.681	4.460	8.425	6.465	136.389
3H22	18	-3.228	5.412	8.265	6.302	120.816
3H3	19	-5.591	6.618	8.626	8.663	130.190
3C3	20	-4.807	6.400	9.323	8.004	126.909
301	21	-3.951	7.530	9.408	8.504	117.688
302	22	-1.492	7.096	9.265	7.251	101.871
303	23	-2.499	9.046	8.047	9.385	105.445
3P	24	-2.639	7.641	8.500	8.084	109.053
304	25	-3.008	6.683	7.273	7.329	114.235
2H51	26	-3.842	8.188	6.127	9.045	115.134
2H52	27	-4.079	6.572	5.508	7.735	121.828
2C5	28	-3.385	7.231	5.987	7.984	115.089
2C4	29	-2.159	7.383	5.120	7.692	106.301
2H4	30	-1.497	8.043	5.644	8.182	100.546
205	31	-1.544	6.070	4.997	6.264	104.273
2DN9	32	-1.728	4.296	3.417	4.630	111.912
2C1	33	-1.424	5.720	3.620	5.894	103.984

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
2H1	34	-0.436	5.917	3.261	5.933	94.214
2C2	35	-2.425	6.570	2.861	7.003	110.256
2H21	36	-3.415	6.161	2.852	7.044	118.995
2H22	37	-2.126	6.787	1.856	7.112	107.391
2H3	38	-3.335	8.375	3.639	9.015	111.714
2C3	39	-2.403	7.850	3.694	8.210	107.016
201	40	-1.368	8.724	3.266	8.831	98.913
202	41	-0.360	10.074	1.417	10.080	92.045
203	42	-1.953	8.207	0.889	8.436	103.383
2P	43	-0.859	8.718	1.750	8.760	95.624
204	44	0.349	7.670	1.814	7.678	87.398
1H51	45	1.732	7.883	3.335	8.071	77.611
1H52	46	1.802	9.112	2.095	9.289	78.815
1C5	47	1.658	8.069	2.284	8.237	78.390
1C4	48	2.721	7.278	1.560	7.770	69.500

TABLE 5.8

Atomic Coordinates for the backbone of the C2' endo intercalation model with 28^{o} unwinding

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
4DC3	1	-7.060	4.229	13.627	8.229	149.079
401	2	-6.544	5.541	13.792	8.575	139.742
402	3	-6.477	7.982	13.248	10.280	129.056
403	4	-8.060	6.463	12.029	10.331	141.273
4P	5	-5.727	6.648	12.652	9.458	135.338
404	6	-5.562	6.258	11.627	8.373	131.633
3H51	7	-5.499	8.173	10.854	9.851	123.933
3H52	8	-5.363	6.894	9.671	8.735	127.878
3C5	9	-5.077	7.211	10.652	8.819	125.149
3C4	10	-3.572	7.298	10.731	8.125	116.078
3H4	11	-3.350	8.076	11.435	8.743	112.532
305	12	-3.086	6.025	11.239	6.769	117.122
3DN9	13	-2.333	4.010	10.213	4.639	120.196
3C1	14	-2.149	5.464	10.322	5.872	111.472
3H1	15	-1.145	5.664	10.633	5.778	101.429
3C2	16	-2.408	6.115	8.977	6.572	111.492
3H21	17	-3.194	5.649	8.419	6.489	119.484
3H22	18	-1.535	6.185	8.361	6.373	103.935
3H3	19	-3.508	7.961	8.702	8.700	113.782
3C3	20	-2.844	7.513	9.414	8.033	110.737
301	21	-1.726	8.369	9.589	8.546	101.651
302	22	0.053	9.962	8.842	9.963	89.695
303	23	-2.102	9.793	7.567	10.017	102.116
3P	24	-1.044	9.101	8.340	9.160	96.544
304	25	-0.459	7.861	7.515	7.874	93.341
2H51	26	-2.326	7.387	6.766	7.745	107.478
2H52	27	-1.223	6.039	6.907	6.162	101.448
2C5	28	-1.310	7.074	6.648	7.194	100.491
2C4	29	-0.891	7.267	5.211	7.322	96.986
2H4	30	-0.014	7.883	5.231	7.883	90.100
205	31	-0.556	5.958	4.673	5.984	95.330
2DN9	32	-1.778	4.293	3.483	4.647	112.500
2C1	33	-1.342	5.697	3.512	5.853	103.255

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
2H1	34	-0.787	5.905	2.621	5.957	97.593
2C2	35	-2.552	6.608	3.587	7.083	111.116
2H21	36	-3.349	6.213	4.183	7.059	118.326
2H22	37	-2.937	6.887	2.627	7.487	113.098
2H3	38	-2.681	8.369	4.841	8.788	107.765
2C3	39	-1.948	7.831	4.275	8.069	103.971
201	40	-1.366	8.717	3.331	8.824	98.905
202	41	-0.674	9.398	1.025	9.423	94.104
203	42	-2.025	7.311	1.370	7.586	105.480
2P	43	-0.959	8.212	1.869	8.268	96.658
204	44	0.383	7.394	2.167	7.404	87.033
1H51	45	1.922	8.156	3.317	8.379	76.739
1H52	46	1.586	9.045	1.851	9.182	80.057
1C5	47	1.658	8.069	2.284	8.237	78.390
1C4	48	2.721	7.278	1.560	7.770	69.500

consists of a section of four base pairs and purports to allow adjacent cohelical BDNA regions. The furanose ring puckering about the intercalation site is C3' exo (5'-3') C3' endo, and the helix is unwound by 13 degrees at this point. The total unwinding over the four residues of the intercalation unit is 18 degrees. It was decided to derive the coordinates of the Alden and Arnott (1975) model in order to use them as a starting model for model building studies with specific drugs.

Since the untwist angle per residue is not known, the first stage is to derive the coordinates of the phosphate backbone using a model building system analogous to that described in section 5.2.1 (method (i)). The final coordinates are derived in the correct coordinate frame in an identical manner to that already described so that the lower sugar atoms took the coordinates of BDNA (Arnott and Hukins, 1972). The stereochemical data of bond angles and bond lengths were obtained from the Arnott and Hukins (1972) model of BDNA; and the torsion angle values from the list quoted in Alden and Arnott (1975). Pendant atom data for the furanose ring atoms of the C3' exo sugars were derived from the sugar coordinates of the BDNA model (Arnott and Hukins, 1972); while the equivalent data for the C3' endo sugar was obtained from the ADNA model by the same authors. As a preliminary check on the data and on the modelling system, the coordinate of BDNA were derived using the torsion angle data provided by Alden and Arnott (1975) for the BDNA conformation. The coordinates agreed with the published set (Arnott and Hukins, 1972).

Finally, the Alden and Arnott (1975) intercalation model was built and the coordinates are presented in table 5.9. Examination of the coordinates shows that the model has the torsion angle values which Alden and Arnott (1975) describe, but appears unsatisfactory in a number of ways and does not, in any case, correspond to the structure determined.

TABLE 5.9

Atomic coordinates for the backbone of the
Alden and Arnott (1975) intercalation model

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
4C5	1	-9.356	1.278	14.883	9.443	172.221
4C4	2	-8.760	2.482	14.193	9.105	164.182
4O5	3	-7.459	2.107	13.641	7.751	164.226
4C2	4	-9.007	2.186	11.839	9.269	166.360
4C1	5	-7.551	1.982	12.235	7.807	165.290
4N9	6	-7.009	0.648	11.861	7.039	174.717
3C3	7	-9.534	3.027	12.991	10.003	162.387
3O1	8	-9.243	4.404	12.739	10.238	154.522
3O2	9	-10.021	6.336	11.341	11.856	147.697
3O3	10	-10.160	3.998	10.443	10.918	158.520
3P	11	-9.428	4.987	11.271	10.665	152.123
3O4	12	-7.920	5.079	10.775	9.409	147.326
3H51	13	-6.857	5.836	12.381	9.005	139.597
3H52	14	-7.536	7.041	11.312	10.313	136.946
3C5	15	-7.055	6.086	11.359	9.317	139.218
3C4	16	-5.752	6.145	10.598	8.417	133.112
3O5	17	-4.411	6.696	10.783	8.018	123.374
3C2	18	-4.396	5.435	8.769	6.991	128.966
3C1	19	-3.591	6.348	9.684	7.293	119.498
3N9	20	-4.417	3.095	10.002	5.393	144.983
3C3	21	-5.825	5.832	9.102	8.243	134.965
2O1	22	-6.188	6.979	8.328	9.327	131.562
2O2	23	-6.576	5.468	6.366	8.552	140.255
2O3	24	-7.138	7.907	6.200	10.653	132.075
2P	25	-6.246	6.865	6.743	9.281	132.300
2O4	26	-4.738	7.177	6.347	8.600	123.435
2H51	27	-4.971	8.286	4.616	9.663	120.962
2H52	28	-4.689	6.580	4.366	8.080	125.478
2C5	29	-4.423	7.432	4.955	8.649	120.755
2C4	30	-2.944	7.698	4.807	8.242	110.925
2O5	31	-2.219	6.453	5.053	6.823	108.975
2C2	32	-2.238	6.820	2.705	7.178	108.164
2C1	33	-1.729	5.932	3.832	6.178	106.252

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
2N9	34	-2.178	4.517	3.730	5.015	115.747
2C3	35	-2.468	8.141	3.422	8.507	106.868
101	36	-1.237	8.867	3.478	8.953	97.945
102	37	0.376	10.435	2.370	10.442	87.936
103	38	-1.530	9.460	1.062	9.583	99.184
1P	39	-0.512	9.282	2.126	9.296	93.157
104	40	0.353	7.984	1.818	7.991	87.472
1H51	41	0.987	6.768	3.368	6.840	81.699
1H52	42	1.605	8.402	3.412	8.554	79.188
1C5	43	1.363	7.577	2.776	7.698	79.805
1C4	44	2.608	7.126	2.046	7.588	69.901
1C3	45	2.818	7.700	0.647	8.199	69.898

The main objections to the model, on the basis of the coordinates derived here are:-

(i) the radial and axial constraints between atoms C1', C2' and O5' of the lower residue and the equivalent atoms in the 4th residue, designed to ensure that the ends of the intercalation unit are cohelical, are clearly not satisfied. Therefore, the required property that adjacent co-axial BDNA regions could follow on from the intercalation model is not attained.

(ii) the two chains of the intercalation model are reported to be dyad related, but examination of the coordinates presented in table 5.9 shows that the dyad related chain, derived from the 5'-3' chain built by the programme, does not form base pairs having standard Watson-Crick hydrogen bonding geometry. In consequence, the base pairs do not have the same disposition to the helix axis as is found in BDNA (Arnott and Hukins, 1972).

(iii) the structure does not possess the twist or rise per residue specified by Alden and Arnott (1975).

In view of the unsatisfactory nature of the above model, an equivalent model building study has been initiated by the author. As in the Alden and Arnott study, the intercalation unit was built so as to have the BDNA model of Arnott and Hukins (1972) as the native conformation, and was, hence, constrained to allow adjacent cohelical BDNA regions in accord with the requirements of an intercalation unit. A series of models were built having C3' exo (BDNA) furanose ring puckering throughout. The mixed sugar puckering (C3' exo (5'-3') C3' endo) about the intercalation site of the Alden and Arnott (1975) model has been reported by other authors (Sobell and Jain, 19720; Tsai et. al. 1975) and, hence, a series of models were built with this combination of sugar puckers. Stereochemical data (bond angles and bond lengths) were obtained as before from the ADNA and BDNA models of

TABLE 5.10

Atomic coordinates of C3' exo/C3' endo intercalation unit

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
4C5	1	-7.303	2.439	16.291	7.699	161.528
4C4	2	-6.677	3.607	15.563	7.589	151.621
4O5	3	-5.258	3.321	15.353	6.219	147.723
4C3	4	-7.213	3.900	14.163	8.200	151.604
4C2	5	-6.375	2.987	13.282	7.041	154.894
3O1	6	-6.997	5.254	13.775	8.750	143.099
H4	7	-6.840	4.418	16.253	8.142	147.140
H3	8	-8.239	3.797	14.123	9.072	155.257
H12	9	-6.375	3.295	12.311	7.176	152.663
H22	10	-6.851	1.996	13.378	7.136	163.755
H1	11	-4.421	3.747	13.510	5.795	139.722
4C1	12	-5.029	3.009	13.992	5.860	149.106
4CN1	13	-4.302	1.711	13.941	4.630	158.309
4CC2	14	-2.918	1.746	13.792	3.400	149.108
4CC6	15	-4.962	0.530	14.038	4.991	173.904
4CC4	16	-2.880	-0.596	13.838	2.941	191.701
4CC5	17	-4.301	-0.658	13.987	4.351	188.702
4CN3	18	-2.238	0.575	13.740	2.310	165.600
D7	19	-0.000	0.000	13.550	0.000	170.328
4DC3	20	-7.209	3.909	14.168	8.201	151.533
3DO1	21	-6.998	5.253	13.727	8.750	143.108
3O2	22	-8.320	6.867	12.335	10.788	140.465
3O3	23	-8.025	4.496	11.571	9.199	150.742
3P	24	-7.452	5.676	12.263	9.367	142.706
3O4	25	-6.059	6.055	11.596	8.566	135.022
3H15	26	-4.650	5.473	12.995	7.182	130.350
3H25	27	-5.206	7.122	13.150	8.822	126.164
3C5	28	-4.932	6.354	12.458	8.044	127.817
3C4	29	-3.762	6.827	11.624	7.795	118.860
3O5	30	-2.632	5.926	11.849	6.484	113.953
3C3	31	-4.152	6.614	10.198	7.809	122.117
3C2	32	-2.853	6.233	9.509	6.855	114.597
2O1	33	-4.799	7.746	9.612	9.112	121.782

		X (Å)	Y (Å)	Z (Å)	R (Å)	φ (°)
H4	34	-3.418	7.837	11.727	8.550	113.564
H3	35	-4.873	5.643	10.281	7.456	130.814
H12	36	-2.444	7.125	9.248	7.533	108.934
H22	37	-3.093	5.611	8.661	6.407	118.860
H1	38	-1.093	5.663	10.511	5.767	100.928
3C1	39	-2.138	5.456	10.609	5.860	111.400
3GN9	40	-2.357	3.985	10.561	4.630	120.600
3GC8	41	-3.536	3.275	10.642	4.820	137.199
3GN7	42	-3.384	1.978	10.563	3.920	149.699
3GC4	43	-1.370	3.035	10.420	3.330	114.301
3GC5	44	-2.007	1.807	10.421	2.700	138.000
3GC6	45	-1.242	0.625	10.291	1.390	153.297
3GC2	46	0.678	2.177	10.170	2.280	72.700
3GN3	47	-0.040	3.290	10.300	3.290	90.699
3GN1	48	0.130	0.911	10.170	0.920	81.900
D6	49	-0.000	-0.000	10.170	0.000	261.849
3DC3	50	-4.152	6.613	10.189	7.808	122.127
2D01	51	-4.761	7.747	9.565	9.093	121.575
202	52	-5.551	9.037	7.562	10.605	121.560
203	53	-5.868	6.550	7.661	8.794	131.860
2P	54	-5.029	7.726	7.997	9.219	123.062
204	55	-3.561	7.506	7.427	8.308	115.378
2H15	56	-4.261	6.827	5.602	8.048	121.968
2H25	57	-3.252	5.696	6.473	6.559	119.728
2C5	58	-3.391	6.726	6.216	7.532	116.755
2C4	59	-2.181	7.218	5.455	7.541	106.810
205	60	-1.288	6.084	5.211	6.219	101.955
203	61	-2.248	7.889	4.027	8.203	105.907
2C2	62	-2.321	6.652	3.146	7.045	109.236
101	63	-1.131	8.680	3.633	8.754	97.423
H4	64	-1.663	7.942	6.146	8.114	101.826
H3	65	-3.089	8.519	4.020	9.062	109.930
H12	66	-2.111	6.880	2.183	7.196	107.058
H22	67	-3.366	6.274	3.238	7.120	118.214
H1	68	-0.408	5.807	3.370	5.821	94.020
2C1	69	-1.360	5.702	3.850	5.862	103.412

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
2CN1	70	-1.781	4.276	3.800	4.632	112.612
2CC2	71	-0.789	3.309	3.650	3.402	103.415
2CC6	72	-3.087	3.923	3.900	4.992	128.199
2CC4	73	-2.438	1.645	3.700	2.941	145.992
2CC5	74	-3.475	2.619	3.850	4.352	142.993
2CN3	75	-1.152	2.003	3.600	2.311	119.900
D5	76	0.000	-0.000	3.410	0.000	287.569
2DC3	77	-2.248	7.889	4.026	8.203	105.907
1DO1	78	-1.123	8.676	3.626	8.748	97.375
1O2	79	-0.194	10.179	1.846	10.181	91.089
1O3	80	-2.029	8.561	1.292	8.798	103.331
1P	81	-0.806	8.857	2.078	8.894	95.200
1O4	82	0.267	7.709	1.838	7.714	88.013
1H15	83	1.115	6.851	3.520	8.941	80.752
1H25	84	1.554	8.519	3.242	8.660	79.661
1C5	85	1.367	7.575	2.774	7.698	79.768
1C4	86	2.611	7.126	2.040	7.590	69.878
1O5	87	2.532	5.681	1.829	6.220	65.975
1C3	88	2.818	7.701	0.641	8.200	69.899
1C2	89	2.034	6.741	-0.239	7.041	73.212
001	90	4.188	7.683	0.248	8.750	61.405
H4	91	3.392	7.403	2.728	8.143	65.385
H3	92	2.568	8.701	0.602	9.072	73.555
H12	93	2.335	6.786	-1.211	7.176	71.008
H22	94	0.984	7.068	-0.139	7.136	82.070
H1	95	3.068	4.917	-0.016	5.795	58.033
1C1	96	2.252	5.410	0.469	5.860	67.400
1GN9	97	1.073	4.504	0.421	4.630	76.600
2GC8	98	-0.269	4.812	0.502	4.820	93.199
1GN7	99	-1.061	3.774	0.423	3.920	105.699
1GC4	100	1.123	3.135	0.280	3.330	70.301
1GC5	101	-0.188	2.693	0.281	2.700	94.000
1GC6	102	-0.459	1.312	0.151	1.390	109.297
1GC2	103	2.000	1.095	0.030	2.280	28.700
1GN3	104	2.256	2.394	0.160	3.290	46.699
1GN1	105	0.726	0.565	0.030	0.920	37.900
D4	106	-0.000	-0.000	0.030	0.000	251.565
D3	107	0.000	0.000	0.000	0.000	90.000

TABLE 5.11

The Worst Non-Bonded Atomic Contacts in the Given Models

Atom Pairs	Distance (Å)	Energy (K cal/mole)
304-H3	1.82	1.10
H12-103	1.90	0.59
H12-204	2.17	0.48
3C2-204	2.54	0.40
H22-204	2.31	0.22

Total Energy = -36.48 K cal/mole

C'3 exo (5'-3') C'3 endo Intercalation Model

Atom Pairs	Distance (Å)	Energy (K cal/mole)
2H12-103	1.66	3.77
4H1-3H15	1.55	1.65
2O4-2H4	2.17	0.48
3H12-204	2.23	0.33
2H12-1P	2.31	0.22

Total Energy = -33.01 K cal/mole

C'3 exo/C'3 exo Intercalation Model

Atom Pairs	Distance (Å)	Energy (K cal/mole)
3H12-203	1.85	0.85
2H12-103	1.92	0.51
4H12-303	1.93	0.48
4H1-3H15	1.85	0.15
4H12-3P	2.39	0.11

Total Energy = -36.82 K cal/mole

BDNA

TABLE 5.12

Final Values of Geometrical Constraints For The Given Models

	(Å)	C'3exo/C'-3endo	C'3exo/C'-3exo	BDNA
	ΔX	0.004	0.532	0.024
4C3-4DC3	ΔY	0.009	0.025	0.017
	ΔZ	0.005	0.311	0.051
	ΔX	<0.001	0.431	0.018
3O1-3DO1	ΔY	0.001	0.013	0.016
	ΔZ	0.048	0.255	0.037
	ΔX	0.001	0.074	0.020
3C3-3DC3	ΔY	0.001	0.060	0.001
	ΔZ	0.009	0.037	0.047
	ΔX	0.038	0.075	0.017
2O1-2DO1	ΔY	0.001	0.138	0.005
	ΔZ	0.047	0.276	0.037
	ΔX	<0.001	0.007	0.008
2C3-2DC3	ΔY	<0.001	0.032	0.014
	ΔZ	0.002	0.069	0.008
	ΔX	0.008	0.075	0.010
1O1-1DO1	ΔY	0.004	0.058	0.010
	ΔZ	0.007	0.078	0.006

(Arnott and Hukins (1972). A skeletal model of the structure was built in each case from which a set of dihedral angles for the starting model could be obtained. Once the data structure had been set up, it was tested by reconstructing the BDNA model from which the stereochemical data was obtained, and was able to reproduce the coordinates accurately.

The subsequent intercalation models were then built by increasing the axial separation between residues constituting the intercalation site, to 6.76Å (the value used for all models) and substituting the appropriate set of dihedral angles. Although alternate binding models could not be built, intercalation units where N=4 were built using this system. Alden and Arnott (1975) report that models having the mixed sugar pucker have a lower value for the conformational energy than do equivalent models having C3' exo sugar pucker throughout. In the present model building study it was not found to be possible to build a stereochemically reasonable model having C3' exo sugar pucker throughout and incorporating features which allowed the constraints imposed to be satisfied. The best model had an untwist of 16^{o} at the intercalation site, and a total untwist over the whole unit of 19^{o}. Some of the final values for the geometrical constraints designed to ensure that the phosphate groups join up correctly to adjacent residues, are unacceptably large (table 5.12), however, and, hence, it cannot be considered to be a satisfactory model.

It was possible to refine a model having the mixed sugar pucker which was much more stereochemically acceptable. Coordinates for the model which has an unwinding of 26^{o} at the intercalation site and 28^{o} overall, are given in table 5.10. A list of the worst short contacts, for this model, the C3' exo intercalation model and for BDNA are given in table 5.11, while the values of the geometrical constraints for the same models are listed in table 5.12. It can be seen that the

TABLE 5.13

Dihedral Angles For Intercalation Models In The Convention Of The IUPAC-IUB Commission on Biochemical Nomenclature (1970)

	C3'exo/C3'endo	C'3exo/C3'exo	BDNA
ϕ	-116.81	-115.40	-95.6
ψ	-19.47	-7.96	-46.1
θ	-174.04	-167.67	-146.5
ξ	-8.2	-1.2	36.4
σ	84.3	156.5	156.5
ω	-190.6	-188.1	154.7
ϕ	-58.67	-100.18	-
ψ	-213.78	-219.15	-
θ	-210.39	-229.22	-
ξ	-115.2	-127.6	-
σ	156.5	156.5	-
ω	-206.4	-217.9	-
ϕ	-95.26	-18.60	-
ψ	-47.27	-312.80	-
θ	-145.54	-180.10	-

geometrical constraints are better satisfied in the model incorporating mixed sugar pucker than is the case for the C3' exo model. The lists of dihedral angles for the models are presented in table 5.13.

5.4.3 Models Built With The Drug Ethidium

A number of models were built involving the drug ethidium, which were based on one of the intercalation units described in previous sections. In the work described here, ethidium was built into intercalation units based on the C2' endo B conformation having values of the unwinding angle of 12^{o} and 28^{o}; and into the intercalation unit having mixed sugar pucker and based on the Arnott and Hukins (1972), BDNA model. In the latter case, the DNA conformation was refined concommitantly with the conformation of the ethidium molecule and its position.

In the C2' endo model having an unwinding angle of 12^{o}, the possibility exists for the formation of hydrogen bonds between the phosphate oxygens and the ethidium methyl groups on the chromophore. Models were built, therefore, in which this hydrogen bonding scheme has been proposed by Fuller and Waring (1964). In models having an untwist angle of 28^{o} at the intercalation site it is not possible to have such a hydrogen bond since the phosphate oxygens protrude out from the helix and not into the intercalation gap.

In the model of Fuller and Waring (1964) the ethyl and phenyl groups of ethidium protrude from the intercalation gap into the large groove; whereas in the alternative model of Tsai et. al. (1975) the groups lie in the small groove. In building models based on the C2' endo DNA conformation each of the possible orientations were considered in all cases. The values of the conformational energy of the complexes are given in table 5.14 to 5.16, and diagrams showing the drug chromophore and adjacent base pairs in helix axis projection are presented in figures 5.7 to 5.14.

TABLE 5.14

The Worst Non-Bonded Atomic Contacts For The Given Models

Atom Pair	Distance (Å)	Energy (K cal/mole)
H37-306	1.40	32.32
H33-2C6	1.50	22.14
H41-306	1.46	19.58
H33-2C5	1.52	18.92
N24-3H22	1.50	18.67

Total Energy = 249.24 (277.68) K cal/mole

12^{o} Unwinding - Groups In Large Groove - No H Bonding

Atom Pair	Distance (Å)	Energy (K cal/mole)
H37-306	1.41	30.05
H33-2C6	1.50	22.33
H33-2C5	1.50	21.96
H33-2C6	1.52	18.56
H41-306	1.46	18.52

Total Energy = 243.09 (271.53) K cal/mole

12^{o} Unwinding - Groups In Large Groove - H Bonds

TABLE 5.15

The Worst Non-Bonded Atomic Contacts For The Given Models

Atom Pair	Distance (Å)	Energy (K cal/mole)
H34-202	1.77	1.65
C22-205	1.86	0.84
H36-1H51	1.64	0.81
C17-202	2.27	0.44
H32-205	1.95	0.40

Total Energy = -23.54 (4.90) K cal/mole

12^{o} Unwinding - Groups in Small Groove - No H Bonds

Atom Pair	Distance (Å)	Energy (K cal/mole)
H46-205	1.47	16.76
N24-204	1.74	12.96
H43-3H22	1.37	8.60
H44-204	1.58	6.95
H46-204	1.65	3.88

Total Energy = 32.24 (60.68) K cal/mole

12^{o} Unwinding - Groups in Small Groove - H Bonds

TABLE 5.16

The Worst Non-Bonded Atomic Contacts For The Given Models

Atom Pairs	Distance (Å)	Energy (K cal/mole)
H37-3O6	1.38	38.43
H33-2C6	1.50	20.79
H36-3TCH3	1.74	8.16
H46-3H22	1.31	15.12
H46-2C6	1.54	15.52
N24-3H22	1.49	19.11

Total Energy = 250.86 (274.32) K cal/mole

28^{o} Untwist - Groups in Large Groove

Atom Pairs	Distance (Å)	Energy (K cal/mole)
H44-3H22	1.65	0.77
H46-2O4	1.86	0.77
H34-2O2	1.88	0.66
H41-$2NH_2$	2.17	0.48
C22-$2NH_2$	2.54	0.40

Total Energy = -19.32 (4.14) K cal/mole

28^{o} Untwist - Groups in Small Groove

The values of the total energies in tables 5.14 to 5.16 includes all van der Waals interactions; the hydrogen bonding interactions between the ethidium and the DNA where appropriate; and the energies of the base pair hydrogen bonds. It is useful to compare the energies of the interactions between the DNA and the drug molecule only (i.e. to exclude from the sum the energies of any interactions where both atoms of the pair are in the DNA portion of the complex) and the approriate summed energy values are indicated in table 5.14 to 5.16 along with details of the worst contacts between non-bonded atoms.

A rather surprising feature of the models is the fact that the best models obtained are those in which the ethyl and phenyl groups of the ethidium molecule project into the narrow groove of the DNA duplex where the relatively close proximity of the two phosphate chains appears likely to lead to short non-bonded atomic contacts. The difficulty in building a model in which the groups project into the large groove is due to the fact that the two outer rings of the ethidium chromophore approach too closely to the atoms of the phosphate chain when the ethidium molecule is in a position to achieve good overlap with the base atoms. In the ethidium intercalation model of Fuller and Waring (1964) the drug molecule is oriented such that the groups project into the large groove so that it is not possible to conclude, on the basis of the above results, that it is impossible to build such a model for ethidium intercalation. In the model presented by Tsai et. al. (1975), however, the groups are visualized as projecting into the small groove as predicted here.

The most energetically favourable model is that in which the DNA is unwound by 12^{o} and in which no hydrogen bonds are formed between the ethidium and the phosphate groups. An increasing amount of evidence (Pulleybank and Morgans, 1975; Tsai et. al., 1975; Wang, 1976) has been published, however, which suggests that the unwinding produced

TABLE 5.17

Ethidium intercalation model with C2' endo intercalation unit

Ethyl and Phenyl groups in small groove, 28^{o} unwinding

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
H42	1	2.847	4.537	5.250	5.356	57.897
H40	2	1.822	3.558	5.007	3.998	62.887
H41	3	3.410	3.115	4.873	4.619	42.408
C22	4	2.632	3.647	5.305	4.498	54.181
H38	5	1.824	4.152	7.406	4.535	66.286
H39	6	3.389	3.621	8.189	4.960	46.890
C21	7	2.577	3.545	6.785	4.383	53.982
H33	8	4.520	0.633	4.474	4.564	7.975
C16	9	5.417	1.216	5.378	5.552	12.648
C17	10	6.810	1.303	5.252	6.934	10.835
C18	11	7.488	1.758	6.398	7.691	13.216
C19	12	6.741	2.155	7.691	7.078	17.730
C20	13	5.240	1.920	7.795	5.580	20.123
H34	14	7.192	1.124	4.510	7.279	8.879
H35	15	8.421	1.673	6.136	8.586	11.239
H36	16	7.218	2.204	8.561	7.547	16.981
H37	17	4.861	2.287	8.570	5.372	25.193
C15	18	4.650	1.581	6.684	4.912	18.779
C9	19	3.172	1.314	6.818	3.433	22.494
N10	20	2.173	2.242	6.907	3.123	45.898
C1	21	-0.265	3.004	7.044	3.016	95.034
C2	22	-1.636	2.771	7.124	3.218	120.559
C4	23	-1.049	0.570	7.071	1.194	151.499
C5	24	1.182	-1.531	6.920	1.934	307.679
C7	25	3.619	-2.120	6.905	4.195	329.640
C8	26	3.909	-0.895	6.855	4.010	347.106
C11	27	0.761	2.005	6.992	2.144	69.208
C12	28	0.392	0.762	6.991	0.857	62.802
C13	29	1.431	-0.241	6.936	1.451	350.439
C14	30	2.846	0.049	6.865	2.846	0.990
N23	31	-2.645	3.717	7.169	4.562	125.436
N24	32	4.663	-3.061	6.909	5.578	326.721
BR25	33	-2.302	7.250	8.518	7.607	107.614

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
O26	34	7.343	-1.664	7.666	7.529	347.229
H27	35	0.015	3.752	6.989	3.752	89.764
H28	36	-2.878	1.354	7.141	3.180	154.807
H29	37	-1.207	-0.191	7.047	1.222	189.005
H30	38	0.268	-1.761	6.952	1.782	278.638
H31	39	2.040	-3.192	6.991	3.789	302.579
H32	40	4.823	-0.748	6.774	4.880	351.185
H43	41	-3.356	3.610	7.253	4.929	132.918
H44	42	-2.323	4.410	7.319	4.985	117.775
H45	43	4.282	-3.468	7.550	5.511	320.996
H46	44	5.408	-2.671	7.208	6.031	333.719
H47	45	8.193	-1.732	7.511	8.374	348.063
H48	46	7.877	-0.838	6.624	7.922	353.924
C6	47	2.236	-2.421	6.913	3.296	312.721
C3	48	-2.019	1.533	7.110	2.535	142.792

by ethidium is between 25^{o} and 30^{o}, hence, the structure incorporating an untwist of 28^{o} degrees is probably a better model of the ethidium binding interaction. This will be discussed more fully in relation to other ethidium models in section 5.6.1. Coordinates for the ethidium molecule in the model are given in table 5.17, and diagram showing a projection of the structure along the helix axis is shown in figure 5.10.

In building ethidium models based upon the C3' exo/C3' endo intercalation unit only models in which the ethyl and phenyl groups of ethidium projected into the small groove of the DNA were considered. Of the models built, two will be described here. The first model (model A) was superior to any other model built in terms of the energy of non-bonded interatomic contacts. It can be seen from table 5.18 that the shortest distance between non-bonded atom pairs in the model are larger than those observed in the models described previously. Nevertheless, the shortest non-bonded contact which does not involve a hydrogen atom is 2.3Å and represents a distance of only 0.72 times the optimum distance of separation. This is rather shorter than is normally observed in molecular structures: in BDNA the shortest non-bonded contact not involving hydrogen is 2.87Å and represents 0.82 times the optimum separation of the constituent atoms. Atomic coordinates for the ethidium molecule in the model are presented in table 5.20, and a helix axis projection in figure 5.13. The worst feature of model A is that the overlap between the drug chromophore and the adjacent base pairs is very poor indeed.

In model B, the overlap is considerably better and, in this respect, is the best model built in the present study. A helix axis projection of the structure is presented in figure 5.14, and a list of the worst non-bonded interatomic contacts in table 5.19. Model B is worse than model A in terms of the energy of van der Waals interactions. Coordinates for the ethidium molecule in the model are

TABLE 5.18

Ethidium C3' exo (5'-3') C3' endo Intercalation Model Groups in Small Groove - Model A

Atom Pairs	Distance (Å)	Energy (K cal/mole)
C20-2CC2	2.30	0.66
H46-2CC5	2.03	0.44
H33-3GC2	2.03	0.44
H33-3GN1	2.04	0.29
2CN3-H40	2.09	0.26

Total Energy = -34.88 (1.61) K cal/mole

TABLE 5.19

Ethidium C3' exo/C3' endo Intercalation Model Groups in Small Groove Model B

Atom Pairs	Distance (Å)	Energy (K cal/mole)
H46-2H25	1.47	3.48
H45-2CC6	1.81	2.12
H33-3GN3	1.81	1.57
H44-2H25	1.65	0.73
N24-2H25	1.87	0.99

Total Energy = -30.89 (5.59) K cal/mole

The Worst Non-Bonded Atomic Contacts For The Given Models

TABLE 5.20

Ethidium intercalation model with C3' exo/C3' endo intercalation unit

Model A

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
H42	1	0.721	3.279	5.759	3.357	77.596
H40	2	-0.424	2.443	5.516	2.479	99.836
H41	3	1.093	1.795	5.378	2.102	58.671
C22	4	0.392	2.424	5.812	2.455	80.815
H38	5	-0.341	3.024	7.916	3.043	96.441
H39	6	1.141	2.292	7.695	2.561	63.540
C21	7	0.325	2.325	7.292	2.347	82.034
H33	8	1.867	-0.809	4.969	2.035	336.569
C16	9	2.834	-0.352	5.874	2.856	352.918
C17	10	4.226	-0.447	5.746	4.249	353.959
C18	11	4.959	-0.089	6.892	4.959	358.977
C19	12	4.272	0.399	8.188	4.291	5.333
C20	13	2.753	0.362	8.293	2.776	7.487
H34	14	4.580	-0.673	5.003	4.629	351.638
H35	15	5.875	-0.294	6.629	5.880	357.131
H36	16	4.752	0.382	9.057	4.767	4.594
H37	17	2.426	0.773	9.070	2.546	17.665
C15	18	2.123	0.107	7.182	2.125	2.880
C9	19	0.622	0.035	7.317	0.623	3.204
N10	20	-0.246	1.086	7.411	1.114	102.753
C1	21	-2.563	2.160	7.554	3.352	139.867
C2	22	-3.953	2.109	7.635	4.480	151.917
C4	23	-3.660	-0.150	7.575	3.663	182.352
C5	24	-1.723	-2.525	7.414	3.057	235.693
C7	25	0.616	-3.428	7.393	3.483	280.187
C8	26	1.064	-2.252	7.346	2.490	295.288
C11	27	-1.677	1.035	7.497	1.971	148.305
C12	28	-2.206	-0.148	7.493	2.211	183.848
C13	29	-1.308	-1.279	7.434	1.829	224.358
C14	30	0.134	-1.176	7.362	1.184	276.475
N23	31	-4.829	3.179	7.684	5.781	146.645
N24	32	1.528	-4.498	7.393	4.750	288.760
H27	33	-2.187	2.866	7.500	3.605	127.349

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
H28	34	-5.369	0.866	7.650	5.439	170.835
H29	35	-3.916	-0.884	7.549	4.014	192.722
H30	36	-2.660	-2.634	7.446	3.743	224.716
H31	37	-1.090	-4.285	7.479	4.421	255.723
H32	38	1.989	-2.225	7.265	2.985	311.786
H43	39	-5.548	3.165	7.769	6.388	150.295
H44	40	-4.418	3.823	7.832	5.843	139.130
H45	41	1.097	-4.854	8.034	4.976	282.738
H46	42	2.317	-4.209	7.692	4.805	298.832
H47	43	5.202	-3.645	7.994	6.352	324.979
H48	44	5.005	-2.715	7.110	5.694	331.521
C6	45	-0.795	-3.545	7.403	3.633	257.359
C3	46	-4.495	0.932	7.618	4.590	168.288

TABLE 5.21

Ethidium intercalation model with C3' exo/C3' endo intercalation unit - Model B

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
H42	1	2.170	4.329	5.586	4.842	63.376
H40	2	1.076	3.428	5.343	3.593	72.579
H41	3	2.627	2.869	5.207	3.890	47.521
C22	4	1.891	3.456	5.640	3.940	61.322
H38	5	1.123	4.016	7.742	4.170	74.379
H39	6	2.645	3.371	7.524	4.285	51.882
C21	7	1.829	3.356	7.120	3.822	61.408
H33	8	3.550	0.313	4.802	3.564	5.035
C16	9	4.489	0.826	5.707	4.564	10.427
C17	10	5.884	0.811	5.580	5.940	7.851
C18	11	6.595	1.213	6.726	6.705	10.425
C19	12	5.880	1.662	8.021	6.110	15.782
C20	13	4.365	1.537	8.125	4.628	19.402
H34	14	6.251	0.605	4.838	6.280	5.529
H35	15	7.519	1.060	6.464	7.593	8.026
H36	16	6.359	1.674	8.890	6.575	14.746
H37	17	4.015	1.930	8.902	4.454	25.671
C15	18	3.751	1.245	7.014	3.952	18.359
C9	19	2.258	1.087	7.148	2.505	25.704
N10	20	1.330	2.085	7.240	2.474	57.478
C1	21	-1.045	3.025	7.381	3.201	109.054
C2	22	-2.430	2.894	7.461	3.779	130.019
C4	23	-2.007	0.655	7.404	2.111	161.921
C5	24	0.064	-1.604	7.248	1.605	272.286
C7	25	2.451	-2.371	7.229	3.410	315.951
C8	26	2.830	-1.170	7.181	3.063	337.533
C11	27	-0.095	1.953	7.326	1.955	92.798
C12	28	-0.556	0.741	7.323	0.926	126.877
C13	29	0.407	-0.336	7.266	0.527	320.445
C14	30	1.839	-0.151	7.194	1.846	355.319
N23	31	-3.366	3.911	7.508	5.160	130.718
N24	32	3.423	-3.386	7.231	4.815	315.310
H27	33	-0.710	3.751	7.326	3.818	100.725

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
H28	34	-3.772	1.571	7.477	4.086	157.385
H29	35	-2.220	-0.092	7.379	2.222	182.377
H30	36	-0.865	-1.767	7.279	1.967	243.906
H31	37	0.797	-3.324	7.315	3.418	283.483
H32	38	3.752	-1.091	7.101	3.908	343.788
H43	39	-4.084	3.856	7.593	5.617	136.638
H44	40	-2.994	4.579	7.659	5.470	123.179
H45	41	3.013	-3.766	7.872	4.823	308.669
H46	42	4.194	-3.052	7.530	5.187	323.956
H47	43	7.042	-2.322	7.833	7.415	341.748
H48	44	6.792	-1.406	6.947	6.936	348.304
C6	45	1.049	-2.569	7.239	2.775	292.210
C3	46	-2.903	1.687	7.445	3.358	149.835

given in table 5.21.

It is difficult to assess the two models quantitatively since it is difficult to calculate the energy of van der Waals contacts and of hydrophobic interactions between the drug molecule and the bases, on an absolute scale. The fact that hydrophobic interactions are of some importance is illustrated by the fact that many intercalating drugs come out of the DNA at low humidity values. On the whole it is felt that model B is a better structure than model A since in the latter the overlap is very poor indeed.

In both the above models, the DNA conformation had been allowed to vary during the refinement so that the atomic coordinates are not identical to those presented in table 5.10, however in neither model were the DNA coordinates significantly different from those given earlier.

5.4.4. Models Involving The Drug Daunomycin

The unwinding angle produced by daunomycin has been estimated as being 0.44 times the value for ethidium on the basis of binding studies involving closed circular DNA, Waring (1970). In the present work the unwinding angle due to ethidium will be taken as 26^{o}, so that the value ascribed to daunomycin is 11.4^{o}. Hence, the daunomycin models have been built using the C2' endo intercalation unit having a 12^{o} untwist. No hydrogen bonds were formed in any of the models considered, so that the complex would be stabilised by van der Waal's interactions. As in the case of ethidium, models were considered in which the side groups of daunomycin projected either into the small groove or into the large groove: the best model obtained having the group in the small groove.

A list of the worst short contacts between the drug atoms and those of the DNA is given in table 5.22 along with the total energy of the structure calculated as before. There are in fact a number of unacceptably short contacts and the model cannot be considered satisfactory from this point of view. Further refinements were not attempted since the calculated transform for the model suggests that it is fundamentally incorrect (see section 5.5.3). Coordinates for the daunomycin molecule in the model are given in table 5.23, while the drug chromophore and adjacent base pairs are given in helix axis projection in Fig. 5.15.

5.5 Results of Fourier Transform Calculations

The cylindrically averaged molecular transform was calculated for all models which included a specific drug. Normally, the calculation was performed for a number of values of the P/D ratio.

5.5.1 The General Nature of the Transforms

The size of the average repeat unit derived is large in all cases, particularly for large values of the P/D ratio. This, combined with the fact that the helix is often non-integral, leads to very large values for the helical pitch, often as large as several hundred angstrom units. The theoretical layer line spacing is correspondingly small, therefore. However, the nature of the repeat unit in practice ensures that the layer lines upon which significant intensity is observed are much more widely spaced than this. If we consider the case of an intercalation unit having a total untwist angle of 12 degrees and the case of P/D = 20, then the repeat unit is such as to give a rotation and translation per residue of 336^{o} and 37.18Å respectively. This gives a helix with 30 units in 29 turns and a pitch of 1,128Å (30 x 37.18). When the calculation of the transform is performed,

TABLE 5.22

The Worst Non-Bonded Atomic Contacts In the Above Model

Daunomycin Model 12^{o} Unwinding Groups in Small Groove

Atom Pair	Distance (Å)	Energy (K cal/mole)
C3-3C3	1.75	14.53
C11-3C3	1.58	11.20
C2-3C3	1.81	10.30
C24-3H3	1.62	8.31
O26-3H3	1.60	6.19

Total Energy = 67.89 (91.35) K cal/mole

TABLE 5.23

Atomic coordinates for daunomycin in the intercalation model described in this chapter

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
C2	1	5.303	-1.783	6.748	5.595	341.417
C1	2	3.868	-1.715	6.899	4.231	336.086
C3	3	6.153	-0.739	6.167	6.197	353.153
C4	4	5.388	0.607	6.198	5.423	6.423
C5	5	4.095	0.762	6.622	4.166	10.539
C6	6	3.434	2.132	6.611	4.042	31.831
C7	7	2.024	2.160	6.756	2.960	46.857
C9	8	-0.214	3.462	7.025	3.468	93.535
C10	9	-0.898	4.669	6.704	4.754	100.887
C11	10	-2.326	4.789	7.124	5.324	115.911
C13	11	-2.336	2.158	7.516	3.180	137.273
C14	12	-0.869	2.141	7.154	2.310	112.085
C15	13	-0.156	0.985	7.396	0.998	99.006
C16	14	1.278	0.919	7.186	1.574	35.733
C17	15	1.889	-0.391	7.198	1.929	348.299
C18	16	3.346	-0.353	7.075	3.365	353.977
O19	17	6.143	1.771	5.984	6.393	16.079
O20	18	7.568	1.770	6.676	7.772	13.164
O21	19	3.941	3.223	6.378	5.091	39.275
C22	20	1.920	4.625	6.428	5.008	67.455
C23	21	-2.972	3.223	5.188	4.384	132.684
C24	22	-4.519	3.438	7.179	5.678	142.735
C25	23	-5.488	4.236	6.290	6.933	142.339
O26	24	-4.808	2.904	8.095	5.617	148.871
O27	25	-0.972	-0.119	7.784	0.979	186.970
O28	26	1.327	-1.457	7.510	1.971	312.319
C12	27	-2.969	3.302	6.755	4.441	131.960
C37	28	3.447	7.307	5.678	8.079	64.746
C33	29	3.237	6.416	5.941	7.187	63.228
O31	30	1.127	6.310	4.231	6.410	79.870
C32	31	2.382	5.480	4.694	5.975	66.507
C34	32	1.870	6.502	6.932	6.766	73.952
C35	33	0.648	7.175	6.330	7.205	84.841

		X(Å)	Y(Å)	Z(Å)	R(Å)	φ(°)
C36	34	3.498	5.542	3.570	6.554	57.741
N38	35	2.304	7.382	8.007	7.734	72.665
C30	36	-0.006	6.480	5.190	6.480	90.056
O29	37	-0.650	5.100	5.390	5.141	97.267
C8	38	1.276	3.353	6.754	3.588	69.161

however, the first layer line upon which significant intensity is observed is layer line 29; corresponding to a measurable pitch of 38.89Å. This figure is in agreement with the equation derived by Fuller (1966) on the basis of an average pitch concept.

At larger distances from the origin in reciprocal space, the intensity observed is spread over three or four adjacent layer lines and is of a lower value on each of them. Calculation of the transform beyond the first meridional at 3.4A shows that this layer line spread is periodic: the first two or three layer lines beyond the meridional being well defined in the ζ direction.

The diffuse component of the scattering has also been calculated in a number of cases. This is a continuous function in two dimensions, unlike the non-diffuse component which is only continuous in the ξ direction. Its value tends to increase, not necessarily monotonically, in the ζ direction up to the first meridional. Since the spread of the intensity peaks of the non-diffuse component increases for higher values of ζ up to the 3.4Å meridional, while the peak value decreases, the contribution of the non-diffuse component becomes less significant at higher ζ values and, hence, any resemblance of layer line structure is lost since it merges with the background.

This loss of layer line detail is in accord with the predictions made in section 3.3.2 and is typical of disorder of the second kind in which details in the diffraction pattern resulting from vectors between repeat units become negligible with increasing distance from the origin of reciprocal space.

5.5.2 Ethidium Models

The transforms calculated for ethidium models are shown for selected layer lines in figures 5.16 to 5.22. Numbers in parenthesis indicate the actual number of the layer line, while the un-bracketed figure gives the "pseudo" layer line number. There is little

difference between these transforms and that of BDNA (figure 5.27), nor is there any significant difference between the models having different values of the unwinding angle.

In view of the error in measuring the intensity data from the diffraction patterns of intercalation complexes it is not possible to do more than make a general comparison of the observed and calculated intensity. Nevertheless, intensity measurements were taken along the first and second layer lines of diffraction pattern taken by Dr. W.J. Pigram from a fibre of an ethidium/DNA complex at P/D = 12.5, and exhibiting a pitch of approximately 38Å. The intensity data, presented in figure 5.24, has been corrected for the effects of cylindrical averaging by multiplying the values by a factor $2\pi\xi$ as prescribed by Franklin and Gosling (1953). The observed and calculated transforms on the first and second layer lines for BDNA are presented for comparison in fig. 5.23. In the data taken from the ethidium pattern there is an enhancement of the first layer line compared to the second, and there is a change in the position and relative heights of the intensity peaks. In particular, the intensity peaks on the first layer lines increase in height with increasing ξ value in the range given, whereas in BDNA the first peak is much more intense than any subsequent ones. It is possible, however, that this change is not due to a structural feature of the complex itself. Some idea of the accuracy of the method of obtaining intensity data in the manner employed here may be obtained from fig. 5.23. Comparison of the observed and calculated BDNA transform shows that the intensity appears to be over-estimated at higher ξ values. This apparent enhancement may, therefore, be due to some other factor such as impurities in the fibre which may give a spherically averaged background scatter.

In fibres involving a drug, such scatter will be produced by any drug molecules which are not complexed with the DNA and which are of random orientation in the fibre. Although the relative peak heights may not represent accurately features of the intercalation complex, the measured intensity values will be largely determined by the nature of the transform of the complex and, hence, may be used to compare the predicted diffraction intensity calculated from different models.

All the ethidium models built give an enhancement of the first layer line, which is in qualitative agreement with the experimental data A surprising feature of the transforms is the fact that the predicted enhancement is greater at P/D = 20 than at P/D = 10. Since more drug is present in the latter case it would seem reasonable that any affects produced by the drug would be more pronounced at the lower P/D value. It may be that the change in the relative layer line intensities is due to the change of position at which the transform is sampled in the two cases which correspond to different pitch values. In all cases, the intensity on the first layer line of the calculated transforms is enhanced to a greater degree than that observed in the intensity measurements. Although the P/D ratio of the complex of plate 5.1 is 12.5, the pitch value ($\simeq 38\text{Å}$) suggests that the drug is bound intercalatively at a P/D around 20 to 25, and, hence, the observed intensity data should be compared with the transforms calculated for the case of P/D = 20.

An important qualitative difference between the calculated transforms and the observed data in fig. 5.24 is that the peak at $0.2A^{-1}$ on the first layer line of the observed data is not apparent in any of the transforms calculated for the models. This is a serious discrepancy between the observed and calculated data. A similar, though less intense peak occurs at a lower ξ value in BDNA, and, hence, the predicted change in the transform in passing from the native form to that of the complex is the opposite of that observed.

5.5.3 Daunomycin Models

The transform of the daunomycin intercalation model with the side groups projecting into the small groove is shown for the first and second layer line in figure 5.25: the calculation was performed for a P/D = 10. A diffraction pattern of a daunomycin intercalation complex (P/D = 12 , pitch = 42Å) taken by Drs. T. Sundaresan and W.J. Pigram is shown in plate 5.2. The general appearance of the diffraction pattern is quite different from that of plate 5.1 from an ethidium complex. In particular, the second layer line is very strong in the daunomycin pattern and there is a particularly strong intensity peak at $0.1 Å^{-1}$ along the layer line. This intensity is not observed in the transform of the present daunomycin model and, hence, even the general nature of the calculated intensity is not in agreement with the observed data.

It is not possible to measure the intensity of the first layer line since it is too close to the arc of the first equatorial spot. However, observation by eye of that part of the first layer line between the equatorial spots suggests that its intensity is low in comparison with that of the second layer line and of the equator. In the calculated transform, however, the first layer line intensity is enhanced in comparison with that on the second, so that this constitutes a further discrepancy between the observed and the calculated data. Methods by which the model might be modified are discussed in section 5.6.3.

5.5.4 Calculation of the Diffuse Scattering

The diffuse scattering component of the transforms, arising because of the substitution disorder, has been calculated for all models. This component is small in magnitude in comparison with the first term of the transform discussed above. For this reason, and in

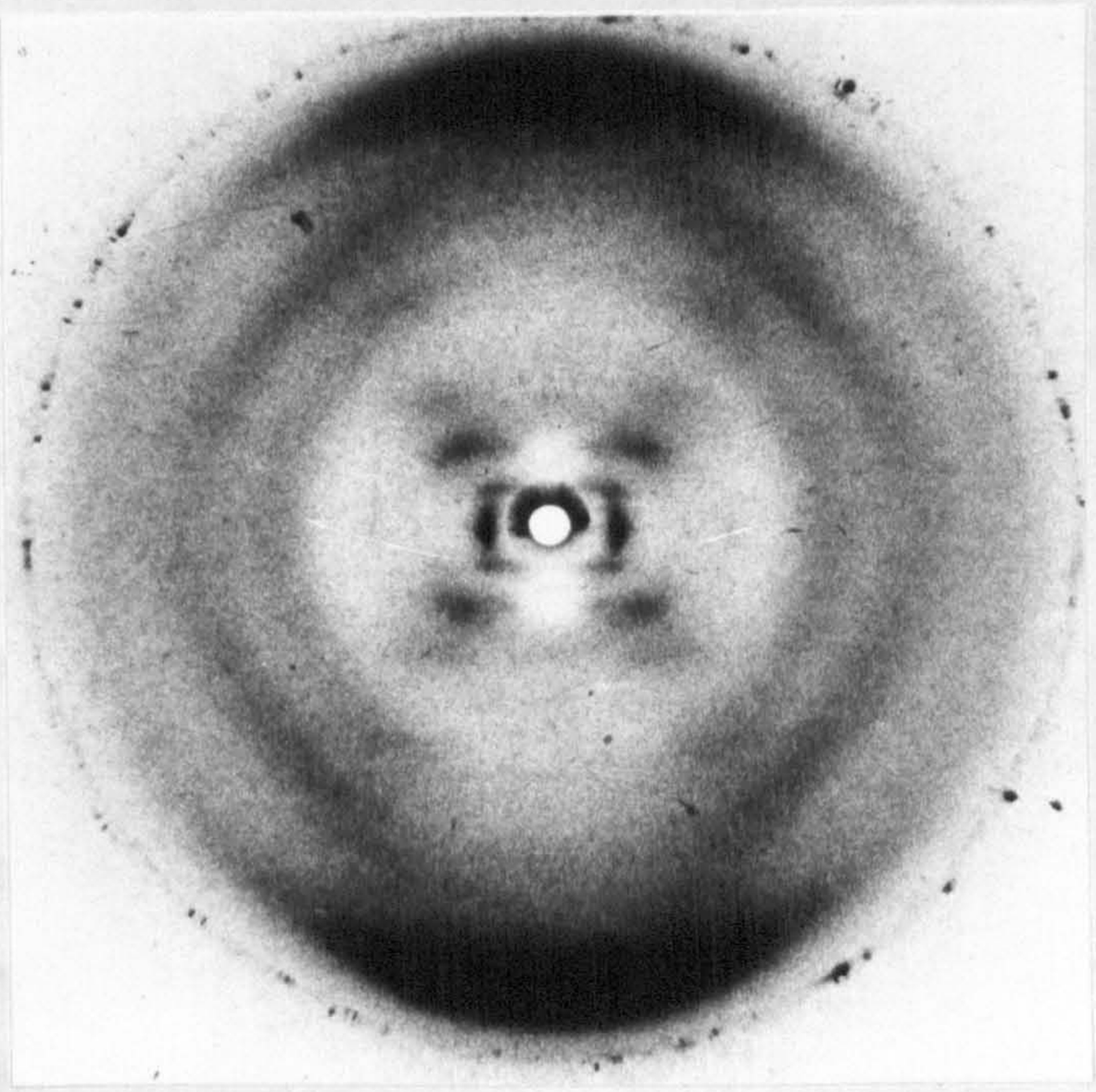


Plate 5.1 DNA-Ethidium P/D = 12.5 92% R.H.
courtesy of Dr. W.J. Pigram

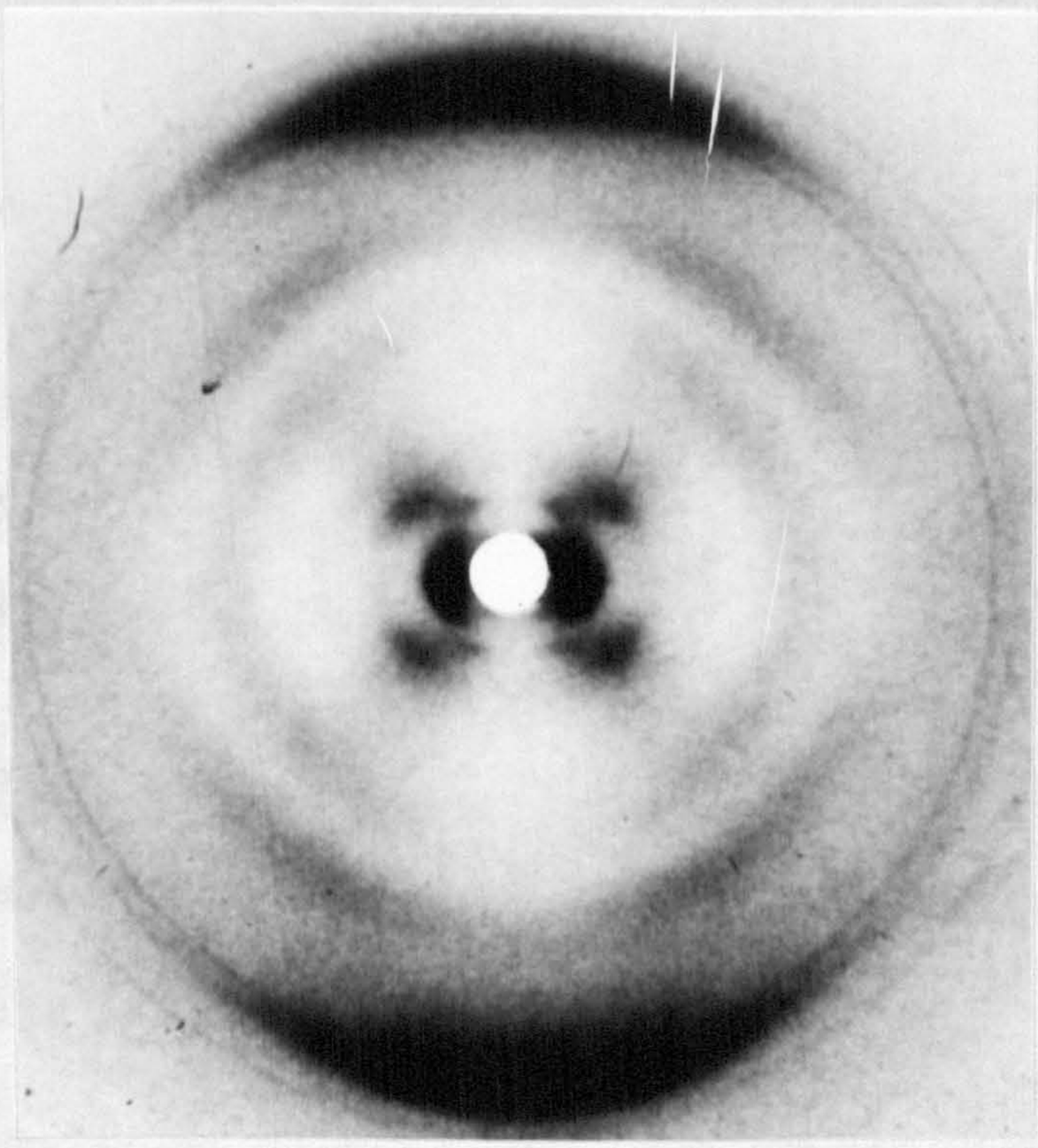


Plate 5.2 DNA-Daunomycin P/D = 12 92% R.H.
courtesy of Drs. T. Sundaresan and W.J. Pigram

view of the limited experimental data, the contribution of the diffuse term has not been included in the transform value presented in figures 5.16 to 5.22. The values obtained for the diffuse scattering component were very similar for all models and, hence, graphs are present (fig. 5.26) for only one of the intercalation models described above (the ethidium model having a 12^{o} unwinding, groups in the small groove, no H-bonds). The graphs presented in fig. 5.26 show the intensity along the equator; along a line at ζ value equal to that of the "first" layer line in figure 5.18; and along a line of constant ζ situated half way between the previous two. The values tend to decrease for increasing ξ value along each layer line and, hence, omission of the diffuse component of the scattering from the values presented in figures 5.16 to 5.22 is not responsible for the increasing discrepancy between the magnitude of the observed and calculated intensity at higher ξ values.

5.6 Discussion

5.6.1 General Intercalation Models

A number of intercalation models have been built which do not involve any specific drug molecule, and which are intended to represent the least modified conformation around the intercalation site for the binding of an intercalating drug into BDNA. When using the BDNA conformation of Arnott and Hukins (1972) as the native form the number of models which could be built was very restricted and satisfactory models were only obtained with a mixed puckering of the furanose rings C3' exo (5'-3') C3' endo about the intercalation site and with a narrow range of values for the total unwinding.

Models based upon a C2' endo BDNA conformation (Arnott et. al., 1969) could be built with a much larger range of unwind angles, although the stereochemistry of the models was less satisfactory than those

based upon the C3' exo B conformation. The values of the stereochemical parameters (bond angles and bond lengths) for the two BDNA models are given in table 5.24, in view of the similarity in these values and of the atomic coordinates for the two models the discrepancy in the range of intercalation models which can be built is surprising. The small change in the furanose ring shape is the most pronounced difference between the two models and the range of possible conformations which can be adopted appears to be a sensitive function of the sugar pucker.

It should be noted that the values of the torsion angles for the models presented are very different from the standard values quoted by Arnott and Hukins (1973). Values of the parameter $\bar{\varepsilon}_o$, the root mean square difference in standard deviations between the standard values and those observed, have been calculated for the models and are all greater than 5.0. Alden and Arnott (1975) adopt the criterion that a value of $\bar{\varepsilon}_o$ greater than 1.65 denotes an overly strained conformation, so that under this convention the models presented here are highly strained. In the present study the dihedral angles were not constrained to adopt standard values since hydrogen atom coordinates were included in the computer refinement and hence, the strain due to unfavourable dihedral angle values should be implicitally calculated and included in the energy minimisation routine. Nevertheless, it is possible that models having more standard values for the torsion angles could allow different ranges of models to be built. So far, a satisfactory model having near standard values for the torsion angles has not been built.

All the intercalation unit models built have some non-bonded contacts which are shorter than are normally found in well refined molecular structures, and, hence, cannot be considered to be entirely satisfactory. Further work is required to determine whether a model

TABLE 5.24

Bond Lengths (Å) and Bond Angles (°) Values For Different Models

	C'3 exo	C'2 endo
C'3-O1	1.43	1.42
O1-P	1.59	1.60
P-O4	1.59	1.60
O4-C'5	1.45	1.45
C'5-C'4	1.51	1.51
C'4-C'3	1.53	1.52
C'3-O1-P	119.45	121.00
O1-P-O4	101.49	101.80
P-O4-C'5	118.70	121.00
O4-C'5-C'4	109.78	116.10
C'5-C'4-C'3	116.35	116.10
C'4-C'3-O1	112.22	110.80

having completely satisfactory stereochemistry can be built with the somewhat rigid constraints imposed in the present study, or whether it will prove necessary to relax some of the imposed conditions to allow a less strained model to be built. The fact that small changes in the nature of certain features of the model can affect strongly the range and type of conformations which can be built is illustrated by the different behaviour of the two BDNA conformations used in the present study.

5.6.2 Ethidium Models

None of the models constructed had completely satisfactory stereochemistry, particularly with regard to non-bonded contact distances. Nevertheless the model utilising the C3' exo/C3' endo intercalation unit has a number of features which are consistent with results obtained by other workers. In particular, the value of the total unwinding angle (26^{o}) is consistent with that predicted by Wang (1974); while the unwinding at the intercalation site (28^{o}) is similar to that obtained by Tsai et. al. (1975), as is the distribution of sugar puckers around this region.

The general nature of the intensity data from diffraction patterns taken from ethidium/DNA complexes (e.g. plate 5.1) resemble closely those observed in BDNA, certainly much more so than do data obtained from complexes involving an anthracycline drug such as daunomycin (plate 5.2). It may well be, therefore, that an intercalation model of the type presented here which can be considered to represent the least modified type of intercalation unit in a BDNA native conformation, is a good approximation to the ethidium intercalation complex. Since it has only been possible to build the intercalation unit based on Arnott and Hukins (1972) BDNA with a narrow range of unwinding angles, it is of interest that the range

coincides with the latest predictions concerning the magnitude of this parameter for ethidium intercalation.

The worst short contacts in this structure (table 5.1g) are somewhat shorter than is acceptable, however, and the sum of the energies of van der Waals interactions between the drug and the DNA is positive. Ideally, if the van der Waals interactions were to stabilise the structure, this value should be negative, however, in the present calculations the effect of solute on the drug has been ignored, and this could markedly affect the relative stability of the free and the bound states.

Having built a model, it would be of considerable interest to compare the predicted diffraction intensity data with that obtained in experiments. Unfortunately, current techniques of intensity measurements do not allow sufficiently accurate data to be extracted from diffuse diffraction streaks such as are obtained from intercalation complexes to enable detailed structural analyses to be carried out. Hence, although Fourier transform calculations have been performed for the models built, it is not possible to use these in conjunction with experimental data to check the validity of the model in detail. The work of Fraser et. al. (1976) opens up the possibility of utilising diffuse diffraction intensity with greater precision so that it may become possible to perform a more detailed analysis in the future.

The current model is probably not correct in detail in view of the unsatisfactory features of its stereochemistry and of the discrepancies in the observed and calculated diffraction data. Nevertheless, the general features of the model are consistent with experimental evidence and only small modifications may be necessary to make it satisfactory.

5.6.3 Daunomycin Models

The best model for daunomycin intercalation is superior from a stereochemical point of view to the models obtained for ethidium. The sum of van der Waal's energies for the non-bonded contacts between the drug and the DNA is not negative as would be the case ideally, although the figure is somewhat artificial as discussed in the previous section. The major objection to the present model is that its intensity transform does not appear to agree with the observed diffraction intensity for the reasons mentioned in section 5.5.3. Hence, the daunomycin intercalation model does not give as good a fit to the observed X-ray data as do the ethidium models.

There is some evidence suggesting that a number of fundamental differences exist between daunomycin and ethidium intercalation complexes. Firstly, the general appearance of diffraction patterns from ethidium/DNA complexes is quite different from the appearance of equivalent patterns involving daunomycin. Secondly, the work of Waring (1970) suggests that the unwinding produced by daunomycin is less than half that produced by ethidium. Drs. T. Sundaresan and W.J. Pigram in this laboratory have obtained patterns from daunomycin/DNA complexes which exhibit pitch values of approximately 50Å, which is larger than any value obtained from ethidium complexes, which suggests a much higher degree of binding in daunomycin complexes. Spectroscopic evidence for this has been obtained by Mr. H. Porumb in this laboratory (Porumb, 1976). It is possible, therefore, that it is this higher degree of intercalative binding which is responsible for the different appearance of the daunomycin diffraction patterns. However, the most striking feature of the daunomycin patterns, the very high intensity on the second layer line is not observed in the transform of the daunomycin intercalation complex calculated for P/D = 6, the saturation value.

The changes to be made to the model are probably in the DNA conformation rather than in the drug position. It is the DNA conformation which appears to dominate the diffraction pattern at least in the central area. This can be seen by comparing the calculated transforms for the daunomycin model and for the ethidium model incorporating a 12^{o} untwist: both models have very similar transforms despite the structural differences between daunomycin and ethidium.

The changes required may well involve a change in the disposition of the base pairs to the helix axis. If the base pairs were moved away from the helix axis in the negative (viz. fig. 1.1) direction as in CDNA, this would result in increased intensity on the second layer line in comparison with the first since the two DNA helices would be separated by a distance more nearly equal to half a pitch length (viz. section 3.2). Data is presently being prepared to construct intercalation models in which the requirement that the base pairs have a BDNA like disposition to the helix axis is relaxed in the hope of improving the stereochemistry of the intercalation models described here, and to improve the fit with the X-ray diffraction data.

Helix Axis Projections

Ethidium Models

C2' Endo Intercalation Unit

5.7 H-bonds, groups in large groove 12° untwist

5.8 No H-bonds, groups in large groove 12° untwist

5.9 H-bonds, groups in small groove 12° untwist

5.10 Groups in small groove 28° untwist

5.11 No H-bonds, groups in small groove 12° untwist

5.12 Groups in large groove 28° untwist

C'3 exo/C3' endo Intercalation Unit

5.13 Groups in small groove, model A

5.14 Groups in small groove, model B

Daunomycin Model

5.15 Groups in small groove

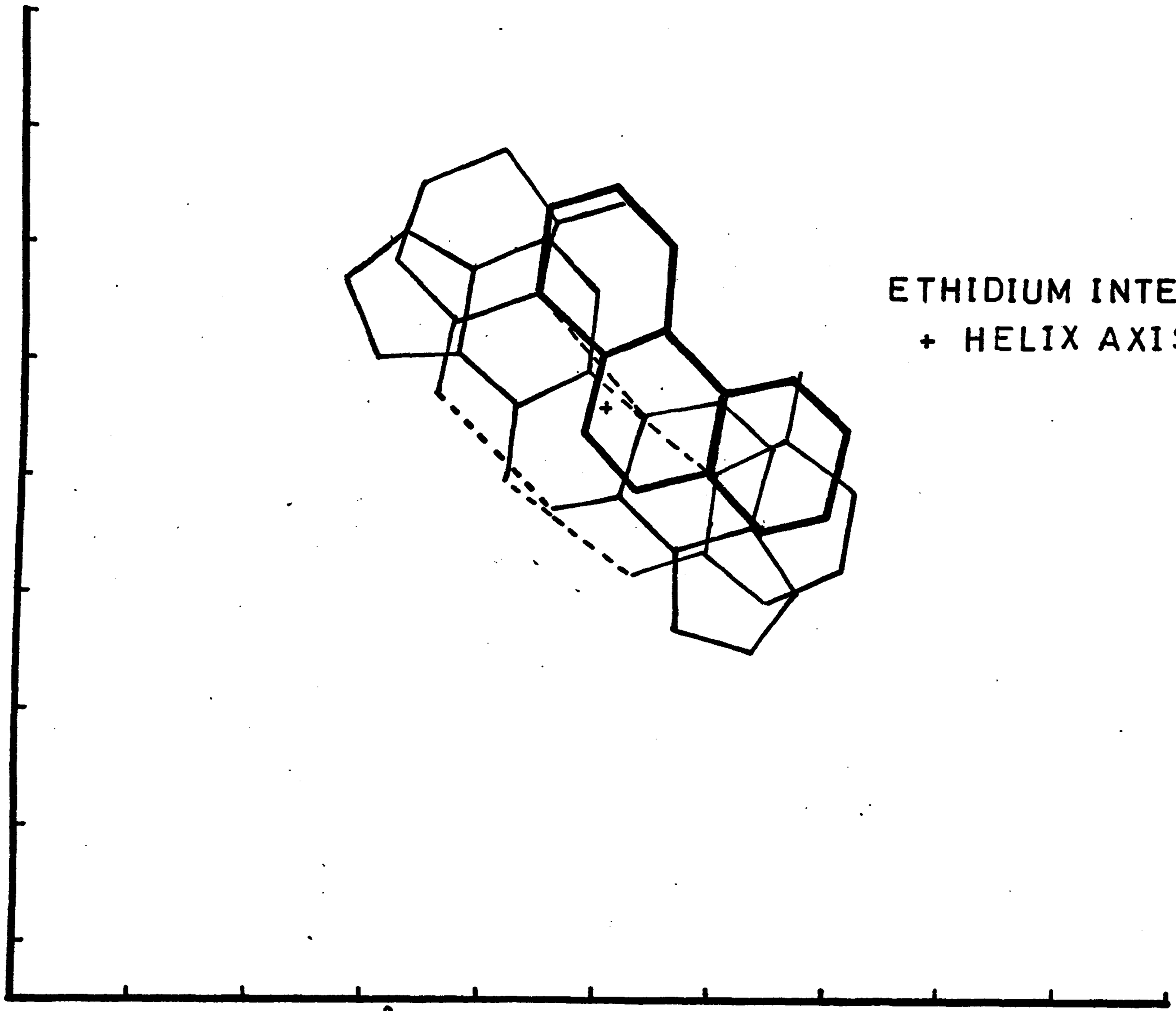
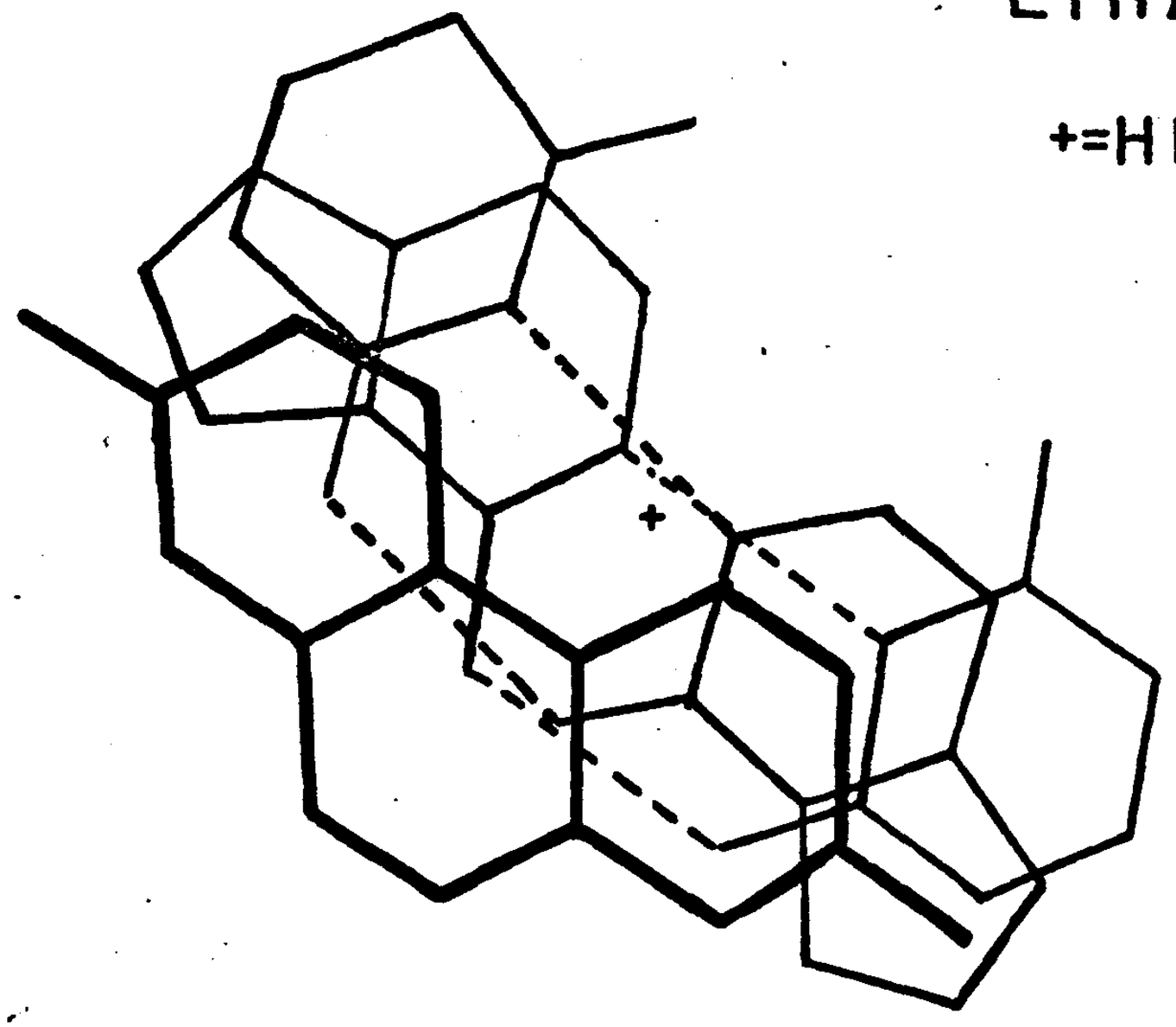


Fig. 5·7

ETHIDIUM INTERCALATION

+=HELIX AXIS



Å

Å

Fig. 5·8

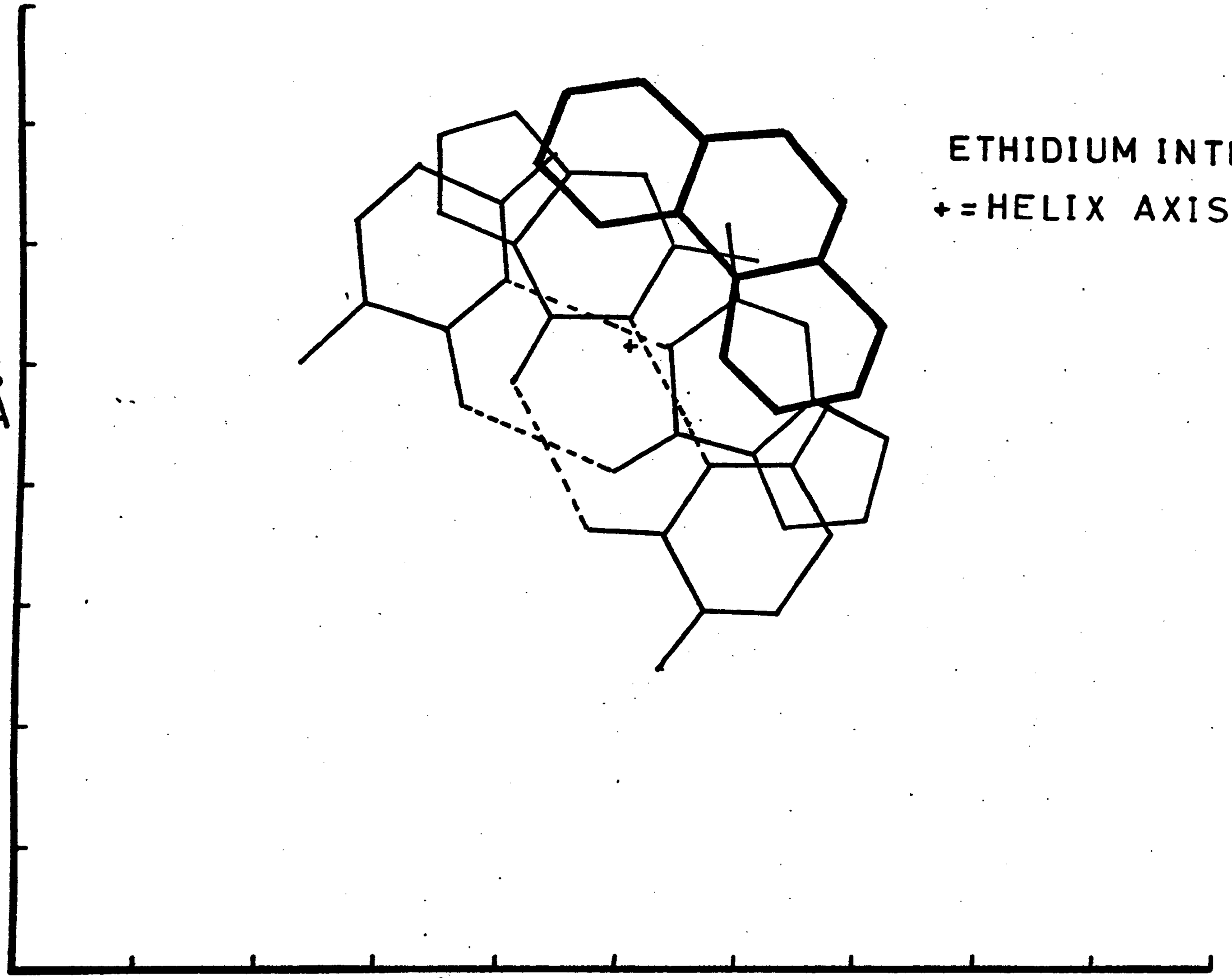


Fig. 5·9

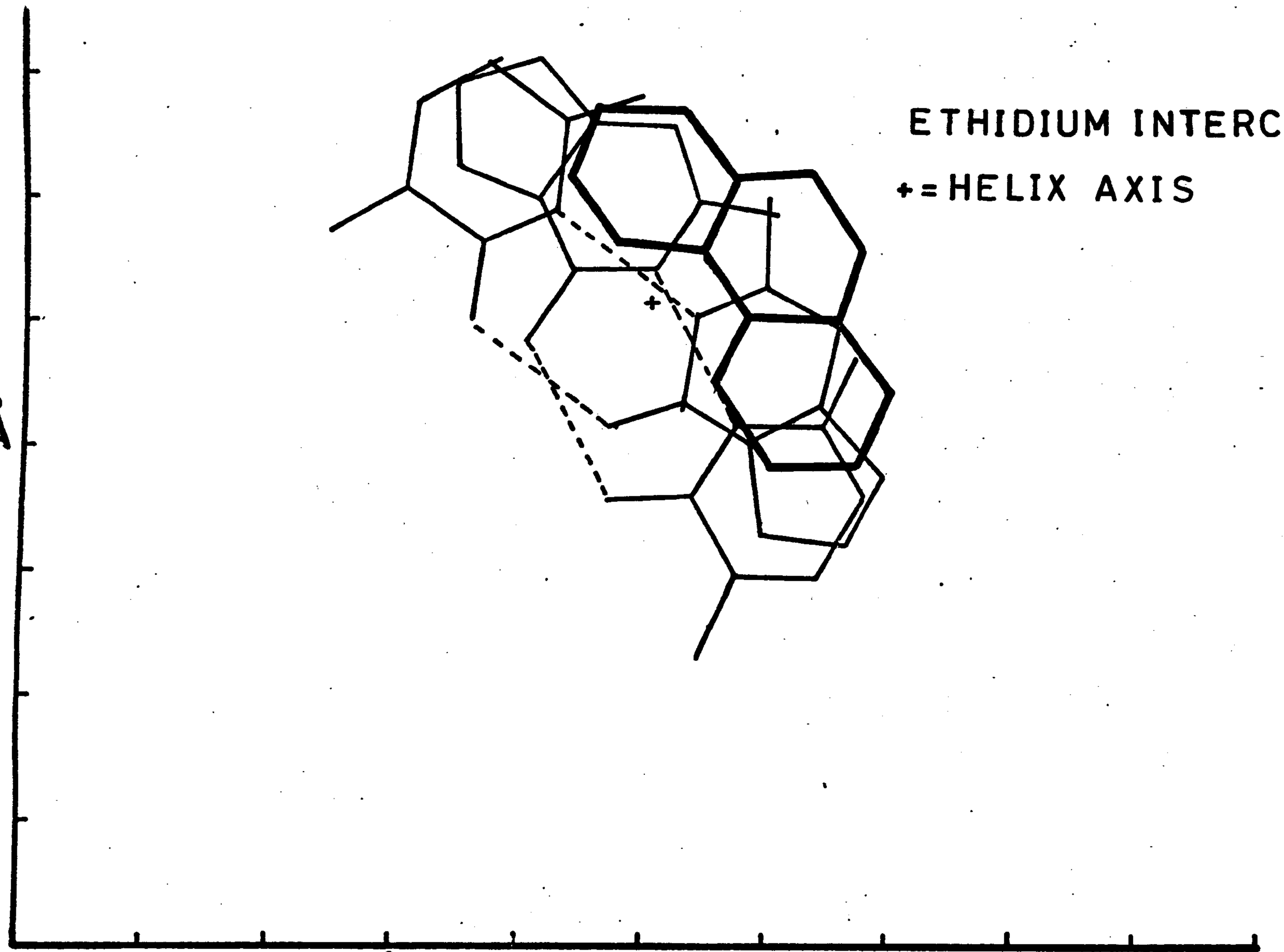


Fig. 5·10

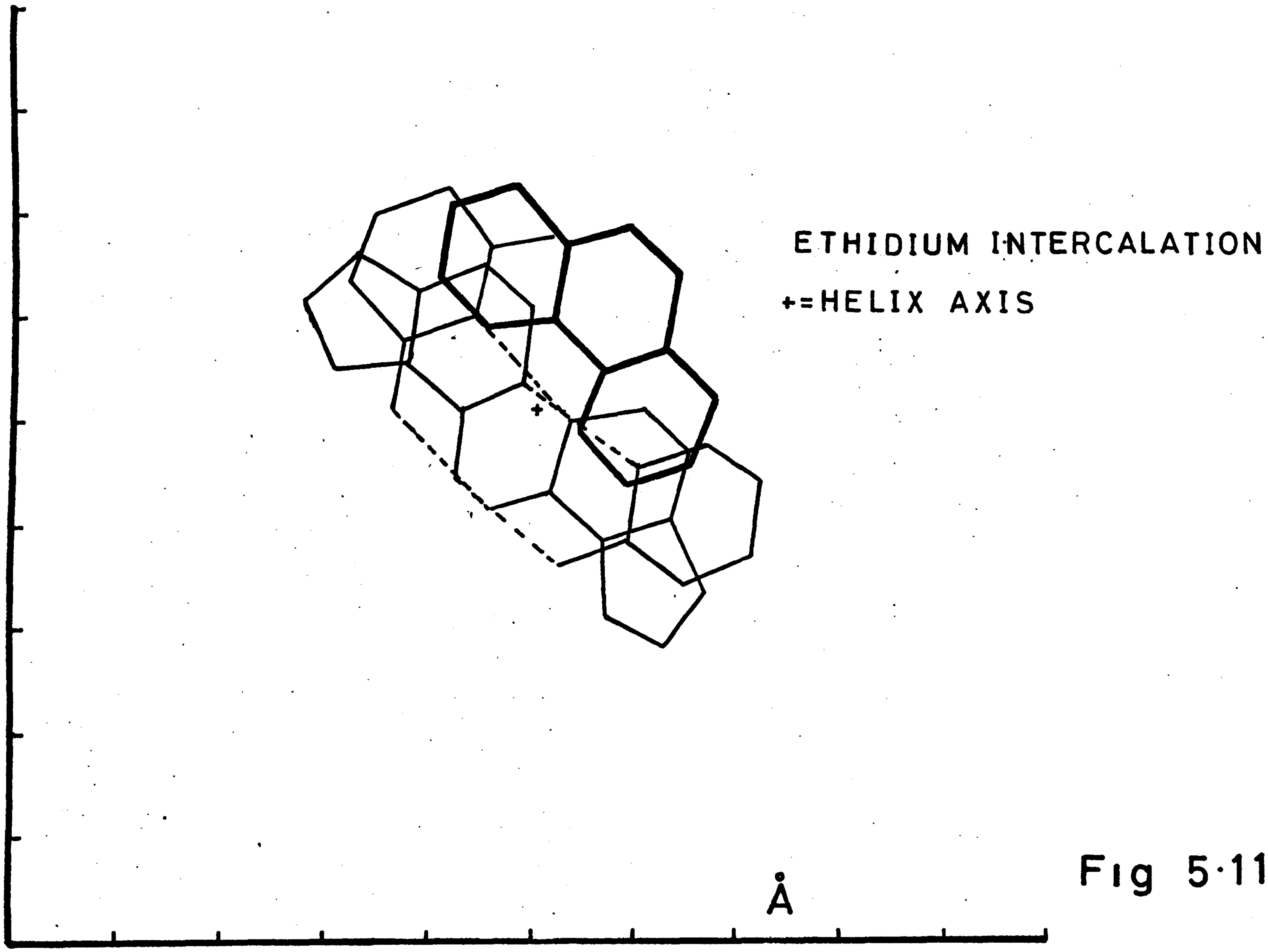


Fig 5·11

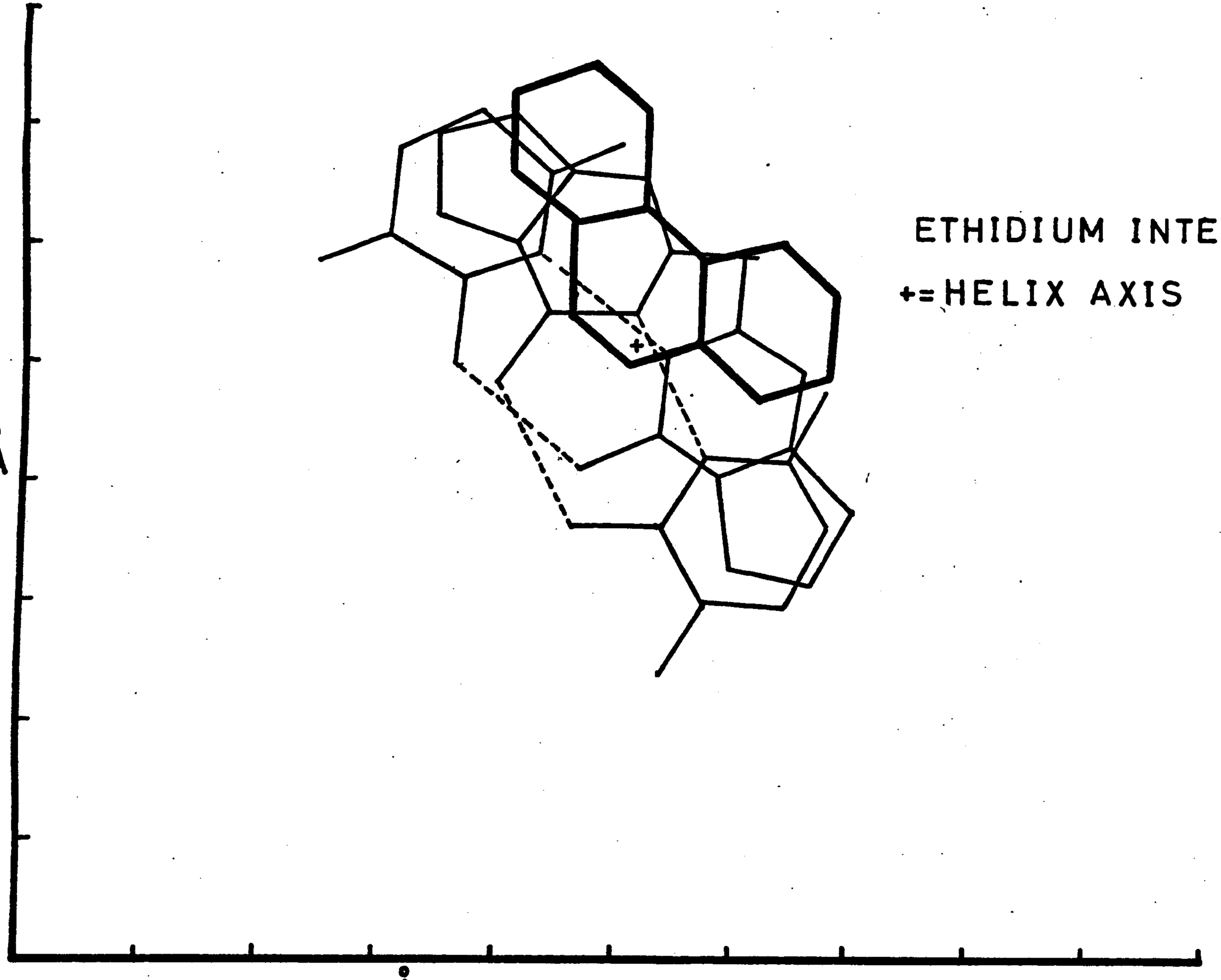
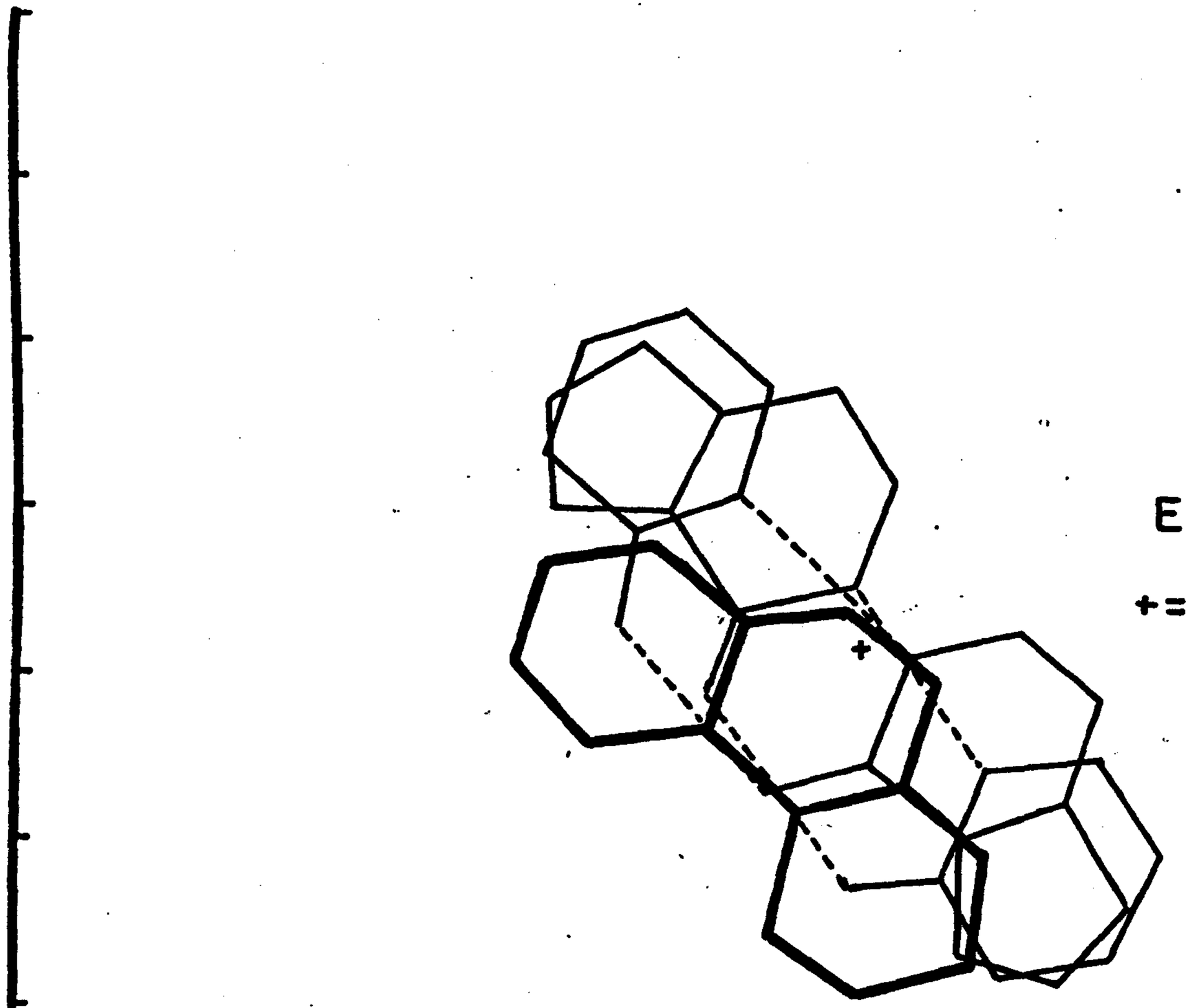


Fig. 5·12



ETHIDIUM INTERCALATION

+=HELIX AXIS

Å

Å

Fig. 5·13

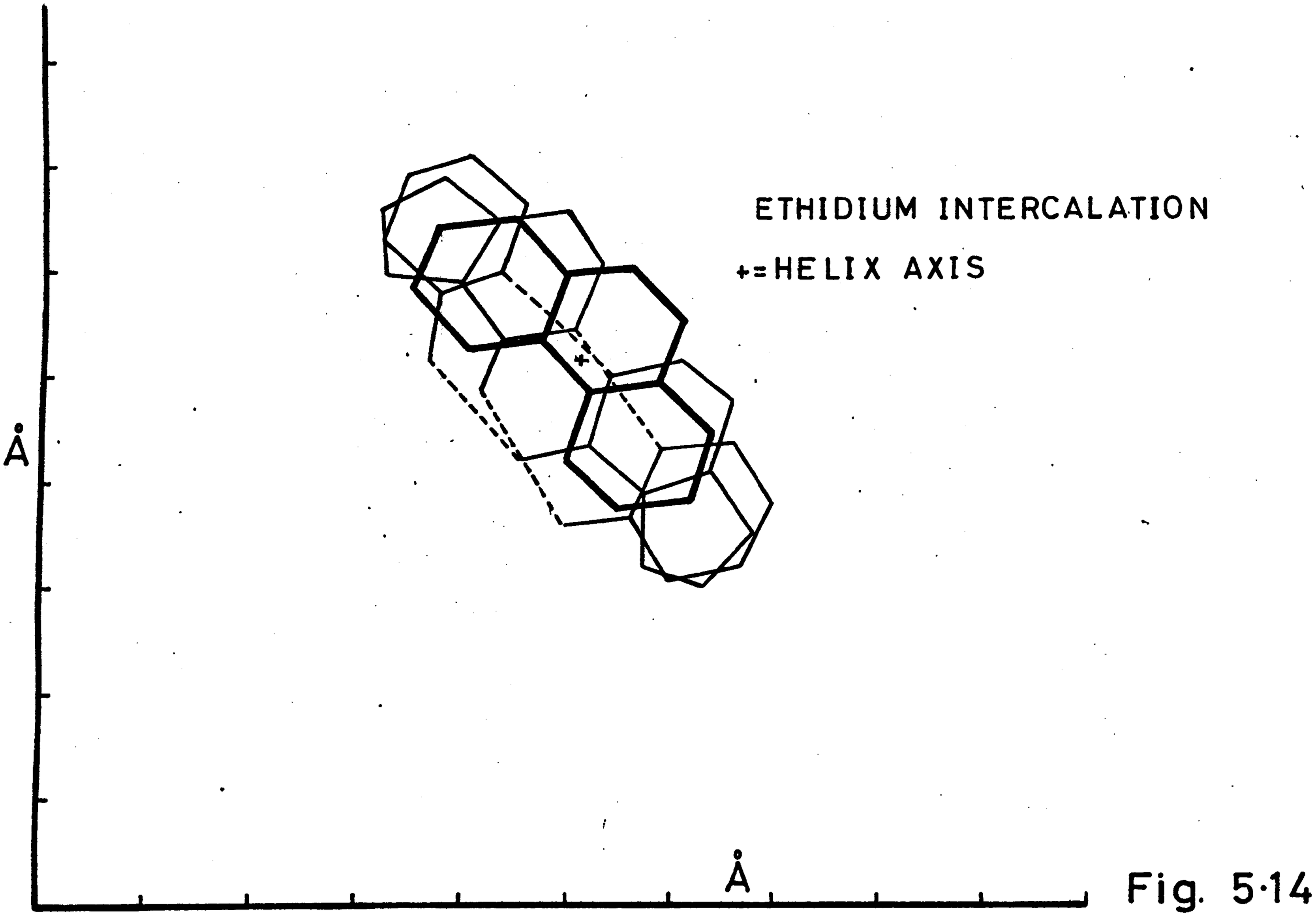


Fig. 5·14

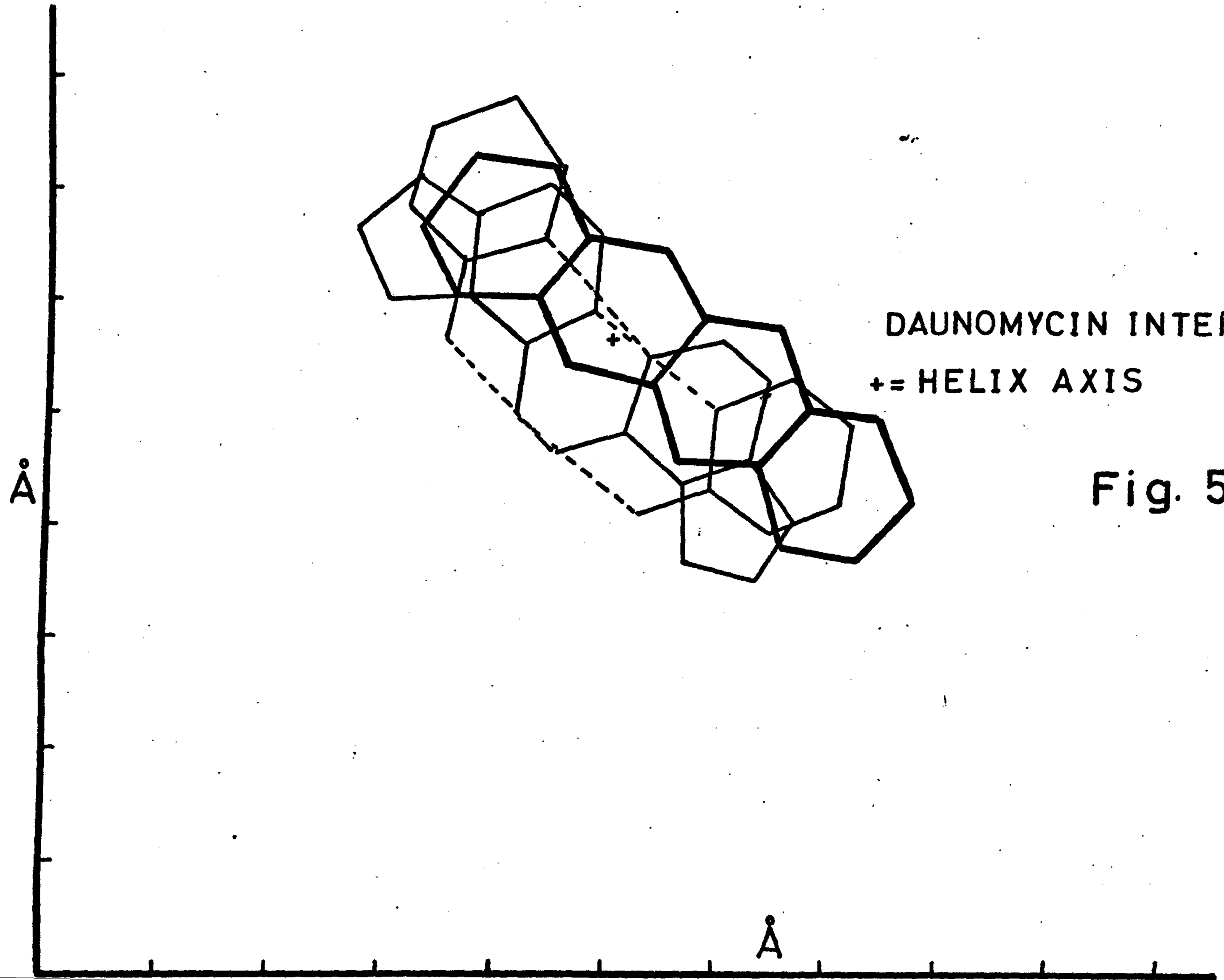


Fig. 5·15

Fourier Transform Calculations

Ethidium Models

C2' endo Intercalation Unit

5.16 Groups in large groove, 12^{o} untwist P/D = 10

5.17 Groups in large groove, 12^{o} untwist P/D = 20

5.18 Groups in small groove, 12^{o} untwist P/D = 10

5.19 Groups in small groove, 12^{o} untwist P/D = 20

5.20 Groups in small groove, 28^{o} untwist P/D = 10

C3' exo/C3' endo Intercalation Unit

26^{o} Total untwist, groups in small groove

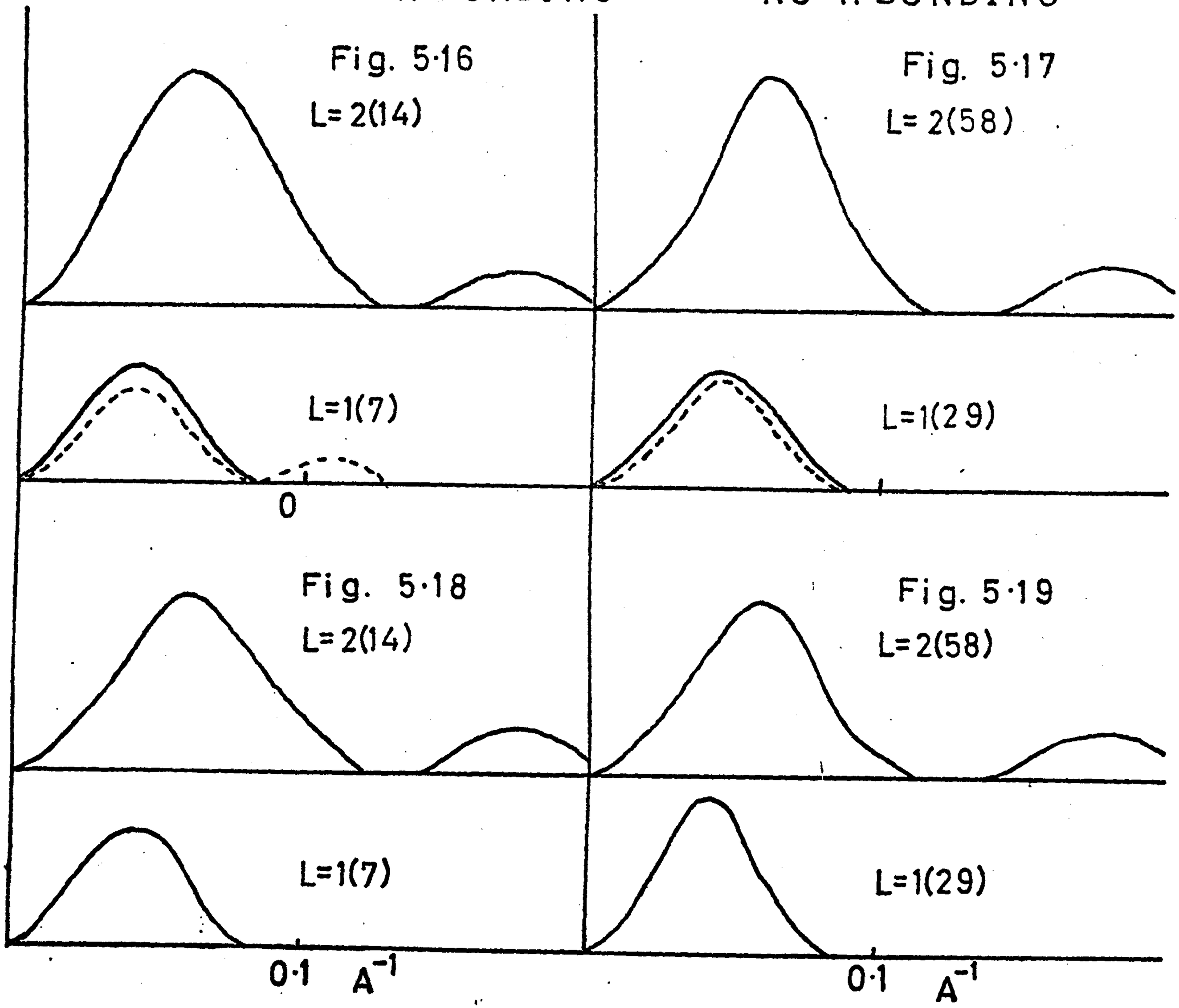
5.21 Model A, P/D = 20

5.22 Model B, P/D = 20

Daunomycin Model

5.25 Groups in small groove, P/D = 10

—H-BONDING ----- NO H-BONDING



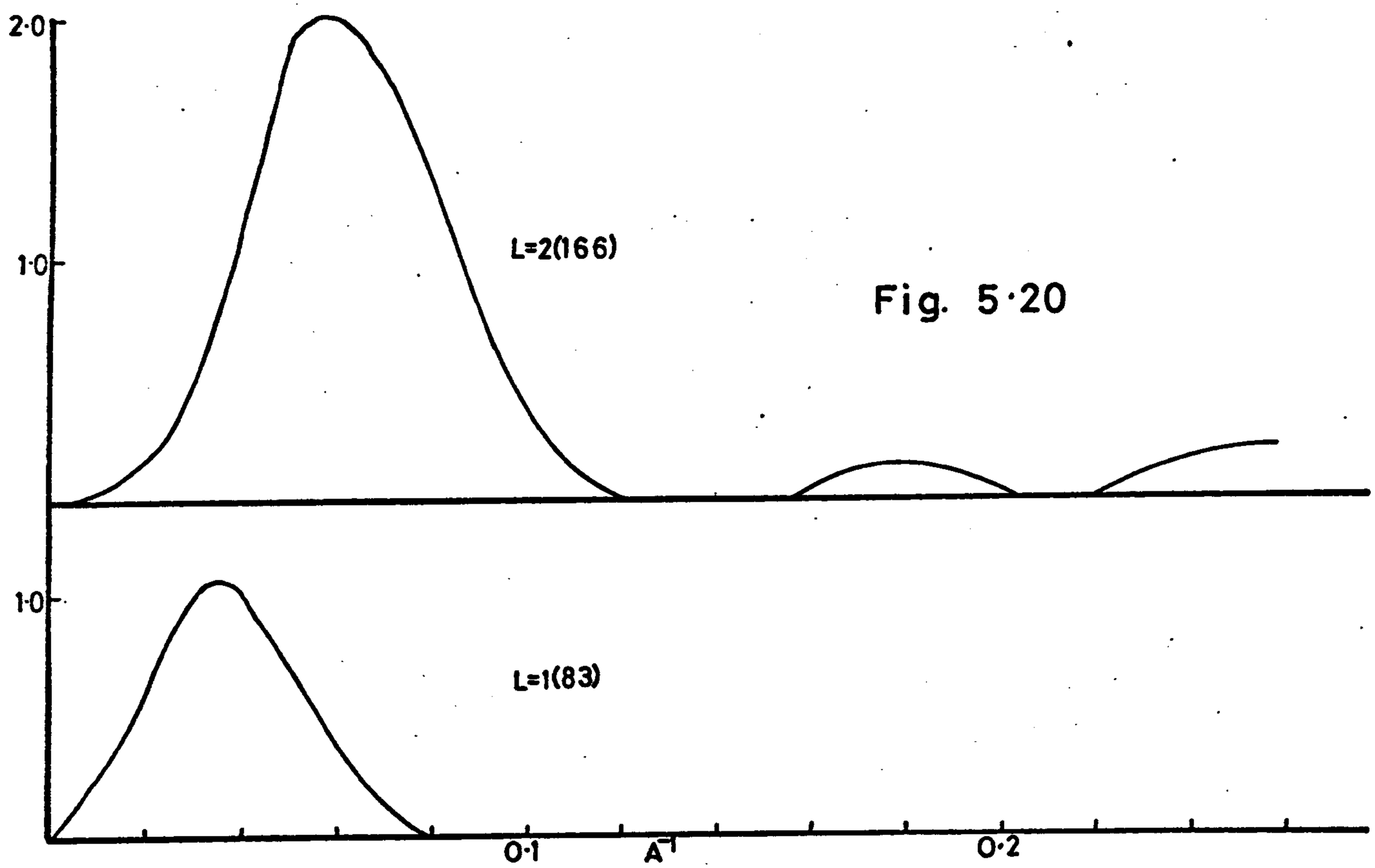


Fig. 5·20

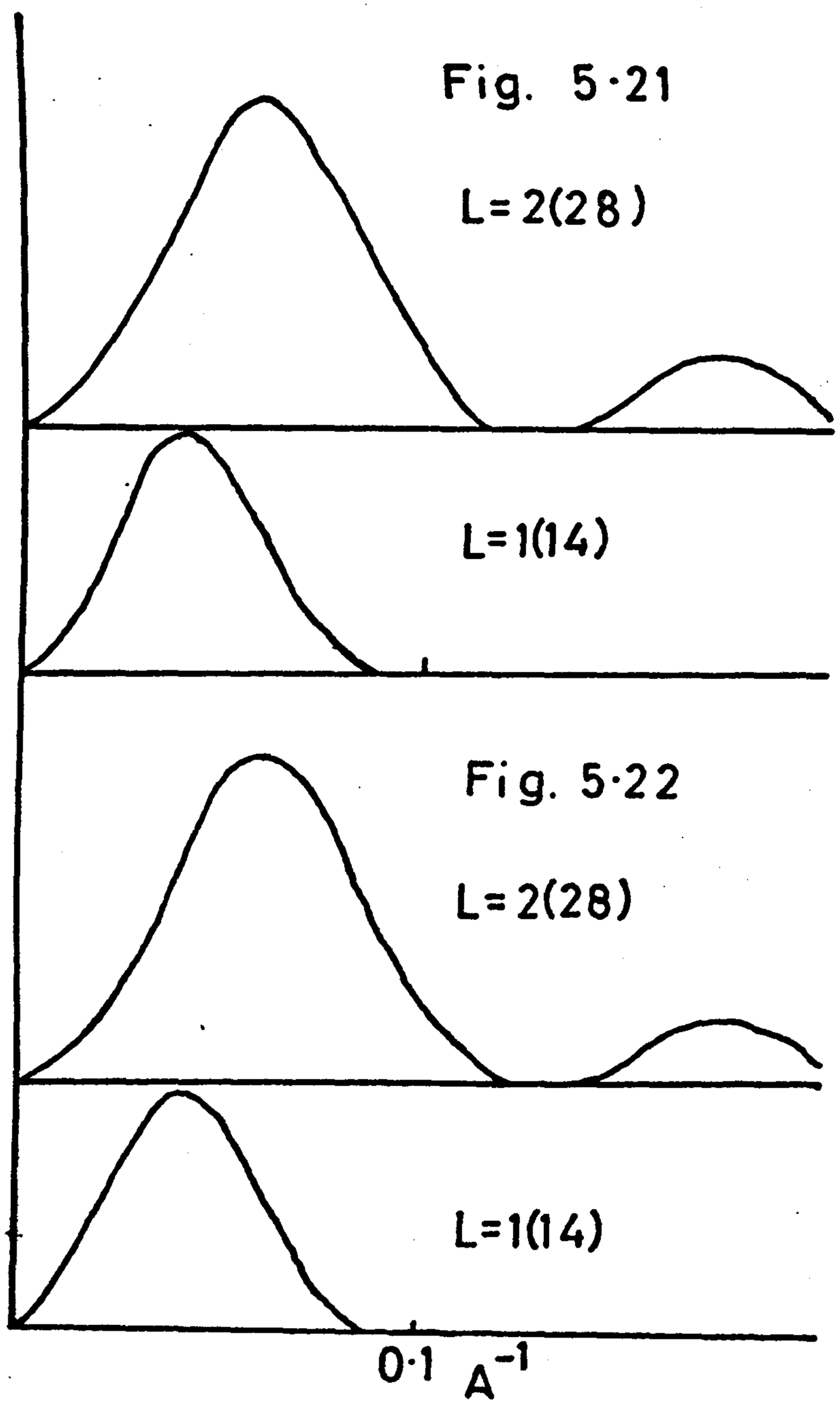


Fig. 5·21

Fig. 5·22

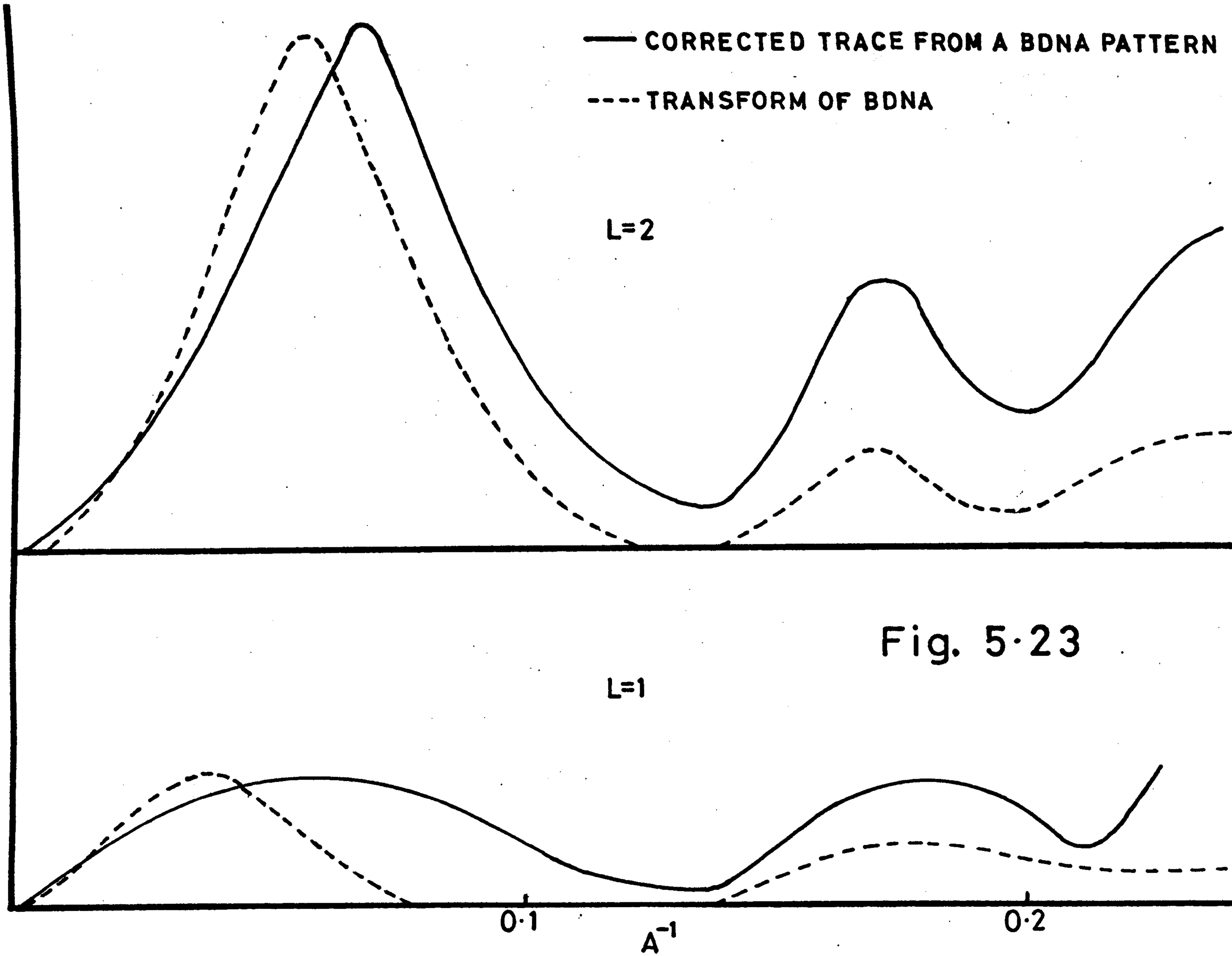
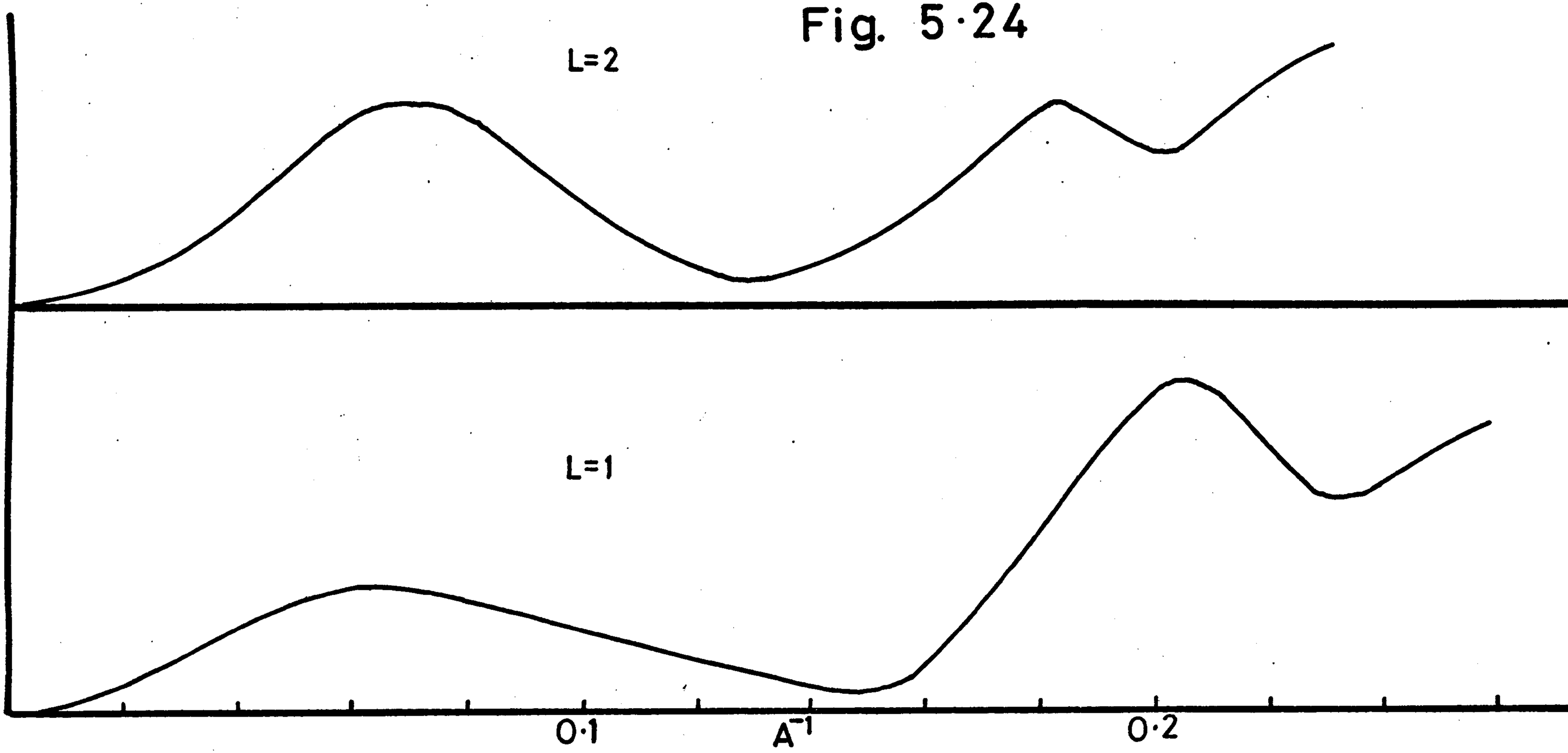

CORRECTED TRACE FROM A BDNA PATTERN
TRANSFORM OF BDNA
L=2
L=1
0·1
0·2
A⁻¹

Fig. 5·23

Fig. 5·24



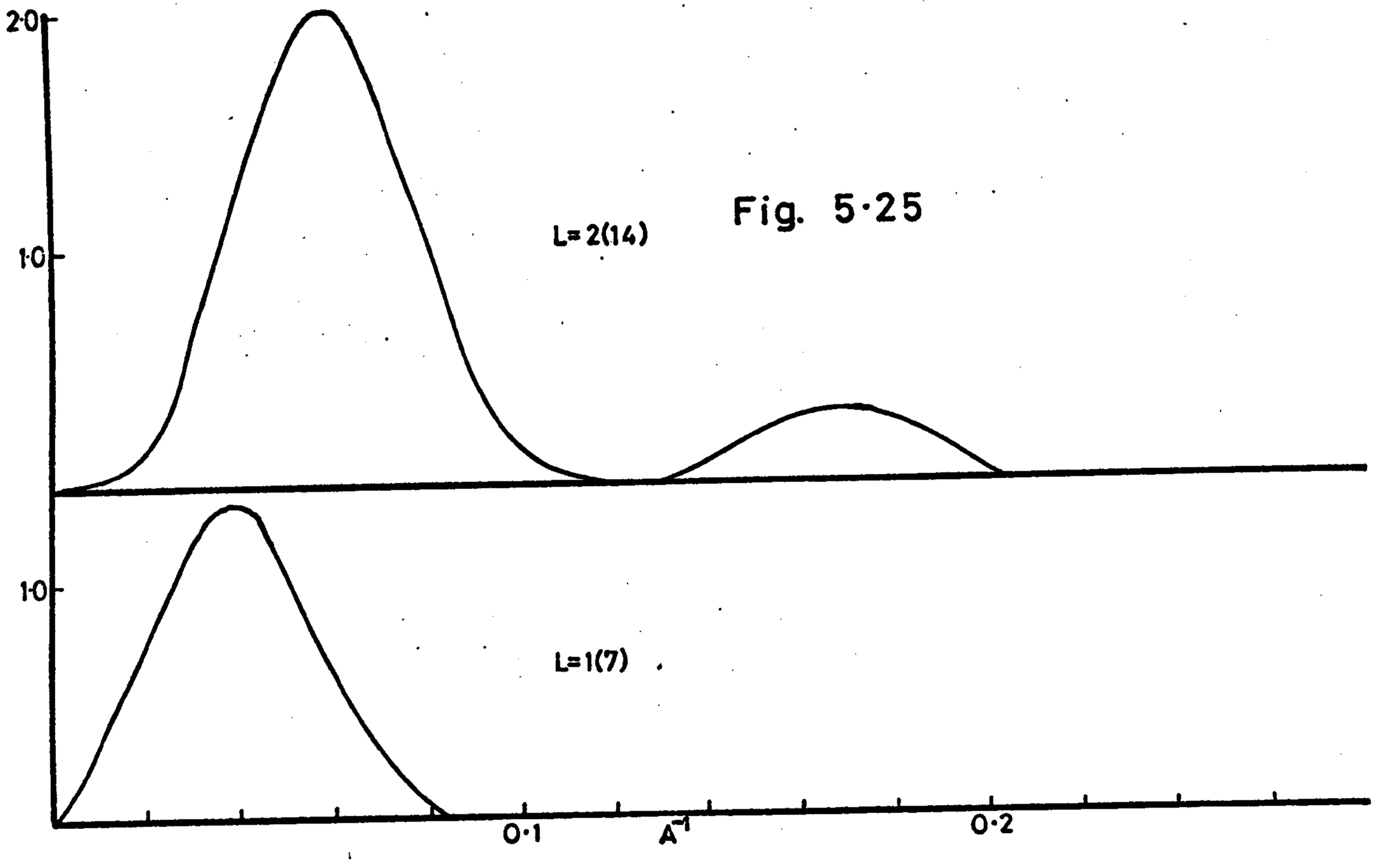


Fig. 5·25

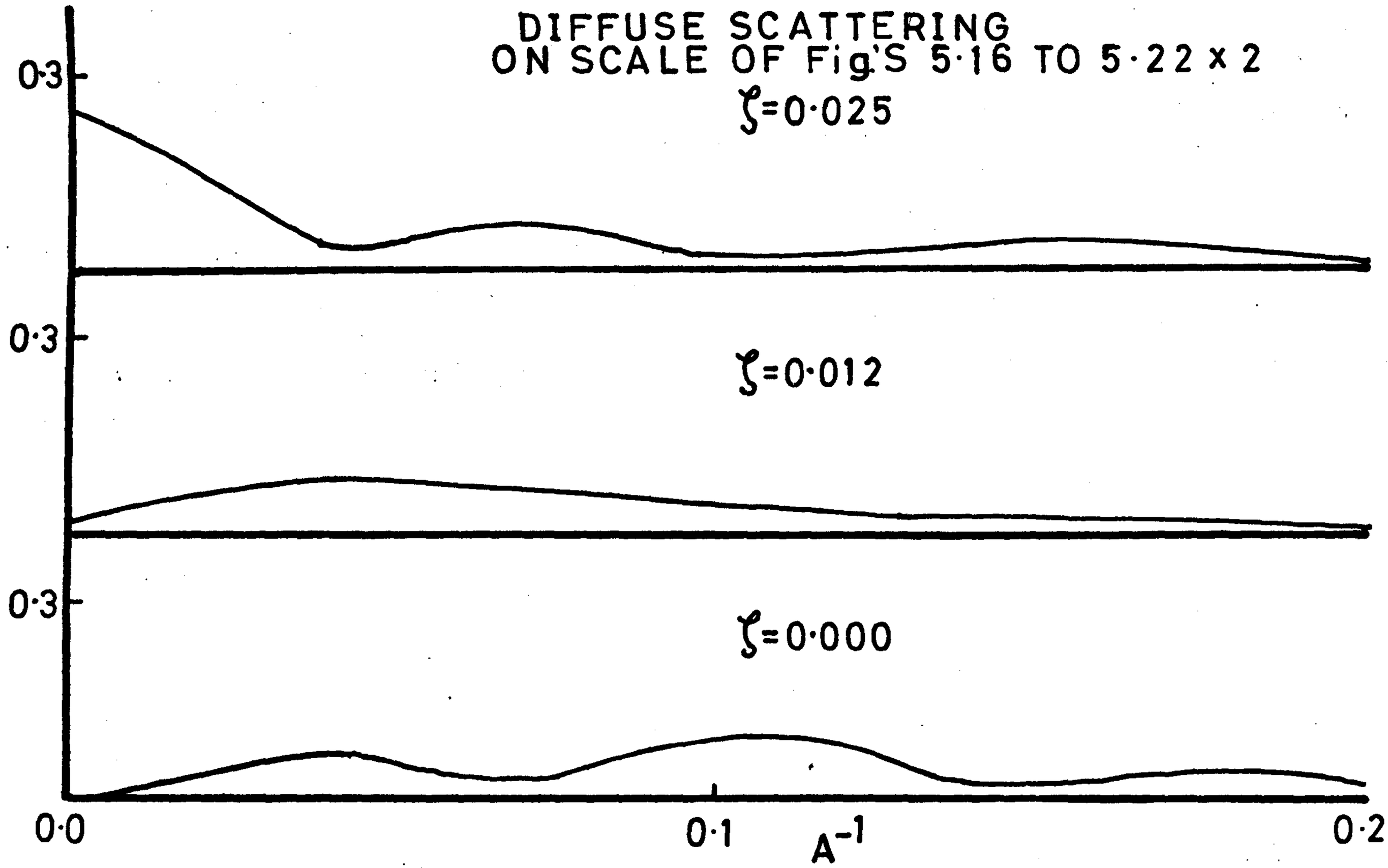

DIFFUSE SCATTERING
ON SCALE OF Fig'S 5·16 TO 5·22 x 2
ζ=0·025
ζ=0·012
ζ=0·000
0·3
0·3
0·3
0·0
0·1
0·2
A⁻¹

Fig. 5·26

CHAPTER VI

DNA CONFORMATION AS A FUNCTION OF BASE COMPOSITION

6.1 Introduction

6.1.1 DNA Conformation in Fibres

It is now widely recognised that the secondary structure of DNA in fibres varies as a function of the relative humidity of the fibre environment and of the ionic strength and type of countercation present. It has recently been suggested by Bram and his co-workers that the secondary structure of natural DNA is a function also of the base composition and sequence. A small number of distinct conformations have been observed for synthetic DNA's of a known repeating base sequence (e.g. Arnott et. al., 1974a; Arnott and Selsing, 1974). However, the suggestion that the secondary structure of natural DNA may be a function of the base composition is a novel one.

A major survey of the diffraction patterns given from DNA's from a wide variety of sources has been made by Hamilton et. al. (1959) and included DNA from prokaryotic cells, eukaryotic cells and from bacteriophages. The cells from tissues of eukaryotes was chosen to include samples in which the cells were rapidly dividing (e.g. human acute leukemia leukocytes) as well as others such as sperm DNA in which the cells were dormant. No new forms of DNA were noted by the above authors, and, generally, all known conformations (A, B and C forms) of DNA were found. They make an interesting comment, however, in regarding the crystalline patterns given by the A form and the lithium B form as being more useful than the sodium B form which gives few well defined Bragg diffraction spots. This is almost certainly indicative of the attitude adopted towards the analysis of diffraction data from nucleic acid fibres at this time, and indeed persists at

present. It is in fact a justifiable attitude to adopt when the elucidation of the conformation of a new structure is sought, since the relevant intensity data is much more easily obtained and analysed from "crystalline" diffraction patterns consisting largely of well defined Bragg diffraction spots. For this reason, the structure of the B conformation has been solved and refined on the basis of data obtained from the crystalline fibres of the lithium salt despite the fact that sodium is more likely than lithium to be the countercation to DNA in vivo.

Although crystalline fibres are useful for structure determination, they are not necessarily the most useful for the study of variations on a conformation. Hence, if base composition and/or sequence is important in determining minor variations in a conformation it is probably the sodium B form which should be studied. The studies of Hamilton et. al. (1959) provide good evidence that no distinct conformations, other than the A, B and C forms are found in natural DNA's, but probably did not include a sufficiently rigorous study of B patterns obtained from the sodium salt to exclude the possibility of a family of structures for the B conformation as suggested by Bram and co-workers (Bram, 1972; Bram and Tougard, 1972). Since analysis of the B conformation of Langridge et. al. (1960) and, eventually to that of Arnott and Hukins (1972), has been carried out using data obtained from fibres of the lithium salt, the possibility arises that the accepted B conformation may not be that found in fibres of the sodium salt.

The A, B and C forms remain the only conformations for natural DNA's which have been obtained reproducibly and for which coordinates have been published. However, it has recently been suggested that there may be several structures of the B type for the

sodium salt of DNA in fibres, and that different members of the set may be adopted by DNA's of differing base composition. The main evidence for the existence of a family of B forms comes from the work of Bram and co-workers. A number of conformations are postulated having pitch values reduced from the value of 33.8Å observed the lithium B form of DNA. The patterns presented (Bram and Tougard, 1972; Bram, 1972; Bram, 1973) are diffuse and the layer line poorly defined, hence, it is probably not possible to measure accurate values for the helical pitch from these patterns. Nevertheless, they may be indicative of the general trend of pitch values in B type conformations in fibres of the sodium salt of natural DNA's. The value of 33.8Å for the pitch of the lithium B form is the largest obtained for any DNA duplex studied in fibres (being exceeded only by the pitch value (39.1Å) for the triple stranded poly (dA) . poly $(dT)^2$ (Arnott and Selsing, 1974), and, hence, the lithium form may be considered as the "extreme case" of BDNA. If this were the case it might explain the greater degree of crystallinity possible with the lithium salt of the B form of natural DNA. One requirement for good crystalline packing is that each molecular unit does not have a conformation which varies in a non-regular manner.

The existence of the two new conformations is claimed for the A-T rich DNA (68% A-T) from the bacterium Clostridium perfringens (Bram and Tougard, 1972). Diffraction patterns from fibres containing 1% excess NaCl indicated that the DNA may be in a new conformation, the 'P' form, having a pitch of 30Å; while preparations having greater than 2% excess salt gave the 'T' form having a pitch of 28Å. A parallel study is also presented (Bram and Tougard, 1972) of DNA from Cytophaga johnsonii (65% A-T) and fibre diffraction patterns are obtained which

are claimed to correspond to two further novel conformations (the J_1 and J_2 forms) to which are ascribed pitch values of 33Å and 31Å respectively. All of the above conformations gave patterns with a strong meridional reflection at 3.4Å and, hence, correspond to forms in which the bases are not significantly tilted. The J_1 and J_2 forms show an intense diffraction ring at 4.3Å which occasionally takes on a slight degree of meridional orientation. This is an interesting feature if it is an effect attributable to the DNA. However, such intensity rings are often associated with the presence organic impurity which does not have helical symmetry and which may pack with varying degrees of order in different fibres. In the absence of a more rigorous analysis it is not reasonable to associate the diffraction ring with a feature of the DNA conformation.

The transitions between the A, B and C forms are sharp and well defined. Bram (1972) has reported obtaining a series of diffraction photographs of fibres of the DNA from the bacterium Staphyloccocus aureous (67% A-T) in which there is a gradual change in the nature of the patterns in passing from 66% to 98% relative humidity: no sharp transition being observed. The patterns obtained at the lower humidity values are ascribed to a new conformation, the 'S' form, having a pitch of 33Å. Although the patterns obtained at higher humidities have a larger line spacing equal to that of BDNA (34Å) it is stated that the intensity on the third layer line is greater than that in BDNA.

The variation in the relative intensity of the first peak on the first, second and third layer line as a function of the A-T/G-C ratio, has also been studied (Bram, 1973). Nine DNA's having A-T/G-C ratios ranging from 0.41 to 4.6 are studied. A progressive decrease is observed in the second layer line intensity

as the A-T/G-C ratio increases: an approximately three fold decrease being observed between diffraction patterns from the G-C rich DNA from Sarcina lutea (0.41 A-T/G-C) and those from the A-T rich DNA from yeast mitochrondria (4.6 A-T/G-C).

It is of interest to survey the published models for B genus DNA structures to see if the enhancement of first and third layer line intensity observed is characteristic of a published structure. The C form of DNA (Marvin et. al 1961) is similar to the B conformation, but does not exhibit enhanced intensity on the first and third layer line when compared to BDNA; while the D form, although a B genus structure, is considerably different from that of BDNA and is not a likely candidate, since the patterns exhibited by Bram (1973) are clearly produced from DNA's having very similar conformations. However, the B' configuration (Arnott et. al 1974) for poly (dA) poly (dT) is very similar indeed to the classical B form and would give a superficially similar diffraction pattern. Calculation of the molecular intensity transform for this structure indeed shows an enhancement of the first and third layer lines in comparison with BDNA. The cylindrically averaged squared transforms for BDNA and B'DNA on layer lines 1, 2 and 3 are given in figures 6.1 and 6.2 respectively, and the relative enhancement of the first peak on layer line 3 in the B'DNA transform can clearly be seen. The magnitude of the change is not identical to that predicted by Bram; in particular, the third layer line is enhanced more than the second in the calculated transforms whereas the measurements made by Bram indicate that it is the first layer line which is primarily enhanced and which, for very A-T rich DNA's, becomes the most intense of the three. Since the patterns

obtained by Bram are non-crystalline and poorly oriented, it is difficult to measure intensity values accurately, and it is in fact stated by Bram that the aim of the study is to present a trend in the nature of the diffraction from DNA of different base composition, and not absolute values.

If the B' conformation is associated only with regions consisting almost entirely of A-T base pairs, then it is unlikely to be found along whole length of a duplex of natural DNA. Hence, if the changes in intensity observed by Bram are due to the B' form or a similar conformation, then it is probably a result of its appearance in segments of the duplex having a localised high concentration of A-T base pairs. In this case the helix would no longer be regular and, hence, diffraction patterns produced would never consist of sharp layer lines.

The results of Bram's work, if correct, not only extend our knowledge of the number of conformations adopted by natural DNA's, but also add a new dimension to the problem in that DNA conformational transitions appear to be a function of the A-T/G-C ratio. Furthermore, if all the forms reported by Bram are confirmed as being distinct conformations, it is strongly implied that conformational changes are also dependant upon base sequence. Some weight is added to the results presented by Bram in that different conformations may be obtained in synthetic DNA's having a simple repeating base sequence, so that the secondary structure appears to be a function of the primary structure in these cases.

It is interesting to speculate whether the conformations observed in these synthetic forms have any relevance for the conformations obtained from natural DNA, particularly those reported by Bram. The Bram structures may represent distinct conformations

in which the secondary structure is regular; or they might be 'statistical structures' in which different segments of the DNA duplex adopt different secondary structures. If the latter were the case, then the local conformation might well be similar to a form found in a synthetic DNA. This explanation also appears to be more significant biologically. If different segments of DNA adopt slightly different conformations under the same environmental conditions, it provides a possible molecular basis for control mechanisms in DNA replication and transcription. Such a series of changes along the DNA duplex would result in a loss of the exact helical symmetry possessed by the regular conformations. Cases in which the quasi-symmetrical nature of a structure arises not merely by defects in the assembly process, but as a vital part of the functioning of the system are observed in other molecular systems such as many viruses.

6.1.2 DNA in solution

It is not possible to use diffraction data from solutions to solve an unknown structure to atomic resolution, and hence fewer reports of solution diffraction data from DNA are available than are equivalent data sets from fibres. Solutions possess disorder of the second kind and this results in diffraction data in which interference effects between molecular units decrease rapidly with increasing distance from the origin in reciprocal space and are effectively zero except at small angles of scatter. At large angles, therefore, a liquid scatters as though it were an ideal gas composed of its molecular constituents, and the measured diffracted intensity corresponds to the spherically averaged intensity transform

of its constituent molecular units. For smaller angles of diffraction, interference effects between molecules are observed and are the dominant feature of the intensity since internal periodicities are too small to contribute significantly. At still smaller angles even intermolecular periodicites are too small to contribute and the dominant features of the diffraction are determined by the shape of the molecular units. In the case of long polymer molecules this enables one to estimate the radius of gyration.

Bram and Beeman (1971) have collected high and low angle X-ray diffraction data from calf thymus DNA in solutions containing either NaCl or CsCl in the concentration range 0.05m to 1.0m. They compared the calculated spherically averaged intensity transform for A and BDNA with the high angle intensity data obtained experimentally and have concluded that a structure similar to the A form could not be present in the solutions studied. Although the calculated transform for the B conformation gave a better fit to the experimental data, there were some discrepancies which were explained as being due to a layer of countercations surrounding the DNA. The thickness of such a layer was estimated from measurements of diffraction data at small angles, from which they showed that it was possible to detect two radii of gyration. This phenomena was interpreted as being due to the layer of countercations which effectively gives a pseudo rod shaped structure which is of greater radius than the "core" formed by the naked DNA duplex.

However, Bram (1971) has extended the work by studying solutions with different monovalent countercations and has concluded that the form of the high angle scattering can best be explained by a modified B-type structure in which the helix is unwound to give a pitch of 37Å. In model building studies performed by the author, it

has not proved possible to build a stereochemically reasonable model for BDNA with a reduced rotation per residue and without a concommitant increase in the separation between residues. Hence, although it is only required to unwind the helix by 3^{o} to attain a pitch of 37Å, the model is considered to be unlikely on stereochemical grounds.

Nucleohistone complexes have also been studied (Bram, 1972a) and gave slightly different high angle scattering curves which are explained in terms of a model having an increase turn angle per residue resulting in a pitch of 32Å. In order to attain this pitch value it is necessary to increase the turn angle per residue to 40^{o}, a value found to be possible on stereochemical grounds for BDNA in model building studies performed by the author.

Bram (1972) has also reported that the A-T rich DNA's (from Cl. perfringens and B. cereus) give high angle solution diffraction data which suggest that the structure in solution is not that of the B form and is different from DNA's having a lower ($\lesssim$1.5) A-T/G-C ratio. The change is relatively slight; consisting of a broadening of the first maximum. At present it has not been explained in terms of a modification to the DNA structure.

Evidence for polymorphism in solutions of the synthetic polynucleotide poly (dG-dC) poly (dG-dC) has been provided by Pohl (1976) on the basis of ultraviolet circular dichroism studies of ethanolic solutions. All solutions were in 1mM sodium phosphate buffer (pH 7.2) in which the polynucleotide concentration varied between 50 → 150μM. A spectrum similar to that of natural DNA is observed in solutions containing between 0 and 45% ethanol (spectrum a). At 50% ethanol a reversible transition to a different spectrum (b) takes place and this gradually changes upon the addition of more ethanol to give a third distinct type of spectrum (c). The spectra a and b can be produced in CsCl solutions without the

addition of ethanol: the transition taking place at 4M CsCl. Upon increasing the CsCl concentration a spectrum (d) is obtained which is distinct from a, b, or c.

The B conformation is attributed to spectrum a; while the similarity of spectrum c to that of double helical RNA conformations suggests that this spectrum may be produced by the A form spectrum d has been explained as being due to the C form of DNA since it resembles spectra obtained for natural Li DNA. The D form has been tentatively put forward to explain spectrum (b). This is consistent with the findings of Arnott et. al. (1974a) that poly (dG-dC) poly (dG-dC) adopts the D form in fibres.

It seems likely, therefore, that structures of the A genus are not adopted in aqueous solutions of DNA. Moreover, conformations adopted in solutions are probably very similar to the Li DNA B form observed in fibres. The D form of DNA is probably not observed in solutions of natural DNA's, nor in synthetic ones at low salt concentration. It appears possible, however, that conformations observed in solution may vary at least in the turn angle per residue, and that this variation may be dependent upon base composition. However, a general point may be made that the nature of diffraction data from solutions is such that it is not easy to extract definite information about a structure (such as the helix pitch in polymers), and, hence, solution diffraction data provide more general information which has to be interpreted. For this reason, solution diffraction of nucleic acids is probably better used as a fingerprint technique to identify the presence of a known conformation. It is probably much less useful to use solution diffraction as a means of obtaining data on which to base a new conformation.

6.2 Method

Three DNA's were included in a comparative study of their behaviour in fibres as a function of ionic strength and relative humidity: DNA from the bacterium Clostridium perfringens which has a high A-T (69%) content: that from the bacterium Microceus Lysodeikticus which is a high G-C (72%) DNA; and calf thymus DNA which has 40% G-C base pairs.

All fibres were pulled from gels obtained by ultracentrifugation of solutions of the appropriate DNA in the manner described in chapter 2. The DNA was dissolved in solutions containing 0.002M Tris pH 7.6 and of varying sodium chloride concentrations (0.05, 0.02, 0.01 or 0.005M).

A number of fibres were also prepared from two other DNA's from eukaryotic cells: that from salmon sperm and from soft roe of pollak; and from DNA from the bacterium E. coli (51% G-C). Fibres were also prepared from the synthetic DNA poly (dA-dT) poly dA-dT) which contains an alternating sequence of adenine and thymine bases in each chain of the duplex.

6.3 Results

The A⇌B transition was obtained from all the natural DNA's studied though Cl. perfringens DNA did not give an A pattern in fibres obtained from solution having the highest (0.05M) NaCl concentration. No distinct new conformations for the DNA's studied was indicated by the diffraction patterns obtained.

6.3.1 The A⇌B Transition

It is interesting to compare the behaviour of the three DNA's under consideration with regard to the relative humidity at which the A⇌B transition takes place. The studies of Pilet and Brahms (1972) indicate that the A form might be more difficult to obtain for DNA's having a high A-T content and may not be found at all if the

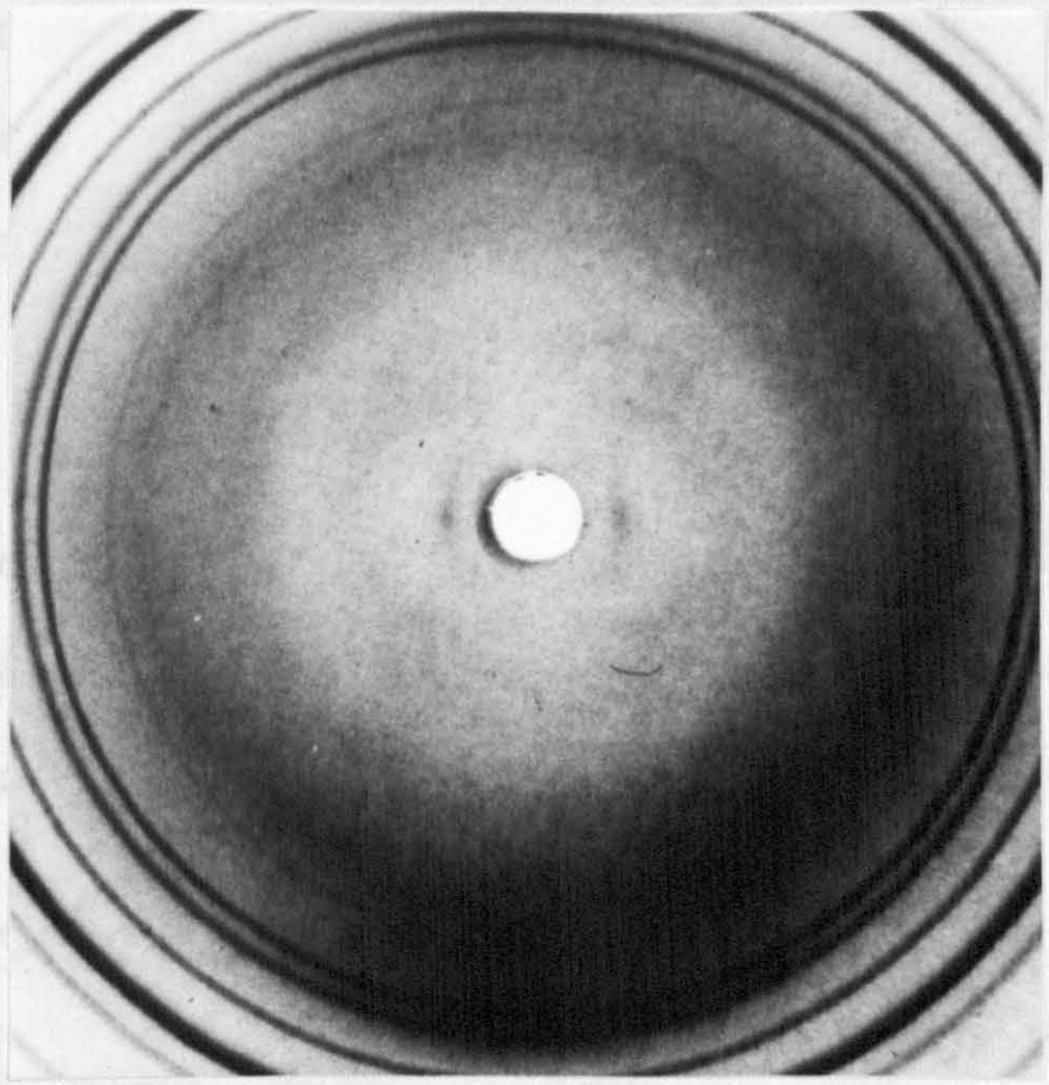


Plate 6.1 Cl. perfringens DNA 44% R.H.
Fibre formed from 0.05M NaCl/tris buffer solution

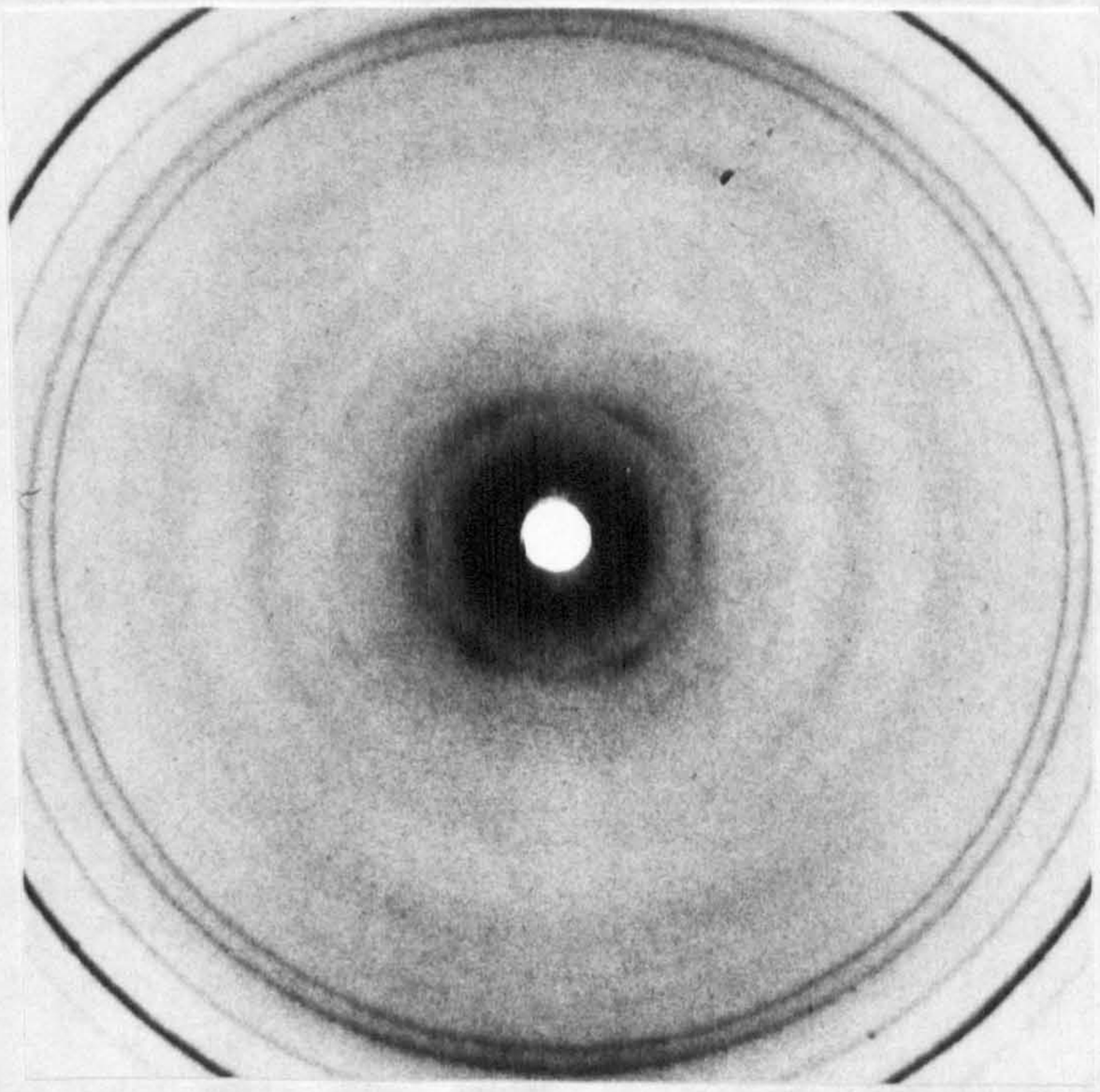


Plate 6.2 Cl. perfringens DNA 75% R.H.
Diffuse A pattern from a salt deficient fibre

G-C content is lower than approximately 30%.

Unfortunately, there is considerable variation in the behaviour of different fibres of the same DNA prepared under apparently identical conditions. This may be due to the fact that the environment in fibres is more difficult to control than is the environment of the DNA in solution or in a film. The results cannot be used, therefore, to give precise data on the variation of the A⇌B transition as a function of the A-T/G-C ratio, though it is possible to give some idea of a general trend.

(a) DNA From Cl. perfringens

Cl. perfringens DNA has been reported by Bram and Tougard (1972) to give two new conformations, the "P" and "T" forms, and to give the A form only in a very narrow range of salt concentrations. In the present study, A patterns were obtained from Cl. perfringens DNA prepared from the solutions containing salt concentrations of 0.005M, 0.01M and 0.02M. A patterns were not obtained, however, from fibres of Cl. perfringens DNA which had been prepared from solutions containing 0.05M NaCl Bram and Tougard (1972) have reported that the A form is not observed in Cl. perfringens DNA at excess salt concentrations below 1%. It is of interest to look at the behaviour of Cl. perfringens DNA in fibres which are salt deficient. In the preliminary stages a number of fibres had been made from calf thymus DNA, purified in the manner described in chapter 2, dissolving some of the dry DNA in a drop of distilled water on the stretching frame. All such fibres gave A patterns at all relative humidities showing that they are highly salt deficient. A number of fibres were also made in this manner from Cl. perfringens DNA. In this case, however, the fibres gave either a diffuse A pattern (e.g. plate 6.2) or a very diffuse B type pattern (e.g. plate 6.3). This behaviour was observed from batches of DNA which gave good fibres

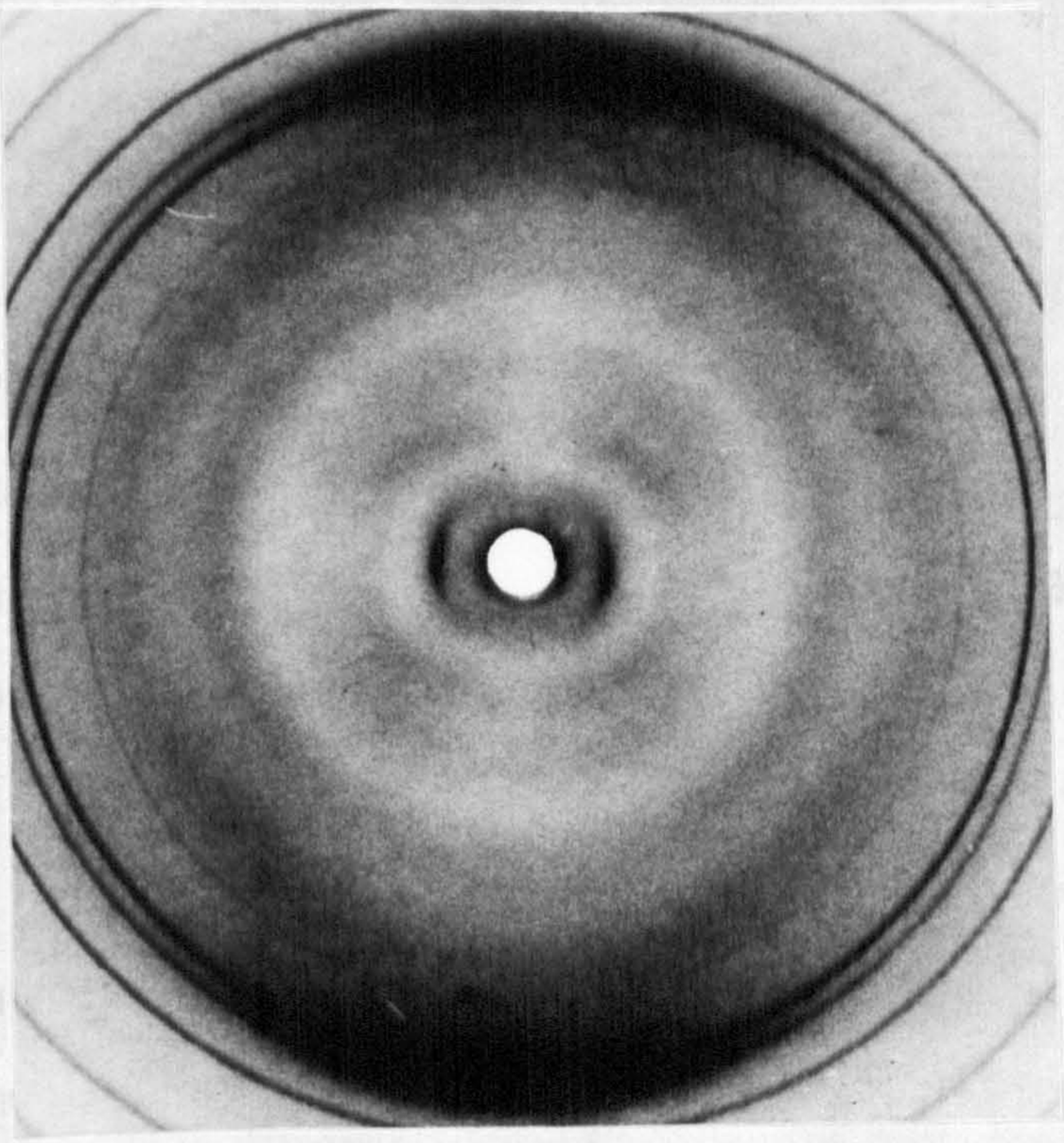


Plate 6.3 Cl. perfringens DNA 66% R.H.
Diffuse B type pattern from a salt deficient fibre

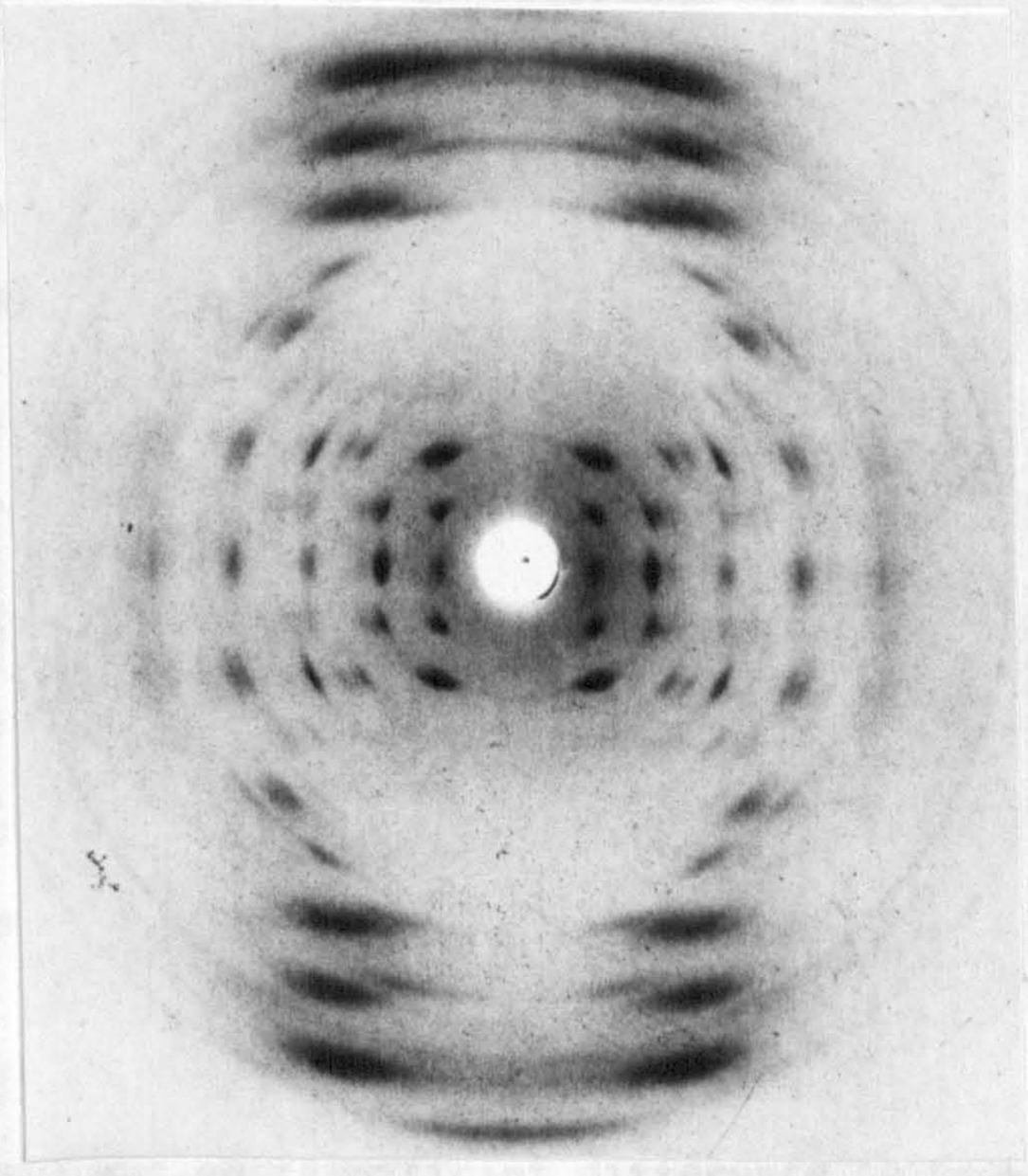


Plate 6.4 Cl. perfringens DNA 66% R.H.

when they were prepared by other methods. The diffuse patterns were obtained from samples of the DNA from which that part of the purification procedure involving phenol had been omitted. Hence, it appears that the A form of Cl. perfringens DNA may be difficult to obtain in fibres which are highly salt deficient.

Some good A patterns were obtained from Cl. perfringens DNA (e.g. plate 6.4) under all other conditions used. However, a large number of patterns were obtained which showed some B character around the central area while not having a significant 3.4A meridional reflection. Selsing and Arnott (1976) report that some difficulty in obtaining the A form is observed in fibres of Cl. perfringens DNA which has not been purified by gel filtration, although A patterns are readily obtained at lower values of the relative humidity, when this process has been carried out. Evidence is quoted that teichoic acid, present in high concentrations in gram positive bacteria such as Cl. perfringens, affects DNA conformation in this way. Hence, it is proposed that residual teichoic acid may be responsible for the difficulty in obtaining the A form reported by Bram and Tougard (1972). Since a gel filtration step is not included in the present purification procedure, teichoic acid, if present, might not be removed, and might explain the rarity with which A patterns have been obtained which are entirely devoid of residual B type character. The results presented here are, nevertheless, in accord with the findings of Selsing and Arnott (1976), that the A form of Cl. perfringens DNA is more readily obtained than is reported by Bram and Tougard (1972).

(b) Calf Thymus and M. Lysodeikticus DNA

There was no significant difference between the behaviour of fibres of M. lysodeikticus DNA and fibre of calf thymus DNA in regard to the A $\rightleftharpoons$ B transition within the limits of the error imposed by

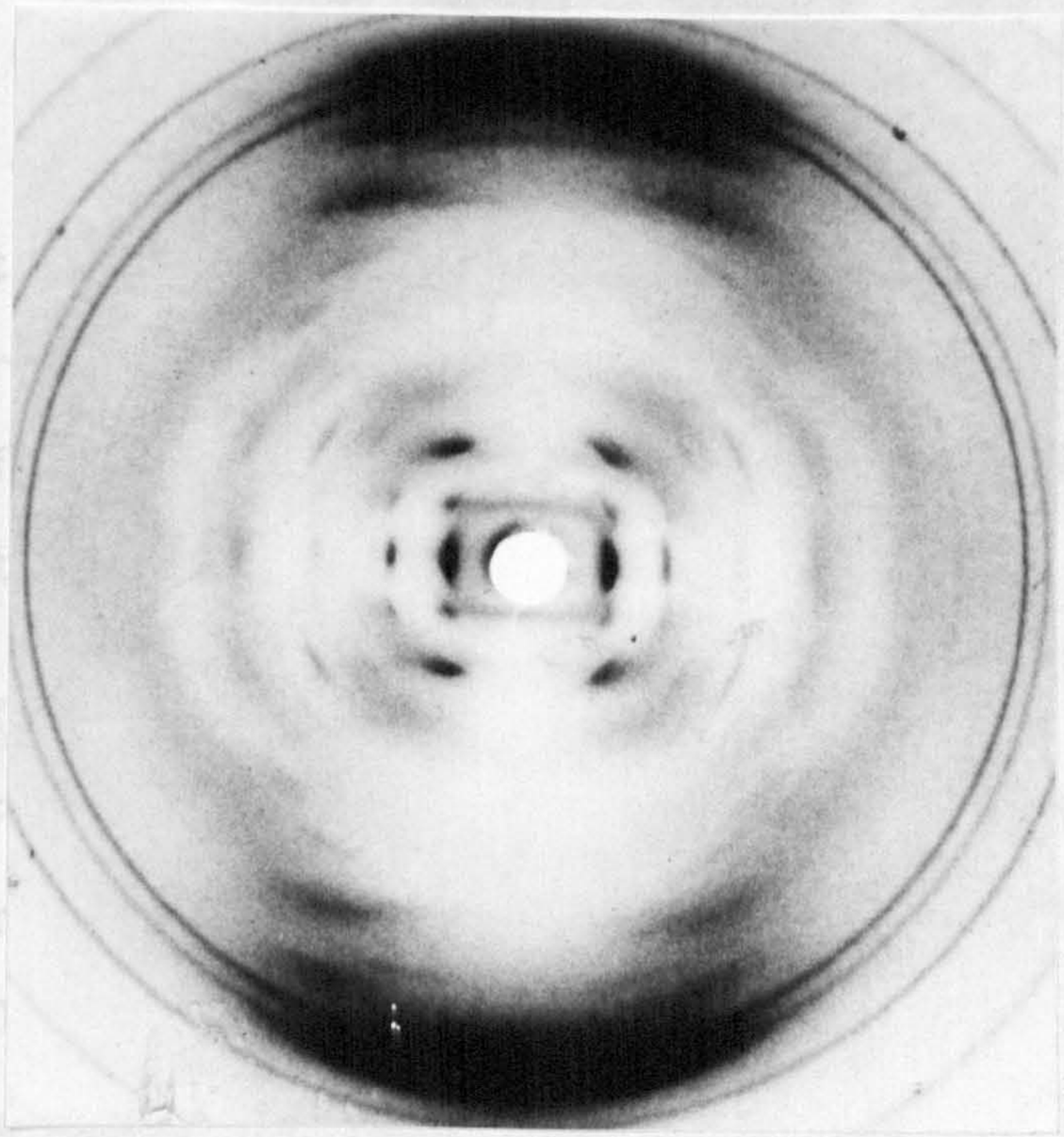


Plate 6.5 Cl. perfringens DNA 66% R.H.
Photograph showing a mixture of the A and B forms

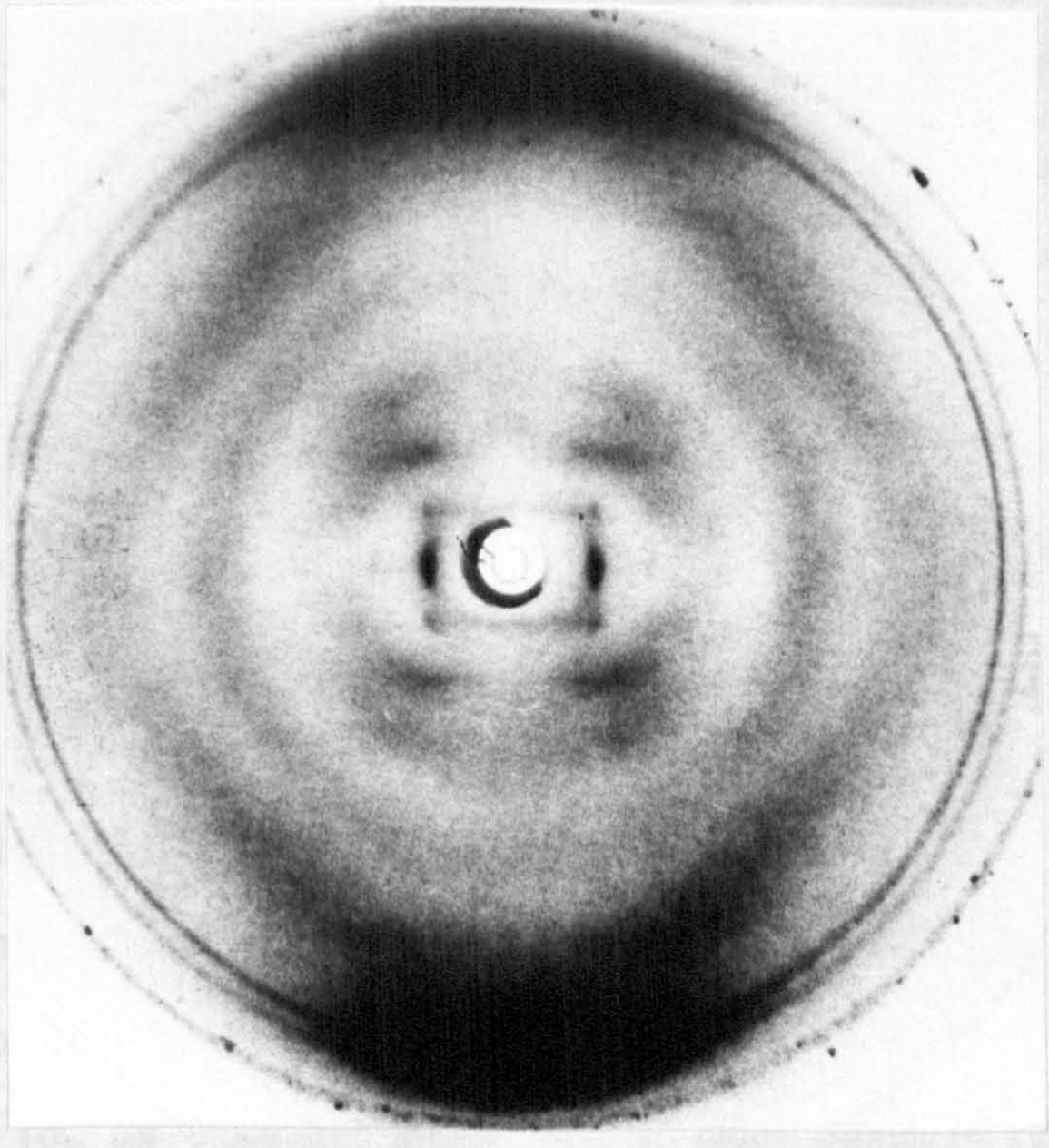


Plate 6.6 Cl. perfringens DNA
"Low pitch" (≈32Å) B type diffraction pattern

variations in behaviour between fibres of the same batch. Generally, the fibres gave the A ⇌ B transition between 75% and 92% relative humidity at all salt concentrations used in the original solution, although a small percentage of fibres gave anomalous results: an increasing number of fibres giving the A form at 92% relative humidity being observed at lower salt contents.

(c) Poly (dA-dT) Poly (dA-dT)

Fibres of poly (dA-dT) poly (dA-dT) were made from gels centrifuged for 14 hrs at 50,000 r.p.m. from solutions containing 0.002m NaCl. The salt concentration is, hence, low and it is under these conditions that poly (dA-dT) poly dA-dT) is reported to give the D form (Davies and Baldwin, 1963; Arnott et. al. 1974). In fact most patterns gave the A form of DNA and this appeared to be stable. This is significant since it is in contrast to the behaviour observed by Arnott et. al. (1974) in that their fibres which gave DDNA were prepared under low salt conditions which normally favour the A conformation, and the addition of salt to the fibre produced the B form without any apparent appearance of the A form. Davies and Baldwin (1963) observed the A form for poly (dA-dT) poly (dA-dT), but it was metastable and subsequently reverted to the D form on rephotographing the unmodified fibre.

One fibre of the batch prepared above gave a pattern (plate 6.10) at 44% relative humidity which appeared to be semi-crystalline and which had a value for the helical pitch of approximately 31Å. Though reduced from the value in BDNA, this pitch measurement is not as low as that observed in the case of DDNA (24.5Å). However, the intermolecular separation is reduced in comparison with values for this parameter measured from patterns of ADNA and BDNA and is equal to the value observed for DDNA (Arnott et. al., 1974).

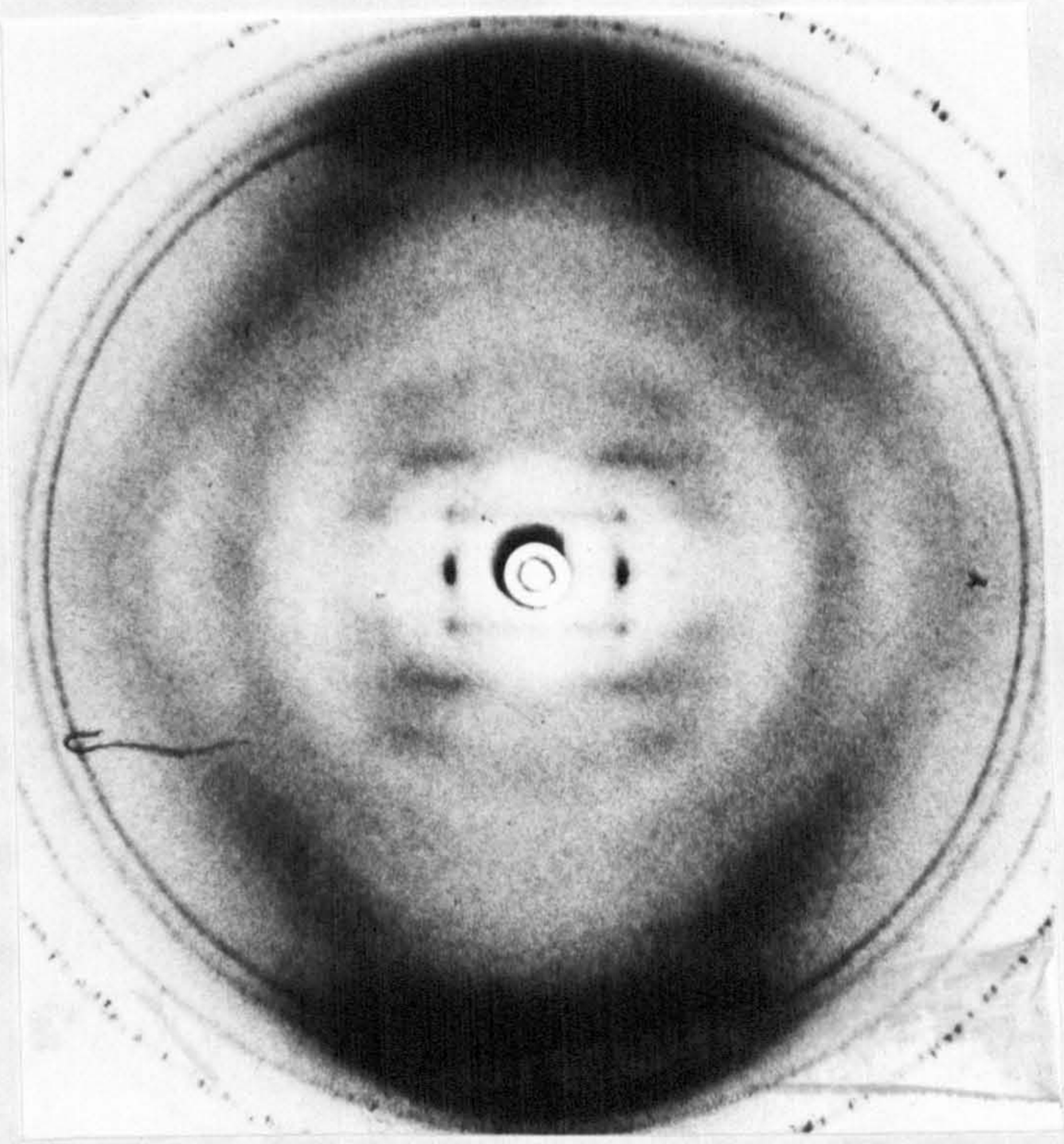


Plate 6.7 Cl. perfringens DNA 66% R.H.
"Low pitch" (≃27.5Å) B type diffraction pattern

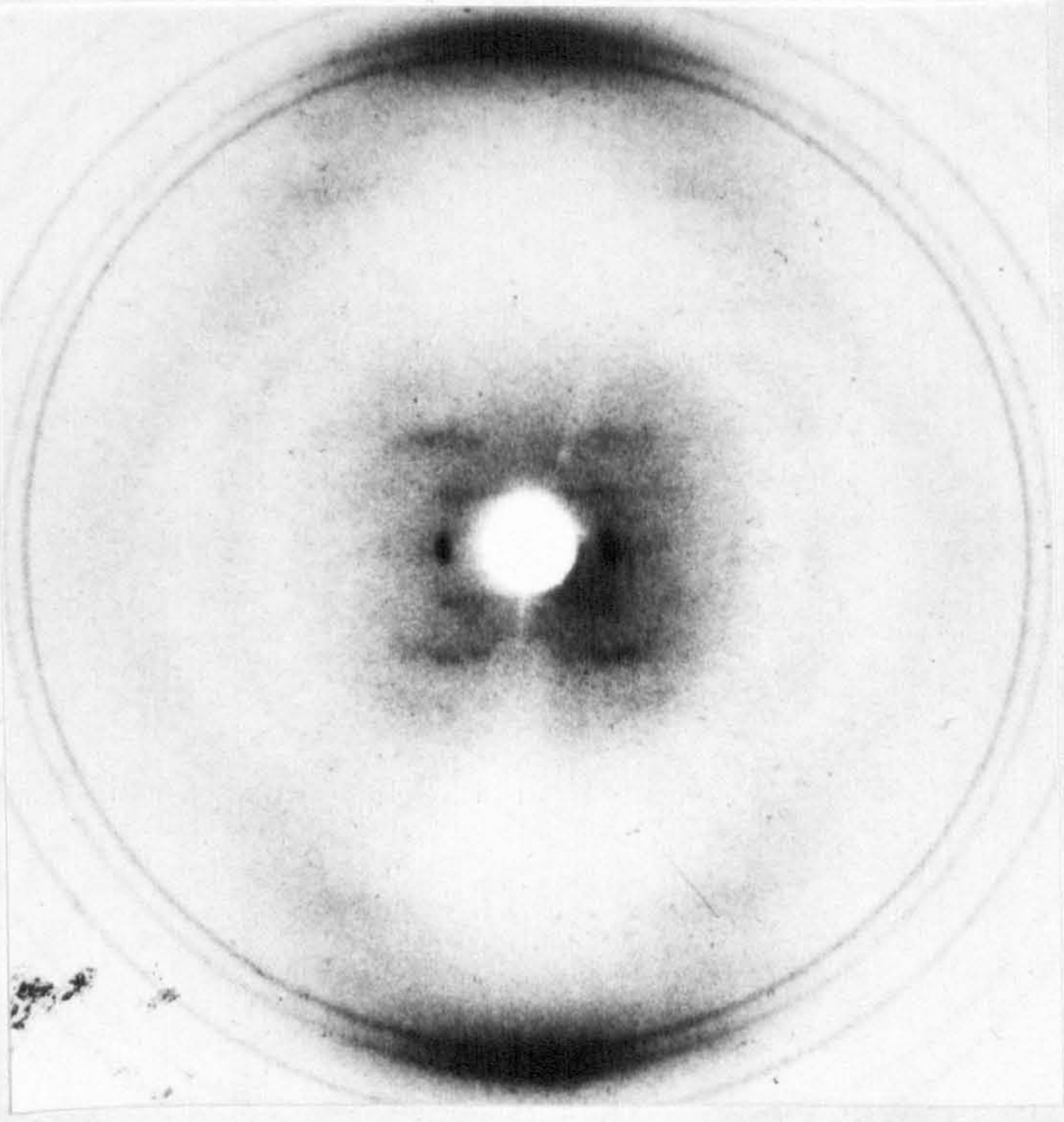


Plate 6.8 Calf Thymus DNA 66% R.H.
"Low pitch" (≃32A) B type pattern

The fibre was rephotographed at 75% relative humidity when a much more diffuse pattern was observed having a similar intermolecular separation: the layer lines being too diffuse, however, to allow an accurate measurement of the helix pitch.

In order to attempt to reproduce the first photograph which was under-exposed, the fibre was photographed yet again at 44% relative humidity but now gave a rather poorly oriented A pattern (plate 6.11). Subsequent rephotographing at 75% and 92% relative humidities respectively, gave similar A patterns; while an exposure at 98% relative humidity yielded a very poorly oriented pattern from which estimates of the pitch value and the intermoecular separation indicated that the DNA was in the B form. The "reduced pitch" form observed in the first photograph was never observed again in this fibre. Thus in the results presented here, it is the A form which is stable and which is reverted to by the "low pitch" form found in the first instance.

This regular appearance of A patterns in poly (dA-dT) poly (dA-dT) fibres has also been observed independently by Dr. C. Nave in this laboratory, and contradicts the general trend that the A-T rich DNA's should not adopt a stable A form.

6.3.2 B Type Patterns Observed

In view of the evidence provided by Bram indicating that the B form of DNA might consist of a family of related structures, an examination was made of the "B" patterns obtained in the present study to search for any significant variation in the nature of the intensity data. A number of patterns were observed which had features different from those attributable to the accepted B structure of DNA (Arnott and Hukins, 1972). These were generally of a very poor quality and did not yield sufficient accurate intensity data to permit a detailed conformational study. Hence, the procedure adopted

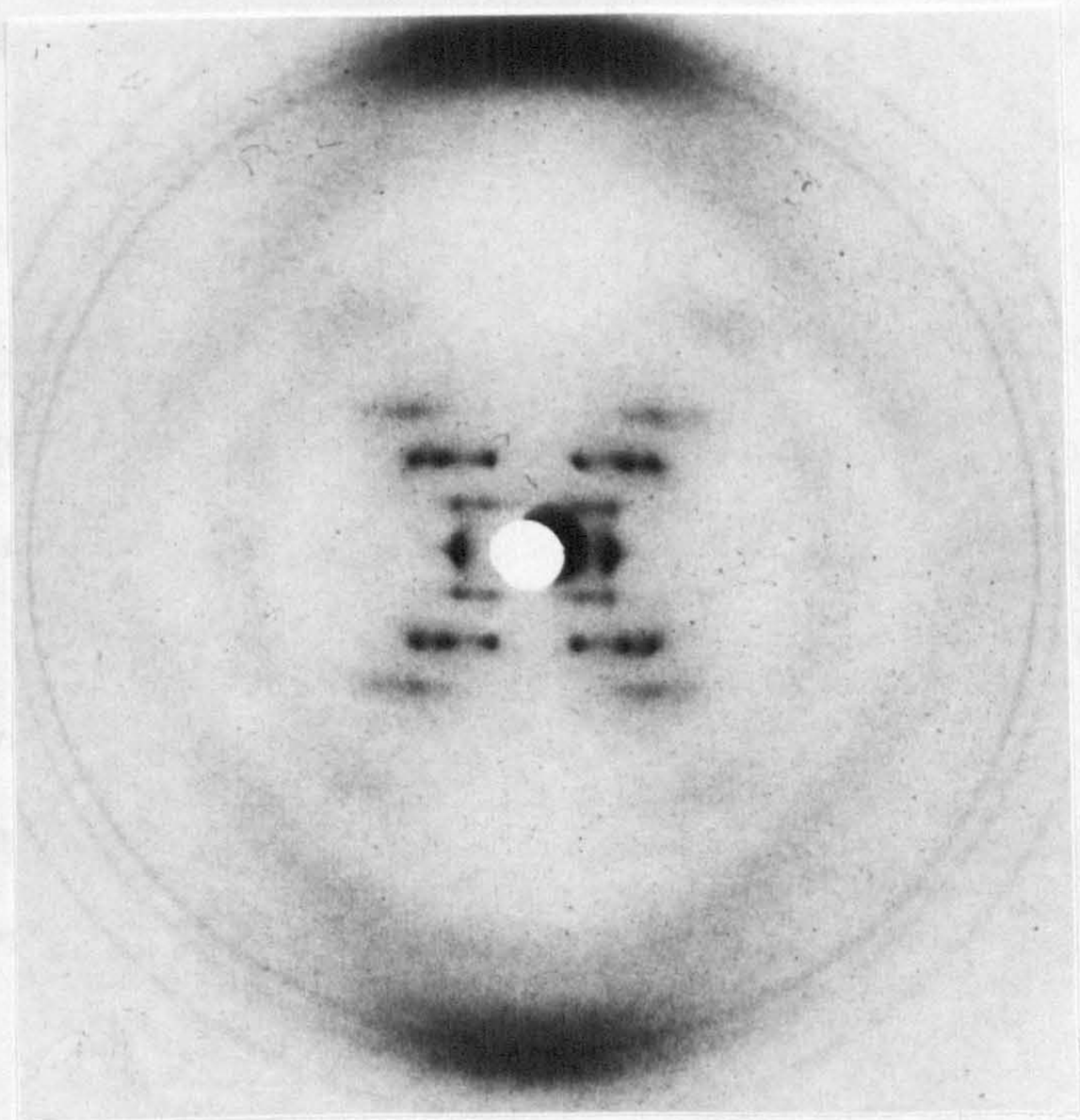


Plate 6.9 M. Lysodeikticus DNA 92% R.H.

is to try and equate the observed diffraction with that produced by a known DNA conformation whose structure has been analysed. Where necessary, possible modifications to this structure have been made to account for special features of the diffraction intensity.

B type patterns were obtained from all DNA's studied. The best B patterns were obtained from M. lysodeikticus DNA (plate 6.9) while those from Cl. perfringens DNA were of a rather poor quality. If the difference is not an artefact produced by a higher level of impurities in the Cl. perfringens DNA, then it is sufficiently marked to be of significance. Both DNA's contain a similar level of protein impurity as measured by the A_{260}/A_{280} ratio, and well oriented A patterns (e.g. plate 6.4) have been obtained from Cl. perfringens fibres which nevertheless gave rather poor B patterns. However, since the A form is highly crystalline, the quality of such patterns may not be such a sensitive function of the level of impurities as is that of the less well packed B conformation.

The layer line spacing and the position of the first equatorial reflection were measured in the patterns obtained from all DNA's, and in each case, patterns having a pitch of 34Å and the normal intermolecular spacing found in sodium BDNA were observed. However, some of the poorer quality patterns from calf thymus and Cl. perfringens DNA had layer line spacings corresponding to reduced values for the pitch length while the position of the first equatorial spot often indicated a reduced value for the intermolecular spacing. The average value of the pitch in these photographs was around 32Å, although considerable variation was observed. Some "low pitch" patterns, with the pitch and intermolecular separation values indicated, are shown in plate 6.6 and 6.8. A similar reduction in the value of these two parameters was observed for the pattern of plate 8.1 which was produced from a gel containing DNA and native chlorpromazine. These effects are probably not directly due to the

chlorpromazine (see chapter 8) and, hence, it may be that it is an example of the type of pattern observed in fibres of pure DNA. Since the layer lines are well defined in this pattern it has been analysed and the measured intensity values along layer lines 1 and 2 are shown in figure 6.5. The graphs represent the integrated intensity along the layer line (Marvin et. al., 1961), which has been corrected to allow for effects of molecular rotation about the fibre axis by mutiplying by the factor $2\pi\xi$ (Franklin and Gosling, 1953). Similar graphs obtained from BDNA are shown in figure 5·23.

The pattern does not give diffraction intensity similar to that observed in either B'DNA or in DDNA. A particular difference between B'DNA and the pattern of plate 8.1 is that while the former predicts an enhancement of the first and third layer lines in comparison with the second; in the pattern it is the second layer line which is enhanced. It does not resemble DDNA in that the pitch value (30.9Å) is much larger than that observed in DDNA (24.5Å). There are a number of similarities between the pattern of plate 8.1 and the diffraction given by CDNA, however. In particular, the pitch of CDNA is approximately the same as that observed in the pattern in plate 8.1, and the position of the first equatorial spot at $0.056Å^{-1}$ is in approximate agreement with the values observed for CDNA, and corresponds to a smaller intermolecular spacing to that observed in the B form. The base pairs in CDNA are tilted to a slightly greater degree (6.0^{o}) than are those of BDNA, and they are moved further away from the helix axis by approximately -2.0Å (see chapter 1 for the definitions of the sign and magnitude of these parameters). The effect of this latter change is to reduce the difference in depth between the large and the small grooves, and, hence, the duplex more closely resembles two helices separated by half a pitch length. Since in such a structure only the

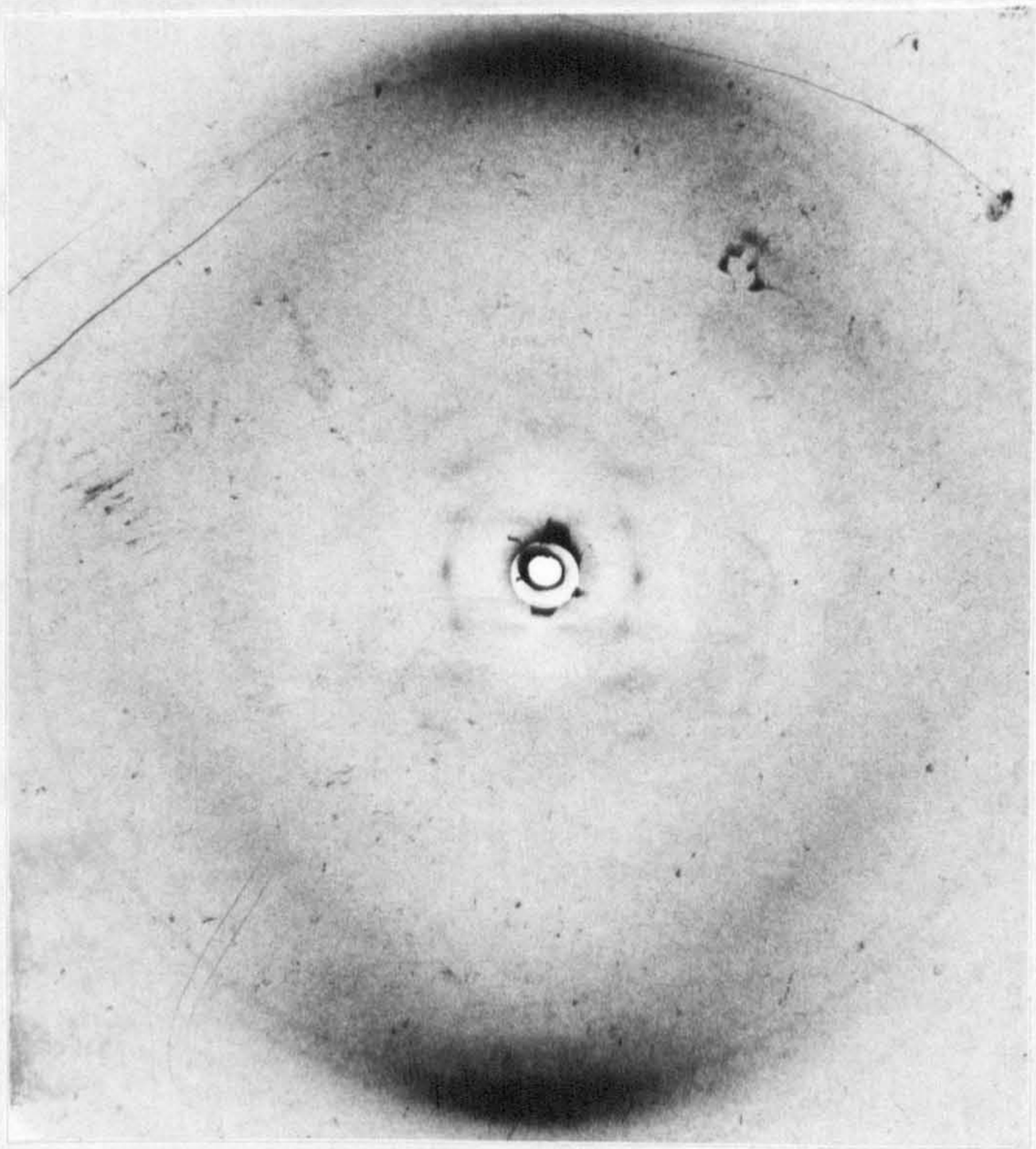


Plate 6.10 Poly(dA-dT) Poly(dA-dT) 44% R.H.
"Low pitch" (≃31Å) B type pattern

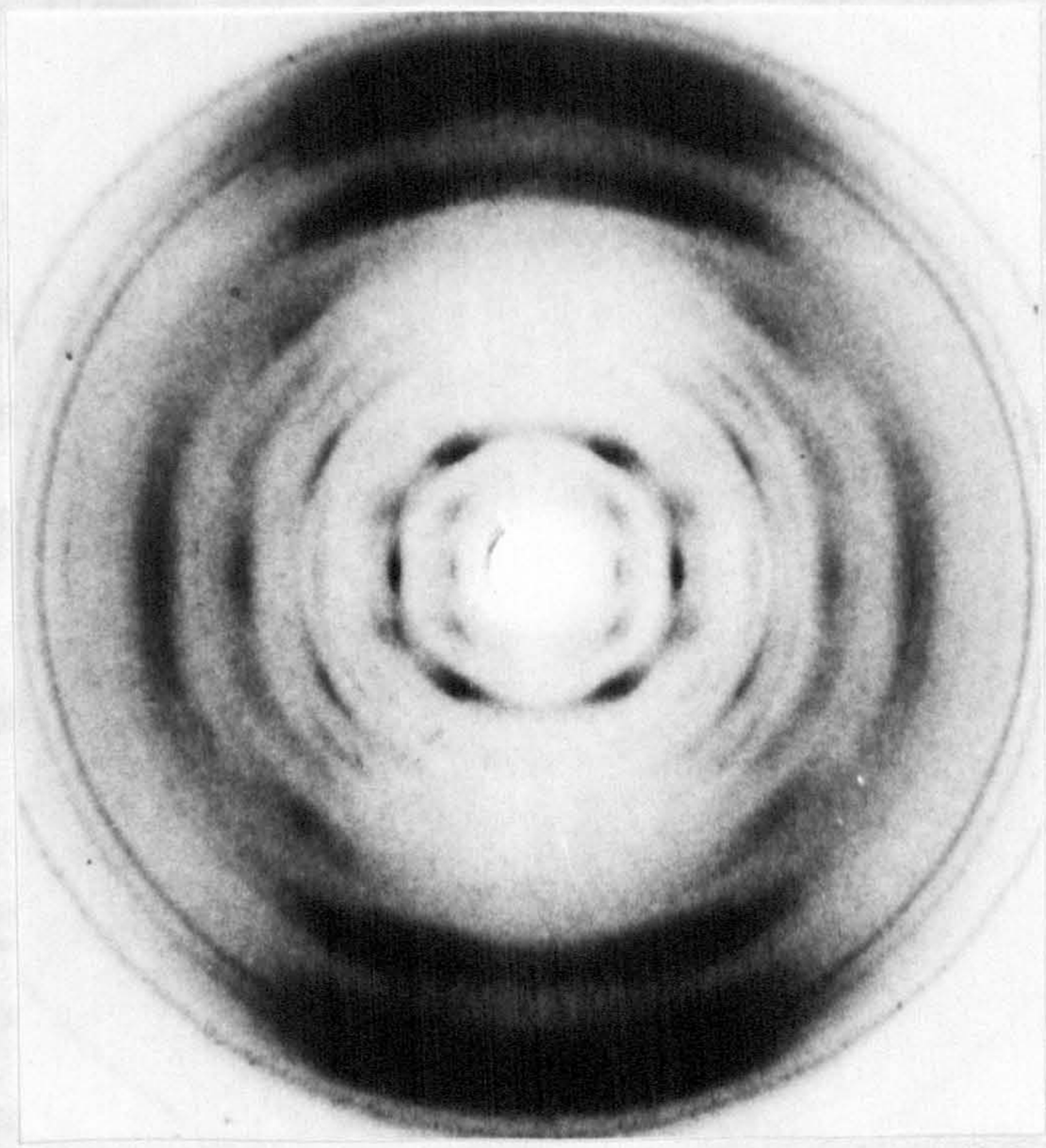


Plate 6.11 Poly(dA-dT) Poly(dA-dT) 44% R.H.
Same fibre as in plate 6.10 above pattern obtained upon rephotographing

even numbered layer lines would be non-zero (see chapter 3), the result of the change which does not remove entirely the difference in depth of the two grooves, is to enhance the second layer line at the expense of the first and third. This is observed in the transform of the CDNA model (Marvin et. al., 1961), and in the pattern of plate 8.1. Finally, a characteristic feature of a C pattern is the appearance of intensity around 0.1Å^{-1} along the first layer line. This is a region of low intensity in BDNA and B'DNA. An intensity peak is observed in this region of the intensity data obtained from the pattern in plate 8.1. There is a marked increase in the height of the peaks on the first layer line with increasing ξ value. The transform of the CDNA model of Marvin et. al., (1961) is given for layer lines 1, 2 and 3 in figure 6.3, where it can be seen that the first layer line peaks are of approximately identical height and, therefore, do not give the same intensity ratio as that observed in figure 6.5. However, examination of figure 5.23 where the calculated intensity transform for BDNA is compared with the observed intensity data, shows that the observed peaks at increasing ξ values become progressively larger than the corresponding peaks in the calculated transform, and, hence, the enhancement of the peaks at increasing distance along a layer line may be an artefact of the method of measurement. It appears likely, therefore, that the patterns in plate 8.1 is produced by the C form of DNA or some similar conformation.

The "low pitch" pattern shown in plate 6.7 is obtained from Cl. perfringens DNA. The intensity of the first and second layer line streaks in relation to that of the background is very low in this pattern and, hence, the layer line intensity is difficult to measure accurately, nevertheless, intensity measurements indicate that it possesses many of the characteristics of the pattern discussed above. The intensity of

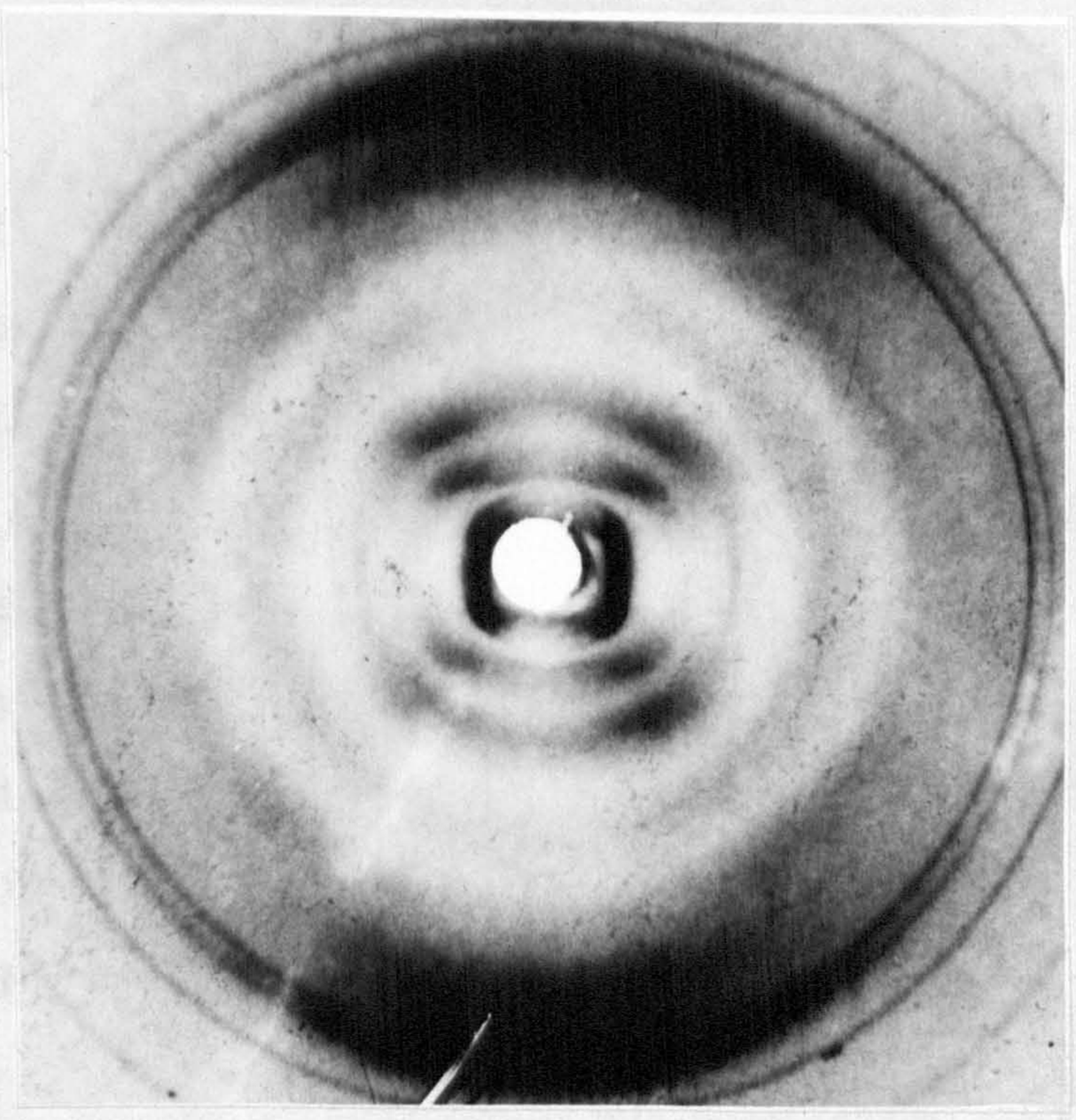


Plate 6.12 Poly(dA-dT) Poly(dA-dT) 98% R.H.
B type pattern exhibiting an unusually strong 3rd layer line

the second layer line is enhanced, while the intensity peak at 0.1$\mathring{A}^{-1}$ on the first layer line is observed. Moreover, the position of the first equatorial reflection is identical to that of the pattern. However, the pitch of the DNA is approximately 27.5$\mathring{A}$, a figure much lower than that observed for any B genus structure other than DDNA. It does not appear likely, however, that the pattern in plate 6.7 is produced by a structure which is similar to DDNA. In particular, the first peak on the first layer line of DDNA is more intense than the equivalent peak on the second layer line: the converse of that observed in the present pattern. Moreover, the pattern exhibits strong meridional reflection around 3.4$\mathring{A}$, whereas the D conformation gives a relatively weak meridional which is at 2.42$\mathring{A}$. Hence, the structure producing plate 6.7 probably does not have a large base tilt as does the D conformation (-16.0^{o}). Hence, it seems likely that the conformation is similar to that of CDNA, but is slightly modified so as to produce a helix of lower pitch. Since the general nature of the intensity is similar to that produced by CDNA, it is likely that the modification required to produce the very low pitch does not involve considerable modification to the C conformation. The required pitch would be produced if the C conformation were to be unwound so as to have $8^{1}/_{3}$ residue with the rise per residue (3.03$\mathring{A}$) of C form, and such a conformation would probably represent the least modified conformation giving the desired pitch value. Work is being carried out by Dr. C. Nave in this laboratory to attempt to build a model having similar base disposition to that observed in CDNA, but having a reduced number of residues per turn to accommodate the requirement of a lower pitch value.

The enhancement of 1st and 3rd layer line intensity reported by Bram has not been observed by the author from fibres of natural DNA, but Dr. C. Nave in this laboratory has reported that such enhancement is indicated in diffraction patterns obtained from gels of Cl. perfringens DNA, and has actually been observed in the diffraction patterns from fibres in which the DNA is poorly oriented. It does not appear in diffraction patterns from well oriented fibres and it may be that packing considerations favour the appearance of the B form in such fibres. If this is the case, then fibres of DNA in which segments of the duplex are in the B' form will not only possess molecular disorder, but will be badly packed. Hence, only poor quality diffraction patterns would exhibit the enhanced layer line intensity reported by Bram if this were the explanation.

An example of a pattern having enhanced third layer line intensity has been obtained by the author from the synthetic DNA, ply (dA-dT). poly (dA-dT) (plate 6.12). This is the alternating polymer, for which the B' form has not been reported, though it seems possible that this, or a similar conformation is present in this case. The fibre is non-crystalline and poorly oriented, which is in contrast to the highly crystalline patterns obtained by Arnott et. al. (1974). In view of the poor quality of the diffraction data obtained it is not possible to determine whether the pattern in plate 6.12 corresponds to B'DNA or to a slightly modified conformation.

The results presented in this section support the idea that there is a family of B genus structures for the sodium salt of natural DNA, and that the base composition and/or sequence might have an effect on the conformation adopted. It is of interest that no variations upon the A form have been noted for any duplex of either natural or synthetic DNA. Arnott and Selsing (1974) have described a structure of the A genus to the triple standard DNA poly (dT) poly (dA)

poly (dT). It should be noted that the general nature of the starting model was derived assuming that the furanose ring pucker was the same (C3' endo) as that observed in the triple stranded RNA structures poly (U) poly (A) poly (U) (Arnott and Bond, 1973a) and poly (I) poly (A) poly (I) (Arnott and Bond, 1973b); while the base arrangement was taken to be similar to that found in the poly (U) poly (A) poly (U) structure. Hence, the nature of the starting model constrains the structure for poly (dT) poly (dA) poly (dT) to be of the A genus; so that the possibility exists that the postulated structure is partially an artefact of the method of analysis.

There are, however, good reasons for supposing that the triple stranded form will adopt a conformation of the A genus. The postulated base pairing scheme is shown in figure 6.6, and since the hydrogen bonding possibilities of the bases are limited, it probably represents a unique solution. If the helix axis is situated in the postulated position of the Arnott and Selsing (1974) structure, then all three chains of the triplex are approximately equivalently situated with respect to the helix axis. The position of the axis as it would be relative to the T_1-A base pair in BDNA is also shown in figure 6.7, and it can be seen that such a disposition of this base pair with respect to the helix axis would result in the position of the two thymine chains relative to the helix axis being quite different. Hence, symmetry considerations dictate that the base pairs T_1-A and A-T_2 should be moved forward (see chapter 1) from the helix axis. The author has performed a model building study to determine the possible effect of changes in the base disposition on the furanose ring puckering. It was not found to be possible to build stereochemically reasonable models including a C3' exo sugar pucker with base pairs moved forward from the

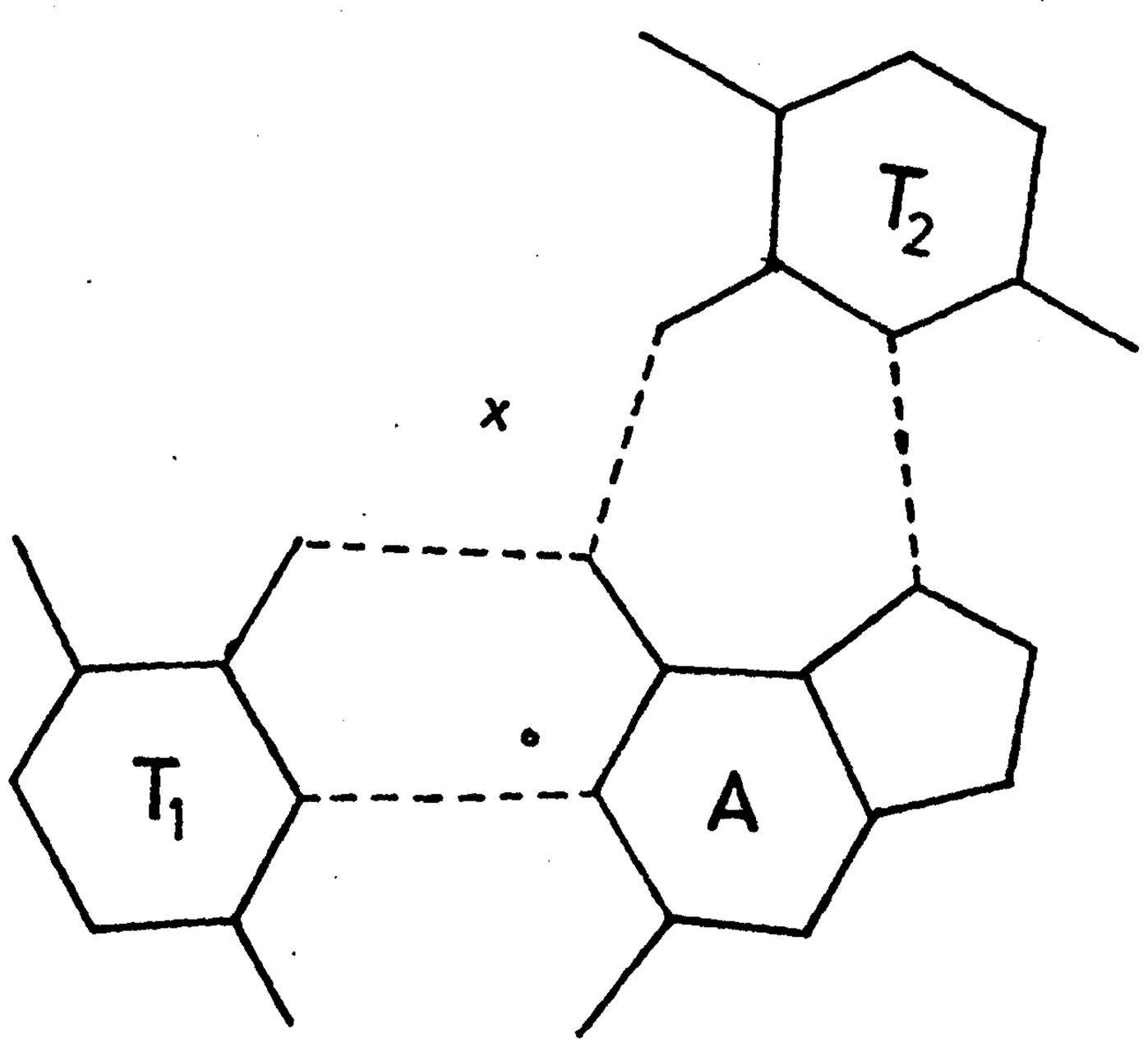


Fig. 6·6 BASE PAIRING IN POLY(dA)·POLY(dT)^2

x=HELIX AXIS

•= " " IN BDNA

helix axis, and incorporating the requirement of a right-handed helical structure. Similar difficulty is found in building models incorporating C3' endo sugar puckers with base pairs moved in a negative direction from the helix axis. The study of possible models was not sufficiently rigorous to eliminate the possibility of models which violate the above rules, although the idea is supported by the observed properties possessed by A and B genus structures. Hence, the C3' endo sugar pucker is probably imposed by the nature of the base disposition to the helix axis, which is, in turn, a consequence of the triple helical nature of the DNA.

For this reasons outlined above, the triple helical structure must be considered as a special case, and hence can be considered as being separate from the main set of double helical structures. If the A form of DNA is considered to be the invariant form, while the sodium "B" form actually consists of a family of related structures, then it is of interest to determine what biological significance, if any, this has for the behaviour of DNA in vivo. This is discussed in chapter 9.

6.4 Discussion

6.4.1 B Type Structures

Evidence presented here indicates that the conformation of B type structures in fibres of sodium DNA may in fact consist of a family of related structures so that the accepted B conformation (Arnott and Hukins, 1972) for lithium DNA may rarely, if ever, be the conformation present in fibres of sodium DNA. It has also been indicated that the appearance of modified forms having reduced pitch values may be related to the base composition and/or sequence. Since the parameter value (pitch and intermolecular separation) which characterise these modified forms are variable it seems likely that they are not distinct conformations in the same sense as the A, B or C forms.

The changes are not sufficiently characteristic in the patterns obtained in this study, to warrant a classification into different conformations such as the 'P' and 'T' forms described to Cl. perfringens DNA by Bram and Tougard (1972). It is a distinct possibility that the duplex is not regular in these forms and that different sections in the length of the DNA are in somewhat different conformations. In relation to this, the patterns exhibiting a distinct mixture of A and B forms are of interest, since there is obviously a possibility of either an alternation of A and B DNA regions along a single DNA molecule, or the existence side by side, of domains of molecules all in the B conformation interdispersed with ADNA crystallites. Since, the ADNA appears well packed in many of these photographs, it does not seem likely that the molecules could not be regular ADNA helices. If it is the case that the two conformations are segregared into different domains in the fibre, then it is apparent that the reduction in pitch observed in the B portion of the pattern results from modification of the B conformation itself and not from a mixture of A and BDNA lengths along one duplex. The reduced pitch of the BDNA portion of these patterns appears to be a feature of all such A/B mixtures and may arise because the B type DNA is in a transition form between the A and B conformations.

The fact that the B type molecules appear to be clustered in domains around ADNA crystallites suggests the possibility that the change from the B to the A conformation in fibres may be a cooperative one amongst groups of molecules within different domains of the fibre. In this case the ADNA features of patterns of A/B mixtures would be given by domains of molecules which had completed the transition to the A form and ADNA type packing. The B type features would be given by molecular domains which were in a transition state between the

two conformations. Such cooperative changes could explain why the first layer line and meridional spot of the BDNA portion of the patterns remains distinct. If the individual molecules undergo the transition from the B to A form independently, then it would be difficult to imagine how distinct features of BDNA could be retained in the pattern.

The question arises as to whether the B type patterns observed which have a reduced pitch and intermolecular separation, but no apparent ADNA component are fibres in which the DNA is in a transition state in all domains, or whether the differences result from a novel conformation, or family of conformations, which may appear under certain environmental conditions and which are not related to any conformation which may be adopted in passing from the B to the A form. This question is of more than merely academic interest, since the appearance of a transition form, and certainly of domains of molecules in the transition form, may well be a function of the fibre state, while the existence of a distinct family of B type conformations which may appear under certain environmental conditions, and perhaps also dependent on the base composition of the DNA, are of more general interest and biological significance since they might appear in solution or in vivo.

6.4.2 The A⇌B Transition

Although the existence of two forms (A and B) for the sodium salt of DNA in fibres, and the transition between them as a function of the relative humidity, was demonstrated by early diffraction experiments (e.g. Franklin and Gosling, 1953), the mechanism of the transition and its biological implications if any have not been rigorously investigated.

It is difficult to obtain evidence relating to the mechanism of the A ⇌ B transition using diffraction techniques since data cannot be collected sufficiently rapidly for the change to be followed. Evidence has been obtained in the present study from fibres exhibiting an A/B mixture indicating that, at least in the fibre state, the transition may be cooperative within domains of adjacent molecules. Patterns exhibiting A/B mixtures have been reported by Franklin and Gosling (1953), though no pitch measurements are given. The B type component of all such patterns obtained by the author exhibited reduced pitch values in comparison with BDNA (Arnott and Hukins, 1972). It appears likely that the ADNA component of the pattern is produced by domains of DNA molecules which have cooperatively undergone the change to the A conformation concommitantly with the adoption of A type packing so as to produce ADNA crystallites inside the fibre. The B type diffraction may be produced by domains of molecules which have not undergone the cooperative transition, and are in some conformation representing a transition state between the A and B forms. The pitch values obtained from the B component of such diffraction patterns are around 32Å and are similar to the values observed in the low pitch patterns described earlier. A conformation similar to that of CDNA (Marvin et. al., 1961) has been ascribed to these structures and it is possible that a conformation similar to that of CDNA might be the transition form. At low values of the relative humidity and salt concentration the A form is observed in fibres of NaDNA; while under equivalent conditions LiDNA gives the C form. This may reflect the fact that LiDNA cannot give the A form, but under extreme conditions can give the transition state; the C form.

If the adoption of the A structure always involves the cooperative change of a group of molecules to form a crystallite, then the possibility arises that the adoption of the A form in fibres is impossible without the concomitant adoption of ADNA packing. If this is the case it implies that isolated molecules do not adopt the A form and, hence, reduces the likelihood that the A structure is biologically significant. Since the sodium B and "C" conformations are not fully crystalline, and, hence, the adoption of exact crystalline packing is not essential to their appearance, it appears more likely that B genus structures are the ones of biological importance.

6.5 Conclusions

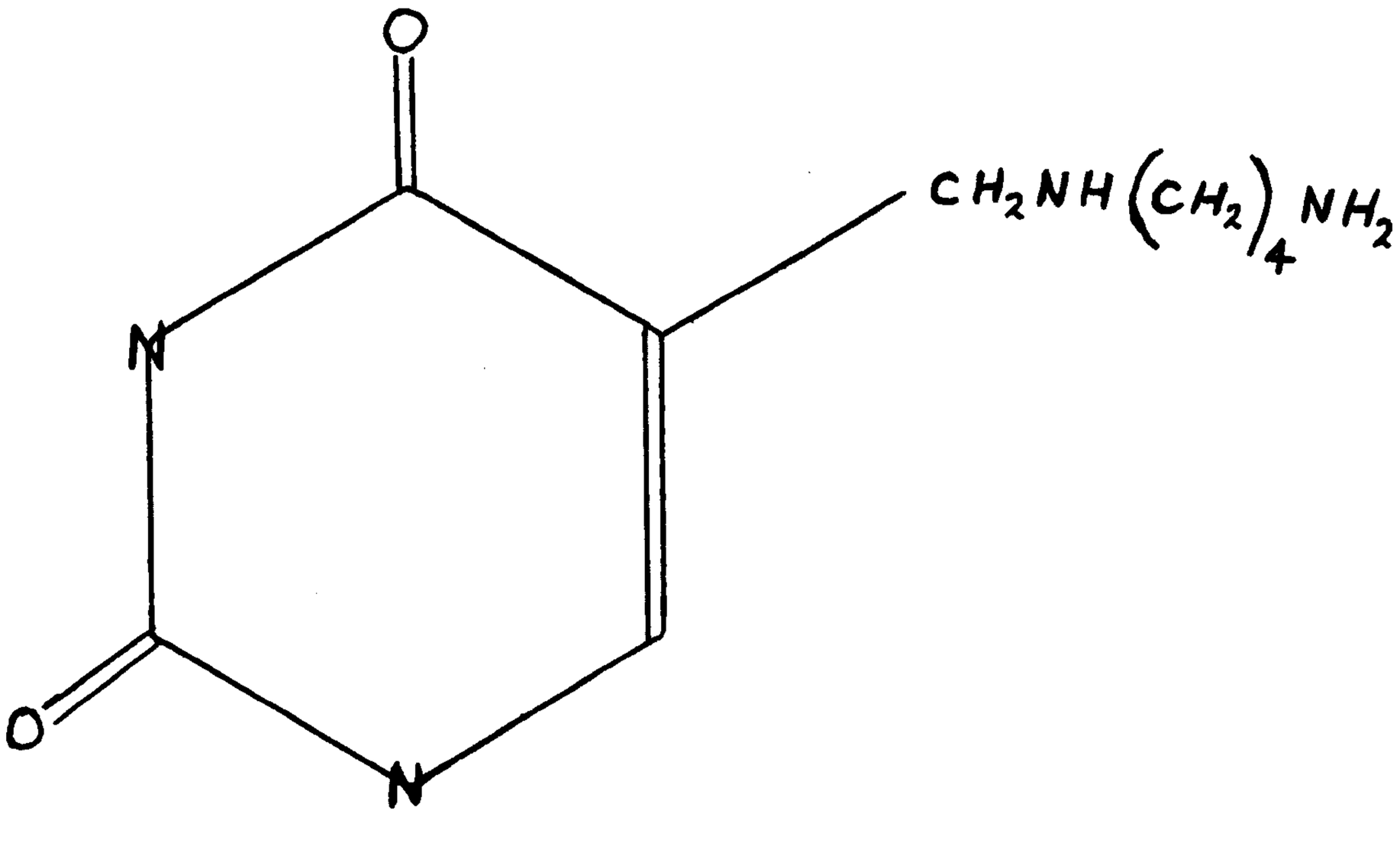
The sodium salt of DNA's from a number of sources have been examined using the technique of X-ray diffraction. Normal A and B forms have been obtained from all DNA's studied. Evidence is presented for the existence of a sodium C form and it is suggested that this may represent a transition state between the A and B forms. No evidence has been obtained which supports the idea put forward by Bram and his co-workers suggesting that specific conformations may be peculiar to particular natural DNA's as a function of their base composition and/or sequence. However, the "low pitch" B type patterns described here may be the forms observed by Bram. It should be noted that evidence for the C form being a transition state in the A$\rightleftharpoons$B transition has been obtained by Brahms et. al. (1973) on the basis of infrared linear dichroism studies.

CHAPTER VII

DNA FROM THE BACTERIOPHAGE ØW-14

7.1 Introduction

ØW-14 is a virus having as host the bacterium Pseudomonas acidovorans (Kropinski and Warren, 1970). Its DNA contains 56.2% G-C base pairs as determined by chemical analysis (Kropinski, Bose and Warren, 1973). Estimates of the G-C content on the basis of buoyant density and melting temperature measurements give widely divergent values (4.5% and 72.9% respectively). The anomalous behaviour has been shown (Kropinski, Bose and Warren, 1973) to be due to the presence of an unusual base in which a molecule of putrescene is covalently bonded to the methyl group of thymine. The structure of the base (α-thyminylputrescene) is given below.



α-Thyminylputrescene

Approximately 50% of thymine bases are modified in this way.

Since the modification is probably of importance in the life cycle of the phage, it is of considerable interest to determine the nature of any structural changes induced in the DNA secondary structure and to elucidate the conformation of the putrescene moety. A series of X-ray diffraction photographs of fibres made from ØW-14 DNA have been taken by the author in conjunction with Dr. C. Nave in this laboratory on material provided by Dr. R.A.J. Warren of the Department of Microbiology, University of British Columbia, Vancouver, Canada.

7.2 Results

7.2.1 Fibre Diffraction Patterns

Fibres were made from DNA gels obtained by ultracentrifugation of a solution of the DNA in tris buffer as described in section 2·3. The buffer solutions contained either 0.01M or 0.02M NaCl and 0.002M tris: the p.H. being adjusted to 7.6 in all cases. Samples of calf thymus DNA were run in parallel under the same conditions to provide a control experiment.

Diffraction photographs were taken at several values of the relative humidity in the range 44% to 98% for fibres of ØW-14 DNA and for the calf thymus DNA used as a control. Both A and B forms of DNA were observed in the ØW-14 DNA fibres, and examples of the patterns obtained are shown in plates 7.1 to 7.4. Measurements of the helix pitch length and of the intermolecular spacing yielded values which were identical, within the limits of experimental error, with those obtained from equivalent measurements of patterns from calf thymus DNA (Table 7.1). The identical values obtained for these parameters in the two DNA's combined with the general appearance of the intensity data indicate that the presence of the modified base does not induce changes in the conformation or packing of ØW-14 DNA in fibres.

TABLE 7.1

Comparison of spot positions in diffraction patterns from calf thymus and ØW-14 DNA

h,k,ℓ	$\rho_C(\text{Å}^{-1})$	$\rho_o(\text{CT})(\text{Å}^{-1})$	$\rho_o(\text{ØW-14})(\text{Å}^{-1})$
1,3,0	0.0866	0.0870	0.0864
1,5,0	0.1312	0.1317	0.1317
0,2,1	0.0609	0.0603	0.0598
-2,0,1	0.0933	0.0927	0.0925
0,2,2	0.0870	0.0846	0.0845
0,4,2	0.1217	0.1207	0.1206

ADNA (Error = ±0.0002)

h,k,ℓ	ρ_C	$\rho_o(\text{CT})$	$\rho_o(\text{ØW-14})$
2,0,0	0.0502	0.0466	0.0482
1,0,2	0.0630	0.0640	0.0645
2,1,2	0.0880	0.0792	0.0808
3,1,2	0.1070	0.0964	0.0937

Sodium BDNA (Error = ±0.001)

h,k,ℓ	ρ_C	$\rho_o(\text{CT})$	$\rho_o(\text{ØW-14})$
1,1,0	0.0545	0.0549	0.0547
2,2,0	0.1090	0.1098	0.1093
0,-1,1	0.0437	0.0442	0.0441
1,0,2	0.0739	0.0694	0.0682
1,2,3	0.0978	0.0985	0.0981

Lithium Crystalline BDNA (Error = ±0.0003)

A quantitative difference was apparent, however, in the induction of the A⇌B transition by changes in the relative humidity. The fibres of calf thymus DNA control gave B patterns at 92% relative humidity in all cases. Fibres made from ØW-14 DNA, however, only gave B patterns regularly at 98% relative humidity; occasionally at 95% and never at 92% or below (Table 7.2).

The B patterns obtained from ØW-14 DNA at 98% were of a high quality for the sodium salt: the first three spots on the second layer line being observable in most cases and were very well resolved in a large number of patterns.

It has been reported (Cooper and Hamilton, 1966) that the A⇌B transition is a function not only of the relative humidity but of the salt content of the fibre: the A form being favoured at low salt concentrations. Indeed fibres highly deficient in salt gave A patterns at all relative humidities. It is important to determine, therefore, whether the salt content in the gels from which the fibres are made have differing salt content for the two different DNA's used.

TABLE 7.2

Relative Humidity (%)	ØW-14 DNA	Calf Thymus DNA
98	B	B
95	B or A/B	B
92	A	B (occasionally A)
75	A	A (occasionally B)
66	A	A

7.2.2 Studies on the B Form of DNA

In the B form of DNA the thyminyl methyl groups lie in the major groove. The phosphate oxygens on either chain are possible hydrogen bond acceptors and, hence, a number of possibilities involving interaction between these oxygens and the free amino group were considered.

According to the model of Arnott and Hukins (1972), the nearest phosphate oxygen of the chain on the opposite side of the major groove lies at a distance of 11.8Å from the carbon atom of the thyminyl methyl group. When the putrescene molecule is fully extended, the distance spanned between the methyl carbon atom and the free amino nitrogen is 7.28Å. Since for hydrogen bond formation the terminal nitrogen should lie approximately 2.9Å from the acceptor atom, the maximum distance allowable between the methyl carbon and the acceptor atom is 10.2Å. Hence, the putrescene is approximately 1.6Å too short to form a satisfactory hydrogen bonding interaction with a phosphate group of the opposite chain.

Other possibilities for models have been tried including cases of inter-molecular hydrogen bonding to a phosphate oxygen of an adjacent DNA duplex. It has not been possible to build a stereochemically reasonable model for this case in accord with the constraints imposed by the molecular packing in DNA fibres. In fibres of sodium BDNA, the intermolecular spacing (between helix axes of adjacent molecules) is approximately 26.5Å. The thymine methyl groups lie on a helix of radius 5.4Å, while the phosphate oxygens of greater radius lie on a helix of radius 10.2Å. Hence, the minimum possible distance of separation between a thymine group on one DNA molecule, and a phosphate oxygen of the adjacent molecule is approximately 11Å. This distance is slightly greater than the maximum acceptable distance between the thymine methyl group and the hydrogen bond acceptor (see above).

Moreover, this is the minimum possible distance of separation between the pair of atoms, and one condition implied by this is that the atoms both have the same z coordinate. However, it would not be possible for the putrescene to form a hydrogen bond with a phosphate oxygen of an adjacent molecule which had the same z coordinate due to steric hindrance from the phosphate backbone of the chain to which it is attached. The requirement that the putrescene must project from the major groove constrains the acceptor oxygen to have a z coordinate substantially different from that of the thymine methyl so that the distance to be spanned is greater than 11Å.

Although specific models have not been built, it is of interest to make some preliminary intensity measurements to compare the diffraction data obtained from ØW-14 with equivalent data taken from calf thymus BDNA patterns. Unfortunately, sodium BDNA does not give fully crystalline fibres so that errors in the intensity measurements are large, so that any effects produced by the putrescene groups are very difficult to detect.

Intensity measurements were taken from a number of sodium B patterns from calf thymus and ØW-14 DNA. In cases where Bragg reflections were well resolved on the second layer line each spot was considered as a separate datum and intensity values measured from them were scaled to be comparable with data obtained from diffuse intensity in other parts of the pattern. The method of scaling and of intensity measurements is that prescribed by Wilkins (unpublished) and described in Marvin et. al. (1961). The form of the equation used to perform the scaling is:-

$$\frac{E_{SP}}{E_{ST}} = \frac{P(h,k,\ell)}{2\pi\xi NA}$$

where,

E_{SP} = The integrated intensity of a spot

E_{ST} = The intensity per Å^{-1} along a streak

N = The number of molecules in the unit cell

A = The area of the C axis projection of the unit cell

$P(h,k,\ell)$ = The packing factor as defined by Fuller et. al (1967)

Intensity measurements were made on the first and second layer lines out to a distance of 0.14Å^{-1}, and the results are given below as a ratio of the second layer line intensity to that on the first.

Calf Thymus	ØW-14
2.34±0.1	2.31±0.1

There is no significant difference between the two values within the limits of experimental error, therefore.

7.2.3 Studies on The A Form

In the A conformations of DNA the thyminyl methyl group project" into the "hollow core" of the DNA helix.

A survey was made, in an analogous way to that for the B conformation, of possible models involving intra and inter-molecular hydrogen bonding. It was not found to be possible to build a stereo-chemically reasonable model involving inter-molecular hydrogen bonding taking account of the constraints imposed by the position of the putrescene on the DNA and the relative positions of DNA molecules in the unit cell.

Eventually three models only were considered for detailed analysis. In the first model the putrescene was constrained to hydrogen bond to a phosphate oxygen on the opposite side of the groove.

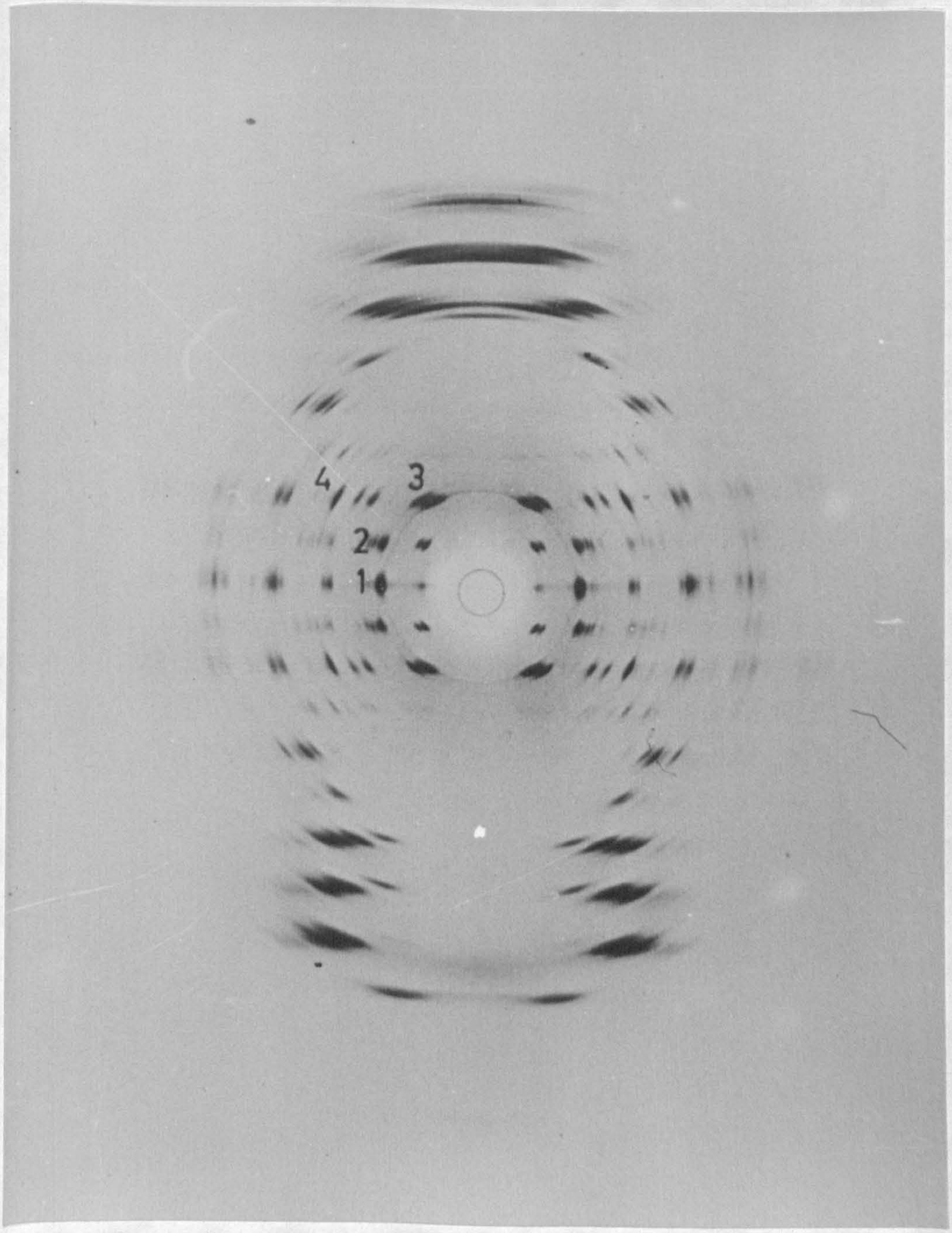


Plate 7.1 ADNA pattern showing the four spots used in the intensity measurements discussed in this chapter

The distance between this phosphate chain and the covalently bonded end of the thymine molecule is less than the equivalent distance in BDNA and hence such a model is feasible in the A conformation.

A second model considered involved the formation of a hydrogen bond between the putrescene and the O6 atom of a thymine or guanine residue on the opposite chain of the DNA; while the third model did not allow hydrogen bond formation at all: the putrescene merely projecting out of the DNA groove approximately centrally between the two phosphate chains. Coordinates were obtained from wire models in each case and are given in table 7.3

The cylindrically averaged molecular transform was calculated for each of the above models. All putrescene atoms were assigned weighting factors of 0.125 to account for the fact that only some 12% of bases have an attached putrescene molecule. Comparison of the transforms for models 1 and 3 with that from ADNA showed small but significant differences in the relative values of the intensity on the equator and first two layer lines with no significant differences in the transform values at higher diffraction angles (Figure 7.1). The intensity changes predicted by model 1 and model 3 are very similar indeed, but model 2 gave much smaller intensity changes further out along layer line 1 and 2.

In order to measure the observed relative intensity on these layer lines, four spots were chosen (indicated on plate 7.1) and their integrated intensities measured on a number of A patterns from calf thymus DNA and from ØW-14 DNA. The four spots chosen are the most intense on these layer lines and, hence, the small differences in relative intensity expected can more easily be detected. Approximately twenty patterns of each type of DNA were measured in arriving at the final averaged results which are presented for each model in Table 7.6.

In order to compare the observed relative intensities for the four spots with the calculated values, structure factors were calculated from the molecular transform data for each of the models and for ADNA (Table 7.4). The equatorial spot which was measured (spot 1 on plate 7.1) is actually a doublet consisting of spots (1,3,0) and (2,0,0); while spots 2,3 and 4 of plate 7.1 are triplets composed respectively of the following sets of spots; (-1,1,1), (0,2,1) and (1,1,1); (-1,1,2), (0,2,2) and (1,1,2); and (-2,4,2), (-1,5,2) plus (-3,1,2). None of these spots have precisely the same value of ξ and ℓ, and are resolved on very good ADNA patterns. Normally, however, they merge into a single spot and this was always the case in the patterns measured. Hence, each of the above groups of spots have been treated as a single datum when taking intensity measurements from the diffraction patterns.

In order to obtain the calculated values of the various ratios from spots 1, 2, 3 and 4, therefore, the calculated intensity values of the appropriate constituent spots were added together to give the total intensity for each of the measured spots. The values for a number of intensity ratios involving pairs of the four spots are given for the three models and for ADNA in table 7.5.

The experimental values of the ratios for calf thymus DNA (table 7.6) are somewhat different from the calculated values for ADNA especially for the ratio of spots 3 to spot 1. This point will be discussed in section 7.3

The changes observed in the measured values of the ratios in ØW-14 DNA compared to those from calf thymus DNA are similar to the equivalent changes in the calculated ratios for models 1 and 3 when compared to values from ADNA. Model 2 does not predict the same intensity changes, however: in two cases (ratio 4/1 and 2/1) the predicted changes are of opposite sign to those observed.

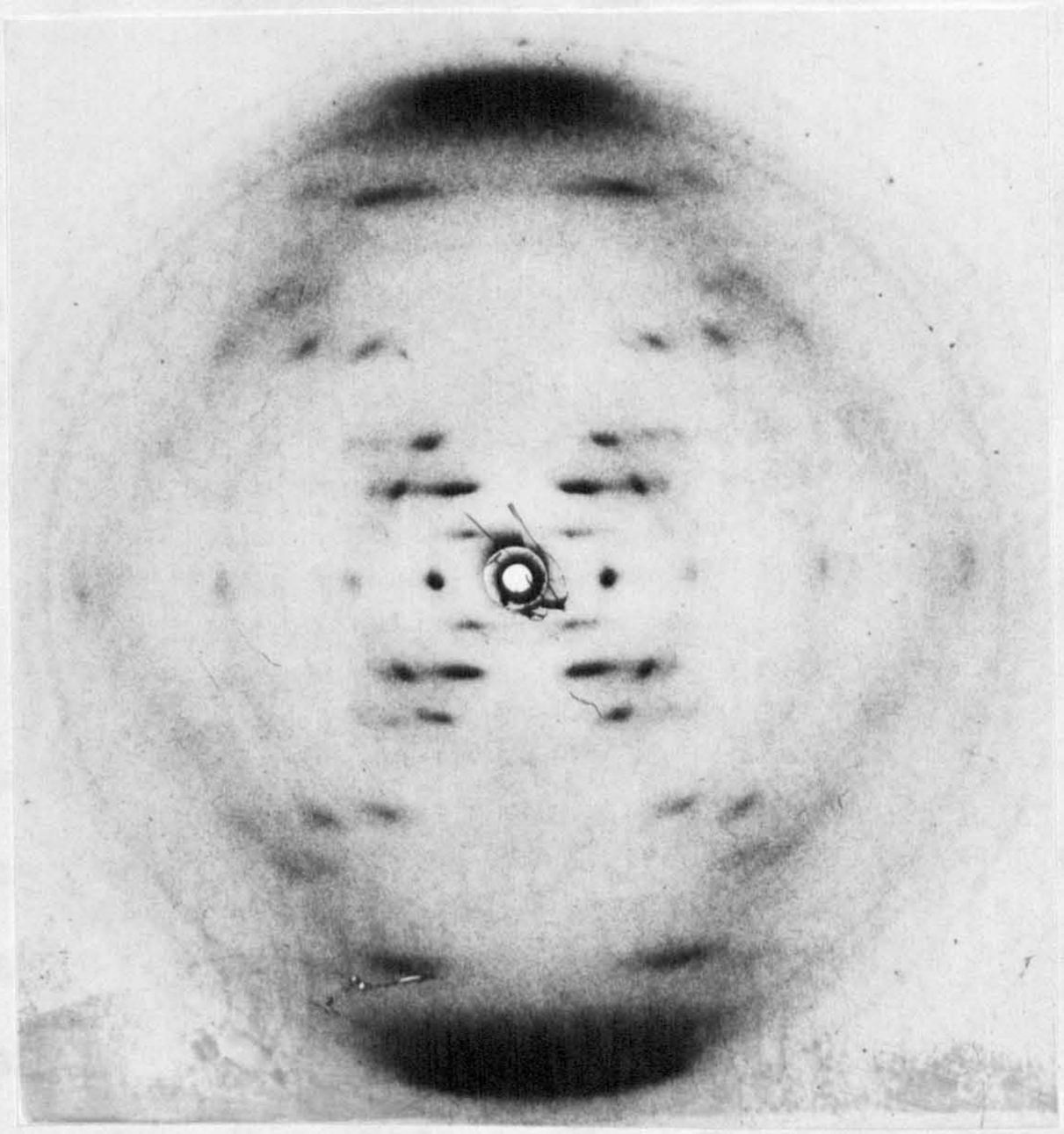


Plate 7.2 ØW-14 DNA 98% R.H.
Crystalline B form from the lithium salt

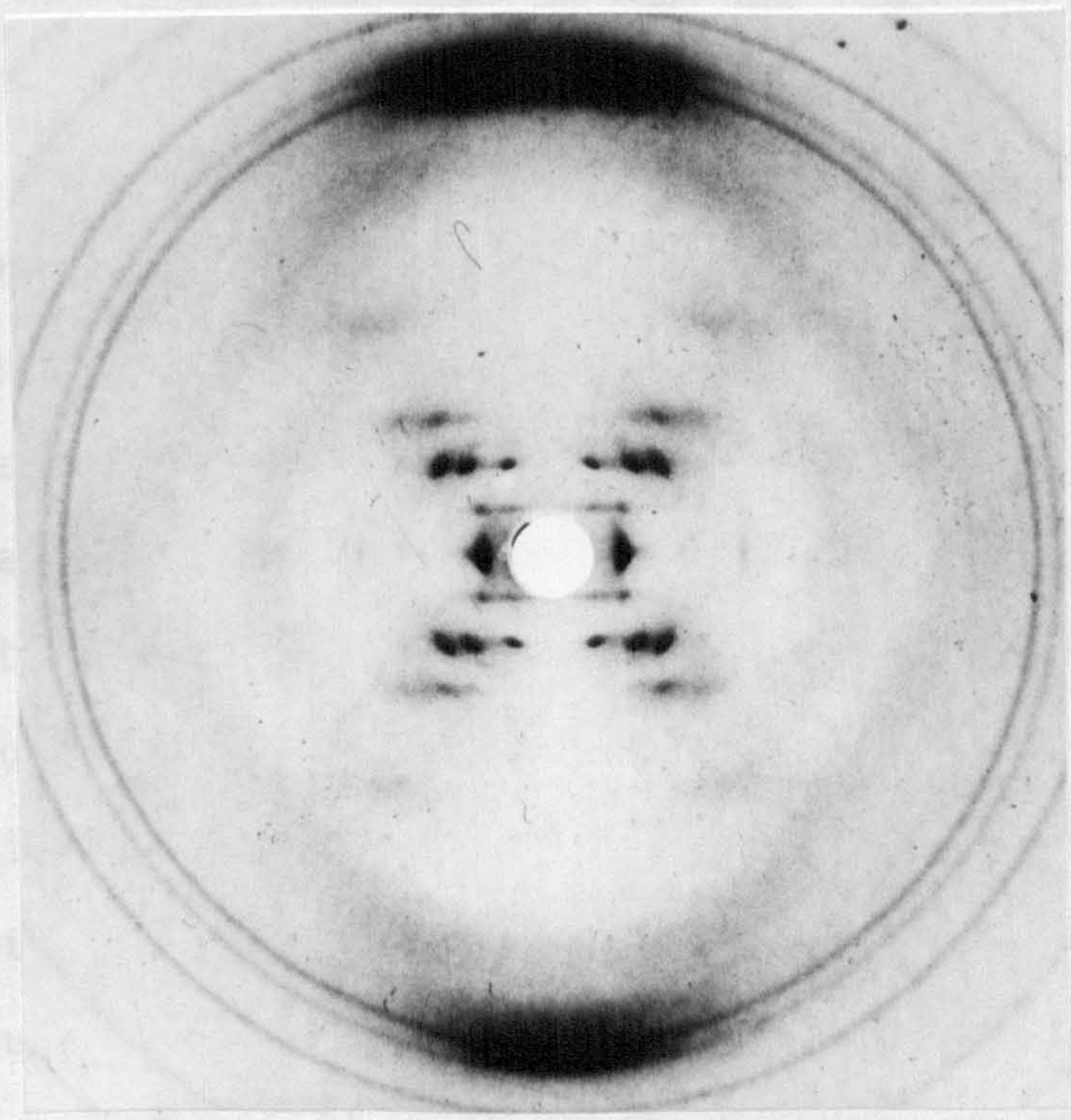


Plate 7.3 ØW-14 DNA 98% R.H.
Semi-crystalline B form from the sodium salt

In order to facilitate comparison of the predicted changes with those observed, it is useful to express the value of a ratio in calf thymus DNA as a fraction of the value of the same ratio in ØW-14 DNA, and to perform analogous calculations with the calculated data: that from ADNA being expressed as a fraction of that from each of the models separately. This has been done and the results are presented in table 7.7.

It can be seen that values of predicted changes from model 2 do not give a good fit with the observed intensity changes. Model 2 is somewhat poorer stereochemically than are models 1 and 3 since in the latter two the putrescene is in a more elongated (more favourable) conformation. On these grounds model 2 was eliminated from further consideration.

It is not possible to distinguish between models 1 and 3 on the grounds of the X-ray diffraction evidence alone. The fit between the observed and calculated data is marginally better for model 1 than for model 3, but the difference is probably not significant considering the errors of measurement (see section 7.3).

Model 1 is preferable stereochemically in that the putrescene is less sterically hindered when it points towards the phosphate backbone of the opposite chain than when it points out of the DNA groove; and because model 1 explicitly allows the hydrogen bonding capacity of the free amino group to be exploited. In model 3 no particular hydrogen bonding interaction is provided for and the only possible interactions which could constrain the molecule to adopt the conformation of model 3 would be those between the amino group and water, but these are unlikely to be as favourable as the interaction between putrescene and part of a stable structure such as the DNA phosphate backbone. Since the X-ray data does not favour either model,

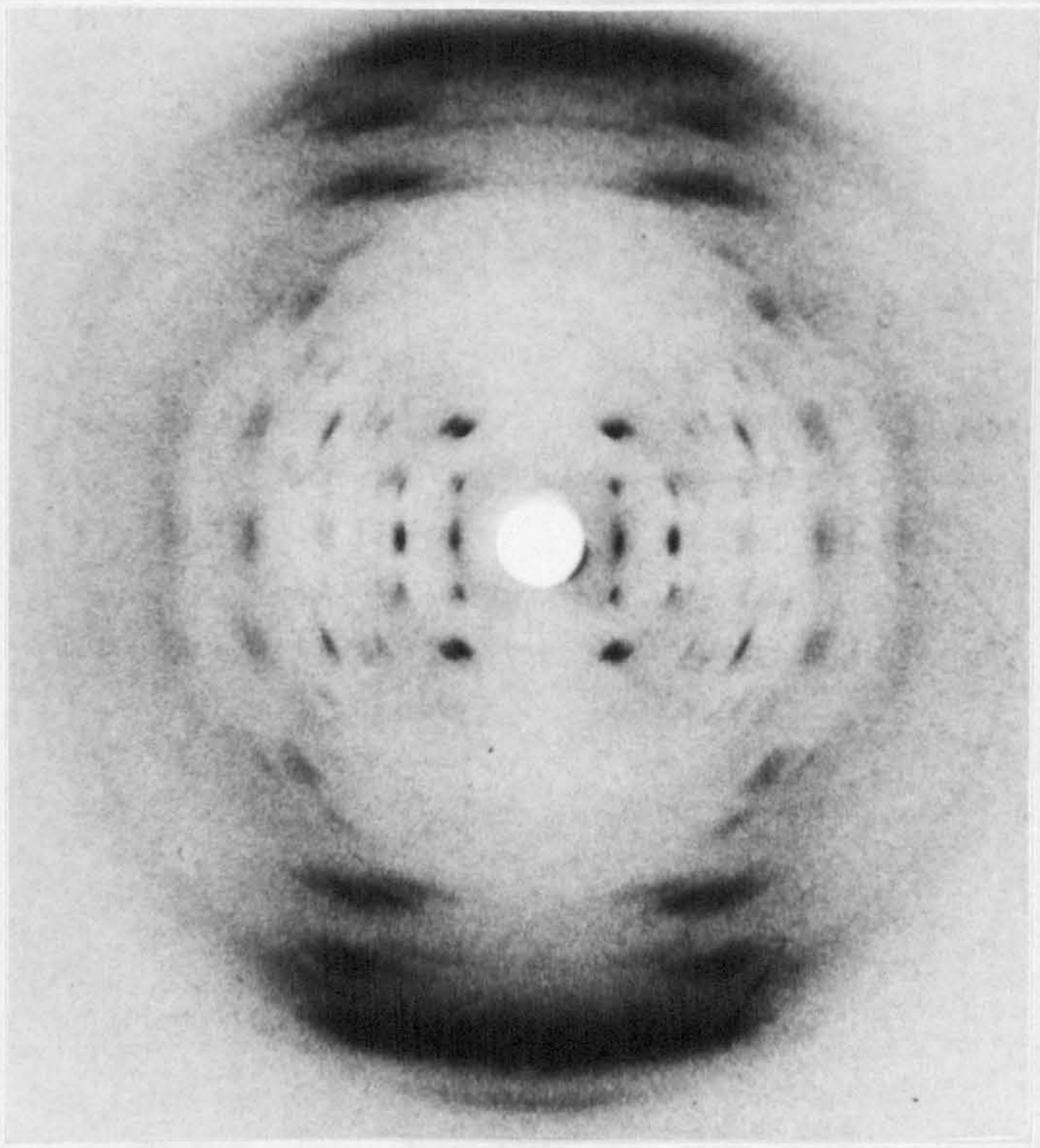


Plate 7.4 ØW-14 DNA
A pattern from the sodium salt

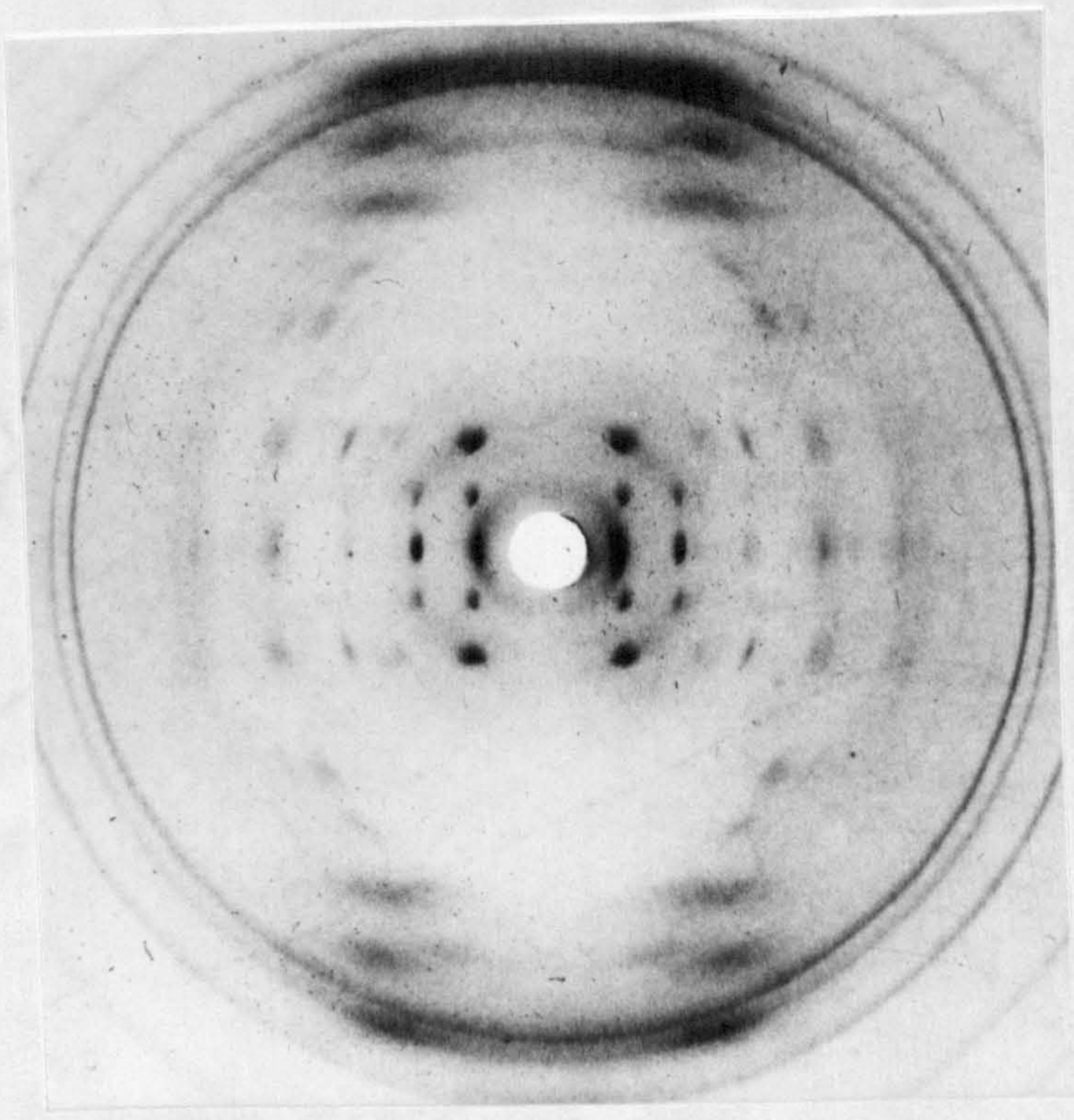


Plate 7.5 Calf thymus DNA 75% R.H.
A pattern from the sodium salt

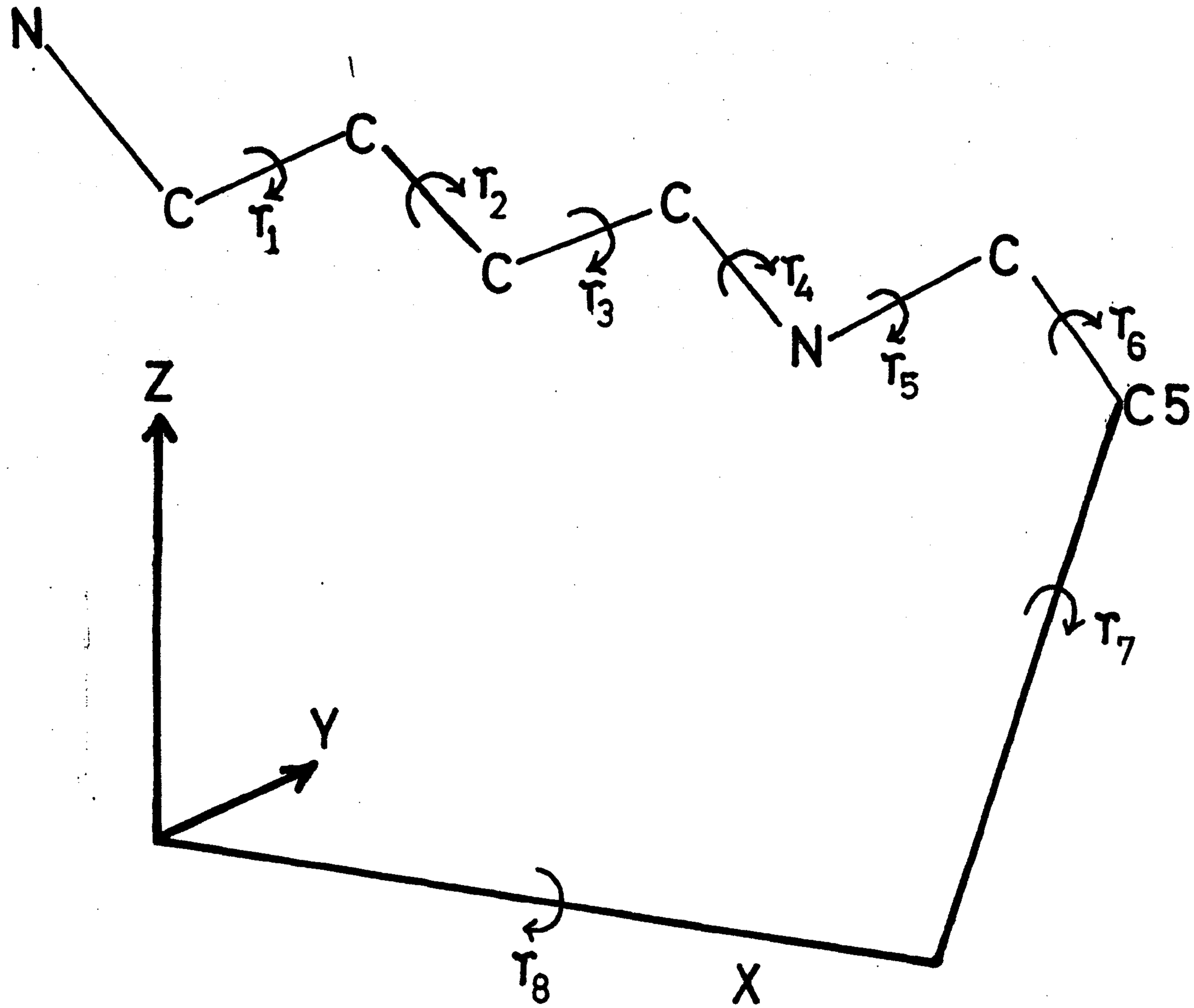


Fig. 7·2 PUTRESCENE
MODEL BUILDING

it was decided to eliminate model 3.

A computerised refinement of model 1 has been performed to improve the stereochemistry of the putrescene conformation and the geometry of the hydrogen bonding interaction. The method of building the putrescene onto the DNA is illustrated in Fig. 7.2. Atomic coordinates for the putrescene molecule in the final model are given in Table 7.8 along with those of the phosphate oxygen to which the putrescene is hydrogen bonded, and the thymine methyl carbon to which it is covalently bonded. Torsion angles for the model are given in table 7.9 along with details of the shortest non-bonded contacts between atom pairs. A CPK model of ADNA with a putrescene molecule attached is shown in plate 7.6

It is unfortunate that the agreement between the predicted intensity changes and those observed is better in the unrefined model (see tables 7.6 and 7.7). The final model is not as satisfactory as it could be, therefore, as far as the present data is concerned. However, the agreement is still considered to be good and the model may be essentially correct.

7.2.4 The Estimation of Na^+ Concentration In Gels

A number of factors may influence the nature of the $A \rightleftharpoons B$ transition in NaDNA. In particular the ionic strength in the fibre is known to influence the relative humidity values at which the different forms are observed (Cooper and Hamilton, 1966). The A form is favoured at low ionic strength, and the B form in fibres possessing excess salt. It is in fact possible to obtain fibres which are permanently in the A form or the B form at any relative humidity provided that extreme conditions of ionic strength pertain in the fibre.

TABLE 7.3

Unrefined wire models for the putrescene conformation in ADNA

	R	ϕ	Z
N	7.30	104.3	-7.88
C	6.88	93.8	-7.19
C	6.22	94.3	-5.85
C	5.73	81.8	-5.47
C	5.05	81.3	-4.00
N	5.10	65.8	-3.64
		MODEL 1	
N	2.3	-97.7	-5.31
C	0.2	-152.7	-5.31
C	0.1	67.3	-5.31
C	1.1	65.3	-4.17
C	3.3	57.3	-4.09
N	3.8	57.3	-2.64
		MODEL 2	
N	8.35	100.3	-6.85
C	6.90	97.8	-6.54
C	6.90	88.3	-5.40
C	5.65	81.3	-5.40
C	6.00	69.3	-4.50
N	4.90	60.3	-4.09
		MODEL 3	

TABLE 7.4

	ADNA	Model 1 (refined)	Model 2	Model 3
1,3,0	9224	9215	7041	9403
2,0,0	6912	7030	5121	7159
0,4,0	2362	2594	1472	2624
2,2,0	918	1111	432	1114
1,5,0	1852	1507	2413	1601
2,4,0	1774	1456	2310	1553
3,1,0	1475	1220	1955	1313
-1,1,1	1887	1400	1887	1409
0,2,1	2085	1566	2082	1583
1,1,1	2919	2283	2904	2335
-1,3,1	3860	3358	3863	3573
-2,0,1	3624	3167	3635	3379
1,3,1	3224	2839	3863	3042
2,0,1	2289	2056	2336	2230
-1,1,2	3995	3630	3828	3588
0,2,2	4485	4059	4310	4028
1,1,2	4816	4322	4651	4322
-2,0,2	469	371	498	425
-1,3,2	413	322	441	375
-2,2,2	472	505	382	406
0,4,2	1090	1108	926	944
2,0,2	1237	1250	1057	1072
-2,4,2	2686	2449	2333	2207
-1,5,2	3274	2984	2866	2698
-3,1,2	3001	2722	2617	2458
-1,5,3	704	666	645	583
2,4,3	813	733	741	638
-3,3,3	2592	2356	2401	976
3,1,3	329	278	291	235
-4,2,3	227	186	210	157

TABLE 7.5

The measured intensity ratios from calf thymus A patterns and from equivalent patterns from ØW-14

Ratio (Spot No.'s)	Calf Thymus	ØW-14
4/1	0.71	0.57
3/1	1.54	1.57
3/2	2.30	2.66
2/1	0.67	0.59

Values ±0.05

TABLE 7.6

Calculated intensity for unmodified A DNA and for the three models involving putrescene

Ratio (Spot No.'s)	A DNA	Model 1	Model 2	Model 3
4/1	0.56	0.50(0.46)*	0.64	0.38
3/1	0.82	0.74(0.73)	1.05	0.74
3/2	1.93	2.29(2.32)	1.86	2.24
2/1	0.43	0.43(0.32)	0.57	0.33

TABLE 7.7

The ratio of the ratios for calf thymus DNA/ØW-14 DNA

Ratio (Spot No.'s)	Measured	Model 1	Model 2	Model 3
4/1	1.25	1.12(1.22)*	0.87	1.47
3/1	0.98	1.18(1.12)	0.78	1.11
3/2	0.86	0.85(0.83)	1.04	0.86
2/1	1.14	1.31(1.35)	0.74	1.32

*Figures in parentheses indicate values obtained from the hand built (unrefined) model 1 conformation.

TABLE 7.8

Atomic coordinates for the refined model 1 conformation for putrescene. The hydrogen bond is between N1 and O2. PC20 represents the carbon atom of the thymine methyl group

	R(Å)	φ(°)	Z(Å)
N1	7.772	95.340	-8.020
C2	6.411	91.273	-7.597
C3	6.370	89.793	-6.097
C4	5.060	83.538	-5.673
C5	4.936	83.828	-4.168
N6	4.536	67.171	-1.443
PC20	4.691	66.551	-2.175
O2	9.800	95.408	-10.280

TABLE 7.9(a)

The worst non-bonded atomic contacts for the refined model 1 conformation for putrescene

Atom pair	Distance (Å)	Energy (K cal/mole)
1H6-U1TCH3	2.40	0.11
A2H6-U1TCH3	2.44	0.11
N6-U1TCH3	2.85	0.07
1H1-3LP	2.57	0.04
1H5-U3O3	2.22	0.04

TABLE 7.9(b)

Dihedral angles for the refined model in the IUPAC-IUB (1970) convention

Torsion Angle	Value (Degrees)
1	180.00
2	204.28
3	188.90
4	217.90
5	167.51
6	190.47

In view of this ionic strength dependence, it is of interest to determine whether the amount of salt present in gels of ØW-14 DNA is different from that fround in equivalent gels of calf thymus DNA. Since the positive charges on the putrescene residues will neutralise some of the phosphate groups it is possible that fewer sodium ions are associated sufficiently strongly with the DNA to be sedimented along with it in an ultracentrifugation run.

A routine for determining the sodium to phosphate ratio in DNA gels using a flame emission spectrophotometer, has been developed in this laboratory by Mr. P.J. Blakeley (Blakeley, 1976). Results presented in this chapter were obtained by the following procedure. Gels prepared by ultracentrifugation as described in chapter 2, were redissolved in 3ml of distilled water (allowed to stand for 72 hrs. to completely dissolve). An aliquote of the resultant solution was taken and the concentration of phosphate groups measured spectrophotometrically, assuming a value for E, the extinction coefficient for DNA at 260nm, of 6600; and that the average nucleotide molecular weight is 330.

The sodium concentration was measured on a Unicam SP1900 Flame emission spectrophotometer, by means of the intensity of radiation emitted at 589nm. It is necessary to calibrate the machine before each run, and a series of sodium chloride solutions of known concentration in the range 5.0×10^{-5}M to 1.0×10^{-2}M were used for this purpose. Normally, values of the sodium ion concentration measured from the dissolved gels were of the order of 10^{-4}M.

These figures for the molar concentration of sodium ions and DNA phosphate groups can be used to determine the sodium to phosphate ratio for each sample. The final results obtained as an average of

six runs are presented below.

Salt concentration of solutions before centrifugation

DNA	0.01m	0.02m
ØW-14	0.42	0.52
Calf Thymus	0.88	1.24

All values are ±0.09

It is evident, therefore, that the gels prepared from ØW-14 DNA contain less imbibed sodium ions than do equivalent gels obtained from calf thymus DNA. The magnitude of the difference between the values for the ratios in the different DNA's is surprising since the putrescene to phosphate ratio is only 0.125, while the average difference between equivalent ratios in the two DNA's is approximately 0.6. It would be necessary to postulate that each putrescene molecule neutralises approximately four phosphate groups if the difference in the sodium to phosphate ratios between the two DNA's were to be explained on this basis.

It is possible that other metal ions are present in the ØW-14 DNA solutions and that these displace some sodium ions from the DNA. The concentrations of selected metal ions in ØW-14 DNA solutions were checked and the results presented in table 7.11. The metal ions tested would in total be equivalent to a sodium to phosphate ratio of 9.9×10^{-2} and 9.0×10^{-2} for calf thymus DNA and ØW-14 DNA respectively. Hence, there is little difference between the levels of the metals tested in the two DNA solutions; while the absolute magnitudes of the values show that sodium ions are in excess in the DNA solutions used. It has not been possible to account for the large reduction in the amount of sodium present in the gels from ØW-14 DNA and further work is required

TABLE 7.10

Metal Ion	Sodium/Phosphate Ratio	
	C.T. DNA	ØW-14 DNA
Li^+	1.4×10^{-3}	2.2×10^{-3}
K^+	8.3×10^{-3}	6.2×10^{-3}
Mg^{2+}	1.1×10^{-3}	3.0×10^{-3}
Ca^{2+}	3.9×10^{-2}	3.6×10^{-2}
Cs^{2+}	$<10^{-3}$	2×10^{-2}

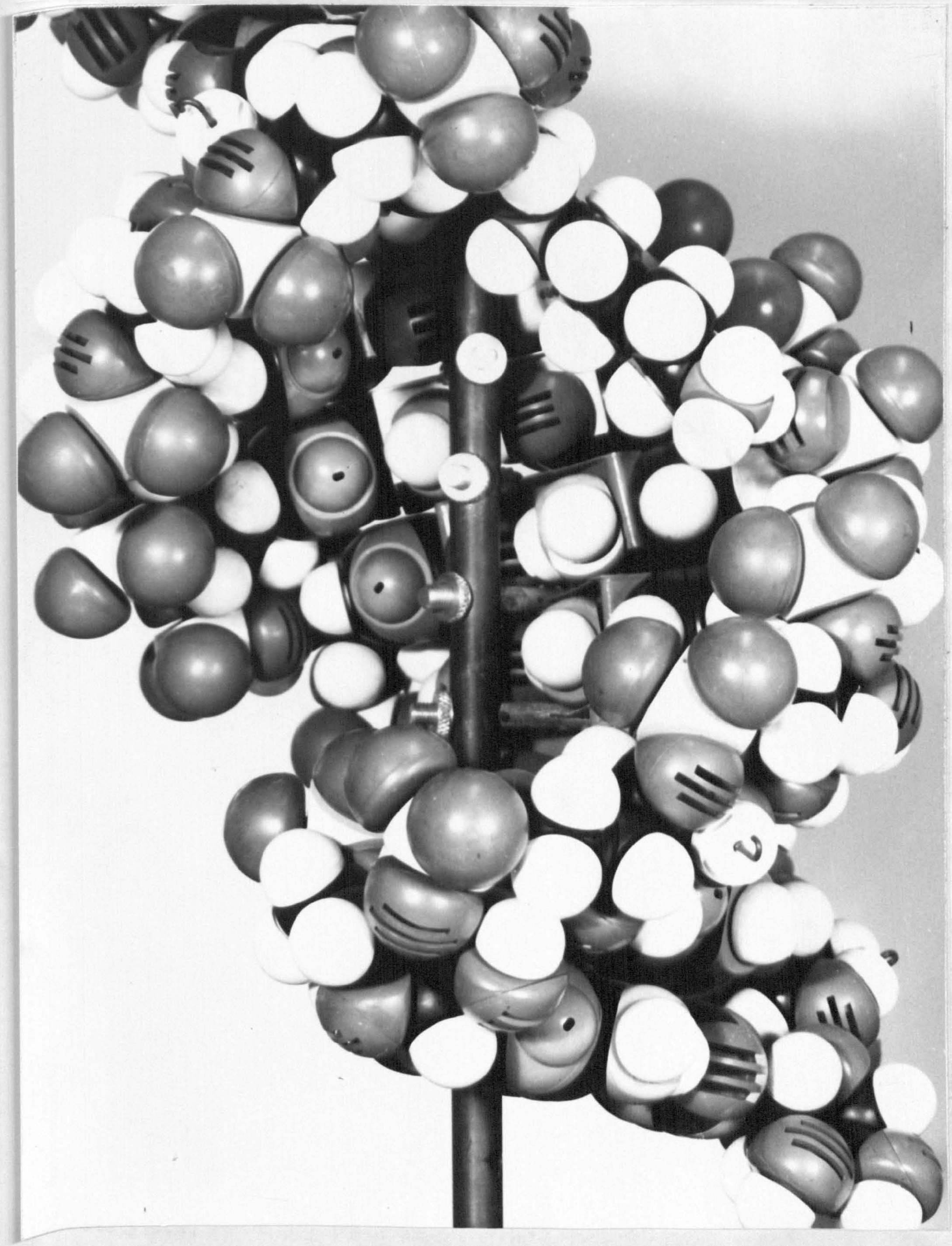


Plate 7.6 CPK model of putrescene attached to the A form of DNA

to determine whether this is a real effect or an artefact of the method of analysis.

It is possible that other charged species such as polyamines other than putrescene are partially responsible for the reduction in sodium levels in ØW-14 DNA. Nevertheless, the effect is probably at least partially caused by the presence of the covalently attached putrescene and, hence, the results will be taken qualitatively to indicate a real effect.

7.2.5 Acetylated ØW-14 DNA

Since it is the free amino group which is evoked as being responsible for the modified behaviour of ØW-14 DNA, it is of interest to study a system in which this group has been blocked by another covalently attached group.

Dr. R.A.J. Warren, has provided material in which the putrescene moety has an attached acetyl group on the free amino end of the putrescene. Fibres have been prepared by Dr. C. Nave from samples of this DNA and diffraction patterns obtained.

Preliminary results appear to indicate that the "acetylated" DNA, does not adopt the A form. This is in direct contrast to the behaviour of ØW-14 DNA in which the A-form is stabilized. The acetylated putrescene acts as a crude analogue of the glucosylated residues found in the DNA from the bacteriophage T_2, which also appears not to adopt the A conformation.

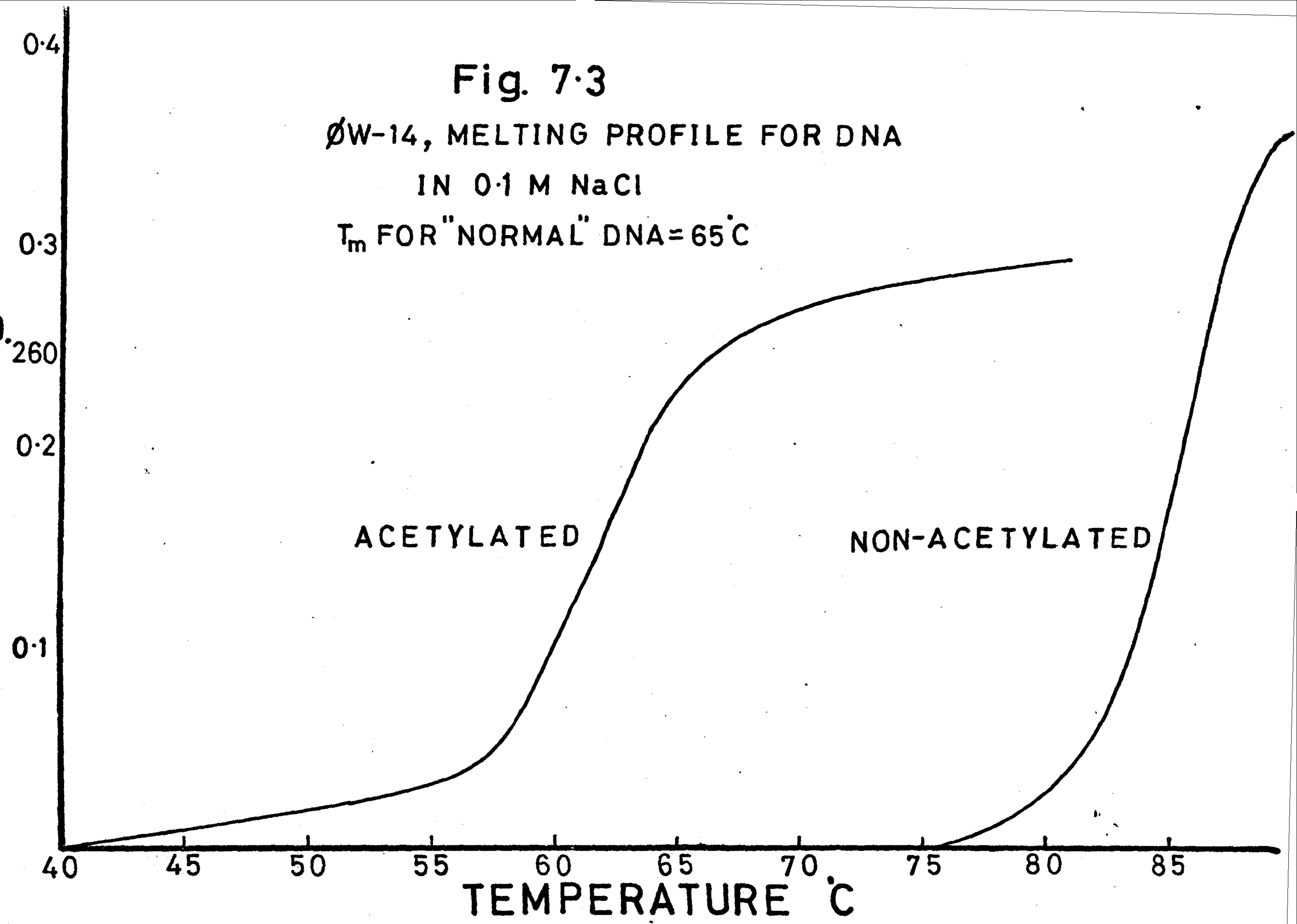
Moreover, the raising of the melting temperature observed in native ØW-14 DNA is reversed when the putrescene is acetylated. Figure 7.3 shows melting temperature curves for native and acetylated ØW-14 DNA obtained by Dr. Warren.

Fig. 7·3

ØW-14, MELTING PROFILE FOR DNA

IN 0·1 M NaCl

T_m FOR "NORMAL" DNA = 65°C



The transition for acetylated DNA is less sharp than that observed for the native form, which may indicate that acetylation is less than 100%.

7.3 Discussion

The results presented here would be of great interest if they could be shown to be of biological significance in the life cycle of the phage. Interpretation of the results of in vivo structural studies in terms of their biological importance is somewhat difficult since the detailed mechanisms of such fundamental processes as DNA transcription and replication have still not been elucidated. In particular, the changes in molecular conformations required by these processes, and hence the importance of established structure such as the A and B conformations of DNA, is not yet clear. Moreover, it is always difficult to extrapolate to the in vivo situation using results of studies of the behaviour of DNA in fibres since the two environments are different.

Nevertheless, the fibre environment is probably more similar to that of the in vivo situation, in terms of the degree of hydration, and localised DNA concentration, than is a dilute ($\simeq$ 1mg/ml) DNA solution, and hence limited conclusions regarding the biological significance of the base modification in ØW-14 DNA may be drawn.

Milman et. al (1967) have shown that a DNA-RNA hybrid has a structure in fibres similar to that of A DNA. It is difficult to build structures of the B genus with RNA since the 2' hydroxyl group cannot be accommodated into the B conformation without considerably distorting the structure (Milman et. al. 1967). All conformations of natural and synthetic RNA's, though not identical to A DNA, are

of the A genus of nucleic acid structures (viz. section 1).

The fact that a conformation of the A type is also always found when RNA is hybridised with DNA suggests that that A form may be the important conformation for transcription. If this is so a DNA, or section of DNA, which is more prone to adopt the A conformation may be preferentially transcribed. If the environment inside the cell is such that the DNA is near the transition point between the A and B conformations, it is possible that ØW-14 DNA may adopt the A conformation when the host DNA is in the B form.

The methods by which phages ensure the transcription of their own DNA is varied, but all appear to involve the production of an RNA polymerase which is specific for the phage genome. T_4 and T_7 phages utilise the host RNA polymerase to transcribe their early genes. In the case of T_7, these code for an RNA polymerase specific for the phage DNA (Chamberlain et. al., 1970), whereas the T_4 early genes produce a modification of the host polymerase which makes it specific for the phage DNA (Haselkorn et. al., 1969). Phage PBS2 of Bacillus subtilis includes its own RNA polymerase inside the virus during assembly and this is injected with the DNA (Clark et. al., 1974).

If the phage polymerase is packaged with the phage during assembly, it is not so important to make the DNA more acceptable to the host polymerase. The methods used by T_4 and T_7, however, both require the host polymerase in the initial stages of the lytic cycle and one could imagine that it is important for the DNA to be susceptible to being transcribed. Unfortunately it is not known (Dr. R.A.J. Warren, private communication) how ØW-14 acts to ensure transcription of its own DNA during infection.

It is equally possible that the presence of putrescene facilitates some other aspect of the phage cycle, such as the packaging of the DNA into the phage head during assembly. It is interesting in this respect that ØW-14 does not appear to be a transducing phage (i.e. does not incorporate sections of host DNA in a small percentage of particles), though ØW-14 infection, unlike that of T_4, does not appear to result in the break down of the host DNA (Dr. R.A.J. Warren, private communication) which is, therefore, still present intact when the virus is assembled.

It would be of interest in this context to determine the conformation adopted by the DNA in the phage head. For this purpose it is desirable to have osmotic shock resistant mutants so that DNA is not released from the phage as a result of lysis of the capsule during centrifugation. Work is going on to prepare such mutant for X-ray analysis.

CHAPTER VIII

BINDING OF CHLORPROMAZINE TO DNA

8.1 Introduction

Chlorpromazine (CPZ) is a drug having therapeutic uses in clinical psychiatry (Delay and Deniker, 1952) and in anaesthesia (e.g. Buxton and Hopkin, 1955): its formula is shown in figure 8.1. It is possible to oxidise chlorpromazine to obtain a positive ion (CPZ^+) and Piette, Bulow and Yamayaki (1964) have suggested that this radical may be responsible for its psychotropic activity.

The drug has a superficial resemblance to many of the intercalating drugs (e.g. Acridine, Proflavine) and this led Ohnishi and McConnell (1965) to attempt to bind the free radical of chlorpromazine to DNA and to take advantage of the unpaired electron system to obtain electron spin resonance (E.S.R.) spectra from the complex in solution. It is possible to determine the orientation of the chlorpromazine chromophore relative to the applied magnetic field in E.S.R. experiments. In order to align the DNA molecules so that they are of known orientation, the solution is flowed through a capillary tube in the resonance cavity. The spectra obtained indicated that the chlorpromazine was oriented with the plane of the chromophore at right angles to the helix axis: a fact which is consistent with an intercalative mode of binding for the drug. Similar conclusions have been reached (Porumb and Slade, 1976) on the basis of E.S.R. spectra obtained from fibres of DNA/CPZ^+ complexes.

It is possible that chlorpromazine binds by some external mechanism. Provided that such a mode of binding involved the drug chromophore being held perpendicular to the helix axis, it would be consistent with the E.S.R. results obtained by the above workers.

It is possible to distinguish between these two types of binding by the technique of X-ray diffraction of fibres of the complex, and hence a series of diffraction photographs have been taken of fibres of complexes between native CPZ or of CPZ^+ and calf thymus DNA.

8.2 Methods

Calf thymus DNA was purified in the manner described in chapter 2. Chlorpromazine hydrochloride (May and Baker Ltd.) was dissolved to make a solution of concentration 5.0×10^{-2} molar, which could be added to a solution of the DNA in the appropriate ratio to obtain the desired value of the P/D ratio. The molar concentration of phosphate groups in the DNA solution is determined by measuring the optical density at 260nm and assuming a value of 20 for the optical density of a 1mg/ml solution and a value of 330 for the molecular weight of the average nucleotide.

The chlorpromazine free radical was produced by reacting the native chlorpromazine solution with an equimolar amount of sodium persulphate. CPZ^+ is very unstable so that it is important to react the radical with the DNA, which stabilises it (Ohnishi and McConnell, 1965), as soon as possible. Since the oxidation reaction takes some time to complete, there is an optimum period after mixing the two solutions at which the

concentration of free radical in the mixture is at a maximum. This occurs 75 seconds after mixing (T. Porumb, private communication) so that this time period is allowed after mixing the sodium persulphate and chlorpromazine solutions before the DNA solution is added. Solutions are kept at 4°C during this process and are shielded from direct light.

Gels are prepared from the resultant mixture by centrifugation at 50,000 r.p.m. for approximately 6 hours in a 10 x 10 ml. angle rotor on an M.S.E. 50 ultracentrifuge.

When using the native chlorpromazine, this buffer solutions were used as described in chapter 2. However, when preparing fibres of the DNA complexed with the chlorpromazine free radical, unbuffered solutions containing 0.04M NaCl were used. These conditions were employed by Mr. T. Porumb in this laboratory when preparing fibres for use in E.S.R. experiments, and result in an acid (pH $\simeq$ 5.0) solution. X-ray diffraction photographs of the fibres were taken, immediately after preparation in the case of complexes involving CPZ^+, from a number of fibres of varying P/D ratios in the P/D range 6 to 12.

8.3 Results

Results of the diffraction studies of fibres of DNA complexed with native CPZ and with CPZ^+ will be discussed separately below, though in neither case was a diffraction photograph obtained which showed the reduction in layer line separation characteristic of intercalation complexes.

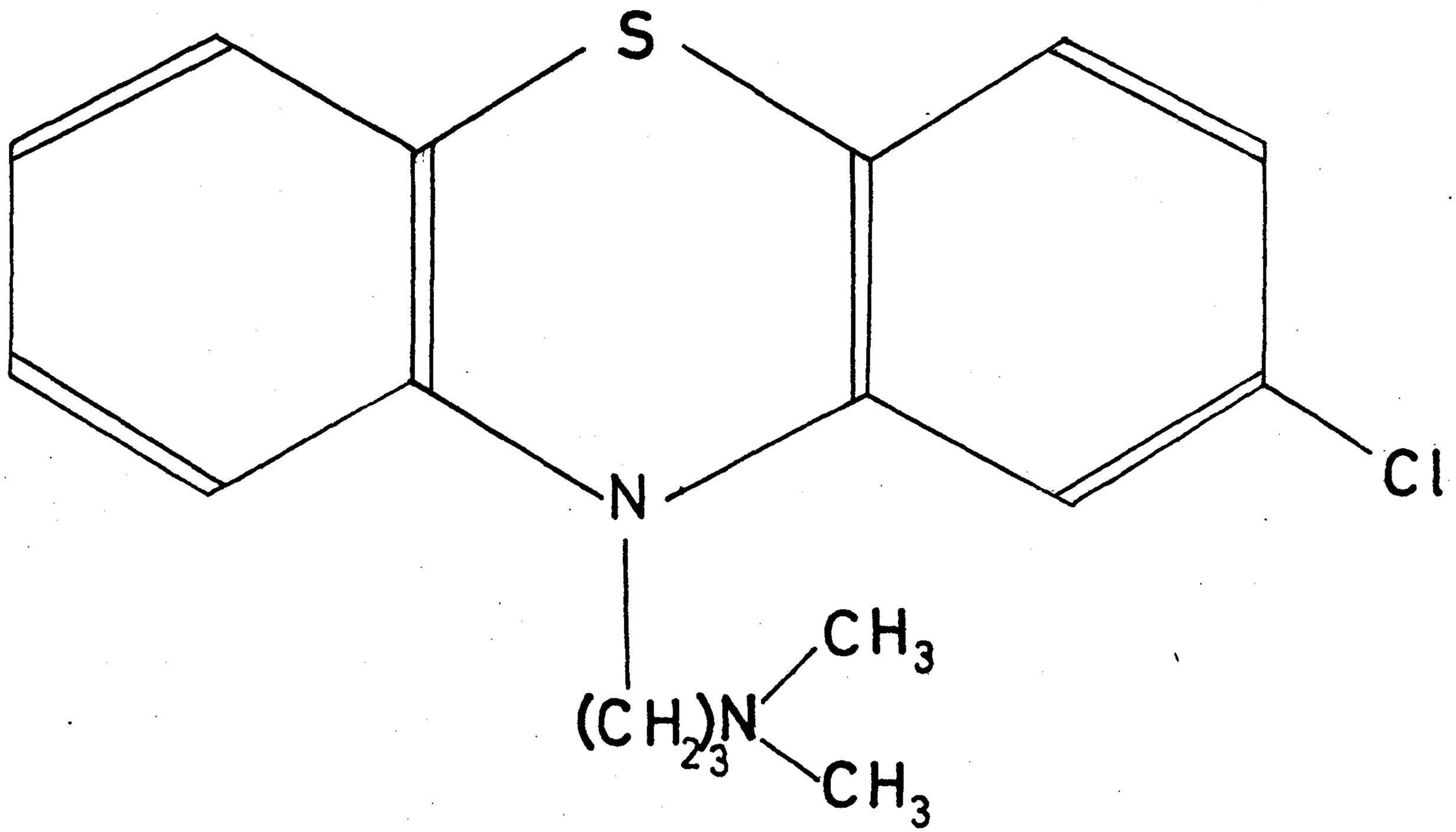


Fig. 8·1 CHLORPROMAZINE

8.3.1 Complexes involving native chlorpromazine

It has been reported (M. Waring, Private Communication) that the native (unoxidised) form of chlorpromazine probably does not bind to DNA. To test this, a number of fibres were pulled from gels made from solutions of DNA/CPZ mixtures. A spectral study of the solutions was also carried out in parallel in an attempt to determine the level of binding of the CPZ to the DNA in solution and in fibres.

CPZ has a strong absorption around 260nm and gives an ultra-violet spectrum which is qualitatively similar in the range 230-300nm to that obtained from DNA: a spectrum from CPZ is shown in figure 8.2. In figure 8.3 a series of spectra from solutions of DNA/CPZ are shown for a variety of values of the P/D ratio in the range 1.0 to ∞ (native DNA). Since the DNA concentration is the same in all the tubes, it is to be expected (and is observed) that the height of the peak around 260nm increases as the P/D decreases (i.e. greater concentration of chlorpomazine in the mixture). Moreover, whereas the peak for native DNA occurs at around 257nm; that from CPZ is observed at 254nm. Therefore, it is to be expected that an increase in the concentration of CPZ will result in a small, though measurable, decrease in the wavelength at which maximum absorption takes place. It is possible to observe such a decrease in the absorption maxima of the spectra in fig. 8.3.

Normally when a drug binds to DNA, the spectrum of the DNA, or of the drug is modified. If the drug does not interact significantly with DNA, it would be expected that the mixture would give spectra which are a linear weighted sum of the DNA and drug spectra. It is of interest to determine whether the spectra in fig. 8.3 can be obtained as a linear sum of the DNA and CPZ spectra

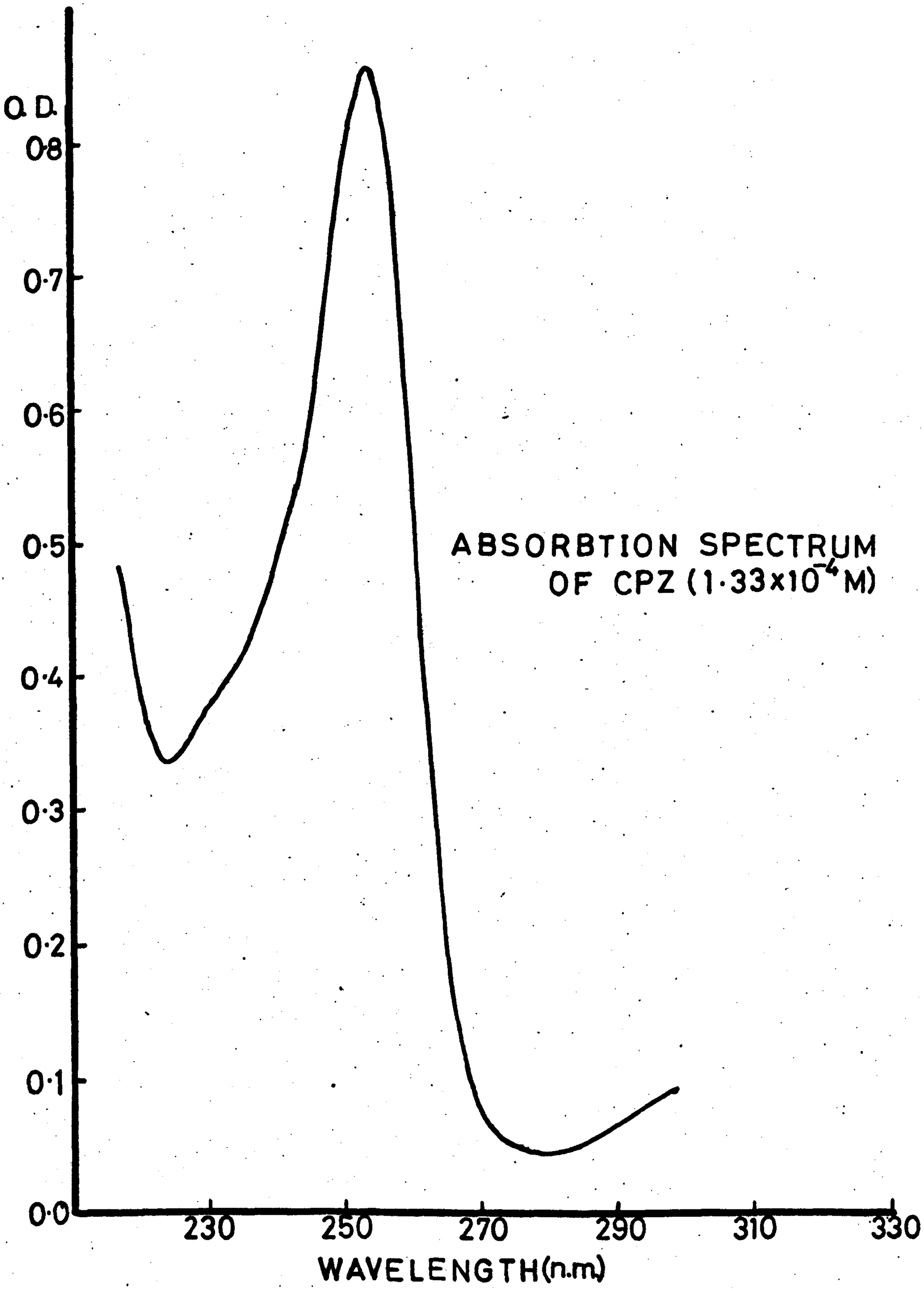

O.D.
0·8
0·7
0·6
0·5
0·4
0·3
0·2
0·1
0·0
230
250
270
290
310
330
WAVELENGTH(n.m)
ABSORBTION SPECTRUM
OF CPZ (1·33x10⁻⁴M)

Fig. 8·2

suitably weighted. Table 8.3 shows the values of the O.D. maxima for the spectra in fig. 8.3 and the values calculated on the basis of a linear sum of the two components. The optical density expected from the CPZ in each solution was determined from a graph of O.D. 260 against concentration for chlorpromazine solution (fig. 8.4). The drug does not appear to obey Beer's Law perfectly over the desired range. The predicted values of the maxima on the basis of a linear sum of the two components is not significantly different from that measured in the solutions. Hence, there is no spectral evidence for interaction between DNA and unoxidised chlorpromazine, although it is not possible to conclude definitely that no binding occurs.

It is of interest to determine the amount of chlorpromazine brought down in the DNA gel obtained from ultracentrifugation. An estimate of this can be obtained by measuring the displacement of the intensity maxima from 257nm (the position of the maxima for native DNA) in a solution obtained by redissolving the gel in buffer. Figure 8.5 shows the variation in the position of the intensity maxima as a function of P/D; while in figure 8.6 we have the spectra of two gels, of original P/D 20 and 10, which have been redissolved in buffer (0.02m NaCl/0.002m Tris): their intensity maxima occurring at 257.0 ± 0.02 and 256.3 ± 0.02 respectively. Since the intensity maximum for native DNA occurs at 257.0nm; there is no evidence of the presence of CPZ in the gel from the solution containing CPZ at a P/D equal to 20. The slight displacement of the maximum in the spectrum from the gel from a solution of P/D equal to 10 indicates that some small amounts of CPZ is present.

However, from the graph in fig. 8.5, it can be seen that this corresponds to a P/D of 42. Hence, there cannot be more than 25% of the CPZ in the original solution which is bound to the DNA. That some chlorpromazine is present in the fibres is shown by the fact that a number of fibres prepared from gels of DNA and CPZ became slightly red in colour after several weeks. This behaviour is characteristic of aqueous solutions of CPZ which, unless stored in the dark, tend to become slightly reddened due to light catalysed oxidation of a small percentage of the CPZ molecules.

If a drop of concentrated (10^{-2}m) CPZ solution is added to a solution of DNA, then some DNA local to the drop of CPZ precipitates and will not redissolve upon leaving in dilute salt solution for 10 days. This effect, which produces a rubbery precipitate, is presumably due to binding of the CPZ to the DNA, so that this gives evidence that such a complex may form; although the local concentration of chlorpromazine is very high in this situation and cannot be considered as being typical conditions for drug binding.

None of the diffraction patterns obtained showed any unusual feature attributable to the presence of chlorpromazine in the fibre. A number of the B patterns exhibited a reduced pitch and intermolecular separation, but since such behaviour has been noted (see chapter 6) for fibres of native DNA, it cannot be interpretted as an effect caused directly by the presence of chlorpromazine. One pattern having a pitch between 31 and 32Å had unusually sharp layer lines (Plate 8.1) and represents the best defined pattern of this type which has been found by the author.

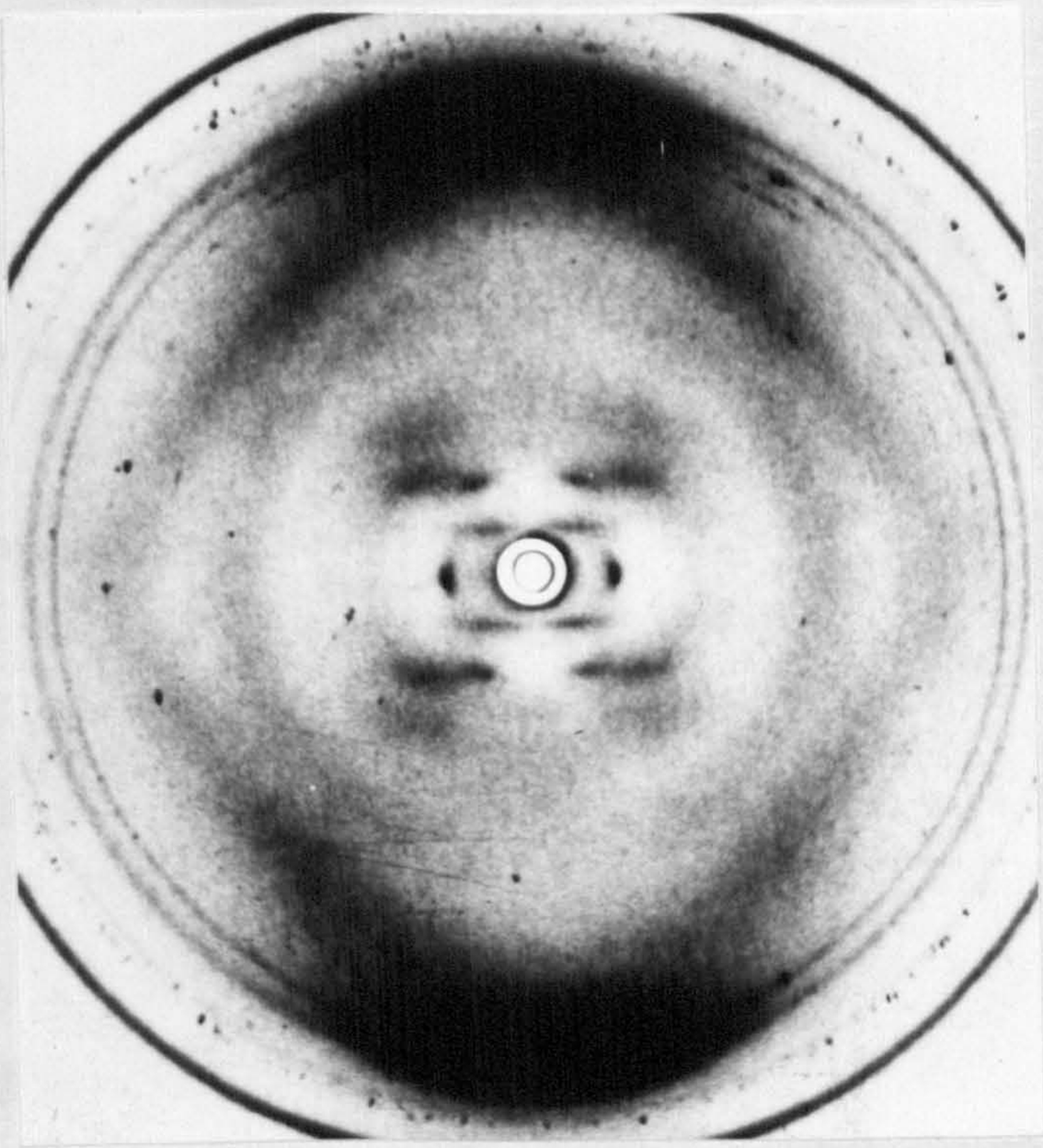


Plate 8.1 DNA-CPZ P/D = 5 95% R.H.

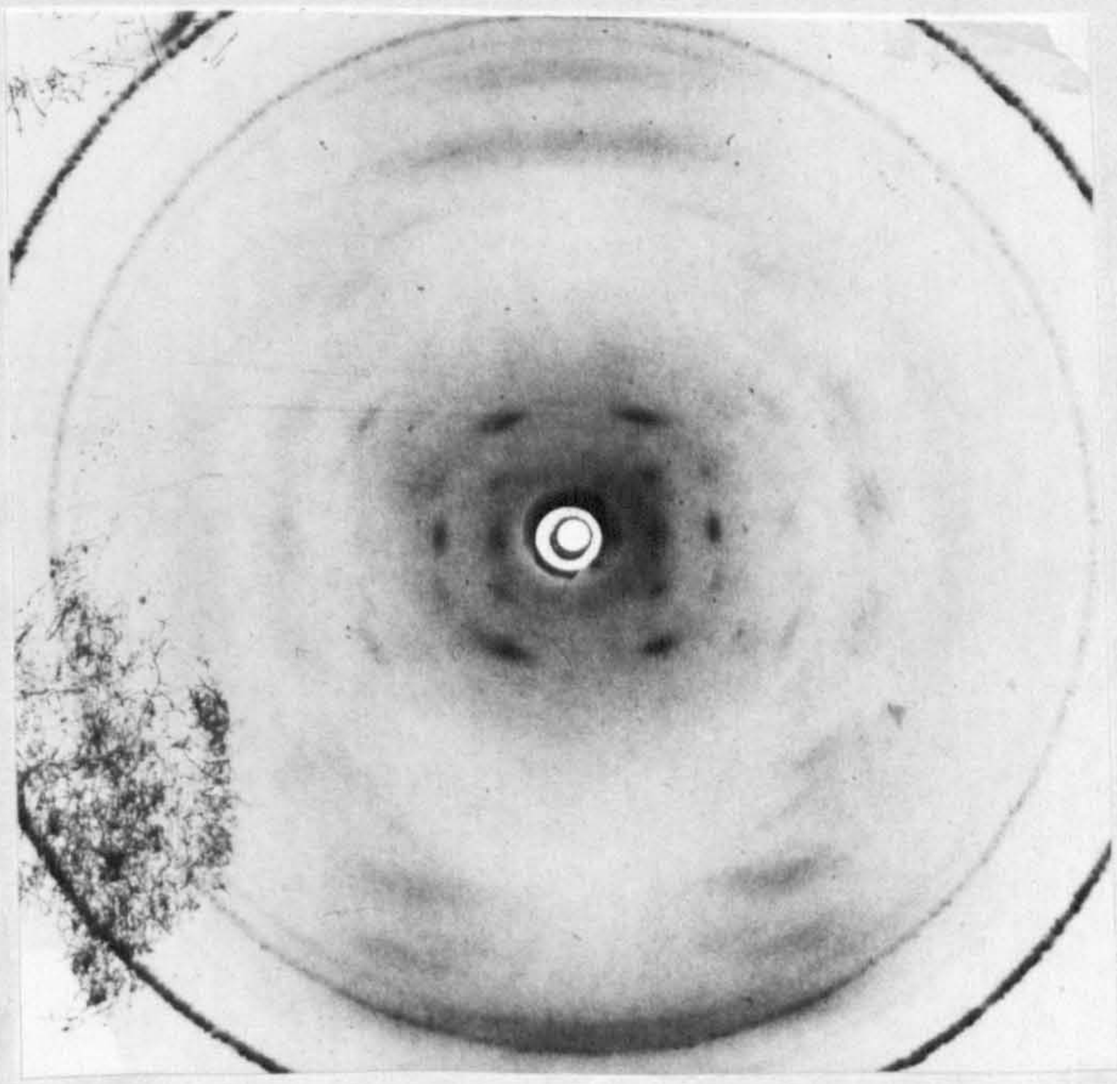


Plate 8.2 DNA-CPZ^{+} P/D = 10 75% R.H.

It was not reproducible since upon rephotograph, after being allowed to equilibrate at 92% relative humidity, the fibre gave a B pattern. Such "reduced pitch" B type patterns appeared to be more frequent in the presence of chlorpromazine than in native calf thymus DNA, and it might be that the small amount of CPZ which is retained in the fibre has a shielding effect on the DNA molecules and thereby inhibits the equilibration process at high relative humidity.

8.3.2. Complexes with the Chlorpromazine Free Radical

Fibres of DNA/CPZ^+ complexes gave much poorer patterns than were obtained with equivalent complexes involving the native form. Although CPZ^+ is more stable when bound to DNA (Ohnishi and McConnell, 1965) particularly in the fibre state, the radical decays slowly and, hence, fibres can normally only be photographed once before appreciable reduction has occurred. The decay is particularly rapid if the fibres are photographed at high $\geq$92% relative humidity, and for this reason most photographs were taken at 75% relative humidity or below. Photographs were generally taken on the rotating anode generator to minimise the exposure time so that photographs could be obtained before significant radical decay had taken place.

Because of the low relative humidities used, most photographs obtained were of the A form, although a small number of B type patterns were obtained. All photographs were of a poor quality; much more so than the patterns obtained from fibres of the native form of DNA, and the B patterns in particular were very poorly oriented indeed. Examples of some patterns obtained from CPZ^+/DNA complexes as illustrated in plates 8.2 to 8.4.

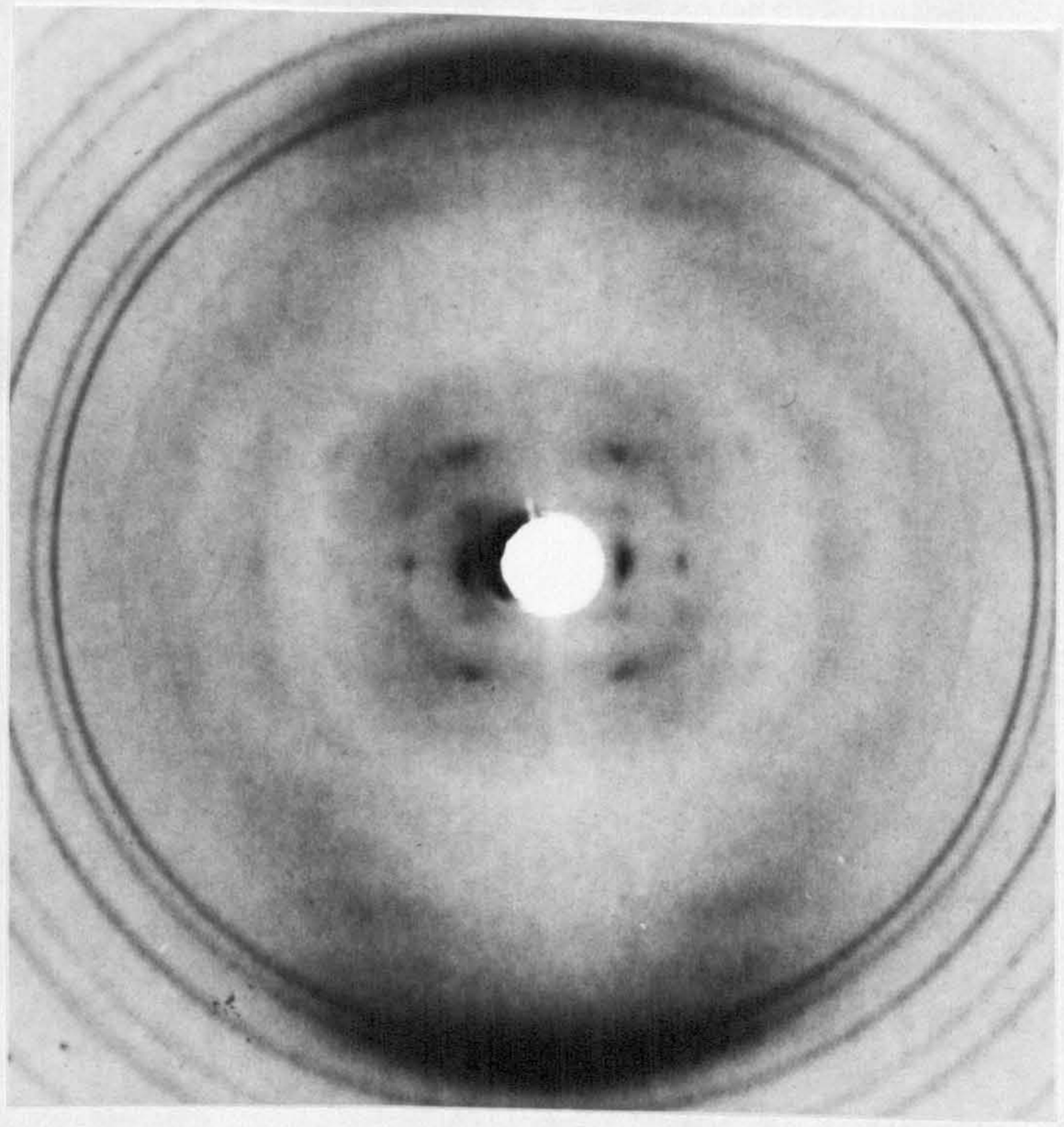


Plate 8.3 DNA-CPZ^+ P/D = 10 75% R.H.

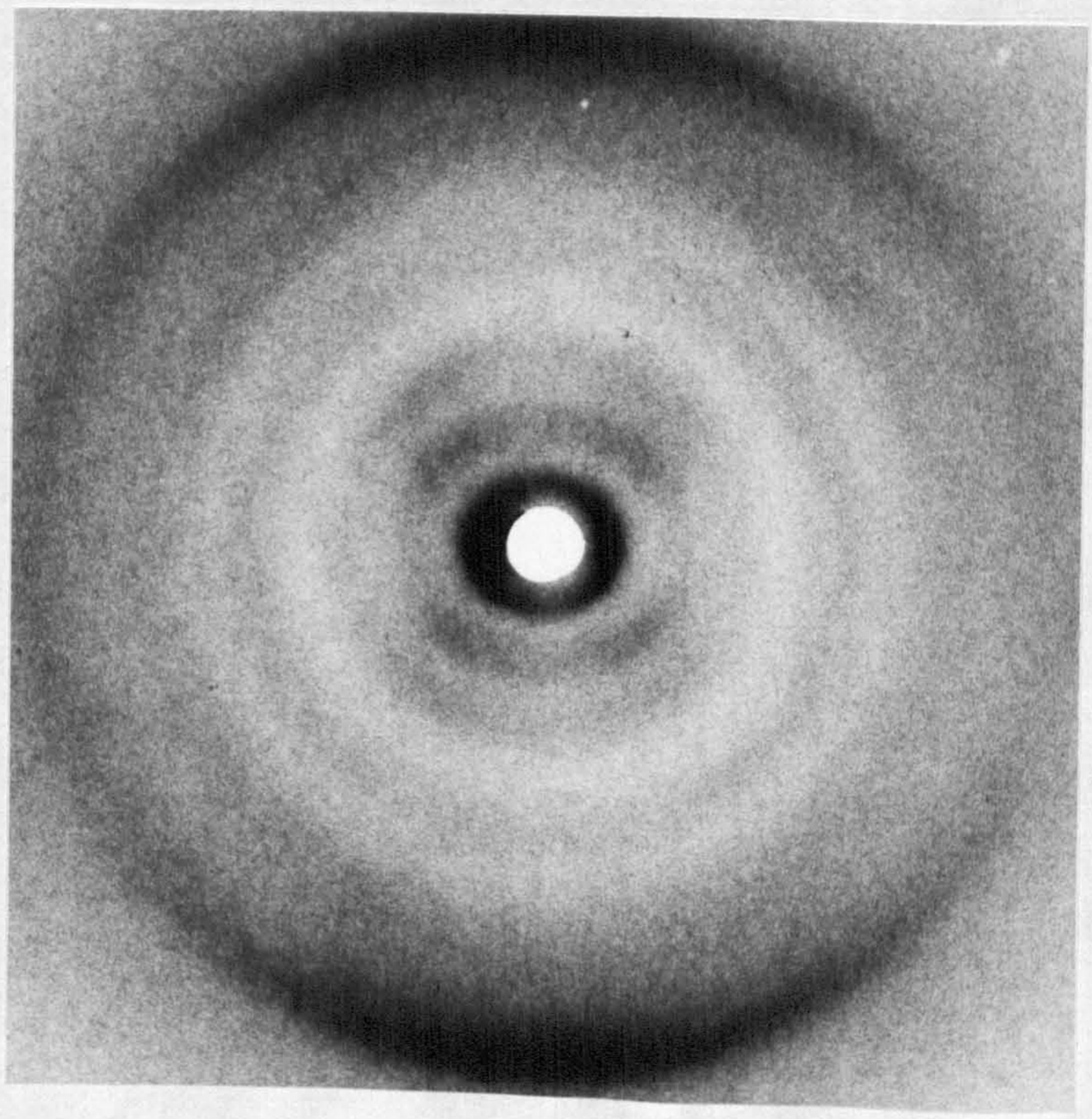


Plate 8.4 DNA-CPZ^+ P/D = 10 98% R.H.

TABLE 8.1

Observed and calculated O.D. maxima for CPZ/DNA solutions of given P/D

Solution (P/D)	O.D. (measured)	O.D. (calculated)
1	0.82	-
5	0.43	0.40
10	0.34	0.32
14	0.31	0.30
20	0.30	0.29
28	0.28	0.28
∞	0.27	0.26

All A patterns and patterns exhibiting an A/B mixture, were measured to determine the helix pitch and intermolecular separation. The spots resulting from the A conformation indexed as for native ADNA in all patterns. In the patterns exhibiting an A/B mixture, the pitch value measured was less than 34Å: the average being approximately 32Å. The intermolecular separation was also reduced in many cases in comparison with that observed in "good" sodium BDNA patterns. However, similar changes in these values for the BDNA component of patterns exhibiting an A/B mixture have been observed for native DNA (see chapter 6) so that they probably do not result from any direct effect of the CPZ^+. Measurements have been made on the B patterns obtained, but no changes in the pattern which could be attributed to intercalation were observed.

A number of measurements have been made to determine the relative intensity for different regions of the above patterns, particularly in the case of the ADNA data. In this case the intensity of the first spot on the second layer line was compared in intensity to that of the second spots on the equator and the first layer line. These correspond to spots numbers 1, 2 and 3 of the analysis of ϕW-14 A patterns discussed in chapter 7. It was not possible to detect a measurable difference, taking account of the large error due to the poor quality of the patterns obtained from the complex, between the relative values observed for CPZ^+/DNA fibres and for native DNA. It is not possible to conclude that no such difference exists however, particularly since the region of the pattern from which data is obtainable is so restricted.

8.4 Discussion

Diffraction patterns from a number of fibre prepared from DNA/CPZ and DNA/CPZ^+ solutions, but none of them have indicated that the chlorpromazine is intercalated. The evidence presented here is in accord with other evidence relating to CPZ^+ binding. Waring (1970) has shown that chlorpromazine does not unwind closed circular DNA as measured by the sedimenation coefficient. Moreover, although chlorpromazine bears a superficial resemblance to the acridine drugs it differs in that the ring system of chlorpromazine is not planar. The known intercalating drugs possess a planar ring system, and a non-planar drug may be more difficult to accommodate in the intercalation gap.

Evidence has also been presented here which suggests that native chlorpromazine does not bind significantly to DNA. In view of the instability of the chlorpromazine radical it appears unlikely that a large percentage of the drug injected *in vivo* exists in the oxidised state. Hence, it is probable that only a small proportion of the drug is bound to the DNA of a recipient when the drug is used therpeutically. Trabbuchi et al. (1974) have presented evidence indicating that chlorpromazine acts as a competitive inhibitor of cholinergic receptors in the brain. This may well be the physiological basis of its therapeutic action which may not, therefore, be related to the binding of DNA to chlorpromazine.

If CPZ^+ does not intercalate, it is necessary to explain the observations of Ohnishi and McConnell (1965) and of Porumb and Slade (1976) that the CPZ^+ ring system is oriented in a plane approximately perpendicular to the helix axis. The work of Ohnischi and McConnell was performed upon oriented solutions

of the complex in which the DNA is probably in the B conformation. In the studies of Porumb and Slade, however, the results were obtained from fibres at room humidity in which the DNA probably adopts the A conformation. Hence, any mechanism of binding should be applicable to either conformation and give similar values for the degree of orientation of the drug ring system.

One possibility is that CPZ^+ may act merely as a counter-cation to the DNA phosphate groups, and that the effect observed by Ohnischi and McConnell (1965) and Porumb and Slade (1976) is merely a self-stacking effect of the chlorpromazine molecule. This would explain the similarity of the results in the above two cases with differing DNA conformations. In view of the resultant proximity of the positive charges of free radical molecules, however, this explanation appears unlikely. Although self-stacking has been observed for the drug ethidium which also possessess a formal positive charge on the ring system so that such a mechanism may not be impossible for chlorpromazine.

It has proved possible to build external binding models which place the chlorpromazine in either of the grooves of BDNA or in the "hollow core" of ADNA. Since these models merely represent one member of a set of numerous similar conformations which cannot be eliminated by means of the diffraction or other data, details of them are not presented here.

8.5 Conclusion

The results of a survey of fibre diffraction patterns obtained from complexes between DNA and native chlorpromazine or the chlorpromazine free radical have shown that the drug probably does not intercalate into DNA after the fashion of Lerman (1961). This conclusion is in accord with the findings of Waring (1970)

that CPZ^+ does not unwind closed circular DNA as measured by the sedimentation coefficient of its complex with the free radical.

CHAPTER IX

CONCLUSIONS

The work presented in this thesis is concerned with the modification of the standard B conformation of natural DNA either as a result of its own primary sequence, or in response to the attachment of specific drug molecules some of which are known to bind to the DNA by intercalation. Modifications of the secondary structure of DNA may be biologically important, particularly with regard to control mechanisms concerned with DNA replication and gene expression. Evidence has been presented that the B conformation of sodium DNA may consist of a family of related structures; while ADNA appears to be defined by a unique conformation. Modified B forms have not been observed from M. lysodeikticus DNA which has a high G-C (72%) content, and it may be that A-T rich regions are more likely to adopt a modified structure.

In view of the small amount of evidence related to the conformations adopted by DNA in vivo, it is difficult to assess the biological importance of the above findings. However, A-T rich regions appeared to be involved in the inition of m-RNA synthesis by forming the binding site for RNA polymerase. The speculation has been made (Watson, 1975) that the lower stability of the A-T base pairs could allow the initiation site to be more easily denatured. Since local denaturation is required by RNA polymerase (and, indeed, by DNA polymerase during DNA replication) a lower stability of the initiation region is desirable. It is possible, though no experimental evidence has been obtained which bears upon the point, that the RNA

polymerase/σ factor complex which recognises the intiation site does so by detecting a particular conformation adopted by the A-T rich initiation region. Hence, the base distribution and composition of natural DNA might well be of greater biological significance than its primary role of coding for polypeptide sequences.

Under certain circumstances it might be advantageous for a DNA to behave in some ways as though its A-T/G-C ratio is different from that which it actually possesses. The DNA from the virus ØW-14 behaves in many ways as though it were very much richer in G-C base pairs than it in fact is. Its melting temperature is anomalously high for a DNA having approximately 50% A-T base pairs (Kropinski et. al. 1973), while the X-ray diffraction patterns taken inthis laboratory showed that ØW-14 DNA adopted the A form more readily than did calf thymus DNA, although the B patterns which were obtained were of a good quality and never exhibited a low pitch form. This behaviour is similar to that exhibited by the G-C rich M. lysodeikticus DNA, although the stabilisation of the A form is not nearly so marked in this DNA as is the case with that from ØW-14. The detailed biological significance of these behavioural changes is difficult to assess because of the difficulty outlined in paragraph 2, but it may effect the relative rate at which ØW-14 and host genes are expressed.

Suppression of m-RNA synthesis and/or DNA replication can be achieved in many in vivo systems by the use of drugs which are known to intercalate into DNA in vitro. All intercalation complexes in fibres give diffraction patterns indicating that the DNA is in B type structure even at lower humidity values at which ADNA would normally be found. If this proved to the mechanism by which intercalative drugs produced their therapeutic effect, then the

suppression of gene expression by these drugs might well be related to the mode of action of a number of proteins (such as repressor proteins) which also inhibit gene expression. However, the inhibition produced by intercalating drugs could equally well be due to the physical presence of the drug chromophore which might represent a pseudo-base pair which, hence, cannot be separated by the relevant enzymes. In this case intercalation may not be, mechanistically, a good analogue of in vivo control processes since proteins probably possess few, if any, groups suitable for intercalation.

Studies of the type presented here have implications outside the field of nucleic acid structure. DNA is rare amongst helical polymers in being able to adopt a number of quite distinct, regular conformations in which the transition between them are controlled by relatively mild changes in environmental conditions. The different DNA conformations may reflect differences in base stacking as a function of the environment, and studies on polymorphism of DNA as a result of differences in base composition or the influence of external agents such as drugs is of significance in relation to quantum mechanical models which predict the properties of the bases and related compounds.

In relation to this, further work needs to be done to increase the extent to which the nature of the environment in fibres can be controlled and determined.

If the biological significance of changes in DNA secondary structure at to be assessed, more information is required which bears upon the conformations adopted by DNA in vivo. From the point of view of the X-ray diffraction studies, it is clear that more research is required to improve the methods by which information is extracted from the diffraction patterns obtained. At present the analysis of patterns consisting of diffuse diffraction data is not

sufficiently precise to allow a rigorous structural analysis of the speciments which produce them.

References

Alden, C.J. and Arnott, S., (1975), Nuc. Acid. Research, 2, 1701.

Anguili, R., Foresti, E., Riva di Sanseverino, L., Isaacs, N.W., Kennard, O., Motherwell, W.D.S., Wampler, D.L., and Arcamone, F., (1971), Nature New Biology, 234, 78.

Arnott, S., (1973), Trans. Am. Cryst. Assoc., 9, 31.

Arnott, S. and Wonacott, A.J., (1966), Polymer, 7, 157.

Arnott, S., Dover, S.D. and Wonacott, A.J., (1969), Acta. Cryst., B25, 2192.

Arnott, S. and Hukins, D.W.L., (1972), Biochem. Biophys. Res. Commun., 47, 1504.

Arnott, S. and Hukins, D.W.L., (1973), J. Mol. Biol., 81, 93.

Arnott, S., Chandrasakaran, R., Hukins, D.W.L., Smith, P.J.C. and Watts, L., (1974a), J. Mol. Biol., 88, 523.

Arnott, S. and Selsing, E., (1974b), J. Mol. Biol., 88, 509.

Bauer, W. and Vinograd, J., (1968), J. Mol. Biol., 33, 141.

Blakeley, P.J., (1976), Ph.D. Thesis, University of Keele.

Bond, P.J., Langridge, R., Jenette, K.W. and Lippard, S.J., (1975), Proc. Nat. Acad. Sci., 72, 4825.

Brahams, J., Pilet, J., Iran Hi Phuong Lan and Hill, L.R., (1973), Proc. Nat. Acad. Sci., 70, 3352.

Bram, S., (1971), Nature New Biology, 232, 174.

Bram, S., (1972), Biochem. Biophys. Res. Commun. 48, 1088.

Bram, S., (1973), Proc. Nat. Acad. Sci., 70, 2167.

Bram, S. and Beeman, W.W., (1971), J. Mol. Biol., 55, 311.

Bram, S. and Tougard, P., (1972), Nature New Biology, 239, 128.

Buxton-Hopkin, D.A., (1955), Pharm. J., 174, 317.

Chamberlain, M., McGrath, J. and Waskell, L., Nature, 228, 227.

Clark, S., Losick, R., Pero, J., (1974), Nature, 252, 21.

Cochran, W., Crick, F.H.C. and Vand, V., (1952), Acta. Cryst., 5, 581.

Cooper, P.J. and Hamilton, L.D., (1966), J. Mol. Biol., 16, 562.

Crick, F.H.C., Barnett, L., Brenner, S. and Watts-Tobin, R.J., (1961), Nature, 192, 1227.

Crawford, L.V. and Waring, M.J., (1967), J. Mol. Biol., 25, 23.

Davies, D.R. and Baldwin, R.L., (1963), J. Mol. Biol., 6, 251.

Elliott, A., (1965), J. Sci. Inst., 42, 312.
Eyring, H., (1932), Phys. Rev., 39, 746.
Frank, A., (1958), Brit. J. Appl. Phys., 9, 349.
Franklin, R.E. and Gosling, R.G., (1953a), Acta. Cryst., 6, 673.
Franklin, R.E. and Gosling, R.G., (1953b), Acta. Cryst., 6, 678.
Fraser, R.D.B., Macrae, T.P., Miller, A. and Rowlands, R.J., (1976), J. Appl., Cryst., 9, 81.
Friedrich-Freska, H., (1940), Naturwissenschaften, 28, 376.
Fuller, W. and Waring, M.J., (1964), Ber. Bunsenges. Phys. Chem., 68, 805.
Fuller, W., Wilkins, M.H.F., Wilson, H.R. and Hamilton, L.D., (1965), J. Mol. Biol., 12, 60.
Fuller, W., Hutchinson, F., Spencer, M. and Wilkins, M.H.F., (1967), 27, 507.
Goldberg, I.H., Rabinowitz, M. and Reich, E., (1962), Proc. Nat. Acad. Sci., Wash., 48, 2094.
Haen, C., Swann, E. and Teller, D.C., (1976), Biopolymers, 15, 1825.
Haselkorn, R., Vogel, M. and Brown, R.D., (1969), Nature, 221, 836.
Hamilton, L.D., Barclay, D.M.R.K., Wilkins, M.H.F., Brown, G.L., Wilson, H.R., Marvin, D.A., Ephrussi-Taylor, H. and Simmons, N.S., (1959), J. Biophys. Biochem. Cytol., 5, 397.
Holmes, K.C. and Barrington-Leigh, J., (1974), Acta. Cryst., A30, 635.
IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains., (1970), Biochemistry, 9, 3471.
Jain, S.C. and Sobell, H.M., (1972), J. Mol. Biol., 68, 1
Kropinski, A.M.B., Bose, R.J. and Warren, R.A.J., (1973), Biochemistry, 12, 151.
Langridge, R., Marvin, D.A., Seeds, W.E., Wilson, H.R., Hooper, C.W., Wilkins, M.H.F. and Hamilton, L.D., (1960), J. Mol. Biol., 2, 38.
Leng, M., Drocourt, J., Helene, C. and Ramstein, J., (1974), Biochemie, 56, 887.
Lerman, L.S., (1961), J. Mol. Biol., 3, 18.
Lerman, L.S., (1963), Proc. Nat. Acad. Sci., 49, 94.
Marvin, D.A., Spencer, M., Wilkins, M.H.F. and Hamilton, L.D. (1961), J. Mol. Biol., 3, 547.
Massie, H.R. and Zimm, B.H., (1965), Proc. Nat. Acad. Sci., 54, 1641.

Meselson, M. and Stahl, F.W., (1958), Proc. Nat. Acad. Sci., 44, 671.
Milman, G., Chamberlain, M. and Langridge, R., (1967), Proc. Nat. Acad. Sci., 57, 1804.
Mitsui, Y., Langridge, R., Shortle, B.E., Cantor, C.R., Grant, R.C., Kodama, M. and Wells, R.D., (1970), Nature, 228, 1166.
Moulton, W.G. and Kromhout, R.A., (1956), J. Chem. Phys., 25, 34.
Muller, H.J., (1947), "The Gene", Proc. Roy. Soc., (London), B134, 1.
Ohnishi, S. and McConnell, H.M., (1966), J. Am. Chem. Soc., 87, 2293.
Paoletti, J. and Le Pecq, J.B., (1971), J. Mol. Biol., 59, 43.
Pauling, L. and Delbruck, M., (1940), Science, 92, 77.
Piette, L.H., Bulow, G. and Yamazaki, I., (1964), Biochem. Biophys. Acta, 88, 120.
Pigram, W.J. (1968), Ph.D. Thesis, University of London.
Pigram, W.J., Fuller, W. and Hamilton, L.D., (1972), Nature New Biology, 235, 17,
Pigram, W.J., Fuller, W. and Davies, M.E., (1973), J. Mol. Biol., 80, 361.
Pilet, J. and Brahms, J., (1972), Nature New Biology, 236, 99.
Pohl, F.M., (1976), Nature, 260, 365.
Porumb, T. and Slade, E.F., (1976), Eur. J. Biochem., 65, 21.
Pulleybank, D.E. and Morgan, A.R., (1975), J. Mol. Biol., 91, 1.
Ramachandran, G.N., (1960), Proc. Nat. Acad. Sci., 52, 240.
Ramachandran, G.N., Venkatachalam, C.M. and Krim, S., (1966), Biophys. J., 6, 849.
Ramachandran, G.N. and Sasisekharan, V., (1968), Advan. Protein. Chem., 23, 283.
Scott, R.A. and Scheraga, H.A., (1966), J. Chem. Phys., 45, 2091.
Selsing, E. and Arnott, S., (1976), Nucleic Acid Research, 3, 2443.
Stockmayer, W.H., (1941), J. Chem. Phys., 9, 398.
Sobell, H.M. and Jain, S.C., (1972), J. Mol. Biol., 68, 21.
Subramanian, Frotter and Bugs, (1971), J. Cryst. Mol. Struct., 1, 3.
Tsai, C., Jain, S.C. and Sobell, H.M., (1975), Proc. Nat. Acad. Sci., 72, 628.
Vainshtein, B.K., (1966), "Diffraction of X-rays by Chain Molecules", English Translation, Elsevier Publishing Company.
Wang, J.C., (1974), J. Mol. Biol., 89, 783.
Waring, M.J., (1965), J. Mol. Biol., 13, 282.
Waring, M.J., (1966), Biochem. Biophys. Acta., 114, 234.

Waring, M.J., (1970), J. Mol. Biol., 54, 247.
Waring, M.J., (1972), "The Molecular Basis of Antibiotic Action" (Ed. Gale, E.F., Cundliffe, E., Reynolds, P.E., Richmond, M.H. and Waring, M.J.) pp.173-277, Wiley, London.
Watson, J.D., (1975), "Molecular Biology of the Gene", W.A. Benjamin, Inc.
Watson, J.D. and Crick, F.H.C., (1953a), Nature, 171, 737.
Watson, J.D. and Crick, F.H.C., (1953b), Nature, 171, 964.
Wells, R.D. and Larson, J.E., (1970), J. Mol. Biol., 49, 319.
Wilkins, M.H.F., (1961), Unpublished. (Details quoted in Marvin et al, (1961)).
Zernicke, F. and Prins, J.A., (1927), Z. Phys., 41, 184.