

**Structural studies of DNA-drug complexes.**

A Thesis Presented for the Degree of

Doctor of Philosophy

of the

University of London

by

David G. Brown. BSc.

Institute of Cancer Research  
Sutton,  
Surrey.

January 1992

## UNIVERSITY OF LONDON THESIS

Degree PhD Year 1992 Name of Author BROWN, D.G.

### COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

### COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

### LOAN

Theses may not be lent to individuals, but the University Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

### REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the University of London Library. Enquiries and orders should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. **Before 1962** Permission granted **only** upon the prior written consent of the author. (The University Library will provide addresses where possible)
- B. **1962-1974** In **many** cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. **1975-1988** **Most** theses may be copied upon completion of a Copyright Declaration.
- D. **1989 onwards** **Most** theses may be copied.

This thesis comes within category D.

This copy has been deposited in the Library of INSTITUTE OF CANCER RESEARCH.

10/013/THESESLABE  
EJR/DW/23/8/88

## ABSTRACT.

This thesis reports on the use of biophysical techniques to elucidate high resolution structural details of the interactions between oligonucleotides and DNA-binding ligands. Single-crystal X-ray diffraction and Nuclear Magnetic Resonance were employed as the main investigative methods.

The research concentrates mainly on the minor groove-binding ligand, berenil [1,3-*bis*(4-phenylamidinium)triazene], which has veterinary application as an anti-trypanosomal agent and also cytotoxic and anti-viral properties. It has previously been shown to bind reversibly to double-helical DNA, primarily at AT-rich regions. The crystal structures of berenil complexed with  $d(\text{CGCGAATTCGCG})_2$  and  $d(\text{CGCAAATTTGCG})_2$  are reported, together with the solution structure of the latter complex as determined by NMR.

These studies confirm the AT preference of the drug and reveal details of the interactions involved. The crystallographic study of the complex with  $d(\text{CGCGAATTCGCG})_2$  shows the molecule bound across the DNA strands, via the minor groove, and spanning three bases (5'-ATT-3'). The drug binds via a hydrogen bond to N3 of adenine at the 3' end whilst at the 5' end hydrogen bond contacts are via a bridging water molecule to the N3 of adenine. The study of the complex with  $d(\text{CGCAAATTTGCG})_2$  revealed a 1:4 binding model with the berenil symmetrically-disposed about the twofold helix axis (5'-AATT-3'). In this complex the ligand is hydrogen-bonded directly to O2 of thymine at both the 5' and 3' ends. The difference in binding between the two structures is explained by the difference in local helical parameters for the two dodecamers. NMR studies revealed at least two binding states which gave rise to two models. One was a 1:2 model disposed symmetrically about the helix dyad and the other a 1:3 cross-strand model (5'-ATT-3') sensing the N3 of adenine at both the 5' and 3' ends, in accordance with the NMR model for the  $d(\text{CGCGAATTCGCG})_2$ -berenil complex. The crystal structure of another minor groove-binding drug, Hoechst 33258, complexed with  $d(\text{CGCAAATTTGCG})_2$  is also reported.

**For my parents.**

## ACKNOWLEDGEMENTS

I would like to take this opportunity to thank all the people who have aided me during my research.

Thanks to my supervisor Professor Stephen Neidle who guided me through three years of research and kept me on course whilst writing my thesis. Special thanks to Dr Mark Sanderson for his friendship, patience and the kindness to assist me in times of difficulty.

I would also like to thank Dr. Terry Jenkins, for his extensive consultations and Drs. Jane Skelly and Charlie Laughton for their help and advice. I would specially like to thank Drs. Andrew Lane and Elspeth Garman for their invaluable assistance during my PhD.

My thanks also go to the Institute of Cancer Research and the Cancer Research Campaign, who funded my research.

Finally I would like to thank my friends and colleagues both inside and outside the laboratory but especially Hilary Hall and Erwin Dorland who have suffered mutely during my research.

## CONTENTS

### CHAPTER 1.

Introduction to DNA structure and nomenclature.	15
<b>1.1 Introduction.</b>	16
<i>1.2 Torsion angle nomenclature.</i>	19
<i>1.3 Sugar conformation.</i>	20
<i>1.4 The phosphodiester linkage.</i>	22
<i>1.5 The glycosyl bond.</i>	22
<i>1.6 Backbone conformation.</i>	25
<i>1.7 Translational and rotational parameters.</i>	25
<i>1.8 Groove widths.</i>	28
<i>1.9 The polynucleotide families.</i>	28
<b>1.10 DNA recognition.</b>	32
<i>1.11 Protein-DNA recognition.</i>	34
<i>1.12 Minor groove interactions with DNA-binding proteins.</i>	35
<b>1.13 Drug-DNA recognition.</b>	37
<i>1.14 Intercalators.</i>	39
<i>1.15 Mode of action of intercalators.</i>	42
<i>1.16 Minor groove binders.</i>	44
<i>1.17 Mode of action of minor groove binders.</i>	46
<i>1.18 Covalent groove binders.</i>	50
<i>1.19 Triple helices.</i>	51

### CHAPTER 2.

The Crystal structure of a complex between the dodecamer $d(\text{CGCGAATTCGCG})_2$ and the minor groove-binding ligand berenil.	54
--	----

<b>2.1 Introduction</b>	55
2.2 <i>Crystallisation and data collection.</i>	56
2.3 <i>Structure solution and refinement.</i>	57
2.4 <i>General structural features.</i>	61
2.5 <i>Hydration in the crystal structure.</i>	66
2.6 <i>Packing in the crystal.</i>	70
2.7 <i>Analysis of helical parameters.</i>	70
2.8 <i>Comparison with the NMR data of the complex of d(CGCGAATTCGCG)<sub>2</sub> with berenil and the complex of d(GCAATTGC)<sub>2</sub> with berenil.</i>	78
2.9 <i>Molecular modelling.</i>	80
<b>2.10 Discussion</b>	81

## CHAPTER 3.

The Crystal structure of a complex between the dodecamer d(CGCAAATTTGCG) <sub>2</sub> and the minor groove-binding ligand berenil.	96
--	----

<b>3.1 Introduction.</b>	97
3.2 <i>Oligonucleotide synthesis.</i>	97
3.3 <i>Crystallisation and data collection.</i>	98
3.4 <i>Structure solution and refinement.</i>	100
3.5 <i>Overall structural features.</i>	104
3.6 <i>Analysis of helical parameters.</i>	109
3.7 <i>Are three-centred hydrogen bonds present?</i>	116
3.8 <i>Hydration of the crystal structure.</i>	121
<b>3.9 Discussion.</b>	124

## CHAPTER 4.

Nuclear magnetic resonance spectroscopic and molecular modelling studies of the interaction of berenil with the dodecamer d(CGCGAATTCGCG)<sub>2</sub>. 137

**4.1 Introduction.** 138

*4.2 Materials.* 140

*4.3 Optical spectroscopy of DNA-berenil complex.* 140

*4.4 NMR spectroscopy.* 141

*4.5 Molecular modelling.* 143

*4.6 Molecular mechanics protocol.* 144

*4.7 Molecular dynamics protocol for DNA-berenil complexes.* 146

*4.8 Optical spectroscopy.* 149

*4.9 NMR spectroscopy of the uncomplexed DNA.* 152

*4.10 NMR spectroscopy of the DNA-berenil complex.* 155

*4.11 Molecular modelling and energy minimisation of the DNA-berenil Complexes.* 162

*4.12 Dynamic simulation of the DNA-berenil complex.* 163

*4.13 Centrosymmetric DNA-berenil model.* 165

*4.14 Non-centrosymmetric DNA-berenil model.* 166

**4.15 Discussion.** 167

## CHAPTER 5.

The Crystal structure of a complex between the dodecamer d(CGCAAATTTGCG)<sub>2</sub> and the minor groove-binding ligand Hoechst 33258. 170

**5.1 Introduction** 171

*5.2 Oligonucleotide synthesis.* 173

5.3	<i>Crystallisation and data collection.</i>	173
5.4	<i>Structure solution and refinement.</i>	174
5.5	<i>Analysis of models detailing their various interactions with the DNA.</i>	177
5.6	<i>Analysis of the “pip-down” model.</i>	179
5.7	<i>Analysis of the “pip-up” model.</i>	180
5.8	<b>Discussion.</b>	180
CHAPTER 6.		
	General conclusions on minor groove-binding ligands.	184
6.1	<i>The factors influencing binding of positively-charged minor groove-binding ligands.</i>	185
APPENDIX I		
	Crystallographic and refinement theory.	190
APPENDIX II		
	Nuclear magnetic resonance and X-PLOR refinement.	206
	REFERENCES	218

## TABLES

### CHAPTER 1

- 1.1 The general helical parameters for A, B & Z forms of DNA. 30
- 1.2 DNA-drug complexes solved by crystallography. 41

### CHAPTER 2

- 2.1 Resolution breakdown of data collected. 57
- 2.2 Refinement statistics including parameters used for  
d(CGCGAATTCGCG)<sub>2</sub> complexed with berenil. 62
- 2.3 Conformational parameters for the B-DNA dodecamer  
d(CGCGAATTCGCG)<sub>2</sub> native and complexed with berenil. 71
- 2.4 Local helix parameters for the B-DNA Dodecamer  
d(CGCGAATTCGCG)<sub>2</sub> native and complexed with berenil. 73
- 2.5 Calculated energies for the octamer d(CGAATTCG)<sub>2</sub>-berenil  
complexes. 81
- 2.6 Final coordinates and temperature factors for the DNA dodecamer  
d(CGCGAATTCGCG)<sub>2</sub> complexed with berenil  
including solvent molecules. 85

### CHAPTER 3

- 3.1 Resolution breakdown of data collected. 100
- 3.2 Refinement parameters used and statistics from the NUCLSQ  
refinement of d(CGCAAATTTGCG)<sub>2</sub> complexed with berenil. 102
- 3.3 Final agreement statistics. 104
- 3.4 Conformational parameters for the B-DNA dodecamer  
d(CGCAAATTTGCG)<sub>2</sub> complexed with berenil and  
d(CGCGAATTCGCG)<sub>2</sub>. 110
- 3.5 Local Helix Parameters for the B-DNA Dodecamer

d(CGCAAATTTGCG) <sub>2</sub> complexed with berenil and d(CGCGAATTCGCG) <sub>2</sub> .	112
3.6 Final coordinates and temperature factors for the DNA dodecamer d(CGCAAATTTGCG) <sub>2</sub> complexed with berenil including solvent molecules.	126
CHAPTER 4	
4.1 Binding properties for various DNA-berenil complexes	142
4.2 Calculated energies for the dodecamer-berenil complexes.	145
4.3 NMR properties of d(CGCAAATTTGCG) <sub>2</sub> in the absence and presence of berenil.	148
4.4 Interproton NOE distances for the A <sub>3</sub> T <sub>3</sub> -berenil complex.	151
4.5 Optical changes on binding of berenil to A <sub>2</sub> T <sub>2</sub> and A <sub>3</sub> T <sub>3</sub> .	151
4.6 NMR assignments for nonexchangeable protons of A <sub>3</sub> T <sub>3</sub> in the absence of ligand.	153
4.7 NMR assignments for nonexchangeable protons of A <sub>3</sub> T <sub>3</sub> in the presence of berenil.	156
4.8 <sup>31</sup> P NMR assignments for A <sub>3</sub> T <sub>3</sub> in presence and absence of berenil.	157
CHAPTER 5.	
5.1 Resolution breakdown of data collected.	174

## FIGURES

### CHAPTER 1.

1.1	The five nucleic acid bases.	17
1.2	The normal Watson-Crick base pairings.	18
1.3	One nucleotide unit showing atomic and torsion angle nomenclature.	21
1.4	The four most common sugar puckers.	21
1.5	The Pseudorotation cycle.	23
1.6	B <sub>I</sub> and B <sub>II</sub> backbone conformation.	24
1.7	The six most favourable conformations of $\alpha$ and $\zeta$ .	24
1.8	The eight translational parameters.	26
1.9	The eight rotational parameters.	27
1.10	The conformations of A, B and Z-DNA.	31
1.11	Watson-Crick base-pairing showing hydrogen bond donors/acceptors.	33
1.12	Comparison of protein-DNA contacts present in the engrailed and $\alpha 2$ homeodomain-DNA complexes.	36
1.13	Cartoon showing possible modes of ligand-DNA interaction.	38
1.14	Chemical structure of three intercalators.	40
1.15	Höogsteen base-pairing.	43
1.16	Chemical structures of well studied minor groove binders.	45
1.17	Covalent binders CC-1065 and a nitrogen mustard.	52

### CHAPTER 2.

2.1	The chemical structure of berenil.	56
2.2	Stereo view of final omit difference fourier map.	60
2.3	Stereo view of structure showing berenil in the minor groove.	60
2.4	Schematic representation of berenil bound to the dodecamer.	63

2.5	Close up view of hydrogen bonding at one end of the bound berenil.	64
2.6	Close up view of hydrogen bonding at the other end of the bound berenil, showing the network involving the bridging water molecule.	64
2.7	Graph of minor groove width based on phosphate-phosphate distances.	66
2.8	Schematic representation of major groove hydration.	67
2.9	Schematic representation of the hydration of the phosphate backbone.	68
2.10	Schematic representation of the minor groove hydration.	69
2.11-22	Newhelix parameters.	75
2.23	Definition of "class I and class II" binding modes.	80

### CHAPTER 3.

3.1	Stereo views of complete structure showing berenil in the minor groove.	105
3.2	Schematic of berenil binding modes.	107
3.3	Detailed interactions between berenil and the dodecamer.	108
3.4	Graph of propellor twist for $A_2T_2$ & $A_3T_3$ native and complexes with berenil.	115
3.5	Graph of base pair roll for $A_2T_2$ . & $A_3T_3$ native and complexes with berenil.	115
3.6	Three-centred vs. bifurcated hydrogen bond.	117
3.7	Hydrogen-bonding geometries for the major groove of the crystal structure of the complex.	119
3.8	Schematic representation of phosphate backbone hydration.	120
3.9	Schematic representation of major groove hydration.	122
3.9a	Schematic representation of the coordination around magnesium cation.	123

3.10 Schematic representation of minor groove hydration showing berenil.	125
--	-----

#### CHAPTER 4.

4.1 Chemical structure <sup>of</sup> berenil including the proton numbering.	140
4.2 Schematic representations of “class I” and “class II” models for berenil binding.	146
4.3 Circular dichroism spectra for $A_3T_3$ in the presence and absence of berenil and the difference plot of the CD curves.	150
4.4 Two-dimensional spectra for $A_3T_3$ in the absence of ligand.	154
4.5 Change in chemical shift and spectral linewidth of H6 (Thy7) with increasing concentration of berenil.	158
4.6 Plots of chemical shift differences for $A_3T_3$ in the presence of berenil.	159
4.7 Plots of linewidth and chemical shift for $A_3T_3$ in the presence of varying concentration of berenil.	161
4.8 Stereo view of non-centrosymmetric berenil model.	168

#### CHAPTER 5.

5.1 Chemical structure of Hoechst 33258.	173
5.2 Photograph of electron density (omit map).	177
5.3 Schematic of interaction with $d(\text{CGCGAATTCGCG})_2$ “pip-down”.	178
5.4 Schematic of interaction with $d(\text{CGCGAATTCGCG})_2$ “pip-up”.	178
5.5 Graph of the minor groove width based on phosphorus-phosphorus distances.	179
5.6 Schematic of interaction with $d(\text{CGCGAATTCGCG})_2$ “pip-down”.	181
5.7 Schematic of interaction with $d(\text{CGCGAATTCGCG})_2$ “pip-down”.	181
5.8 Schematic of interaction with $d(\text{CGCGATATCGCG})_2$ “pip-up”.	182
5.9 Schematic of interaction with $d(\text{CGCGAATTCGCG})_2$ “pip-down”.	182

## APPENDIX II.

II.1 A generalised representation of the 2D-NMR experiment.	209
II.2 Pulse schemes for NOESY and TOCSY experiments.	209
II.3 A simple two-dimensional minimisation.	214

In this thesis the *EcoRI* recognition sequence  $d(\text{CGCGAATTCGCG})_2$  is sometimes referred to as A2T2 and the sequence  $d(\text{CGCAAATTTGCG})_2$  is sometimes referred to as A3T3.

## **CHAPTER 1.**

Introduction to DNA structure and nomenclature.

## 1.1 Introduction.

DNA (deoxyribonucleic acid) is both the vector for genetic inheritance and the polymer that governs the processes of growth, differentiation and regulation in the cell. It performs all these tasks by virtue of its molecular structure.

DNA is a linear polymer built up of monomeric subunits, the nucleotides. A nucleotide consists of three fragments, a furanose sugar connected by a  $\beta$ -glycosyl linkage to a nitrogenous heterocyclic base with a phosphoric acid ester that acts as the linkage unit between nucleotides via 5' and 3' phosphodiester bonds.

The five common bases fall into two chemical groups, the purines and the pyrimidines. The purine bases are adenine (A) and guanine (G) and the pyrimidine bases are cytosine (C), thymine (T) and uracil (U), which takes the place of thymine in RNA (ribonucleic acid) (figure 1.1).

In normal polymeric DNA adenine pairs with thymine and cytosine pairs with guanine to form the two complementary strands of the classical double helical DNA. The strands run in opposite directions and are always defined in the 5' - 3' direction. The base pairs are stabilised by the formation of hydrogen bonds between the two nucleotides. For a G•C base pair with normal Watson-Crick base-pairing three such hydrogen bonds are formed between atoms O6...H-N4, N1-H...N3 and N2-H...O2 and for an A•T base pair only two hydrogen bonds are formed: N1...H-N3 and N6-H...O4. (See figure 1.2). These base pairs are isomorphous in shape and have practically identical C1'...C1' distances. In classical B-DNA they lie essentially perpendicular to the helix long axis with regular inter base pair spacing. These features, together with the fact that the N-C1' glycosyl bonds between each base and its respective  $\beta$ -D-2 deoxyribofuranose sugar branch off with approximately the

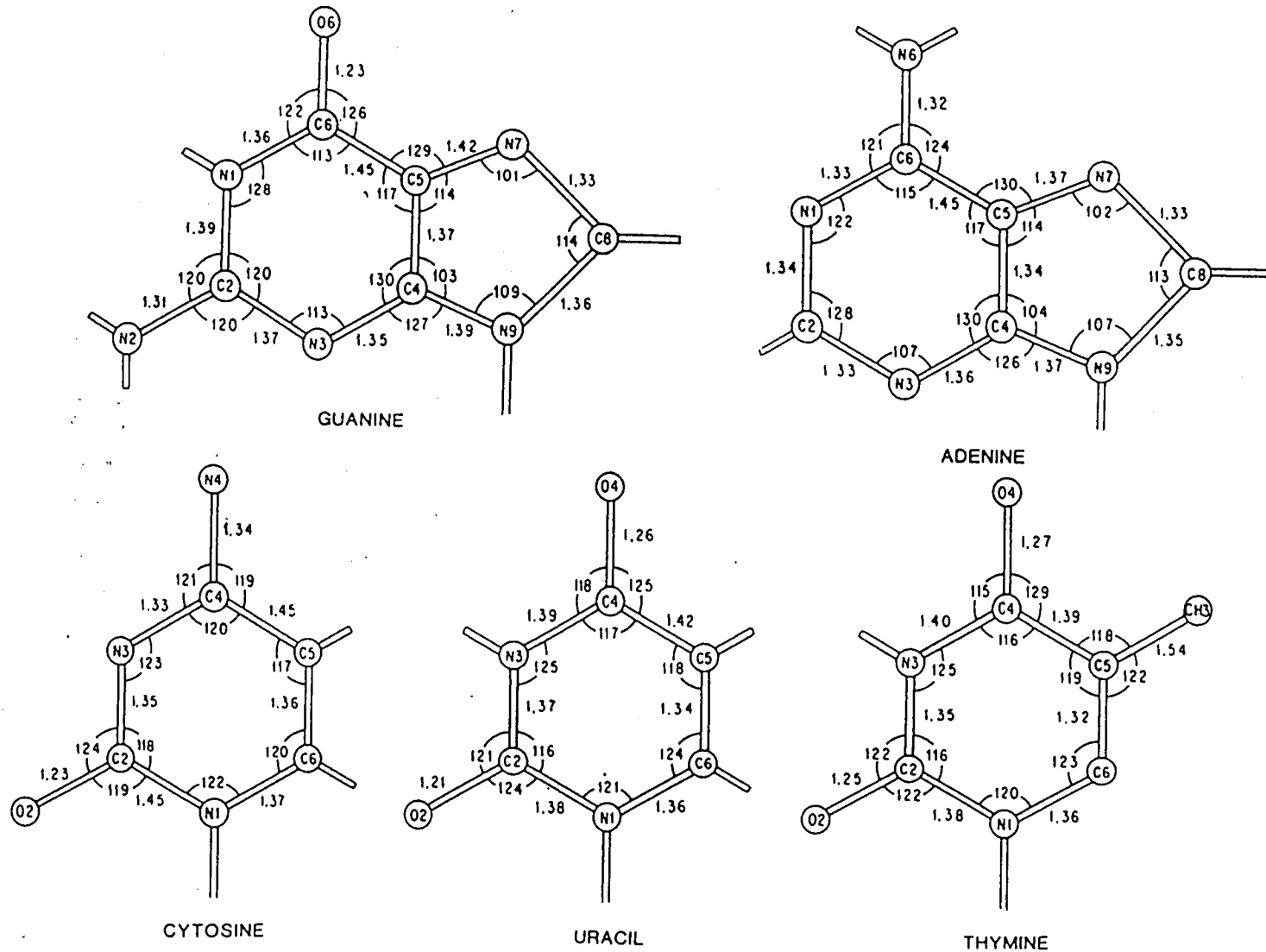


Fig. 1.1. The 5 nucleic acid bases.

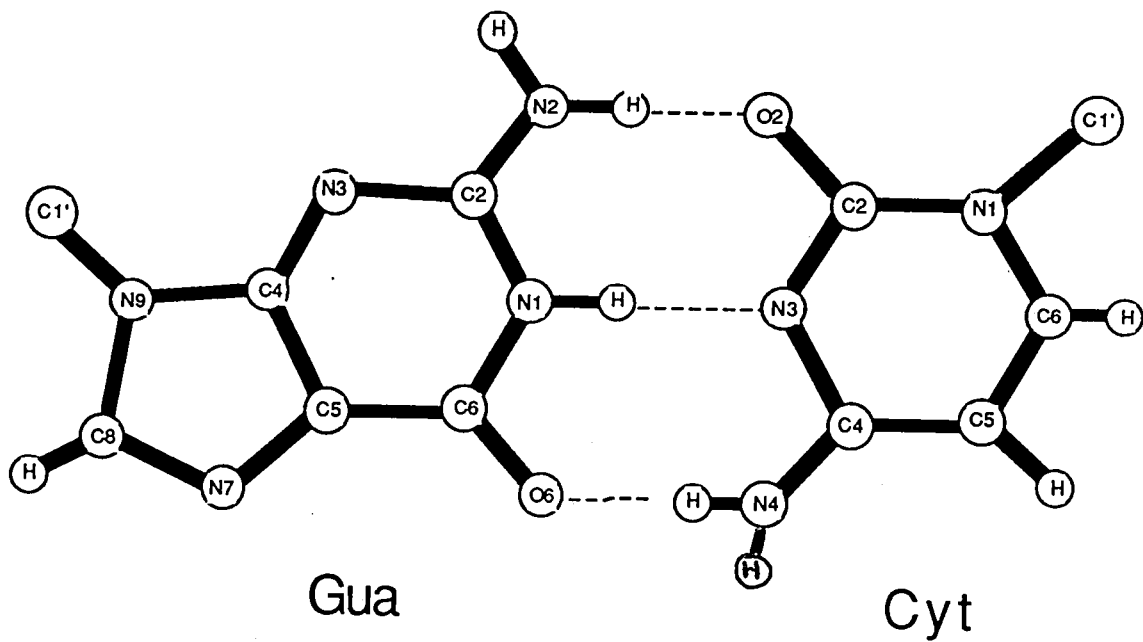
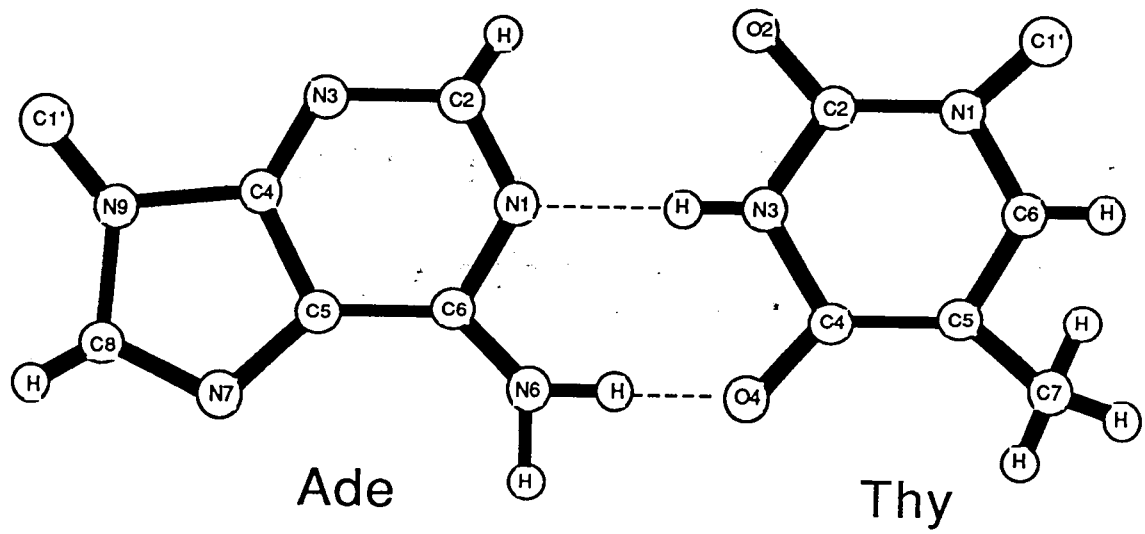


Fig. 1.2. The normal Watson-Crick base pairings.

same angle from the same side of the base pair, gives rise to two grooves of different dimensions which run along the helix. One groove is known as the major groove the other, the minor groove. In classical B-DNA the minor groove is at the O2 (pyrimidine) or N3 (purine) side of the base pair with the major groove, so called due to its greater width and depth, lying on the opposite side of the base pair.

### 1.2 Torsion angle nomenclature.

The nomenclature used to describe nucleic acids is extensive as although the individual bases can be thought of as essentially rigid there are a large number of backbone torsion angles that lend a great deal of conformational flexibility to DNA.

For the sugar and phosphodiester linkage the nomenclature is as follows:

Torsion angle	Atoms involved
$\alpha$	${}_{(n-1)}\text{O3}'\text{-P-O5}'\text{-C5}'$
$\beta$	$\text{P-O5}'\text{-C5}'\text{-C4}'$
$\gamma$	$\text{O5}'\text{-C5}'\text{-C4}'\text{-C3}'$
$\delta$	$\text{C5}'\text{-C4}'\text{-C3}'\text{-O3}'$
$\epsilon$	$\text{C4}'\text{-C3}'\text{-O3}'\text{-P}$
$\zeta$	$\text{C3}'\text{-O3}'\text{-P-O5}'_{(n+1)}$
$\nu_0$	$\text{C4}'\text{-O4}'\text{-C1}'\text{-C2}'$
$\nu_1$	$\text{O4}'\text{-C1}'\text{-C2}'\text{-C3}'$
$\nu_2$	$\text{C1}'\text{-C2}'\text{-C3}'\text{-C4}'$
$\nu_3$	$\text{C2}'\text{-C3}'\text{-C4}'\text{-O4}'$
$\nu_4$	$\text{C3}'\text{-C4}'\text{-O4}'\text{-C1}'$

Atoms designated (n-1) and (n+1) belong to adjacent nucleotide units.

The remaining torsion angle is the glycosidic torsion angle  $\chi$  which defines the rotation between the sugar and base of the nucleotide. The atoms involved vary depending on whether the base is a purine or pyrimidine.

Purines	O4'-C1'-N9-C4
Pyrimidines	O4'-C1'-N1-C2

Although there are a large number of torsion angles that can adopt any value from  $0\pm 180^\circ$  sterically based preferred values exist, to give favourable conformations. In crystallographic and spectroscopic publications within the DNA field the notation used historically is that of a combination of the two standard conventions. The usual terms used are *cis* ( $\sim 0^\circ$ ), *trans* ( $\sim 180^\circ$ ) and  $\pm$ *gauche* ( $\sim \pm 60^\circ$ ) and part of the, IUPAC-IUB recommended, Klyne-Prelog convention using *syn* ( $\sim 0^\circ$ ) and *anti* ( $\sim 180^\circ$ ), but not the suggested  $\pm$ *synclinal* ( $\sim \pm 60^\circ$ ) and  $\pm$ *anticlinal* ( $\sim 120^\circ$ ). Figure 1.3 shows each of the above torsion angles together with the atom numbering for the non-hydrogen atoms.

### 1.3 Sugar conformation.

The sugar conformation (the sugar pucker) in polynucleotides also adopts preferred modes due to the unsymmetrical substitution of the five-membered ring. Again there is standard nomenclature for describing the sugar pucker.

There are principally two different ways the five-membered ring can be puckered, one with four atoms in a plane with the fifth atom being out of plane by  $0.5\text{\AA}$  (known as the envelope (E) form) or secondly with three atoms being coplanar and the other two atoms in the ring lying on opposite sides of this plane (the twist (T) form). In the envelope form, if the out of plane atom is on the same side as C5' then it is termed *endo*; if on the other side of the plane

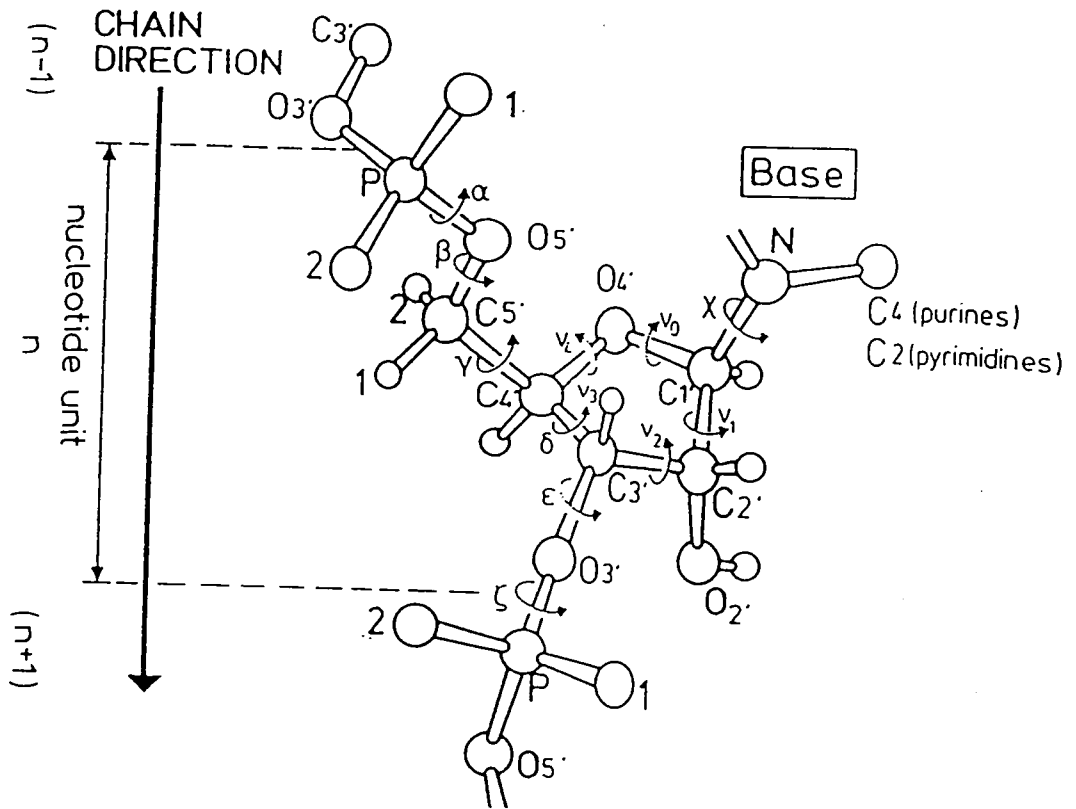


Fig. 1.3. One nucleotide unit showing atomic and torsion angle nomenclature.

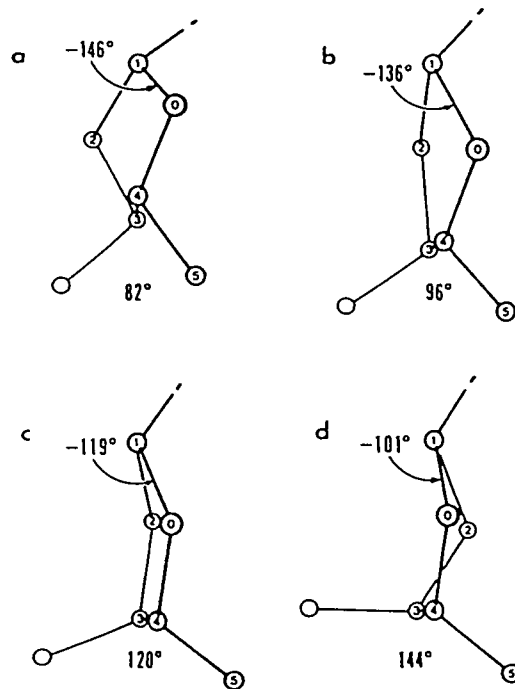


Figure. 1.4. The 4 most common sugar pucker.

a) C3'-endo b) O1'-endo c) C1'-exo d) C2'-endo

they are called *exo*. Figure 1.4 shows some of the more commonly observed sugar pucker conformations.

Obviously this notation is only approximate as intermediate modes can occur. A more satisfactory notation that lends itself more conveniently to mathematical interpretation is that of the pseudorotation cycle. This notation describes the continuum of possible sugar puckers by means of two parameters,  $P$ , the pseudorotation angle and  $v_{\max}$  the maximum torsion angle (degree of pucker).

Where

$$\tan P = \frac{(v_1 + v_4) - (v_0 + v_3)}{2 \cdot v_2 \cdot (\sin 36^\circ + \sin 72^\circ)}$$

and  $v_{\max} = v_2 / \cos P$

Figure 1.5 displays the pseudorotation cycle and also gives the corresponding envelope and twist nomenclature.

#### 1.4 The phosphodiester linkage.

Energy calculations involving rotations about the two torsion angles, describing the phosphodiester linkage between nucleotide subunits,  $\alpha$  ( $_{(n-1)}$  O3'-P-O5'-C5') and  $\zeta$  (C3'-O3'-P-O5' $_{(n+1)}$ ), result in six potential energy minima corresponding to six favourable conformations. Using the IUPAC-IUB nomenclature these are shown in figure 1.7.

#### 1.5 The glycosyl bond.

There are two main orientations that the base can adopt relative to the sugar moiety, by rotation about the glycosidic torsion angle  $\chi$ . The conformation is said to be *syn* if the six-membered ring of the purines or the O2 of pyrimidines points towards the sugar. The more commonly occurring

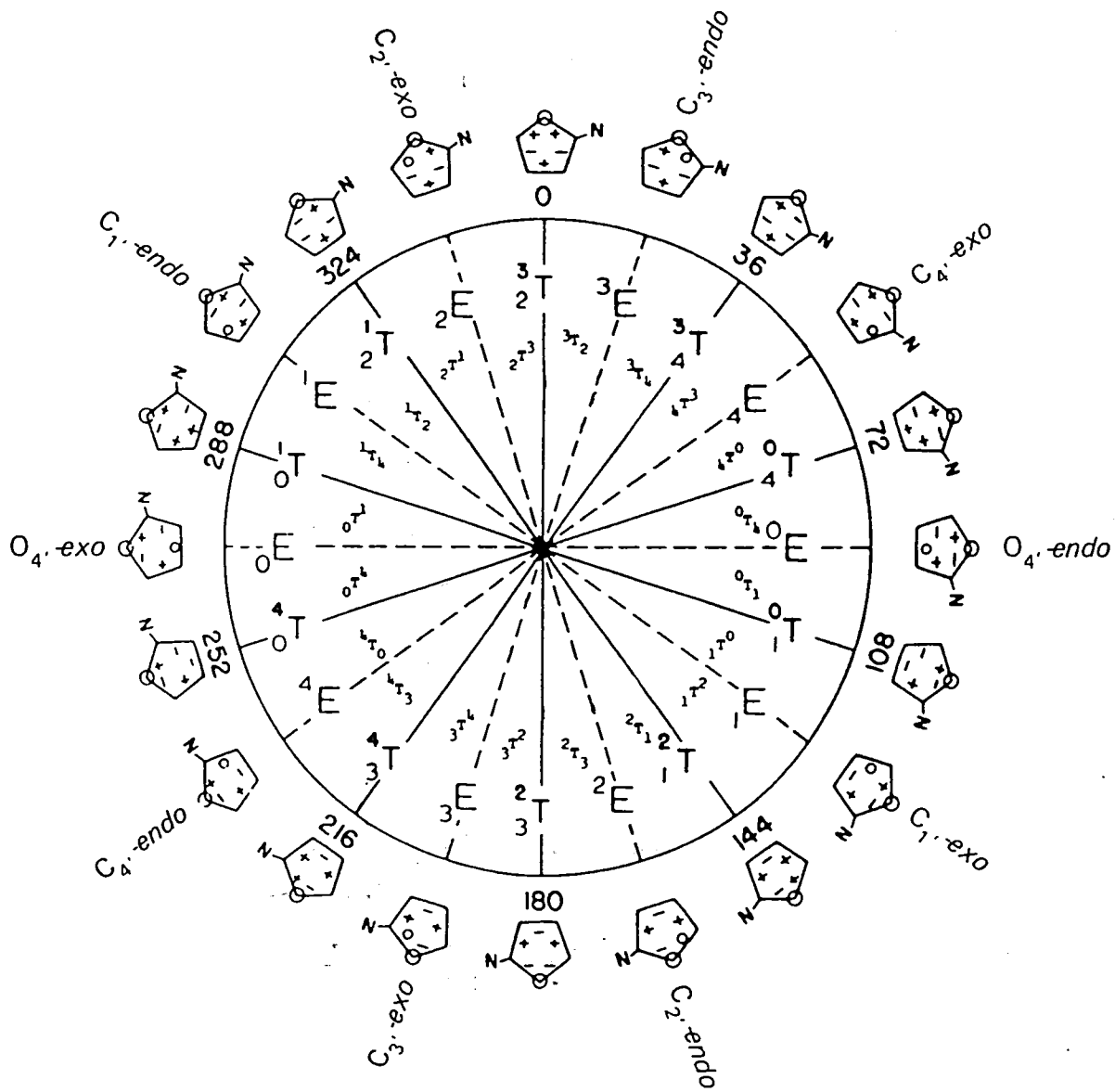


Figure. 1.5. The Pseudorotation cycle.

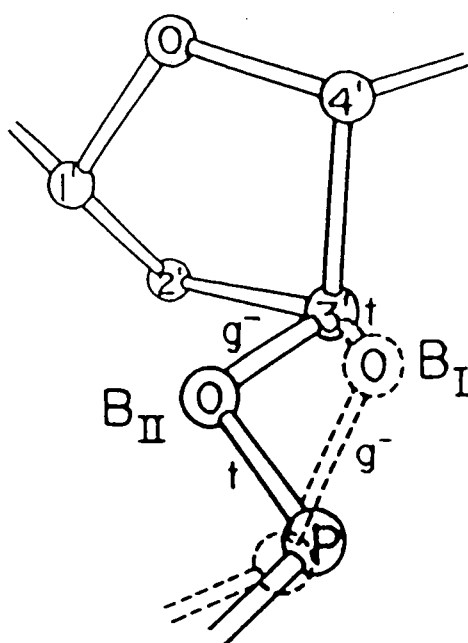


Figure. 1.6. B<sub>I</sub> and B<sub>II</sub> backbone conformation.

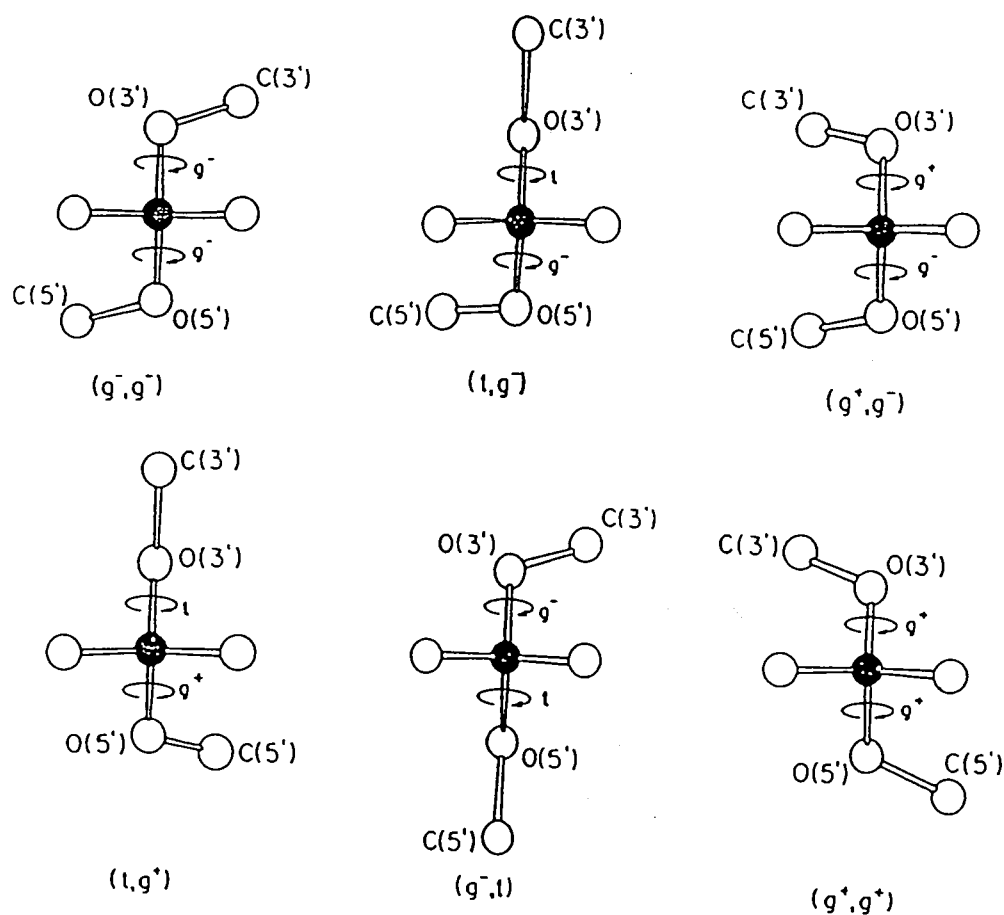


Figure. 1.7. The 6 most favourable conformations of  $\alpha$  and  $\zeta$ .

situation is with the six-membered ring of the purines or the O2 of pyrimidines pointing away from the sugar, in this case the conformation is *anti*. In terms of  $\chi$  this means anti ranges from  $-70$  through  $0$  to  $+90^\circ$  and *syn* ranges from  $+90$  through  $180$  to  $-90^\circ$ . The region  $-70$  to  $-90^\circ$  is termed *high anti*.

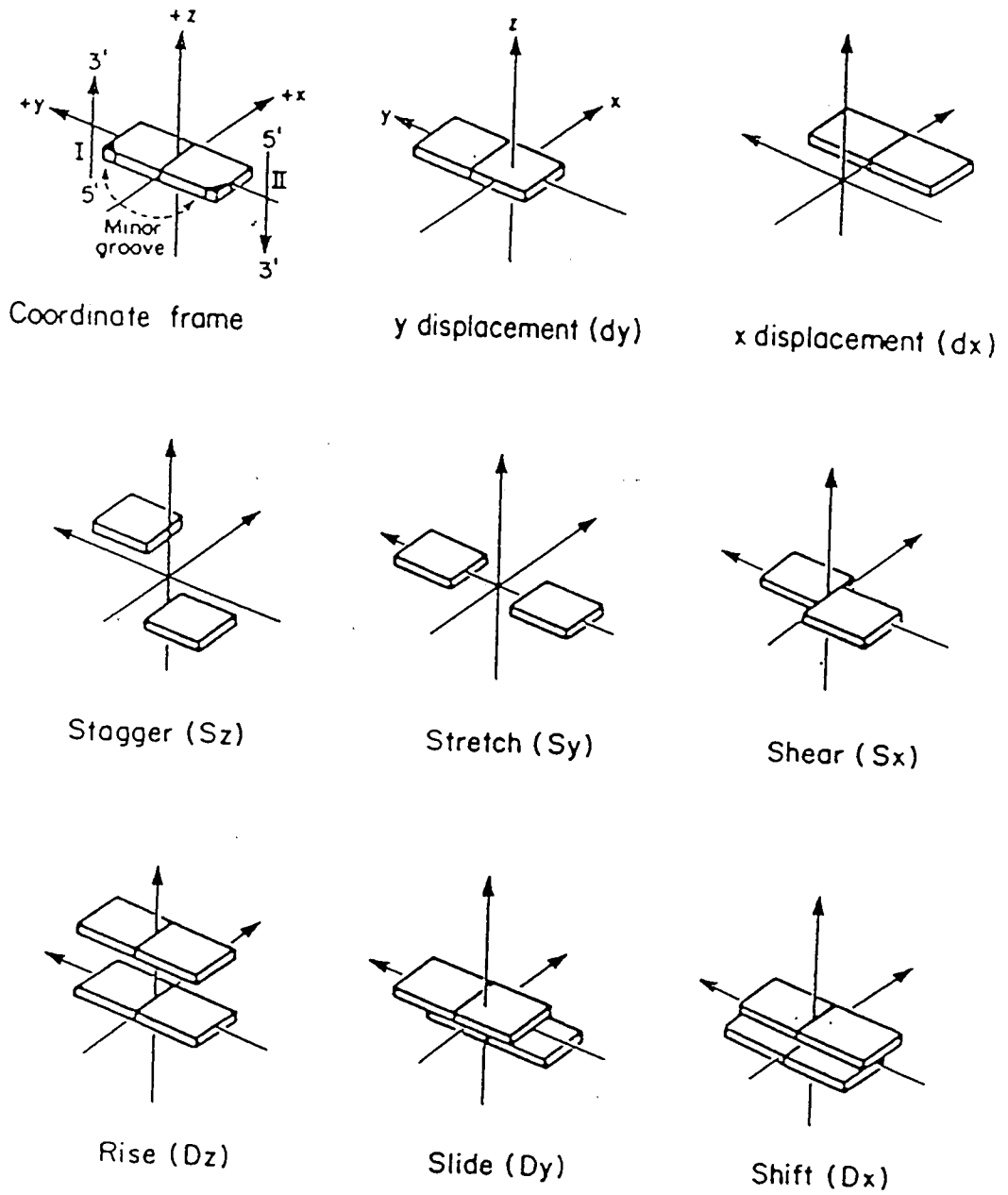
### 1.6 Backbone conformation.

A further terminology used in the description of polynucleotides, particularly DNA is that used to describe the overall appearance of the phosphate backbone. Phosphate backbone conformation has been grouped into two forms, that of the  $B_I$  and  $B_{II}$ . In B-DNA the predominant conformation is  $B_I$  which has the main chain torsion angles  $\epsilon$  and  $\zeta$  adopting *trans* and *gauche*<sup>-</sup> values respectively. In  $B_I$ ,  $\delta$  is free to assume a wide range of values and the general appearance of the C3'–O3'–P elbow is pointing outwards from the helix (slightly towards the minor groove), to lie on the surface of the helix. In the less common  $B_{II}$  conformation,  $\epsilon$  and  $\zeta$  adopting *gauche*<sup>-</sup> and *trans* swings the C3'–O3'–P elbow inward to point towards the helix axis (see figure 1.6). The torsion angle  $\delta$  is opened and tends to a limit of  $160^\circ$  set by the sugar ring geometry. The rotation about the C3'–C4' bond also enforces a rotation of the C2' atom resulting in a change in sugar pucker to C2'-*endo* or even drive the pucker as far as C3'-*exo*. The sugar pucker in the  $B_I$  conformation allows greater flexibility in the sugar pucker.

### 1.7 Translational and rotational parameters.

Whilst the torsion angles previously mentioned can be used to give a full geometrical description of a polynucleotide, it is almost impossible to form a mental image of the overall appearance of the structure being studied. From the point of view of understanding polynucleotide sequence recognition and specificity, it is useful to use a set of parameters that describe the general

## TRANSLATION



**Figure. 1.8. The 8 translational parameters as described at the EMBO workshop.**

## ROTATION

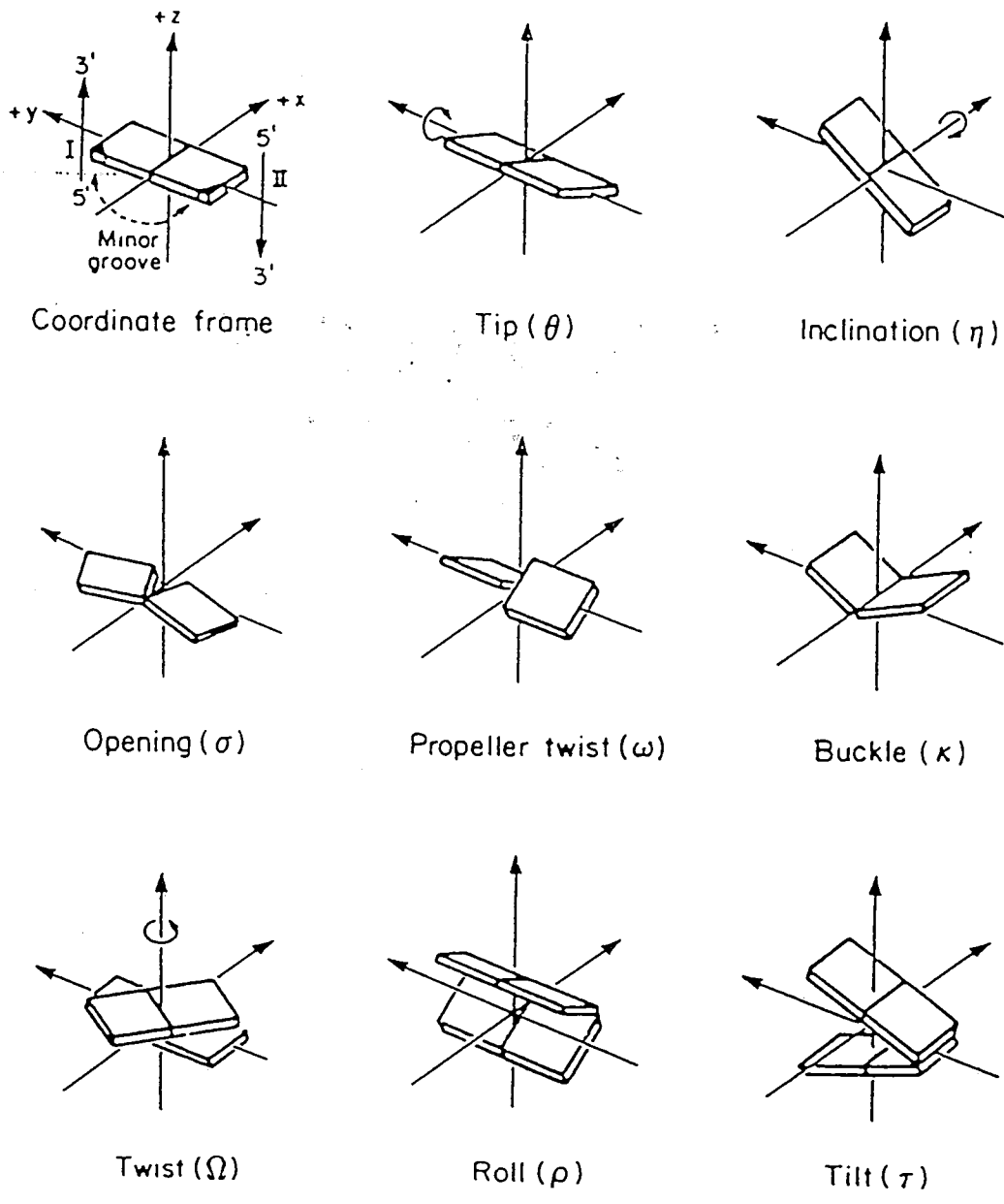


Figure. 1.9. The 8 rotational parameters as described at the EMBO workshop.

orientation of bases with respect to one another and to the helix axis as a whole. There are sixteen basic parameters which were agreed upon at an EMBO workshop in 1989. These are described diagrammatically in figures 1.8 and 1.9.

The most commonly used parameters in analysis of polynucleotides are those listed below.

Rotational Parameters	Translational Parameters		
Between two successive base pairs.			
Twist	$\Omega$	Rise	$D_z$
Roll	$\rho$	Slide	$D_y$
Tilt	$\tau$	Shift	$D_x$
Involving two bases			
Buckle	$\kappa$	y displacement	dy
Inclination	$\eta$	x displacement	dx
Tip	$\theta$		
Propeller Twist	$\omega$		

### 1.8 Groove widths.

Groove widths are usually defined as the shortest P–P vector across the groove less the sum of the van der Waals radii (5.8Å). Minor groove depth is measured as the distance between P and N2 of guanine or O2 of thymine and major groove depth as the distance between P and O6 of guanine or N6 of thymine, less the sum of their van der Waals radii in each case.

### 1.9 The polynucleotide families.

Polynucleotides are highly polymorphic and can display a wide range of structures that can be grouped into three families, A, B and Z-DNA (figure 1.10). Further there are structural variations within these families A and A' (RNA is restricted to the A family) forming the A family. B, C, C', C'', D, E, B'

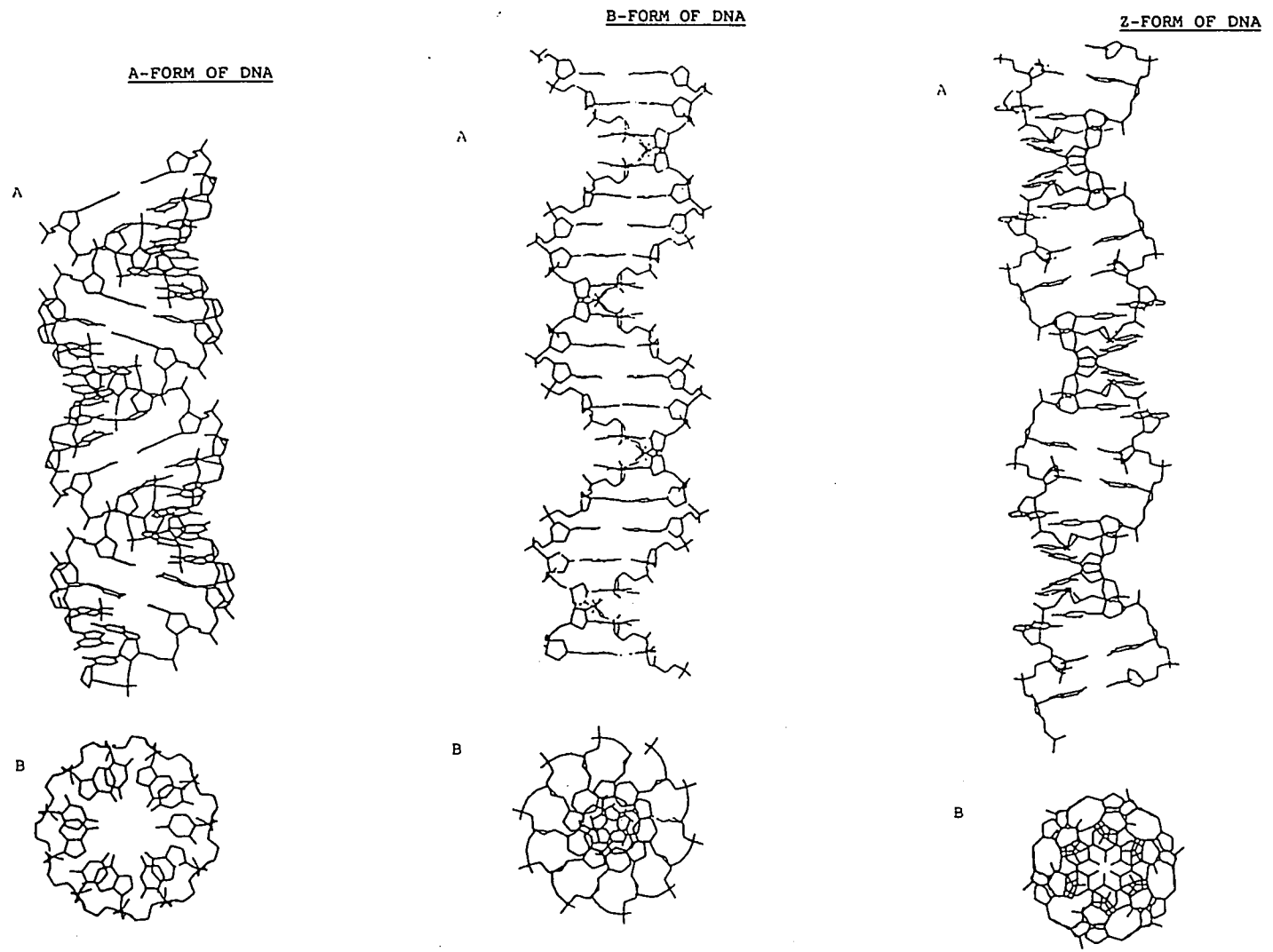
$\alpha$ -B'  $\beta$ -B' and T form the B family and Z, Z', Z<sub>F</sub>, Z<sub>T</sub>, Z<sub>II</sub> constitute the unusual left-handed Z family. For DNA with random sequence A, B, and C forms of DNA have been observed whilst specific sequence repeats are required for the formation of D, E and Z-DNA. Conversions within and between families are induced by change in relative humidity and salt content of the sample, or by change in ionic strength or solvent polarity. Although there are many classes of polynucleotide given above the main polymorphs of interest for DNA are A, B, C and D-DNA of which A and B are the major forms that will be discussed in this thesis.

The essential differences between A and B-DNA are concerned with the sugar pucker, base-pair tilt and displacement with respect to the helix axis and rise per residue. The sugar pucker for B-DNA is usually C2'-*endo* whilst in A-DNA the sugar pucker is C3'-*endo*. These puckers result in a distance between adjacent phosphates, of the same chain, of  $\sim 7.0\text{\AA}$  in B-DNA and  $5.8\text{\AA}$  in A-DNA. The tilt angle is also correlated with these different sugar puckers giving rise to a small negative tilt of base-pairs with respect to the global helix axis in B-DNA, but a larger positive tilt for A-DNA. The relative rotations of one base-pair to the next are also different for the two forms being approximately  $32.7^\circ$  in A-DNA and  $36^\circ$  in B-DNA. A striking difference between the two forms is that in B-DNA the base-pairs are displaced by less than  $0.5\text{\AA}$  from the helix axis towards the minor groove, whilst in A-DNA there is a large displacement of typically more than  $4.5\text{\AA}$  towards the major groove. The consequence of this large displacement in A-DNA is that the two chains of the DNA wind themselves around the helix axis. The relative displacements of the base pairs result in a great difference in the grooves in the two forms of DNA. In A-DNA the major groove is very deep but narrow whilst the minor groove is wide and shallow. In B-DNA the major groove is wide and the minor groove narrow with both grooves being of approximately the same depth. These features are highlighted in the table 1.1.

Table 1.1. General helical parameters for A, B & Z forms of DNA.

	A-DNA	B-DNA	Z-DNA
Helix sense	RH	RH	LH
Base-pairs/turn	11	10	12 (6 CG dimers)
Helix twist/base pair	32.7°	36.0°	CpG -15° GpC -45°
Rise/base-pair	2.9Å	3.4Å	3.7Å
Base-pair tilt	10°⇒20.2°	-5.9°⇒-16.4°	-7°
Glycosidic orientation	<i>anti</i>	<i>anti</i>	C <i>anti</i> G <i>syn</i>
Sugar pucker	C3'- <i>endo</i>	C2'- <i>endo</i>	G C3'- <i>endo</i> C C2'- <i>endo</i>
Base-pair displacement			
from helix axis	4.4°⇒4.9Å	0.2Å	
Minor groove width	11.0Å	5.7Å	
Major groove width	2.7Å	11.7Å	
Minor groove depth	2.8Å	7.5Å	
Major groove depth	13.5Å	8.5Å	

Figure 1.10. The conformations of A, B and Z-DNA.



A) A view perpendicular to helix axis, showing the high base tilt.  
 B) A view down the helix axis, showing the close proximity of the bases to the helix exterior.

A) A view perpendicular to the helix axis, showing the small base tilt.  
 B) A view down the helix axis, showing the central position of the bases. For clarity, only one strand is shown.

A) A view perpendicular to the helix axis, showing the zig-zag pattern of the phosphate backbone.  
 B) A view down the helix axis. Purine bases actually form part of the helix exterior.

## 1.10 DNA recognition.

The sequence specific binding of proteins to nucleic acids is involved in many fundamental processes such as gene control and expression, that occur within cells. These interactions govern packaging, replication, recombination, restriction and transcription. Consequently the determination and detailed analysis of the specific interactions, by use of X-ray crystallography and two dimensional N.M.R., is currently the focus of considerable interest. Considering helical antiparallel double-stranded B-DNA as the target for recognition there are particular features presented to a potential binding molecule.

As the sugar-phosphate backbone contains little selective information the edges of the base-pairs in the major and minor grooves act as the conduits for intrinsic information readout by proteins and ligands. The main recognition features of DNA are:

- (i) the positions of hydrogen-bond donors and acceptors which vary for A•T and G•C and their reversals T•A and C•G. It is apparent from figure 1.11 that although both major and minor grooves have the same number of hydrogen bond donor/acceptor groups, the major groove carries the potential for a greater variation of donor/acceptor patterns since the reversal of an A•T to T•A or G•C to C•G changes the pattern in the major groove whilst the minor groove pattern remains essentially unchanged.
- (ii) the steric considerations of the groups on the bases particularly the bulky methyl group of thymine and the N2 amine of guanine.
- (iii) the electrostatic potential of the DNA as a whole which varies considerably between the two grooves and has also been shown to show sequence determined variations. i.e. poly (dA•dT) is more negative in the minor groove than mixed or homopoly (dG•dC).
- (iv) the overall shape and flexibility of the DNA which may vary locally due to

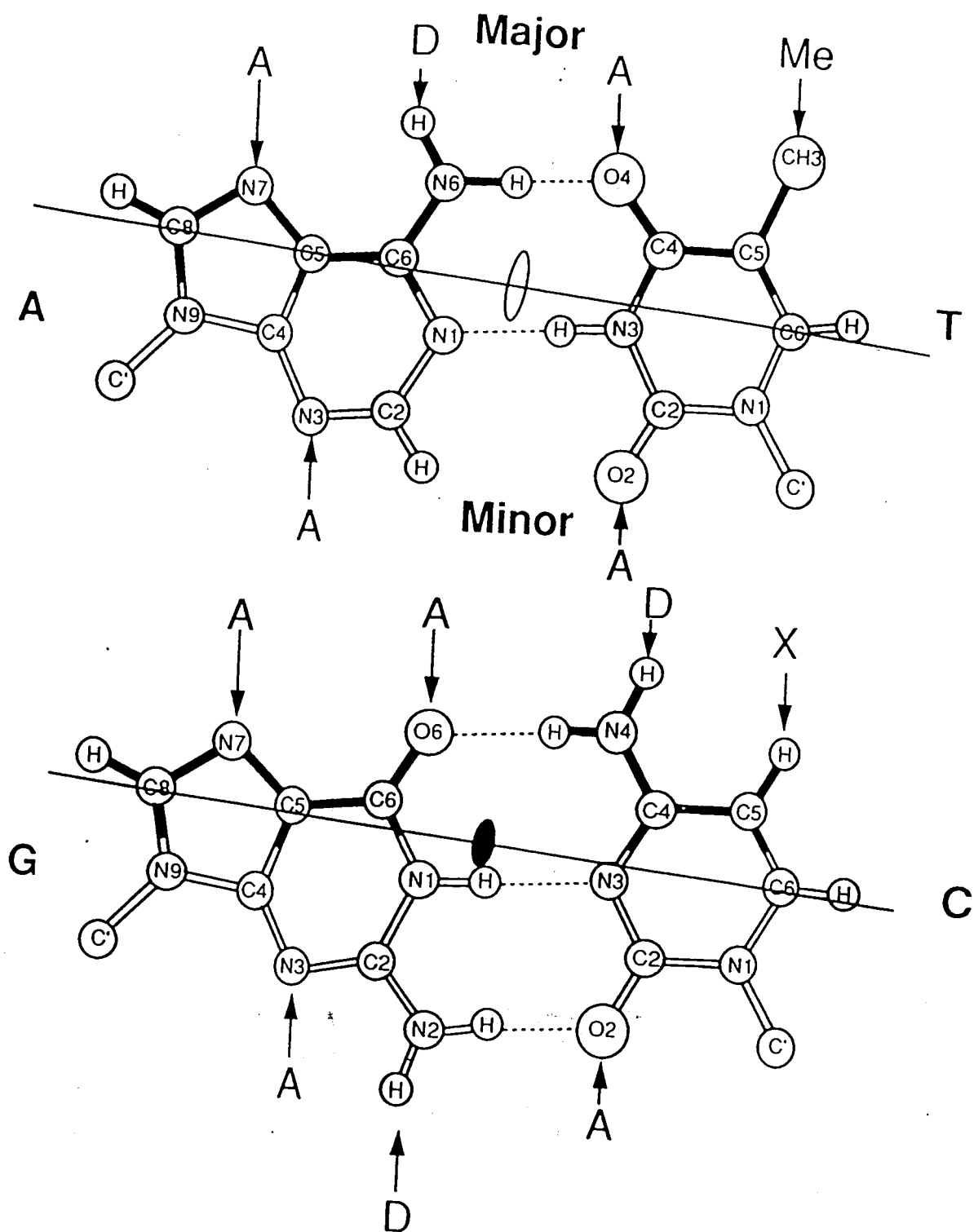


Fig. 1.11. Watson-Crick base-pairing showing hydrogen bond donors/acceptors.

specific sequence. G•C base pairs form three hydrogen bonds, as opposed to the two present in normal Watson-Crick base pairing for A•T base pairs, and is consequently more rigid. It has been postulated that long stretches of poly dA•dT can form three-centred hydrogen bonds that stabilises them making them more rigid. Local helical bends have been observed for phased AT tracts.

Thus the varying width and depth, surface contours, electrostatic potential, hydrogen-bond donors/acceptor patterns, in the major and minor grooves together with the overall charge, shape and helix flexibility gives DNA the ability to exhibit a multitude of potential binding sites for a ligand. As the human genome is comprised of some 2.5 billion bases, the four bases adenine, guanine, cytosine and thymine a ligand would have to recognise a 16 base-pair stretch of DNA to identify a unique site.

### 1.11 Protein-DNA recognition.

Although it is unnecessary to identify a unique binding site for proteins such as histones involved in packaging, or the DNA polymerases involved in replication and repair, it is essential that repressors and transcriptional activators are highly selective. These highly sequence specific proteins such as *cro* and CI repressors from  $\lambda$  phage, P434 repressor and *cro* repressor, and *metJ* and catabolite activator protein (CAP) of *Escherichia coli* recognise sequences of 10-20 base pairs (Anderson *et al.*, 1981; Brennan *et al.*, 1990; Pabo and Lewis, 1982; Jordan and Pabo, 1988; Clarke *et al.*, 1991; Mondragon *et al.*, 1989a; Anderson *et al.*, 1987; Aggarwal *et al.*, 1988; Mondragon *et al.* 1989b; Wolberger *et al.*, 1988; Rafferty *et al.*, 1989; Phillips, 1991; McKay and Steitz, 1981; Weber and Steitz, 1987, Schultz *et al.*, 1990 Steitz, 1990). Although these highly sequence specific proteins recognise segments of approximately the same length they do so by different means. DNA-binding proteins have now been characterised into groups based on the structural motif that senses the DNA. At present there are four main structural motifs that are known to

specifically recognise DNA sequences. These are the helix-turn-helix, zinc finger, leucine zipper and beta (beta-zip) proteins.

These highly sequence specific DNA-binding proteins use the major groove as the main recognition target. Sequence recognition is sensed within the major groove by the appropriate motif, which forms the majority of the hydrogen bond, electrostatic and van der Waals interactions. Additional favourable interactions are provided by non-specific hydrogen bonds to the phosphate backbone. In some cases such as  $\lambda$ CI repressor and possibly *trp* repressor long flexible terminal extensions, far from the motif region are also used to form specific contacts to the DNA. In many cases where the crystal structures of the complex have been determined, the DNA is considerably bent or unwound. This has led to the conclusion that the structural motif itself is insufficient to impart the sequence specificity required for DNA-binding and that distortion<sup>of</sup> the DNA or other regions of the protein providing additional specific contacts, must play an important role. It has been noted that those proteins which distort the DNA most appear not to use terminal extensions to make additional specific groove contacts (Freemont *et al.*, 1991).

### 1.12 Minor groove interactions with DNA-binding proteins.

Minor groove contacts play an important role in the recognition of specific DNA sequences in the case of the *engrailed* (Kissinger *et al.*, 1990) and *antennapedia* (NMR data, Otting *et al.*, 1990) homeodomain proteins of *Drosophila* ; and the MAT $\alpha$ 2 homeodomain from yeast (Wolberger *et al.*, 1991). Although helix 3 of the helix-turn-helix motif seems to be the 'recognition helix' in homeodomain proteins, a significant number of minor groove contacts are made via the flexible N-terminal residues. The contacts are from Arg 3 and 5 in *engrailed* Arg 5 in *antennapedia* (implied by an observed NOE to H1' of G12) to AT base pairs (figure 1.12). The fact that Arg 5 is a strongly

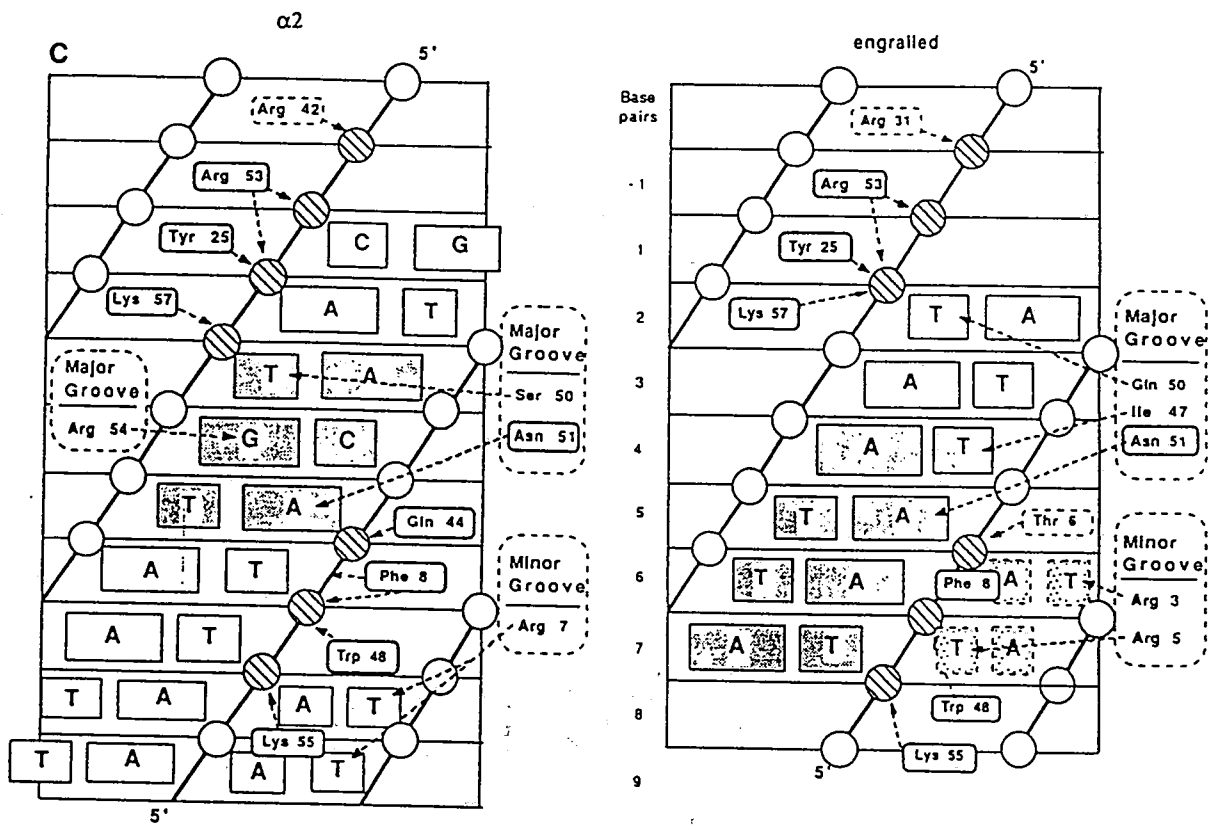


Figure. 1.12. Comparison of protein-DNA contacts present in the engrailed and  $\alpha 2$  homeodomain-DNA complexes.

conserved residue in the different homeodomains characterised so far (Scott *et al.*, 1989) suggests that these minor groove contacts are a significant additional recognition feature. In the case of MAT $\alpha$ 2 minor groove recognition is by Arg 7 to two AT base pairs (Arg 5 is not conserved in this protein). Phage 434 repressor also make use of an arginine, Arg 43 which makes contacts on either side of the minor groove which is thought to stabilise the minor groove compression found in the complex with DNA (Harrison and Aggarwal, 1990).

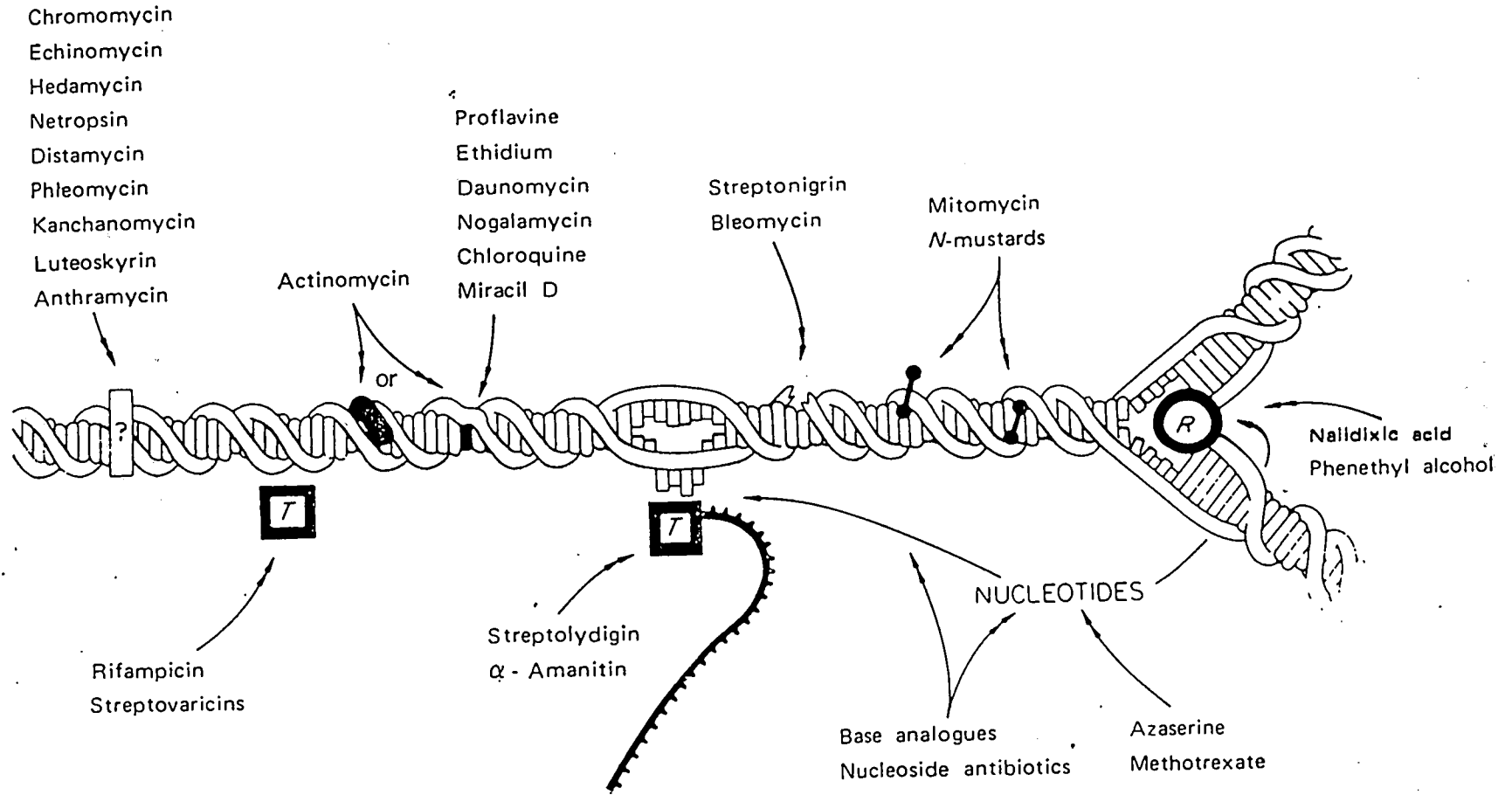
The structure of non-specific DNAase I endonuclease complexed with a nicked octanucleotide revealed contacts with the minor groove of the DNA with Arg 38 and Arg 9 forming hydrogen bonds in the minor groove and Tyr 73 either hydrogen bonding or involved in stacking interactions with adjacent pyrimidines (Suck *et al.*, 1988). However the bulk of the DNA contacts are with the phosphate backbone supporting the argument that local variation in minor groove width, determines binding affinity. This would also explain the relatively low cleavage rates for runs of Ade's or Thy's, which are known to have narrower minor groove width.

Churchill and Suzuki (1989) have provided evidence that the SPKK (Ser-Pro-Lys-Lys) motif, found in the tails of certain histones, binds preferentially to the narrow minor groove of DNA characteristic of oligo(dA).(dT) tracts. This recognition could be an important aspect of stabilising DNA bends that are induced on histone binding such as in the nucleosome core particle.

### **1.13 Drug-DNA recognition.**

Synthetic attempts to achieve DNA sequence specificity are somewhat hindered by the problem that, to be useful, the ligand must be able to get into the cell and be resilient enough to chemical breakdown or modification to reach its target intact (as in some cases such as the tumour forming benzopyrenes, it is the final metabolite, that covalently binds to DNA). There are four broad classes of non-protein ligand-DNA interaction; intercalators,

Figure. 1.13. Cartoon showing possible modes of ligand-DNA interaction.

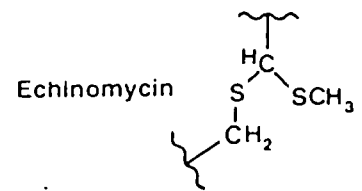
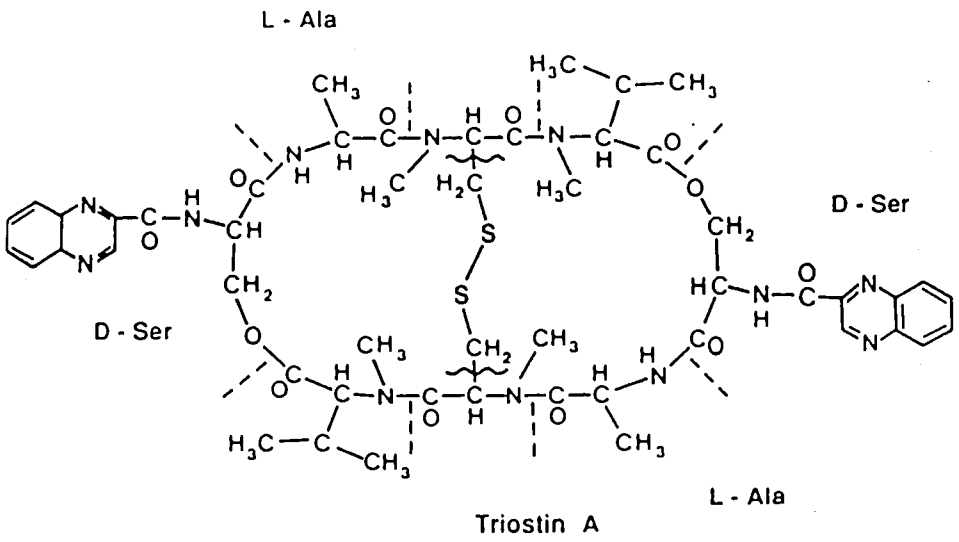
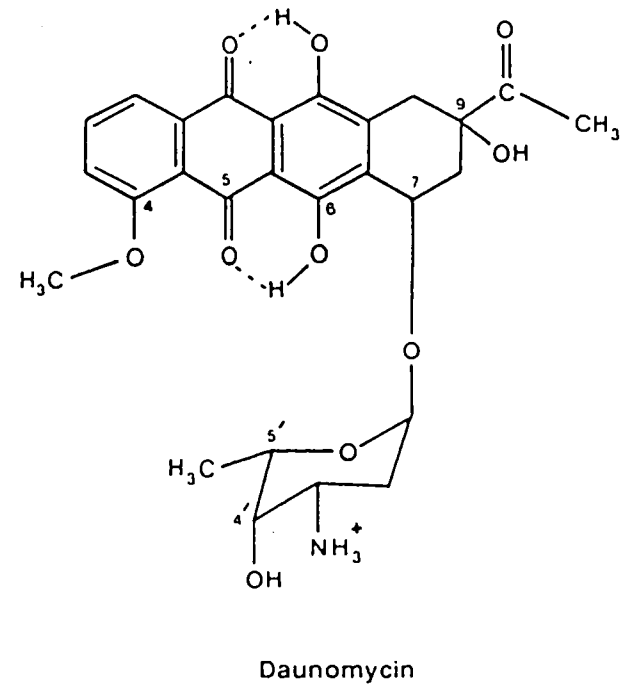


non-covalent minor groove binders, covalent groove binders and those that form triple helices (figure 1.13).

#### 1.14 Intercalators.

Intercalation is a mode of drug-DNA interaction first proposed by Lerman (1961) on the basis of binding studies and X-ray fibre diffraction. Intercalators are usually planar molecules consisting of three or four aromatic rings. The aromatic portion of the drug molecule is positioned between two adjacent base pairs, increasing their separation to 6.8 Å and resulting in an unwinding of the double helix. The complex is stabilised by aromatic stacking interactions between the bases and the chromophore of the intercalating molecule. The interaction may be further stabilised by side chains with functional moieties, that interact with the major and minor grooves of the double helical DNA. The effects of intercalating agents include inhibition of cell growth, cell death and induction of cellular transformation. These effects are particularly pronounced in rapidly proliferating cells and consequently they are often used as antibacterial, anti-parasitic and anti-tumour agents. Anthracyclines, which are widely used in cancer therapy, are a class of intercalators that includes the antibiotics daunomycin and adriamycin (figure 1.14). The detailed interactions between DNA and both these mono-functional intercalators, which consist of a planar chromophore with a sugar residue, have been determined by X-ray crystallography, (table 1.2) (Quigley, *et al.*, 1980; Wang, *et al.*, 1987; Moore *et al.*, 1989). The chromophore is inserted at right angles to the long axis of the adjacent base pairs with the remainder of the ligand occupying the minor groove of the distorted helix. Footprinting methods show that both ligands exhibit a higher affinity for a CpG step with adjacent A•T base pairs but adriamycin with its additional hydroxyl group displays enhanced binding.

Another group of intercalators are the bifunctional or bis-intercalators



Intercalators

Figure. 1.14. Chemical structure of three intercalators.

Table 1.2. DNA-drug complexes solved by crystallography.

Sequence	Ligand	Temp.	Space Group	Cell dimensions			Res.	Solvents	R-Fac	Ref.
d(CG TACG)	Triostin A	-16	P222	31.4	62.4	61.3	1.7	137	19	1
d(CG TACG)	Echinomycin	-16	P222	30.7	62.6	61.7	1.8	88	20	2
d(GCG TACGC)	Triostin A	12	P6 <sub>3</sub> 22	40.9	40.9	80.7	2.2	91	20	3
<sup>a</sup> d(m <sup>5</sup> GCTsAm <sup>5</sup> CG)	Nogalamycin	RT	P6 <sub>1</sub> 22	26.3	26.3	100.0	2.0	39	21	4
d(CG TACG)	Daunomycin	-15	P4 <sub>1</sub> 2 <sub>1</sub> 2	27.9	27.9	52.7	1.2	83	18	5
d(CG TACG)	Daunomycin	4	P4 <sub>1</sub> 2 <sub>1</sub> 2	28.0	28.0	52.9	1.5	40	25	6
d(CGCGAATTCGCG)	Hoechst 33258	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	25.0	40.3	65.9	2.2	175	14	7
d(CGCGAATTCGCG)	Hoechst 33258	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	25.2	40.6	66.1	2.2	126	16	8
<sup>b</sup> d(CGCGAATTCGCG)	Hoechst 33258	-100,-25,0	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	23.9	39.0	65.2	2.0	90	15	9
d(CGCGATATCGCG)	Hoechst 33258	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	25.6	40.6	67.1	2.3	112	14	10
d(CGCGAATTBrCGCG)	Netropsin	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	24.3	39.6	63.6	2.2	75	25	11
d(CGCGATATCGCG)	Netropsin	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	25.5	41.3	66.9	2.4	60	20	12
d(CGCGAATTCGCG)	DAPI	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	25.4	40.7	66.5	2.4	25	22	13
d(CGCAAATTTGCG)	Distamycin A	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	25.2	41.1	64.7	2.2	70	20	14
d(CGCGAATTCGCG)	cis-Platinum	RT	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	24.2	39.9	66.1	2.6	128	17	15

1. Ughetto et al., (1985); Wang et al., (1984)
2. Ughetto et al., (1985)
3. Wang et al., (1986); Quigley et al., (1986)
4. Quigley et al., (1980); Wang et al., (1987)
5. Williams et al., (1990)

6. Moore et al., (1989)
7. Pjura et al., (1987)
8. Teng et al., (1988)
9. Quintana et al., (1991)
10. Carrondo et al., (1989)

11. Kopka et al., (1985)
12. Coll et al., (1989)
13. Larsen et al., (1989)
14. Coll et al., (1987)
15. Wing et al., (1984)

<sup>a</sup>Cytosines methylated at 5 position and TpA step is a phosphorothiolate linkage.

<sup>b</sup>Crystal structure solved at three different temperatures. (Figures in °C averaged).

RT= Room temperature

consisting of two intercalating planar chromophores separated by a linker. The linker, of defined length, aligns the chromophores allowing them to intercalate at two sites with the linker accommodated in the minor groove of the DNA. Examples of bis-intercalators that have both been co-crystallised with DNA are the quinoxaline antibiotics echinomycin and triostin A (Ughetto, *et al.*, 1985; Wang, *et al.*, 1984, 1986). The binding of triostin A induces a change in the geometry of the base pairs adjacent to the quinoxaline ring such that they adopt the alternative Hoogsteen geometry (figure 1.15).

### 1.15 Mode of action of intercalators.

It has been shown repeatedly that intercalation is a necessary condition (though by itself, an insufficient property) for anti-tumour activity for all the classes of intercalating compounds, although they may have many other biological effects. DNA binding affinity was at first thought to be directly correlated with anti-tumour activity, hence the development of bis-intercalators such as diacridines with association constants comparable with those of regulatory proteins ( $>10^9 \text{ M}^{-1}$ ). Many exceptions to this general rule have been found and Feigon *et al.* (1984) concluded that it was the dissociation rate (the period for which the drug was bound) which correlated with biological activity for a broad variety of intercalating agents. The same was found to be true for a series of diacridines (Denny *et al.*, 1984). How then does this intercalation manifest itself in terms of cytotoxicity? There are three main ways in which DNA intercalating agents cause cell death, inhibition of DNA (and/or RNA) synthesis, DNA damage and through their interaction with DNA type II topoisomerase. Inhibition of nucleic acid synthesis is thought to be due to inhibition of the action of polymerase I. DNA damage is in the form of strand breakage by oxygen radicals, either intrinsic in the drug or due to modification of the intercalators by metabolic enzymes in the cell. Detection

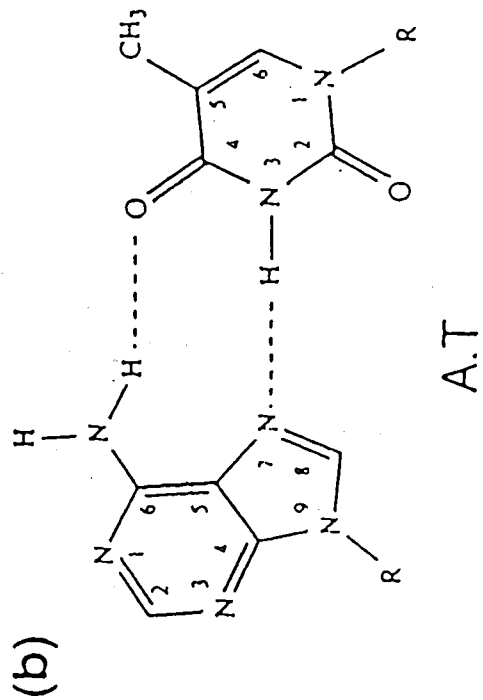
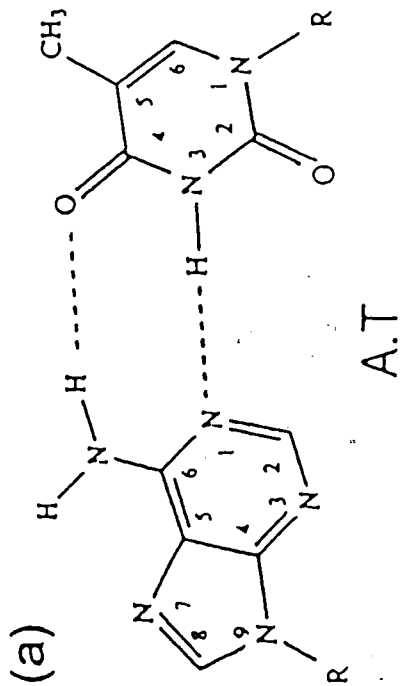
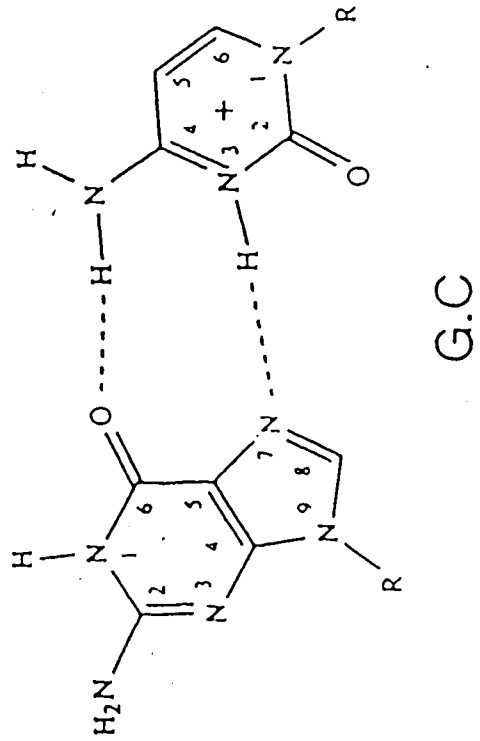
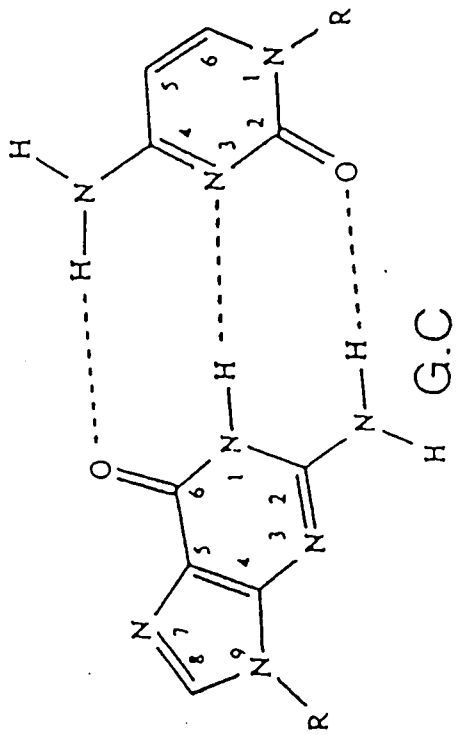


Figure. 1.15. Höogsteen base-pairing.

of this form of DNA strand cleavage is dependent upon high drug concentrations. At low drug concentrations in anaerobic condition (i.e. oxygen-independent) another mechanism of strand breakage has been detected. In this mechanism protein was always detected at the 5' end of the strand break and hence the inference of a mechanism involving a type II topoisomerase. Topoisomerase II is a homodimeric protein of molecular weight 168 kD, its role being to maintain and interconvert the topological isomers of DNA found in the cell (Liu, 1984). Topoisomerase II binds reversibly to DNA forming a 'non-cleavable complex', then it creates a double-stranded break, the cuts occur four base pairs apart with the 5' ends of each broken strand being linked to the protein via the phosphates to tyrosines (Nelson *et al.*, 1984). The other broken ends of the DNA are held 'in register' by non-covalent association with the protein to form a 'cleavable complex'. The enzyme then passes a second piece of double-stranded DNA through the break, which is subsequently resealed. The result of topoisomerase II activity then is to open the DNA and pass one double strand through another, if dealing with circular, or a closed loop of DNA a twist has been inserted. To prevent potentially lethal breakdown of the fragile 'cleavable' complex the equilibrium in the cell lies far from its formation. Intercalators such as amsacrine seem to bind preferentially to the 'cleavable' complex altering the equilibrium in this direction resulting in DNA strand breakage and cell death.

### 1.16 Minor groove binders.

A number of antitumour antibiotics such as netropsin, distamycin and Hoechst 33258 (figure 1.16) are known to interact via the minor groove of B-DNA. Footprinting experiments (Taylor, *et al.*, 1984; Dervan *et al.*, 1986; Portugal and Waring, 1987; Ward *et al.*, 1988), NMR (Patel, 1978; Patel and Shapiro, 1985; Klevitt, *et al.*, 1986; Leupin, *et al.*, 1986; Pelton and Wemmer, 1988, 1990; Parkinson, *et al.*, 1990; Fede, *et al.*, 1990) and

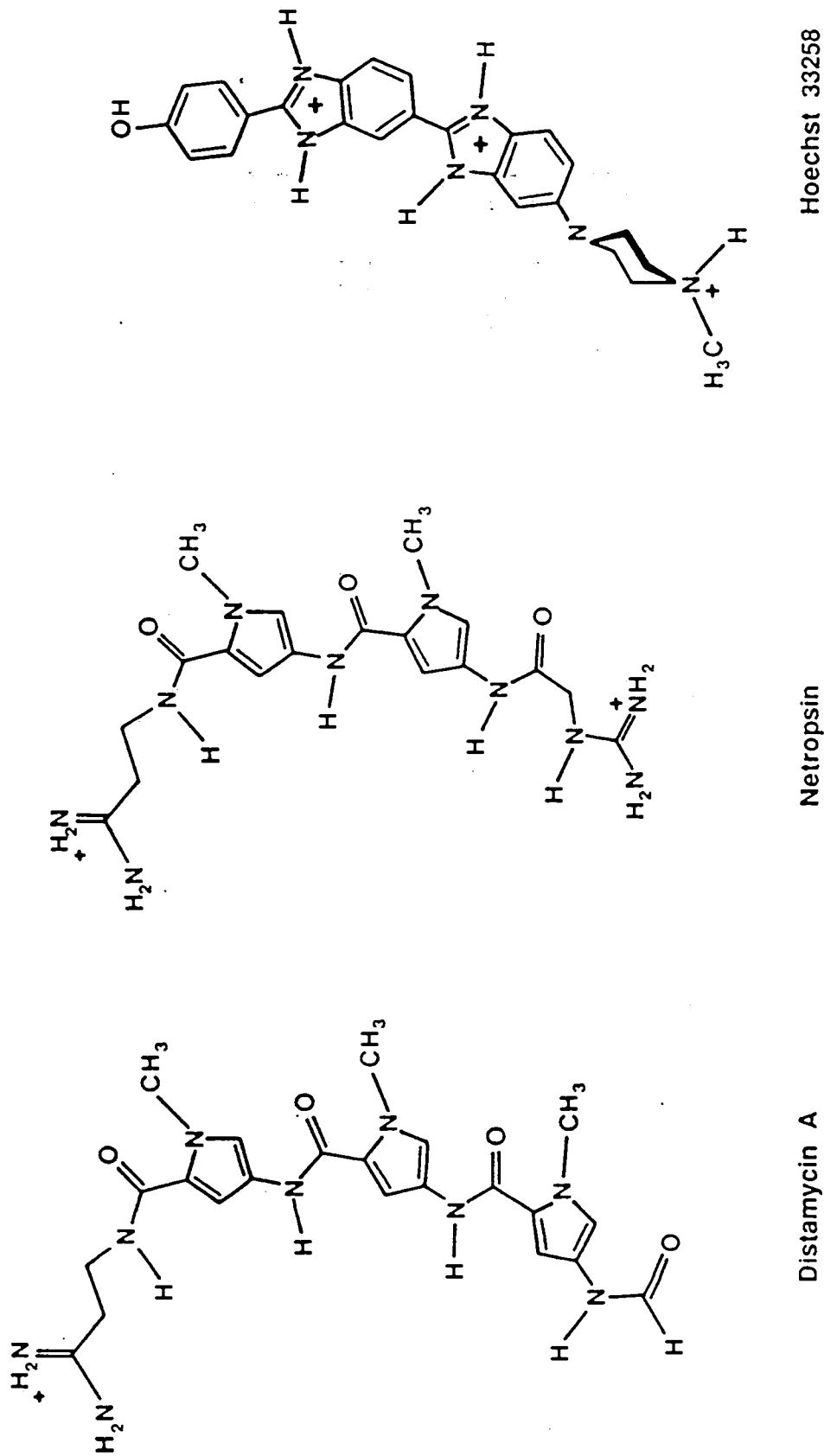


Figure. 1.16. Chemical structures of well studied minor groove binders.

crystallographic studies (Kopka, *et al.*, 1985; Coll, *et al.*, 1987, 1989; Pjura, *et al.* 1987; Teng, *et al.*, 1988; Carrondo, *et al.*, 1989; Larsen *et al.*, 1990; Quintana *et al.*, 1991) of these three compounds has shown that they have a preference for AT-rich sequences. Further evidence for AT preference has come from chemical footprinting and temperature melting experiments (Zimmer and Wahnert, 1986). These data show that there is a preference for polydA•polydT rather than polydA•dT stretches of DNA.

The AT-rich minor groove preference of these ligands can be explained by four arguments. First all three ligands are positively charged and will therefore be attracted to negatively charged DNA, also this positive charge may impart some degree of AT selectivity as it has been suggested that AT-rich minor groove regions of DNA exhibit the deepest negative charge potential (Lavery and Pullman, 1981; Zarkezewska, *et al.*, 1983). A third factor is that crystal structures of dodecamers have shown that AATT and AAATTT core sequences, of the DNA to which the ligands have been co-crystallised, have a narrow minor groove which can provide many van der Waals interactions with the ligand. It is interesting to note that a dodecamer containing the core ATAT has not been crystallised alone, and although it does have a narrow minor groove on drug binding it has been suggested by Wang and Teng, 1990; that this ATAT core will not tolerate high propeller twist which results in an intrinsically narrower minor groove. Finally each of the ligands is essentially planar and crescent-shaped so that it can form a close isohelical fit with the minor groove floor allowing hydrogen bonds to be formed between the ligand and the DNA. Presence of a guanine would result in an amino  $\text{NH}_2$  group in the DNA target site that would cause steric hindrance to ligand binding.

### 1.17 Mode of action of minor groove binders.

Like the intercalators there are many postulated mechanisms by which

minor groove binders are thought to exert their cytotoxic effects.

These include:- interactions with parasitic protozoa mitochondrial kinetoplast DNA and interference with the catalytic activity of type II topoisomerases, disruption of the spermine and spermidine biosynthetic pathway via inhibition of S-adenosyl-L-methionine decarboxylase in trypanosomes, induction of strand breakages at rare fragile sites in chromosomal DNA resulting in chromosomal rearrangements such as sister chromatid exchanges in neoplastic cells, interference with the catalytic activity of type I topoisomerases, rotation of DNA in nucleosome core particles, disturbance of binding of DNA to the nuclear scaffold and histone H1, polyploidisation of chromosomal DNA via impairment of the G2 phase of the cell cycle, and inhibition of the high mobility group protein (HMG-I)-DNA complex which is involved in either replication or recombination, where the HMG protein has an AT-rich recognition site.

The bis-benzamidine compounds berenil and pentamidine have therapeutic applications for the treatment of trypanosomiasis and Leishmaniasis (De Clercq and Dann, 1980; Sands *et al.*, 1985; Walzer *et al.*, 1988) In addition pentamidine has widespread therapeutic use with AIDS-related *P. carinii* pneumonia (Sands *et al.*, 1985; Wispelwey and Pearson, 1991). Trypanosomiasis and Leishmaniasis are caused by the related flagellated parasitic protozoa of the family *Leishmania* and *Trypanosoma*. These and related parasitic protozoa have unusual mitochondrial DNA known as kinetoplast DNA (kDNA) which is thought to be a target for the mode of action of berenil and pentamidine (Luck *et al.*, 1988).

kDNA consists of a network of thousands of interlocked (catenated) DNA circles of two types, maxicircles and minicircles, which have a very low degree of superhelicity (Bernard and Riou, 1980). There are some 20-50 maxicircles in the network which contain the genes essential for biogenesis and appear similar to the genes in the mitochondrial DNA in other eukaryotes. There are

approximately 1,000 minicircles within the network that contain one to four conserved regions depending on the species and a variable region (Ray, 1989). They range in size from 700-2500 bp depending upon the species, but within a network they are of uniform size (Marini *et al.*, 1984). Maxicircle transcripts have been identified as guide RNAs (gRNAs) (Blum *et al.*, 1990) and code for components of the mitochondrial respiratory system. gRNAs are small RNA molecules which play a crucial role in the process of RNA editing trypanosome mitochondria (Strum and Simpson, 1991). The process of RNA editing (for a review see Benne, 1989) alters mitochondrial transcripts by insertion of uridines not encoded in the genomic sequence and deletion of uridines encoded by the DNA. The editing process produces functional transcripts with homology to genes in numerous other organisms whose transcripts are not edited. The editing appears to regulate gene expression at the RNA level creating inhibition codons and extending open reading frames, creating termination codons, eliminating internal frameshifts and modifying nucleotide sequence within the 3' untranslated region and the poly(A) tail of the transcripts. Strum and Simpson have shown that all minicircles of *Leishmania tarentolae* encode single guide RNAs within their variable region.

kDNA is double-stranded B-form DNA, which contains phased adenine tracts with respect to the helix screw (Burkhoff & Tullius, 1987). Each tract induces a small local bend in the DNA, which combine in minicircle DNA to provide the 360° bend necessary for circularisation of the DNA (Griffith *et al.*, 1986). Hydroxyl radical footprinting of kDNA has shown that there is a progressive decrease in the width of the minor groove in the 5'→3' direction in the adenine tracts (Burkhoff and Tullius, 1987).

Berenil and pentamidine bind preferentially to AT-rich DNA (Braithwaite and Baguley, 1980; Newton 1975; Zimmer and Wahnert, 1986; Luck *et al.*, 1988) and has been shown by recent hydroxyl radical footprinting

studies on linear DNA (Laughton *et al.*, 1990; Fox *et al.*, 1990) to exhibit a preference for at least three contiguous AT base pairs. Berenil and pentamidine have been shown to bind to A-tracts, such as those present in kDNA, inducing mitochondrial type II topoisomerase double stranded breaks by stabilising the fragile topoisomerase II-DNA cleavable complex and linearising minicircle DNA (Chen *et al.*, 1984; Shapiro & Englund, 1990), resulting in dyskinetoplasmic cells (Riou & Bernard, 1980). As mentioned above, type II topoisomerases mediate a transient double-stranded DNA break to form a gate through which the same, or another, DNA molecule can be passed, facilitating interconversion between topological isomers of DNA. The same mode of action is suspected for other AT-selective minor groove binding ligands, such as Hoechst 33258 distamycin and DAPI (Woynaroski *et al.*, 1989). Thus topoisomerases can alter the degree of superhelicity which is known to be a factor in the replication and transcription of DNA (Fairfield *et al.*, 1985). Neither berenil or pentamidine have been shown to induce detectable cleavage of nuclear DNA (Shapiro & Englund, 1990).

Another mode of action which may contribute to the overall efficacy of berenil and pentamidine as anti-trypanosomal agents is by inhibition of S-adenosyl-L-methionine decarboxylase which is an obligatory enzyme in the biosynthetic pathways of spermine and spermidine. Berenil binds irreversibly whilst pentamidine inhibition is reversible. At higher concentrations berenil has also been shown to inhibit ornithine decarboxylase which is the first enzyme in the polyamine biosynthetic pathway (Bitonti *et al.*, 1986).

In the same way that berenil is thought to alter the equilibrium of topoisomerase II-DNA complexes to cause strand breakage, Fairfield *et al.*, (1990) have postulated that berenil has a similar effect on a type I topoisomerase in rat liver mitochondria, although the classification of the topoisomerase is somewhat unclear.

There is less evidence for the involvement of topoisomerases in the

effects of minor groove binding ligands on nuclear DNA. However, Portugal and Waring (1986; 1987) have shown, via footprinting, that berenil and netropsin have a marked effect on the nucleosome core particle.

The nucleosome core particle is a fundamental structural subunit of eukaryotic chromatin. It consists of four types of histones (H2A, H2B, H3 and H4) containing two units of each, tightly associated with a 146 base pair stretch of DNA. The structure was first solved to a resolution of 7.0 Å by X-ray crystallography (Richmond *et al.*, 1984), then to higher resolution 3.3 Å by Burlingame *et al.*, (1985) and recently to 3.1 Å (Arents *et al.*, 1991). The DNA was shown to be wound round the histone octamer in a superhelix of 1.8 turns. Berenil and netropsin were found, on binding, to rotate the DNA approximately half a helical turn with respect to the the histone octamer. This could lead to the inhibition or enhancement of the transcription of particular genes.

Another possible mode of of minor groove binding activity in the nucleus is binding to what have been termed "scaffold associated regions" (SAR's) of DNA (Gasser and Laemmli, 1986). These AT-rich regions of DNA are thought to mediate the specific binding to the non-histone like proteins of the nuclear scaffold, the major of which has been identified as topoisomerase II (Gasser *et al.*, 1986), to form the base of DNA loops on metaphase chromosomes. The SAR's are located near important control elements of several regulated genes (Gasser and Laemmli, 1987). They nucleate preferential cooperative binding of histone H1 (Izaurrealde *et al.*, 1989) which has been implicated in transcriptional repression (Wolffe, 1989). Käs *et al.* (1989) have shown that distamycin inhibits the association of these AT-rich SAR's with both histone H1 and the nuclear scaffold.

### 1.18 Covalent groove binders.

Electrophilic functional groups such as epoxides, aziridines,

carbinolamines, imines and cyclopropanes are found in a variety of synthetic and natural products capable of covalent interaction with DNA. The nitrogen mustards such as mechlorethamine represent one of the simplest classes of covalent groove binders that are highly efficacious antitumour agents, widely used in the clinic. They react as aziridinium cation intermediates with a preference for the nucleophilic N7 atoms of guanines in the major groove of B-DNA forming covalent bonds. The preference of nitrogen mustards for N7 of guanine imparts some sequence selectivity to the drug, which shows a preference for runs of guanines.

Another covalent groove binder that is a highly potent antitumour antibiotic is that of CC-1065 which binds covalently to N3 atoms of adenines in the minor groove of double helical B-DNA (figure 1.17). CC-1065 displays a high degree of sequence selectivity in terms of both the covalent adduct that it forms and the sequence in which that adduct occurs. CC-1065 is similar in shape to other minor groove binders with its crescent shape that can form a close isohelical fit with B-DNA spanning four base pairs. It is therefore not unreasonable that there will be some sort of sequence selectivity against the exocyclic  $\text{NH}_2$  groups of guanine protruding into the minor groove, and towards AT-rich regions that favour close van der Waals contacts between the narrow minor groove and the drug. Covalent binding of CC-1065 to either N2 of guanine or N3 of adenine then leads to a depurination (figure 1.17) or potentially strand breakage.

### 1.19 Triple helices.

Homopyrimidine sequences were reported as early as 1957 (Felsenfield *et al.*, 1957) to form stable triple helices to the homopurine strand of a Watson-Crick homopurine-homopyrimidine sequence. It binds via the major groove in a parallel orientation by means of two specific hydrogen bonds in Hoogsteen geometry (see figure 1.15). Sequence specificity of these triple helices is

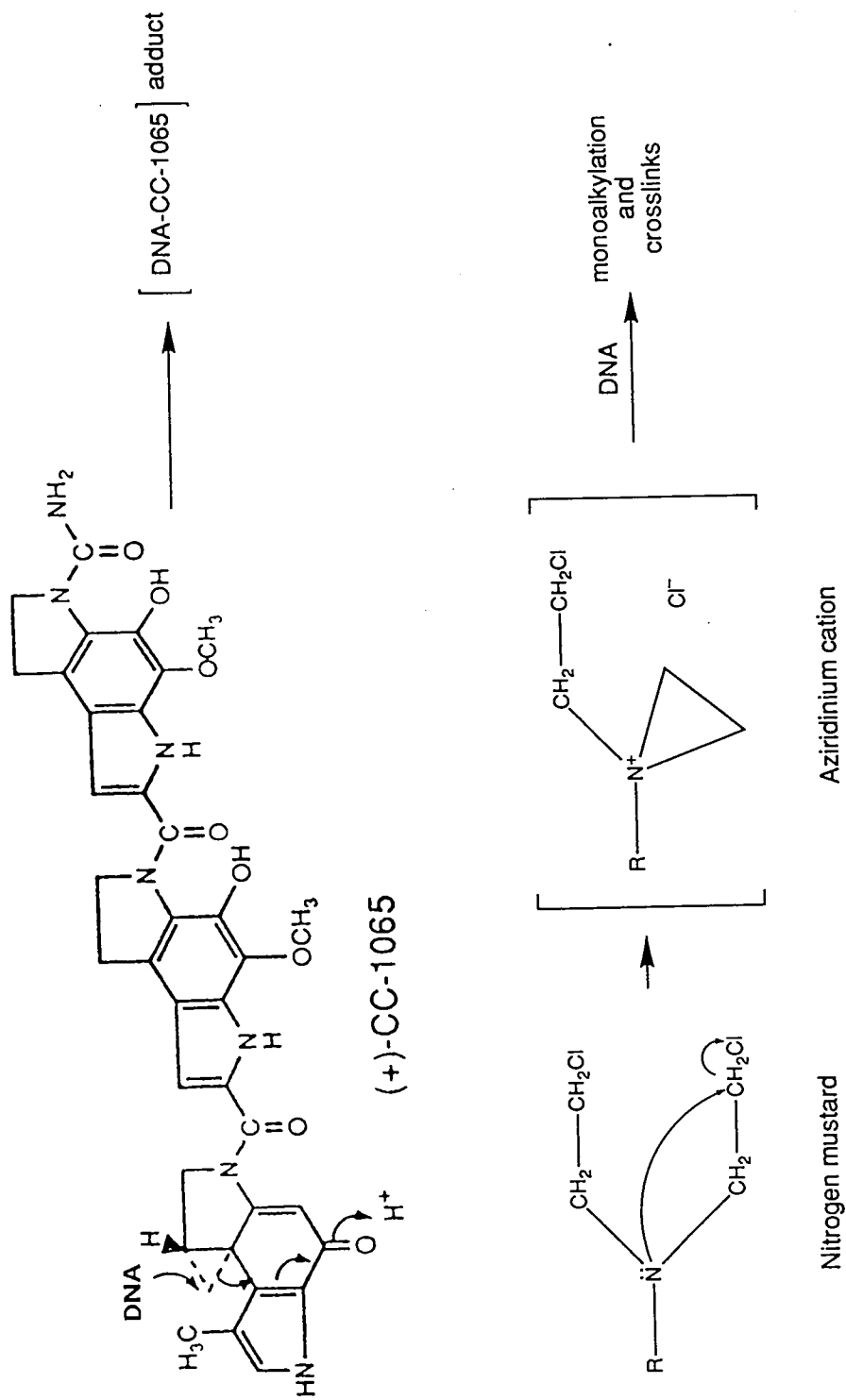


Figure 1.17. Reactive species of two covalent groove binders: CC-1065 and a Nitrogen mustard.

provided by thymine binding to an adenine-thymine base pair forming T•AT, and protonated (at N3) cytosine binding to a guanine-cytosine base pair forming C<sup>+</sup>•GC (Sklenar and Feigon, 1990). Further sequence-specific selectivity has been reported by Griffin and Dervan (1989) with the recognition of a T-A base pair by guanine forming a G•TA triplet. Interest in the area of triple helices has been rekindled due to their potential use as gene regulators (Maher *et al.*, 1989; Postel *et al.*, 1991). Also by attaching chemical moieties such as EDTA•Fe at the 5' and/or 3' ends their sequence-specific cleavage of double stranded DNA (Moser and Dervan, 1987; Strobel *et al.*, 1988) makes them powerful tools for chromosome mapping, gene isolation and DNA sequencing (Strobel and Dervan, 1990).

## **CHAPTER 2.**

The crystal structure of a complex between the dodecamer  $d(\text{CGCGAATTCGCG})_2$  and the minor groove-binding ligand berenil.

## 2.1 Introduction

The 1,3-*bis*(4-phenylamidinium)triazene compound berenil (figure 2.1) has veterinary application as an anti-trypanosomal agent and has cytotoxic and anti-viral properties (De Clercq and Dann, 1980). It binds reversibly to double-helical DNA primarily at AT-rich regions (Newton, 1975; Braithwaite and Baguley, 1980; Baguley, 1982), as shown most recently by a footprinting study with DNA fragments (Portugal and Waring, 1987). Berenil interacts with the kinetoplast DNA of *Trypanosoma cruzi* cells (Bernard and Riou, 1980) which has been found to have bending at its phased A tracts (Marini *et al.*, 1982; Burkhoff and Tullius, 1987). The footprinting revealed that the strong-binding sequences are at least three base pairs long, with a possible preference for alternating AT sites. The drug also binds to DNA on reconstituted nucleosome core particles, in common with netropsin and other minor groove binders. It changes the rotational orientation of the DNA by exposing otherwise hidden AT sequences (Portugal and Waring, 1986). Molecular modelling studies have suggested that berenil interacts via the minor groove of B-DNA (Gresh and Pullman, 1984; Pearl *et al.*, 1987; Gago *et al.*, 1988), with the drug appearing to span two base pairs and hydrogen-bonding preferentially to adjacent thymine bases via their O2 atoms. Berenil was co-crystallised with the well-studied self-complementary duplex sequence d(CGCGAATTCGCG) (Drew and Dickerson, 1981).

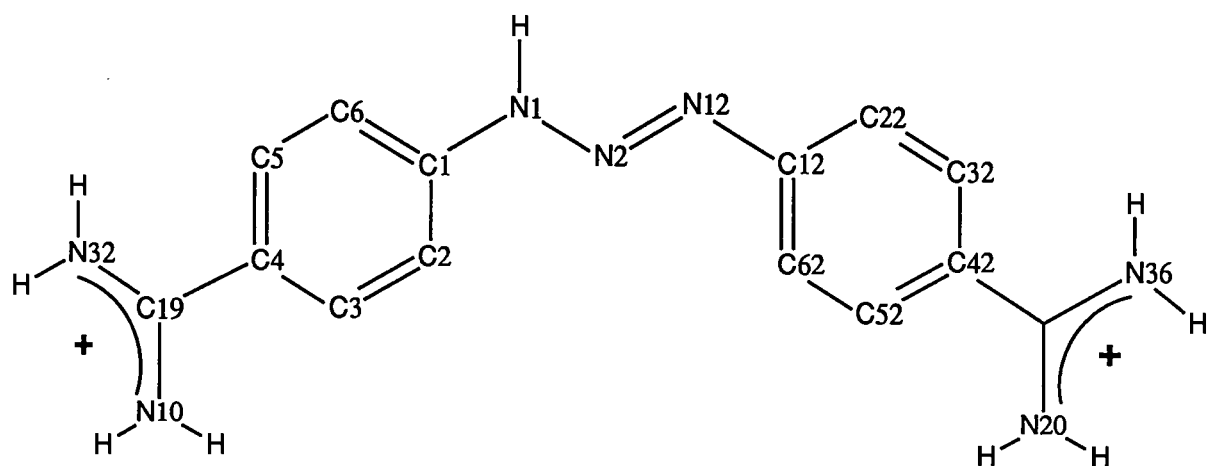


Figure 2.1. The chemical structure of berenil.

## Materials and Methods.

### 2.2 Crystallisation and data collection.

The dodecanucleotide sequence was synthesised on an Applied Biosystems synthesiser using phosphoramidite chemistry and purified by anion exchange and reverse-phase HPLC courtesy of Dr. Tom Brown (University of Edinburgh). Berenil was synthesised at the Institute of Cancer Research by Dr. Terry Jenkins. Crystals of the complex were obtained by the vapour diffusion method at 278K from a solution containing 3 mM DNA dodecamer, 10 mM sodium cacodylate (pH 7.0), 30 mM  $MgCl_2$ , 2 mM berenil and 20% v/v 2-methylpentan-2,4-diol (MPD) against a 50% MPD reservoir. The

crystals of the complex grow in approximately 21–28 days, are yellow in colour with typical dimensions of 0.1 mm x 0.1 mm x 0.5 mm

and form highly elongated needles. They have unit cell dimensions of  $a = 24.51$ ,  $b = 39.98$ ,  $c = 66.23$  Å and are in the same space group ( $P2_12_12_1$ ) as the native dodecamer which has dimensions of  $a = 24.87$ ,  $b = 40.39$ ,  $c = 66.20$  Å; the two are therefore isomorphous. X-ray crystallographic data were collected on a Xentronics area detector mounted on a Rigaku RU200 rotating anode X-ray source, with the aid of Dr. Elspeth Garman at the Laboratory of Molecular Biophysics at Oxford. The data collected extend to a resolution of 2.35 Å. A

Using two 100 degree omega scans, rotating phi 60 degrees between scans and 200 s frames, 0.2 degree oscillation and power of 40 kV 60 mA, a total of 5762 reflections were collected, which merged to 2195 unique reflections with an overall merging agreement of 3.7% for the useful data to 2.5Å resolution, of which 1759 reflections had an intensity above the 2σ level.

Table 2.1. Resolution breakdown of data collected.

Resolution limit Å	Number of Reflections		R-Factor <sup>a</sup>
	Possible	Collected	
4.28	535	398	3.67
3.40	498	396	4.11
2.97	493	405	6.25
2.70	480	404	13.94
2.50	477	410	17.37
2.35	474	182	21.55
	—	—	—
	2957	2195	5.89

<sup>a</sup>. Unweighted absolute R-factor on F\*100

### 2.3 Structure solution and refinement.

The structure was solved by molecular replacement, using the native d(CGCGAATTCGCG) duplex coordinates as a starting model. The coordinates were extracted from the Brookhaven Data Bank (Bernstein *et al.*, 1977) and reformatted to carry out initial refinement with the constrained/restrained least squares refinement program CORELS (Sussman *et al.*, 1977). Restraints restrict the features of the model to a realistic range of possibilities, thereby increasing the number of observations by adding information such as bond lengths, bond angles and torsion angles. Constraints confine such features of the model to specific values thus reducing

the number of variables. By use of both restraints and constraints CORELS allows one to perform rigid body refinement, thereby minimising the number of variable parameters in the refinement to eight (x,y,z, three angles, temperature and scale factor). Initially the data used for rigid body refinement was from 8-3 Å to obtain a large radius of conversion to ensure correct orientation of the molecule within the unit cell. At this early stage  $|2F_o - F_c| \alpha_{calc}$  and difference maps,  $|F_o - F_c| \alpha_{calc}$  were calculated using <sup>the</sup> ~~PROTEIN~~ package (Steigemann, 1974) and viewed on a Silicon Graphics Iris 3130 using the graphics package TOM (Cambillau, 1988). The map implied that the structure was essentially correct, by virtue of the fact that all the bases and phosphate groups were positioned in density of greater than  $2.5\sigma$  in the  $|2F_o - F_c| \alpha_{calc}$ . The sugar carbons were less well defined. Ten cycles of rigid body refinement, extending the data used to 2.5 Å, yielded an R factor of 36.2%, at which stage the model was divided into rigid groups consisting of base pairs, sugars and phosphate groups. Further cycles of both positional and temperature factor refinement resulted in an R factor of 30.8%. The bases were then treated individually with normal Watson-Crick hydrogen bonding patterns, which reduced the R factor to 28.5% with further rounds of refinement. Refinement was then switched to the restrained-refinement program NUCLSQ (Westhof *et al*, 1985). NUCLSQ treats the atomic positions and their associated temperature factors individually, not as groups. A consequence of this is greater variation in sugar pucker (the sugars do not have to be targeted to the standard C2' *endo* associated with regular B-DNA) which is known to greatly influence the overall conformation of an oligonucleotide. NUCLSQ uses slightly different weighting schemes which allow individual weights on the various restrained factors to be altered more easily than is the case with CORELS. A further difference between refinement using CORELS and NUCLSQ is that CORELS used all data input in the 8-2.5 Å

range, whereas a sigma ( $\sigma$ ) cutoff is inherent in the NUCLSQ program. For this refinement a  $2\sigma$  cutoff was used. Successive rounds of refinement reduced the number of bad distances and torsion angle contacts and gave an R factor of 21.6%, using data from 8-2.5 Å. At this point further  $|2F_o - F_c|/\alpha_{calc}$  difference and 'omit' maps were calculated. These showed numerous putative water peaks and an extended continuous volume of electron density in the minor groove of the duplex. Using the "peakpk" option of PROTEIN to output density peaks of  $\geq 2\sigma$  in difference maps, the search for associated solvent molecules was initiated. Dummy atoms were generated and viewed on the graphics, to assess their possible hydrogen bond geometries and contacts with the oligonucleotide structure. Care was taken not to include any solvent molecules in the continuous density of the minor groove in an attempt to further improve the density in this region, to allow fitting of the berenil molecule. The inclusion of 31 waters and subsequent refinement of both positions and temperature factors yield an R factor of 18.9% and a slight improvement in the minor groove density. At this stage this density was unambiguously fitted to a berenil molecule in terms of size and overall shape. The addition of the ligand reduced the R factor to 17.1%. During the refinement the 5' end of the drug moved out of the groove slightly to leave a discrete, isolated, intense peak of density (figure 2.2). Efforts to rerefine with slightly different placing of the drug continued to yield the same result. The peak was then assigned as a water molecule bridging the drug and the DNA. This bridging water molecule subsequently refined satisfactorily as the overall R factor was lowered and its isotropic temperature factor (B) fell. (New putative solvent molecules were arbitrarily assigned temperature factors of  $32 \text{ \AA}^2$ ). Continued refinement together with the addition of 26 more waters reduced the R factor to 14.7% Detailed analysis of the berenil-DNA

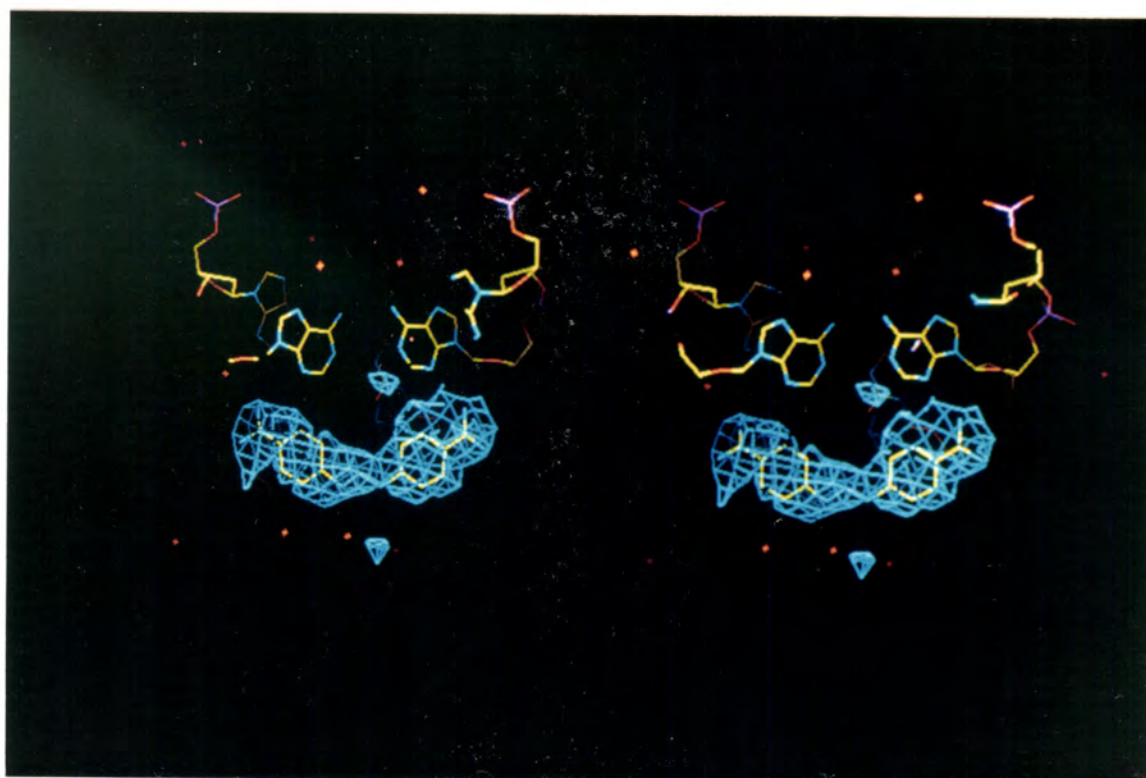


Figure 2.2. Stereo view of final omit difference fourier map.

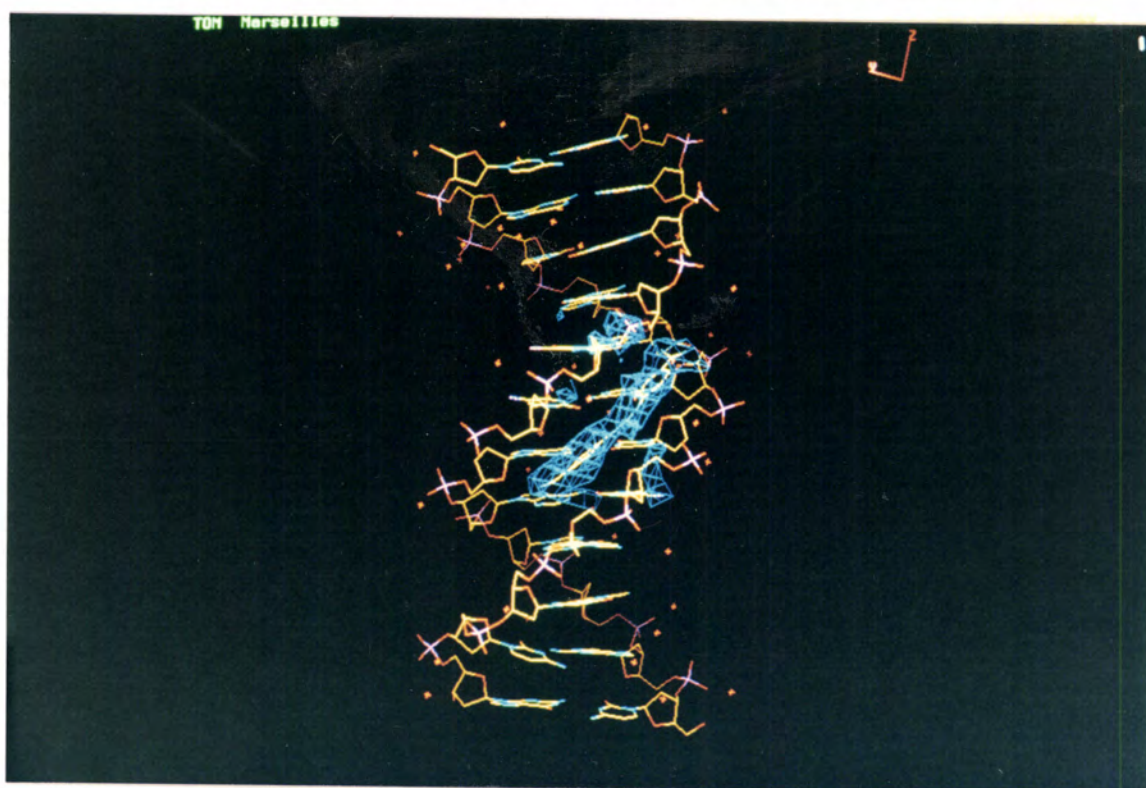


Figure 2.3 View of structure showing berenil in the minor groove

interactions, which included the generation of ideal hydrogen atoms, using GEMINI an in-house graphics package (Beveridge, 1990), highlighted two clashes at the 3' end of the molecule. NUCLSQ makes no allowance for hydrogen atoms on ligands and there is no way to accommodate this short of rewriting the entire program and creating new atom types and idealised bond distances, angles and geometries etc. Therefore the weighting scheme was stiffened by increasing the weights on distances and to regularise geometries and remove bad contacts (see table 2.2 for a summary of final refinement parameters used). This resulted in rejecting certain water molecules and an associated increase in the R factor to 17.7% for the DNA, ligand and 49 remaining water molecules, using data from 8-2.5 Å. The temperature factors (B) for the water molecules ranged from 10 to 71 Å<sup>2</sup> with a mean of 32 Å<sup>2</sup>. The bridging water molecule has a B of 30 Å<sup>2</sup>.

## Results

### 2.4 General structural features.

The observation of continuous electron density (figure 2.2) in the minor groove of the d(CGCGAATTTCGCG) double helix, together with the subsequent satisfactory refinement of a berenil molecule in this position, unequivocally places the drug at the AT region (figure 2.3). It is asymmetrically disposed with respect to the diad axis of the duplex and is bound to the 5'-AAT-3' sequence, reading down strand 1. The berenil molecule spans some three base pairs, interacting primarily with the two adenines, A5 and A18 on opposite strands (figure 2.4), in agreement with the estimate of the size from DNAase I footprinting (Portugal and Waring, 1987). The overall disposition of the drug is quite distinct from previous proposals based on molecular modelling (Gresh and Pullman, 1984; Pearl *et al.*, 1987; Gago *et al.*, 1988). The footprinting data strongly disfavour a binding-site model with just two contiguous AT base

Table 2.2. Refinement statistics including parameters used for d(CGCGAATTCGCG)<sub>2</sub> complexed with berenil.

Resolution range	8-2.5 Å	
Number of reflections [I>2σ(I)]	1759	
Temperature of study	-18°C	
Final R <sup>a</sup> factor	17.7%	
distances > 2σ	57	
	rms dev	σ value
sugar-base bond distances	0.011	0.025 Å
sugar base bond angle distances	0.032	0.050 Å
phosphate bond distances	0.183	0.050 Å
phosphate angle and H-bond distances	0.056	0.075 Å
planar groups	0.021	0.030 Å
chiral volumes	0.064	0.100 Å <sup>3</sup>
single torsion contacts	0.122	0.063 Å
multiple torsion contacts	0.149	0.063 Å
Isotropic temperature factors		
sugar-base bonds	5.079	10.5 Å <sup>2</sup>
sugar-base angles	5.905	10.5 Å <sup>2</sup>
phosphate bonds	6.551	10.5 Å <sup>2</sup>
phosphate angles, H bonds	5.770	10.5 Å <sup>2</sup>
weighting scheme applied to the structure factors	[1/(SIGAPP) <sup>2</sup> ] <sup>b</sup>	
AFSIG	1.6	
BFSIG	-34.6	

$$R^a = \sum |F_o - F_c| / \sum F_o$$

$$\text{SIGAPP}^b = \text{AFSIG} + \text{BFSIG} (\text{STHOL} - 0.1666667)$$

$$\text{STHOL} = \sin \theta / \lambda$$

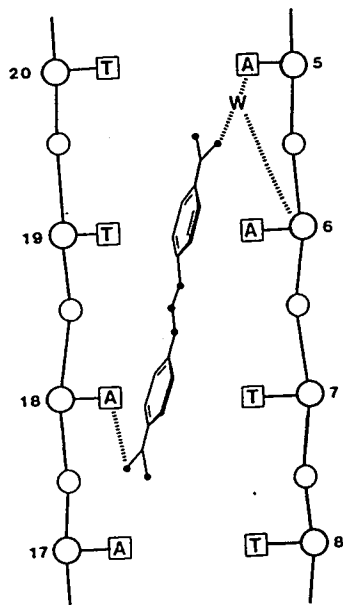


Figure 2.4. Schematic representation of berenil bound to the dodecamer.  
(Dashed lines indicate hydrogen bonds)

pairs, and generally favour binding sequences with three or more A and T residues, suggestive of flanking sequence effects.

When the structure is examined in detail the phenyl rings of berenil lie parallel to the hydrophobic minor groove walls whilst the terminal amidinium groups are in close contact with the adenine bases A5 and A18. The more polar triazene group of the drug is between the two phosphodiester groups of the backbone. The N3 atom of A18 is linked by a hydrogen bond of length 2.98 Å to a nitrogen atom of one amidinium, which is located between the two TA base pairs, T7-A18 and T8-A17, (figure 2.5) and is therefore significantly out of the base pair plane at this point. At the other end of the ligand, the amidinium does not directly contact N3 and A5. Instead, a water molecule mediates between this charged group and N3 of A5, with hydrogen-bond distances of 2.80 and 3.10 Å respectively (figure 2.6). This water molecule, which is in the plane of the A5-T20 base pair, thus effectively bridges the respective hydrogen bond donor/acceptor groups and clearly plays a crucial role in the interaction of

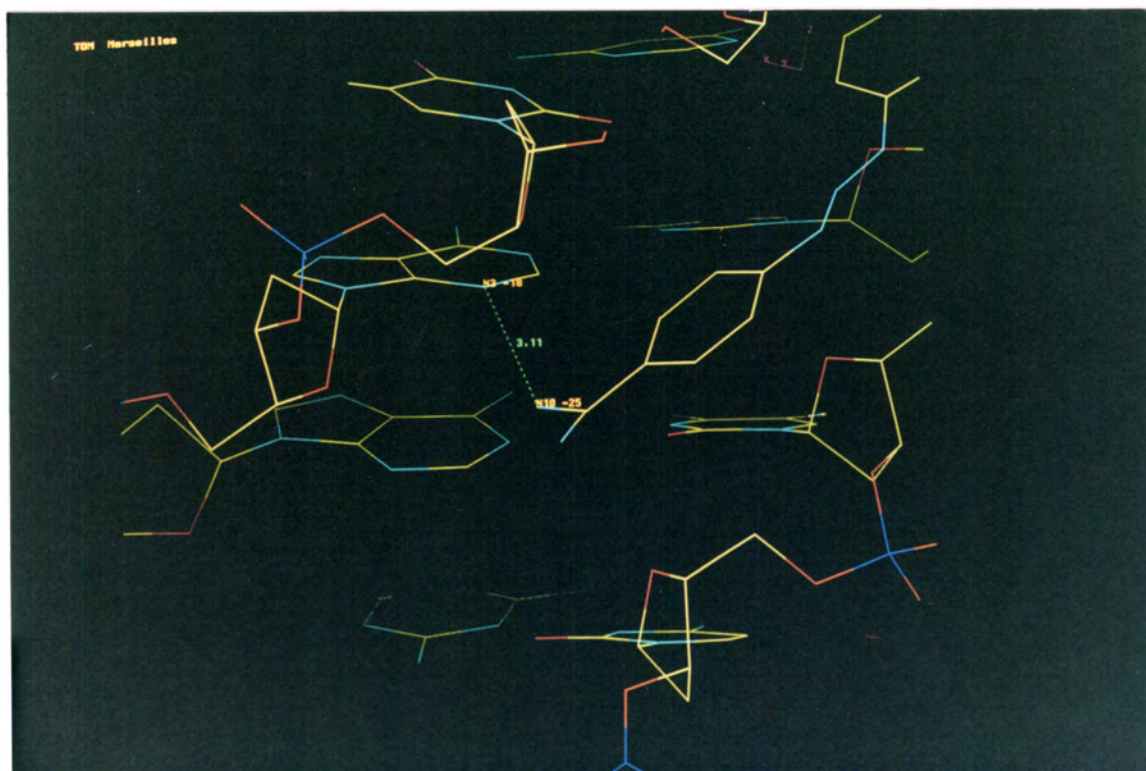


Figure 2.5 Close up view of hydrogen bonding at one end of the bound berenil.

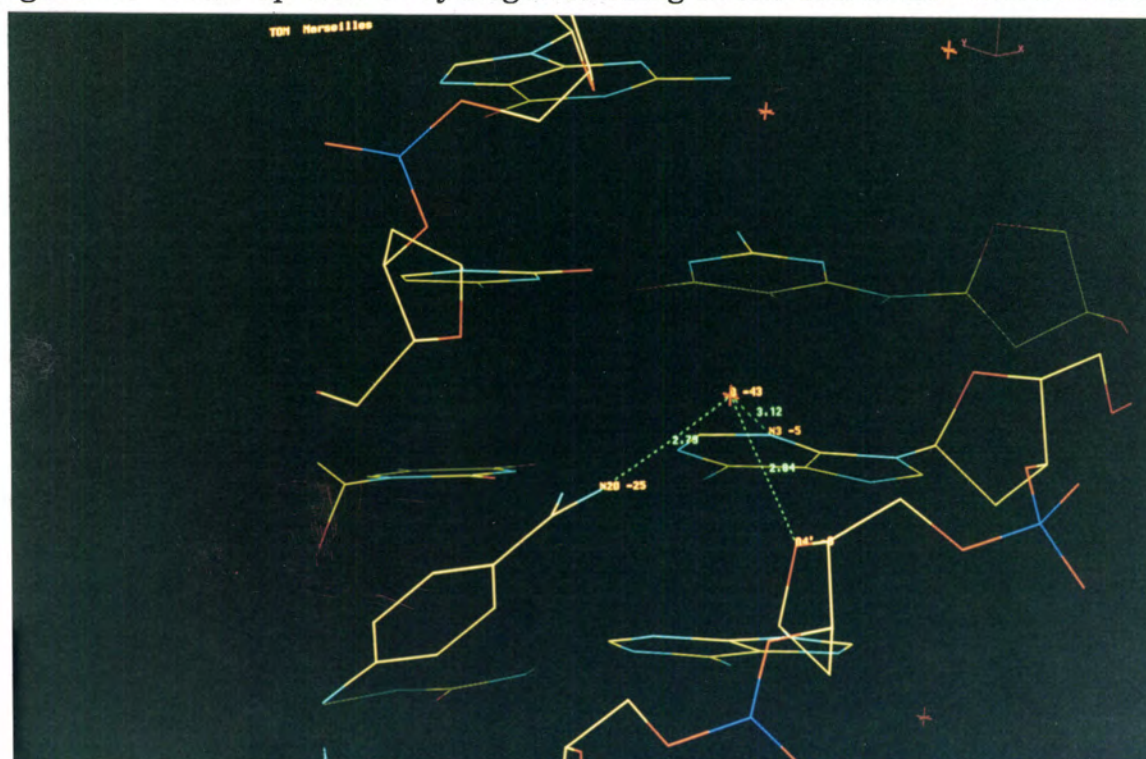


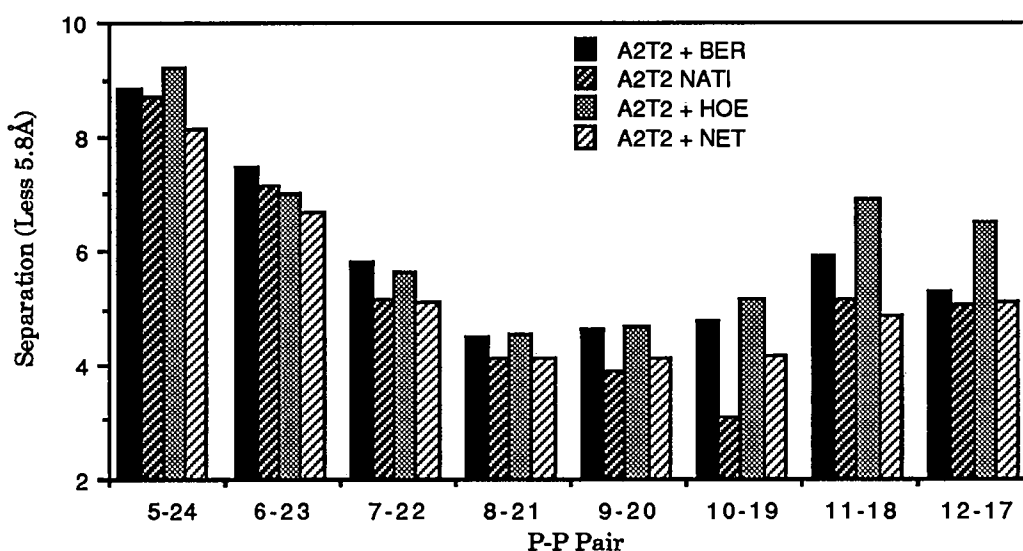
Figure 2.6 Close up view of hydrogen bonding at the other end of the bound berenil, showing the network involving the bridging water molecule.

berenil. The water molecule also forms a hydrogen bond with the O4' ribose ring atom of A6, at a distance of 2.82 Å. This water molecule does not occupy the same position as any of those found in the "spine of hydration" which are found in this region in the crystal structure of the native dodecamer (Drew and Dickerson, 1981). In that case the water in the most similar position is weakly hydrogen-bonded to both the N3 atom of adenine 5 and O2 of cytosine 21.

If idealised hydrogens are generated based on the crystallographic coordinates one can examine the close van der Waals contacts between the ligand and the DNA. These hydrogens are generated using the in-house graphics package GEMINI (Beveridge, 1990). The closest contacts to the minor groove floor are from H2 Ade18...HC5 Ber 1.91Å and H2 Ade6...C5' Ber 2.10Å. These close contacts are equal to the sum of the van der Waals radii of two hydrogens given the experimental error in the 2.5Å structure.

It is apparent then that this minor groove-binding drug effectively displaces any ordered water network in the minor groove, at least in the AT region. Figure 2.7 shows that the berenil molecule has widened the minor groove compared to the native structure by approximately 1.5 Å at its maximum. This is similar to changes observed for the crystal structures of the same dodecamer with the minor groove-binding ligands Hoechst 33258 and netropsin (Pjura *et al.*, 1987; Teng *et al.*, 1988; Quintana *et al.*, 1991; Kopka *et al.*, 1985 a,b). The significant widening extends for approximately 2 base-pairs.

Figure 2.7 Graph of minor groove width based on phosphate-phosphate distances.



### 2.5 Hydration in the crystal structure.

Examination of the general pattern of water molecules in the present structure suggests that berenil considerably changes the hydration pattern of the DNA. The readily located waters are now in the major groove and around the phosphate groups, many of them actually bridging phosphate oxygens (figures 2.8 and 2.9), in striking contrast to the native structure where the waters are found mainly in the grooves. A specialised case of bridging involves the triazene group of berenil, which does not directly participate in base recognition, where the hydrogen atom present in the triazene bridge must point outward from the minor groove. We find that the triazene nitrogen at the 5' end of the binding site is close to a water molecule at the mouth of the minor groove (3.17 Å separation). There is little evidence for the location of other solvent molecules within the minor groove (figure 2.10). It should be noted however that due to the limited resolution of the data care was taken too only include the most probable solvent molecules. The major groove waters are mostly hydrogen-bonded to polar atoms of the bases in a monodentate manner, with O6 and N7 of guanines being preferred. There are several localised base-water-base interactions, especially in the GC regions.

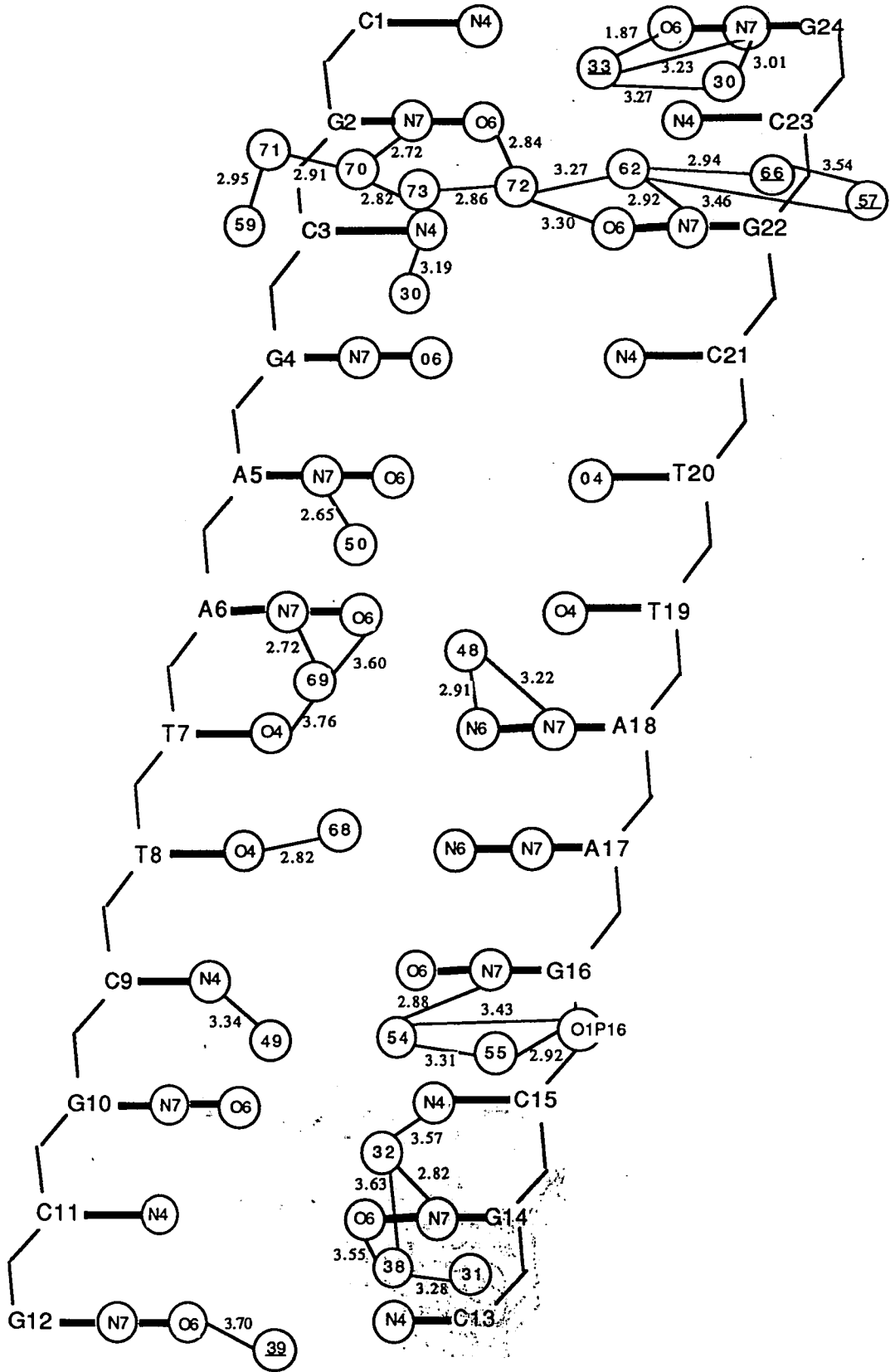


Figure 2.8 Schematic representation of major groove hydration



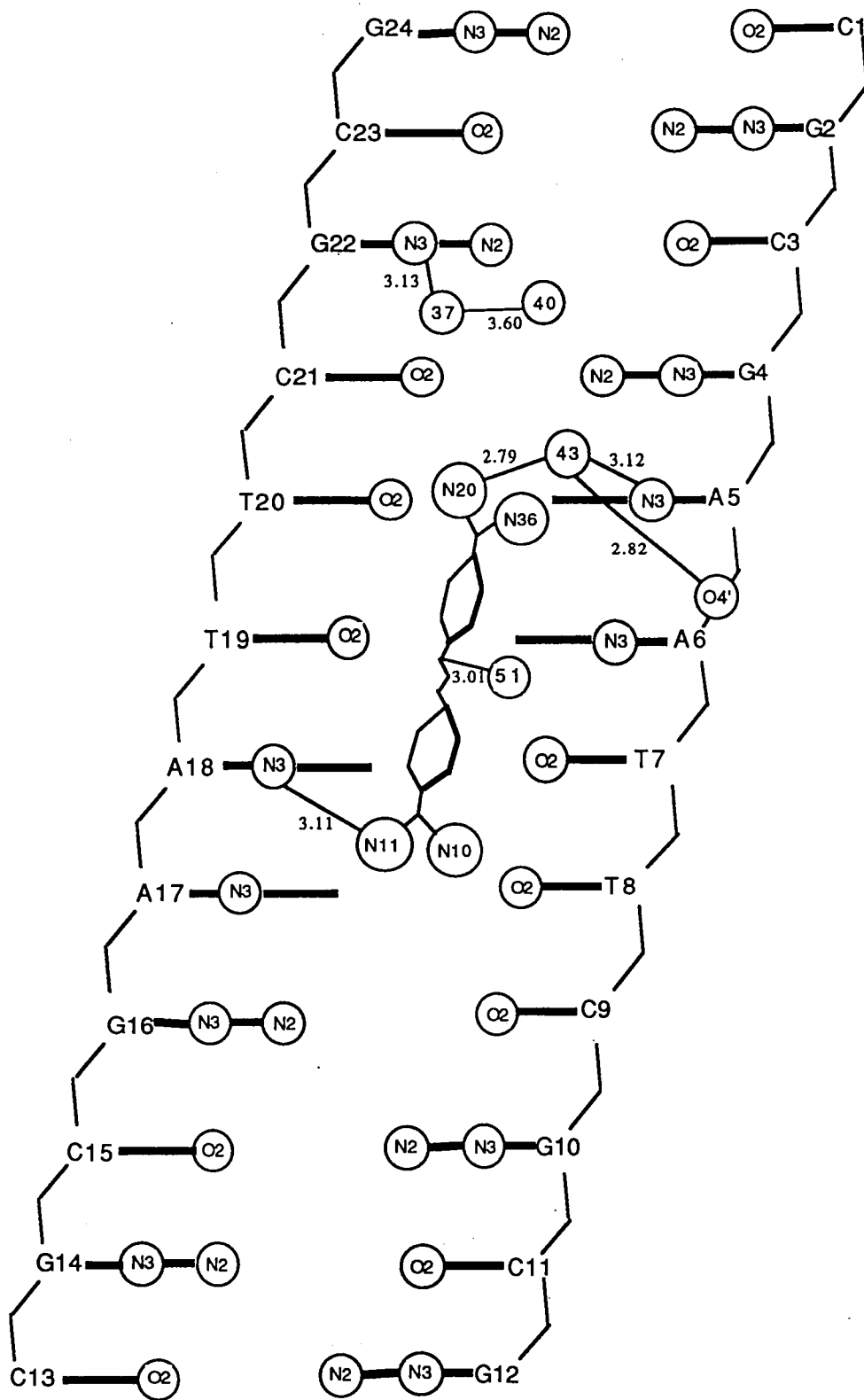


Figure 2.10. Schematic representation of minor groove hydration.

## 2.6 Packing in the crystal.

As is the case in other dodecamers packing in the space group  $P2_12_12_1$  the helix packs with the minor groove edges of the top two base-pairs of one helix overlap the minor groove edges of the bottom two base-pairs of the helix related by a  $2_1$  screw axis. The base-pairs intersect obliquely, enabling each guanine N2 to donate a proton to a hydrogen bond with the N3 atom of the guanine of the symmetry related helix. In all the end overlap is maintained by a network of four N2-H...N3 hydrogen bonds with an average N2...N3 distance of 3.12 Å and good hydrogen bond geometry. The helix overlap is further stabilised by hydrogen bonds involving of each of the 3'-OH groups, which interact directly to the N2 of Gua (22) and O2 of Cyt (23). Both N3 and N2 of Gua(12) are also in hydrogen bonding distance from O4' of Gua (2).

## 2.7 Analysis of helical parameters.

Analysis of the helical parameters for the DNA was carried out using NEWHELIX91. Comparison of the values for each of these parameters with those obtained for the native DNA (Drew and Dickerson, 1981) and for the same sequence with two other minor groove binding ligands, netropsin (Kopka *et al.*, 1985a) and Hoechst 33258 (Pjura *et al.*, 1987), are highlighted in tables 2.3 and 2.4 and figures 2.10–22.

The helix in the complex with berenil is general B-type DNA with 10 base pairs per turn, and exhibits an overall bend of approximately 14° using the normal vector plot method of Dickerson *et al.*, (1983).

It is immediately obvious from these figures that none of the minor groove binding ligands grossly distort the DNA and berenil perturbs the DNA less than either of the other ligands plotted. The perturbations induced in the DNA by berenil are concentrated in the central AAT binding site where a combination of small differences in tip, roll and tilt together with a reduced base pair buckle allow optimisation of the interactions between the drug and

Table 2.3.: Conformational parameters for the B-DNA dodecamer  $d(\text{CGCGAATTCGCG})_2$  complexed with berenil.

Residue	Angles, Degrees							Sugar Pucker
	$\chi$	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	
Cyt1	-108	—	—	251	146	-153	-136	$C_2'$ -endo
Gua2	-91	-32	164	21	141	-174	-114	$C_2'$ -endo
Cyt3	-106	-40	174	27	137	-164	-124	$C_2'$ -endo
Ade4	-95	-45	168	43	140	-172	-112	$C_2'$ -endo
Ade5	-108	-65	168	47	141	-158	-123	$C_2'$ -endo
Ade6	-102	-28	161	11	139	-174	-123	$C_2'$ -endo
Thy7	-101	-2	159	0	136	-161	-130	$C_2'$ -endo
Thy8	-99	-30	154	34	133	-170	-138	$C_2'$ -endo
Thy9	-104	-8	157	8	144	-152	-129	$C_2'$ -endo
Gua10	-89	-30	149	28	145	-114	-206	$C_2'$ -endo
Cyt11	-113	-58	139	47	144	-161	-93	$C_2'$ -endo
Gua12	-87	-65	188	31	142	—	—	$C_2'$ -endo
Cyt13	-121	—	—	-24	144	-134	-169	$C_2'$ -endo
Gua14	-104	-5	127	6	140	-153	-132	$C_2'$ -endo
Cyt15	-106	-50	149	51	135	-155	-155	$C_2'$ -endo
Ade16	-100	-34	142	48	140	-183	-108	$C_2'$ -endo
Ade17	-104	-63	190	52	140	-197	-78	$C_2'$ -endo
Ade18	-97	-83	208	45	141	-183	-103	$C_2'$ -endo
Thy19	-108	-55	182	42	140	-162	-108	$C_2'$ -endo
Thy20	-107	-61	174	50	136	-160	-147	$C_2'$ -endo
Thy21	-90	+20	132	-12	142	-161	-120	$C_2'$ -endo
Gua22	-93	-52	167	37	149	-109	-188	$C_2'$ -endo
Cyt23	-111	-52	140	26	140	-160	-128	$C_2'$ -endo
Gua24	-115	-52	162	47	137	—	—	$C_2'$ -endo
Mean	-103	-40	161	29 <sup>1</sup>	141	-160	-130	
$\pm$ SD	8	24	20	21	4	20	29	

Sugar puckers are an artefact due to targeting during refinement.

<sup>1</sup> Cyt1 value omitted because it represents end effects.

Table 2.3. (continued): Conformational parameters for the B-DNA dodecamer d(CGCGAATTCGCG)<sub>2</sub>.

Sugar residue	Angles, Degrees							Pucker
	$\chi$	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	
Cyt1	-105	—	—	174	157	-141	-144	C <sub>2'</sub> -endo
Gua2	-111	-66	170	40	128	-186	-98	C <sub>1'</sub> -exo
Cyt3	-135	-63	172	59	98	-177	-88	O <sub>4'</sub> -exo
Gua4	-93	-63	180	57	156	-155	-153	C <sub>2'</sub> -endo
Ade5	-126	-43	143	52	120	-180	-92	C <sub>1'</sub> -exo
Ade6	-122	-73	180	66	121	-186	-89	C <sub>1'</sub> -exo
Thy7	-127	-57	181	52	99	-186	-86	O <sub>4'</sub> -endo
Thy8	-126	-59	173	64	109	-189	-89	C <sub>1'</sub> -exo
Cyt9	-120	-58	180	60	129	-157	-94	C <sub>1'</sub> -exo
Gua10	-90	-67	169	47	143	-103	-210	C <sub>2'</sub> -endo
Cyt11	-125	-74	139	56	136	-162	-90	C <sub>2'</sub> -endo
Gua12	-112	-82	176	57	111	—	—	C <sub>1'</sub> -exo
Cyt13	-128	—	—	56	137	-159	-125	C <sub>2'</sub> -endo
Gua14	-116	-51	164	49	122	-182	-93	C <sub>1'</sub> -exo
Cyt15	-134	-63	169	60	86	-185	-86	O <sub>4'</sub> -endo
Gua16	-115	-69	171	73	136	-186	-98	C <sub>2'</sub> -endo
Ade17	-106	-57	190	54	147	-183	-97	C <sub>2'</sub> -endo
Ade18	-108	-57	186	48	130	-186	-101	C <sub>2'</sub> -endo
Thy19	-131	-58	174	60	109	-181	-88	C <sub>1'</sub> -exo
Thy20	-120	-59	179	55	122	-181	-94	C <sub>1'</sub> -exo
Cyt21	-114	-59	185	45	110	-177	-86	C <sub>1'</sub> -exo
Gua22	-88	-67	179	50	150	-100	-188	C <sub>2'</sub> -endo
Cyt23	-125	-72	139	45	113	-174	-97	C <sub>1'</sub> -exo
Gua24	-135	-65	171	47	79	—	—	C <sub>3'</sub> -endo
Mean	-117	-63	171	54 <sup>2</sup>	123	-169	-108	
±SD	14	8	14	8	21	25	34	

<sup>2</sup>Cyt1 value omitted because it represents end effect.

Table 2.4. Local helix parameters for the B-DNA Dodecamer  
d(CGCGAATTCGCG)<sub>2</sub> complexed with berenil.

Base-pairs	Propeller twist, ° $\theta_p$	Helix twist angle, ° $t^3$	Base-pairs per turn ( $n$ )	Rise per base pair, Å ( $h$ )
C1/G24	-13.9	38.6	9.32	3.47
G2/C23	-11.2	36.8	9.78	3.06
C3/G22	-1.1	30.4	11.84	3.84
A4/T21	-8.4	36.4	9.89	3.16
A5/T20	-15.7	35.2	10.23	3.35
A6/T19	-17.8	32.6	11.04	3.23
T7/A18	-13.2	34.6	10.40	3.28
T8/A17	-17.0	40.6	8.87	3.10
T9/A16	-11.7	33.0	10.91	3.98
G10/C15	-3.0	39.0	9.23	2.92
C11/G14	-22.8	36.5	9.86	3.70
G12/C13	-9.3			
Mean	-12.1	35.8	10.12	3.37
SD	6.1	3.0	0.84	0.34

<sup>3</sup> Helical parameters obtained by using vectors between atoms C<sub>1'</sub> and attached N of one base and the equivalent atoms of the next nucleotide along the chain.

Table 2.4.(continued ) Local helix parameters for the B-DNA dodecamer  
d(CGCGAATTCGCG)<sub>2</sub>.

Base-pairs	Propeller twist, ° $\theta_p$	Helix twist angle, ° $t^*$	Base-pairs per turn ( $n$ )	Rise per base pair, Å ( $h$ )
C1/G24	-13.2	38.3	9.40	3.36
G2/C23	-11.7	39.6	9.09	3.38
C3/G22	-7.2	33.5	10.75	3.26
G4/C21	-13.2	37.4	9.63	3.30
A5/T20	-17.1	37.5	9.60	3.27
A6/T19	-17.8	32.2	11.18	3.31
T7/A18	-17.1	36.0	10.00	3.29
T8/A17	-17.1	41.4	8.70	3.14
C9/G16	-18.6	32.3	11.11	3.56
G10/C15	-4.9	44.7	8.05	3.21
C11/G14	-17.2	37.0	9.73	3.54
G12/C13	-6.2			
Mean	-13.4	37.3	9.75	3.33
SD	4.9	3.8	0.98	0.13

\* Helical parameters obtained by using vectors between atoms C<sub>1'</sub> and attached N of one base and the equivalent atoms of the next nucleotide along the chain.

Figure 2.11. Base pair twist.

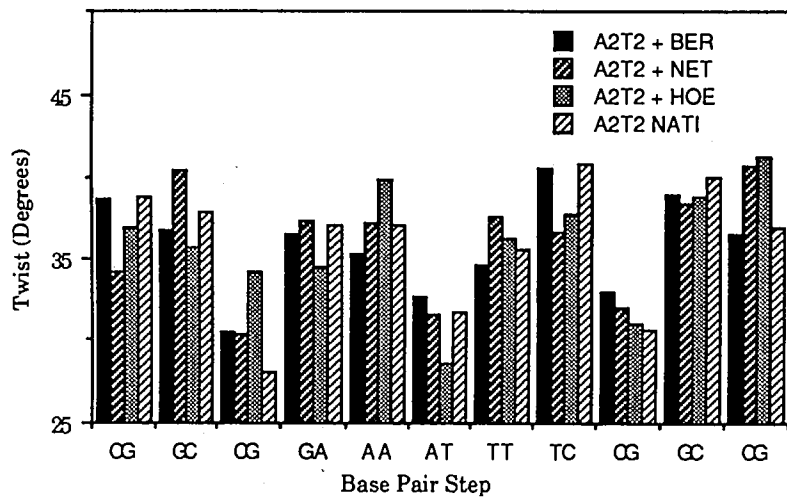


Figure 2.12. Base pair roll.

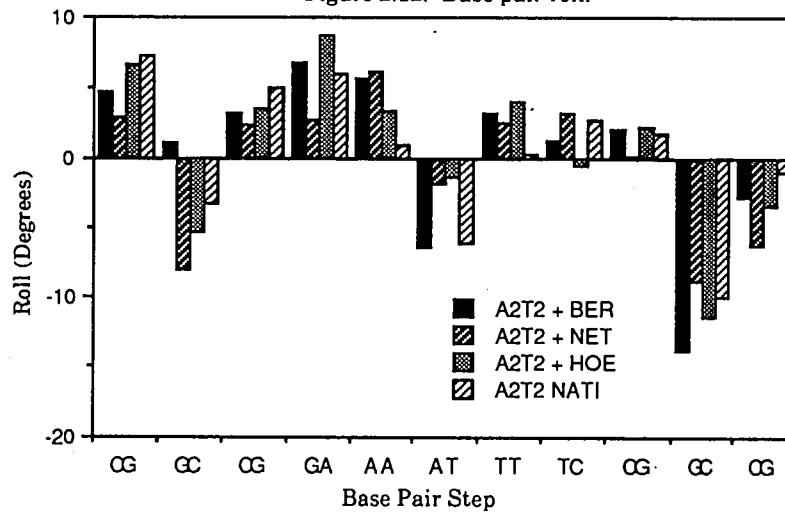


Figure 2.13. Base pair tilt.

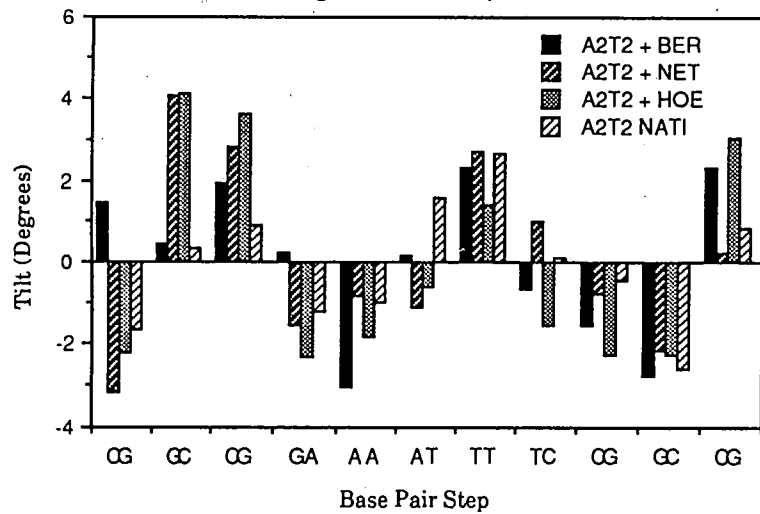


Figure 2.14. Base pair rise.

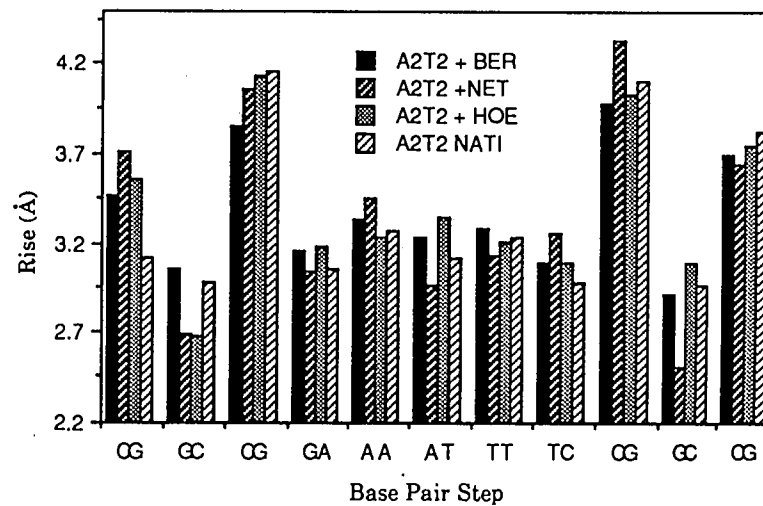


Figure 2.15. Base pair buckle.

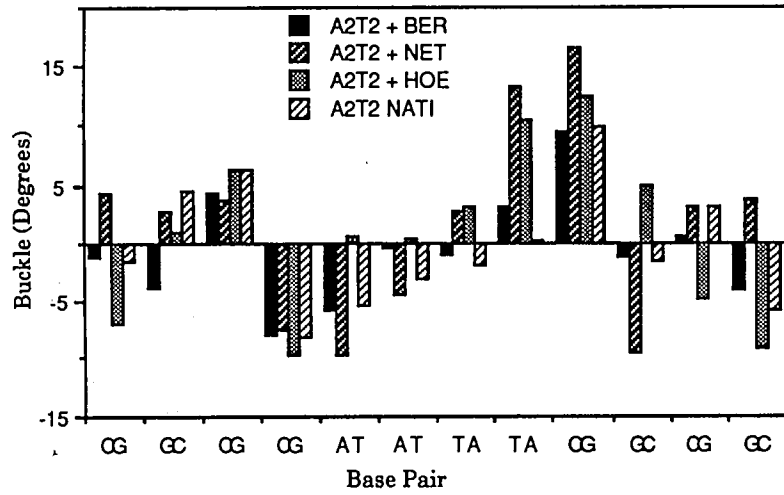


Figure 2.16. Base pair inclination.

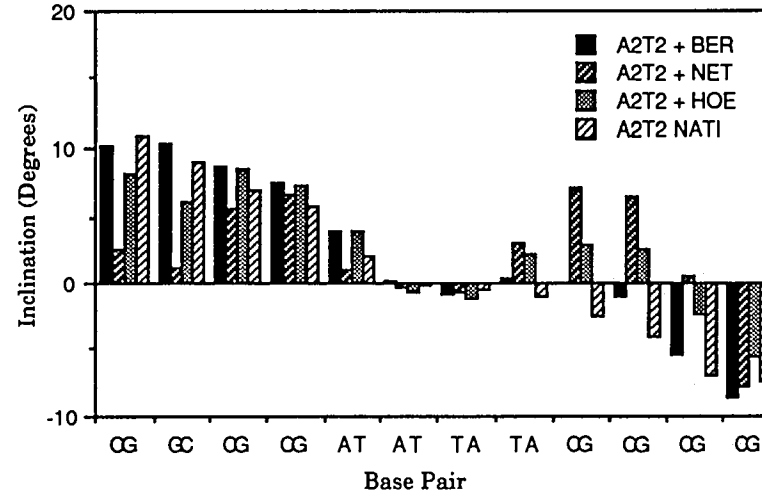


Figure 2.17. Base pair tip.

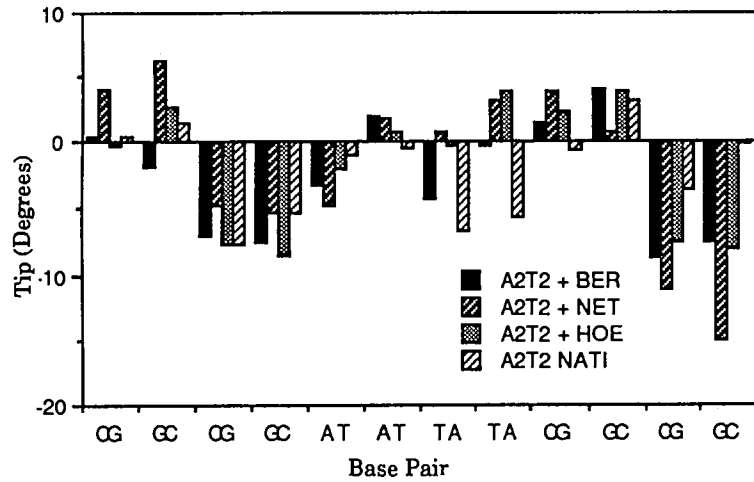
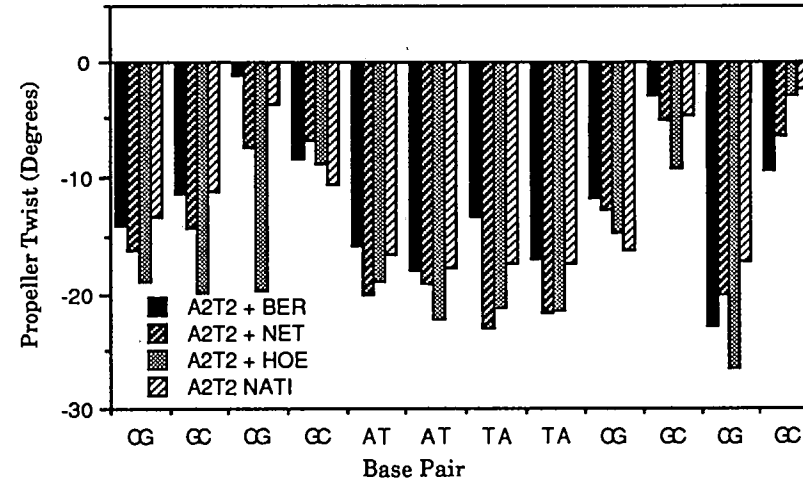
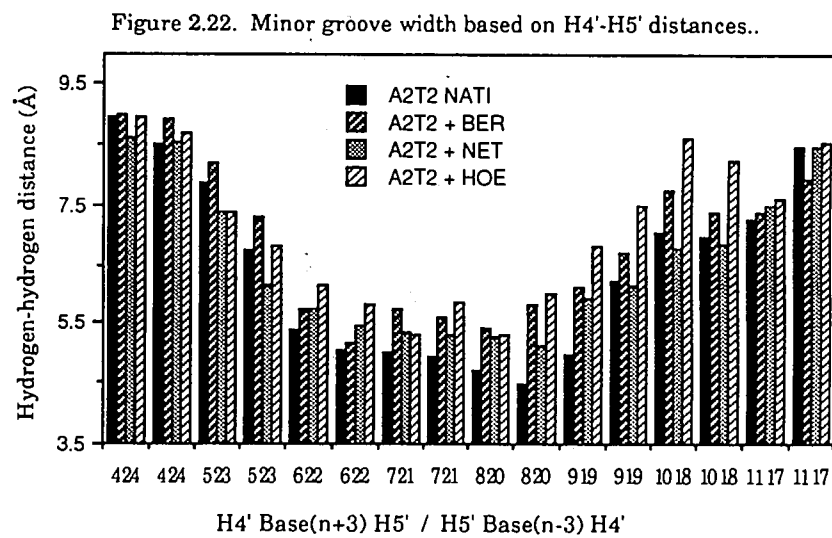
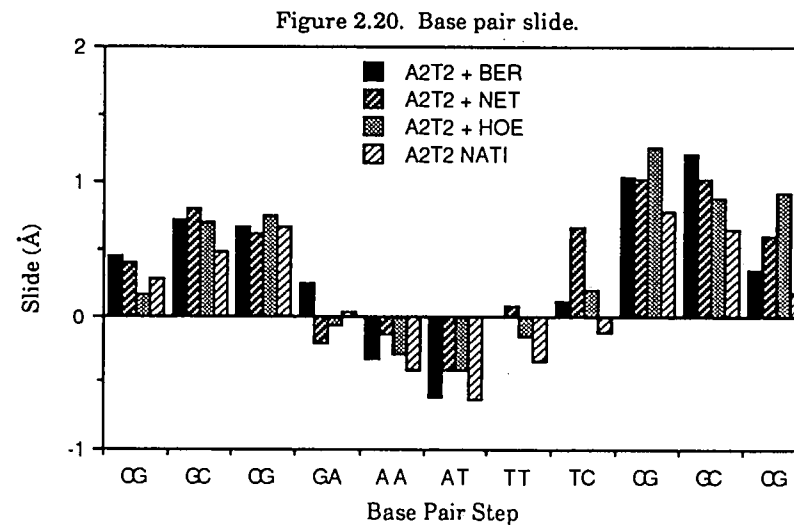
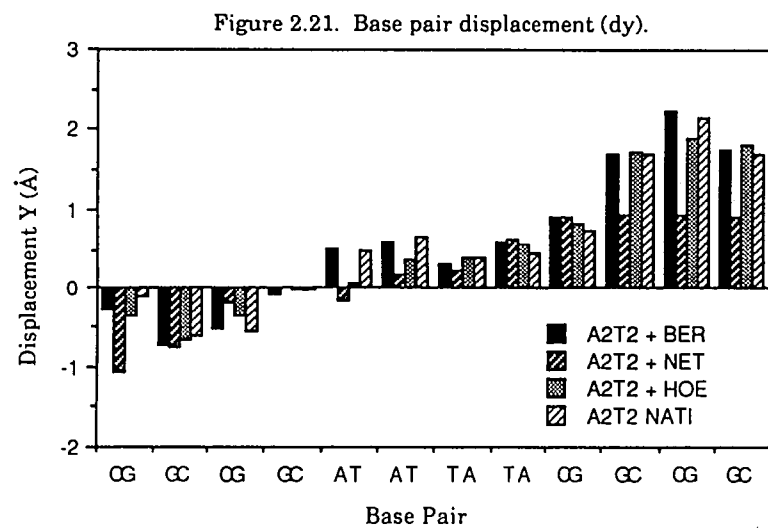
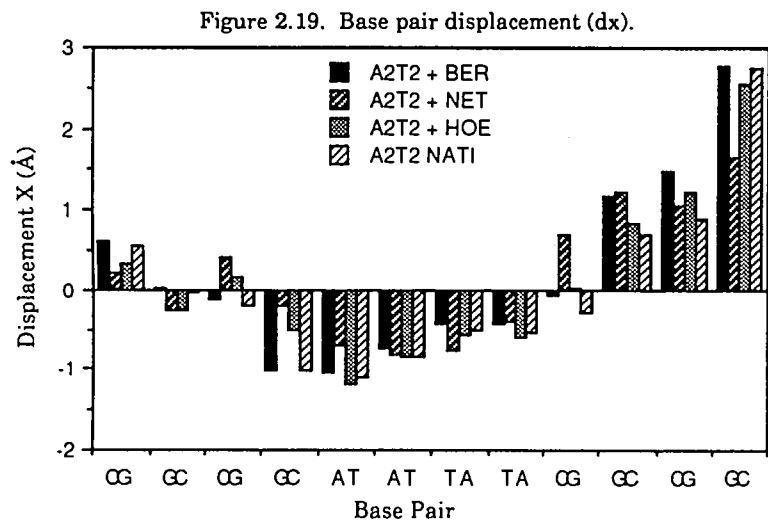


Figure 2.18. Base pair propeller twist.





DNA, whilst maintaining favourable base stacking and avoiding steric clashes in the minor groove.

The most striking feature is the small, but significant, increase in minor groove width in the central region of the dodecamer on drug binding. The importance of this increase in groove width is further highlighted by examination of the idealised hydrogens generated for the complex. Measurement of groove width in the region of the bound drug, using H4' and H5' cross-strand differences ( $H4-H5'_{n+3}$ ), shows that the minor groove width (less the sum of the van der Waals radii) is  $\approx 3.4$  Å, which is the thickness of the aromatic phenyl rings of the berenil (figure 2.22). The outer hydrogens of the backbone sugars seem to "pinch" inwards clamping the drug into the groove, H5' Ade5...HC3' Ber 2.15Å, H5' Cyt9...C3 Ber 2.60Å and H4' Thy8...C1 Ber 2.34Å.

### *2.8 Comparison with the NMR data of the complex of $d(CGCGAATTCGCG)_2$ with berenil and the complex of $d(GCAATTGC)_2$ with berenil.*

Subsequent to the publication of the crystal structure of the complex of berenil with  $d(CGCGAATTCGCG)_2$  Yoshida and coworkers (1990) published their findings of a complex of berenil with  $d(GCAATTGC)_2$  from solution studies by NMR. This was closely followed by a more extensive and detailed study by Lane and coworkers (1991) with the same EcoRI dodecamer as the crystal structure. The study with the shorter octamer concluded that there were two possible models each of which could account for all the observed NOE's. Their first model spans four base-pairs with the amidinium moieties of each end of the drug forming hydrogen bond contacts to the O2 of thymines 14 and 6 respectively. The second model was a 1:2 model spanning the central A•T step. In this model the amidinium moieties are in hydrogen-bonding contact with the N3 of adenines 4 and 12. The significant difference in the two models as far as the NOE data is concerned is that model I predicts a strong

contact between the amidine protons and H1' of the external thymidine (Thy 3/14), whereas model II would give rise to a strong NOE between the amidine protons and H1' of the internal adenosine (Ade 4/12). However, the authors were unfortunately not able to detect either the amidine-thymidine H1' or amidine-guanosine H1' cross-peaks in the NOESY spectrum which made the two models indistinguishable on the basis of the predicted contacts.

The more extensive study by Lane *et al.*, 1991, resulted in a more definitive answer. Initial analysis of the chemical shift differences between the free EcoRI site and the EcoRI-berenil complex for H8/H6, H1',H2', H2'' and phosphates (see figure 2.6 of Lane *et al.*, 1991) give a strong indication that the binding site spans three base pairs 5' ATT. Subsequent refinement using molecular dynamics with NOE restraints and glycosidic torsion angle constraints confirmed the 1:3 model, which is in general agreement with the crystal structure. The main difference between the crystal structure and the NMR derived solution structure is that the drug is displaced approximately half a base pair towards the diad axis of the DNA. Also of course in this NMR model the amidinium binds directly to N3 of adenine at the 5' end and does not involve a bridging water molecule, although the study did not attempt to refine a structure with such a water molecule. It is interesting to note that there was some evidence for local ligand-induced perturbation on DNA conformation. On binding there was an increase in the roll and tilt angles for the Thy8/Ade17 and Cyt9/Gua16 base pairs. Globally the DNA is not significantly altered by the minor groove binding ligand. The DNA is B-type with sugars around the C2' endo conformation and B<sub>I</sub> phosphates. However, <sup>31</sup>P relaxation data and NOE intensities imply conformational anomalies at Cyt3-Gua4/Cyt9-Gua10 and Ade5-Ade6/Thy7-Thy8. Ott and Eckstein (1985) and Gornstein *et al.*, (1988) noted that these positions have anomalous chemical shifts consistent with sequence-dependant variations in the C-O-P backbone angles. This evidence

supports the argument that these regions are somewhat more flexible, (the sites of 'neo kinks' in the EcoRI-DNA complex (M<sup>c</sup>Clarín *et al.*, 1986), and may be sequence related regions of local bending (Fratini *et al.*, 1982). Recently extensive molecular dynamics calculations based on the GROMOS force field (with an additional restraint potential for maintaining Watson-Crick base pairing) including water and counterions, (Swaminathan *et al.*, 1991) have found that C3 and G10 are sites of intrinsic bend within the EcoRI dodecamer.

### 2.9 Molecular modelling.

Binding and perturbation energies (which measure conformational energy change) were calculated by Dr. Jenkins for each of the modelled complexes in table 2.5 based on the core octamer d(CGAATTTCG)<sub>2</sub>. The latter reflect the distortions included by the drug during binding. The binding energies and individual perturbation energies for each drug-DNA complex are given in table 2.5. These show that berenil binds preferentially at a d(..AAAT..) site which involves spanning three base pairs and hence hydrogen bonds are formed with non-adjacent bases (adenines 1, 3).]

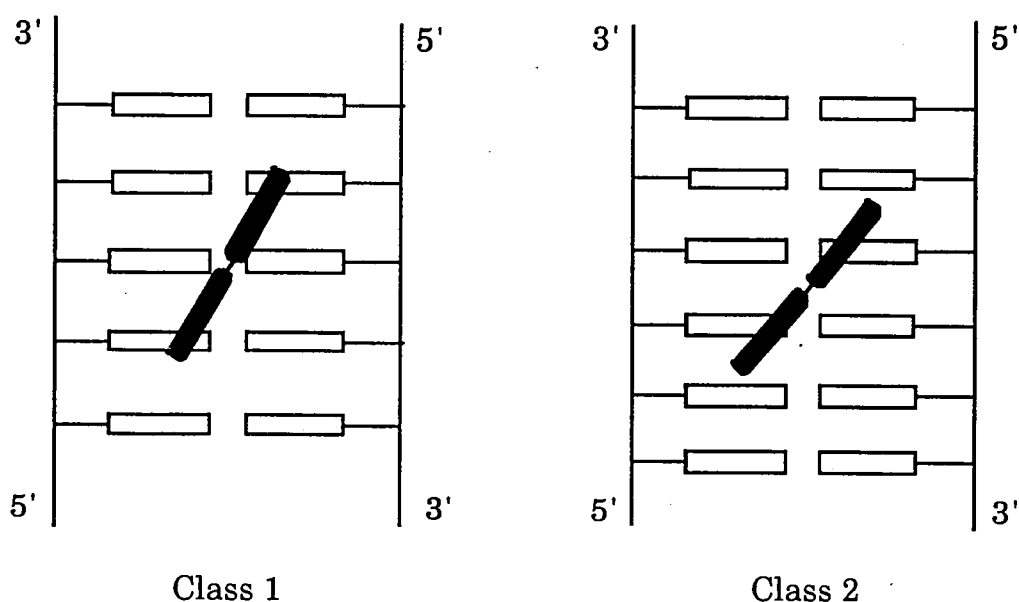


Figure 2.23. Schematic showing "class I" and "class II" binding modes

Table 2.5: Calculated energies for the octamer-berenil complexes.

sequence <sup>a</sup>	$\Delta\Delta H^b$	perturbation <sup>c</sup>	
		DNA	ligand
<u>A</u> ATT	-44.6	+1.0	+1.0
A <u>A</u> T	-44.1	0.0	+1.6
<u>ATT</u>	-43.8	+1.1	+0.7
A <u>T</u> T	42.3	+1.1	+0.9
<u>A</u> ATT	41.1	+2.0	+0.7
A <u>A</u> T	40.8	+3.9	+0.6
A <u>T</u> T	34.6	+5.6	+4.3

<sup>a</sup> Base sequence for binding site in d(CGAATTCG)<sub>2</sub>, where the underlined bases represent the possible “class I” or “class II” complexes (see figure 2.23), respectively.

<sup>b</sup> Energies in kcal mol<sup>-1</sup>, where

$$\Delta\Delta H = \Delta H_{\text{complex}} - [\Delta H_{\text{DNA}} + \Delta H_{\text{ligand}}]$$

<sup>c</sup> Perturbation energies for complexed versus free molecules, where

$$\Delta H_{\text{DNA}} = \underline{E}_{\text{DNA}(\text{bound})} - \underline{E}_{\text{DNA}}; \Delta H_{\text{ligand}} = \underline{E}_{\text{ligand}(\text{bound})} - \underline{E}_{\text{ligand}}$$

and  $\underline{E}_{\text{DNA}} = 380.1$  and  $\underline{E}_{\text{ligand}} = -0.5$  kcal mol<sup>-1</sup>, respectively.

## 2.10 Discussion

It is apparent that both the length of the berenil ligand and the nature of the interactions with the adenines do not permit symmetrical binding across the diad axis of this sequence. Why does berenil not bind between A6 and A17, since this appears to be an equivalent AT binding site? (There is no indication of electron density in this region.) The exocyclic N2 of guanine G16

presumably prevents such binding, due to steric hindrance in the minor groove region where interaction with N3 of A17 would occur. Flanking sequences thus appear to play an important role in discriminating between different binding sites. Interactions of this type have been examined in detail by us in a molecular mechanics study using a large number of sequences (Jenkins and Neidle, unpublished data). This study also indicates that a replacement of A5 by Thymine would not remove the necessity for a bridging water molecule, since a thymine O2 would not come sufficiently close to the amidinium group for hydrogen bond formation. There is no opportunity in the present dodecamer sequence for such an interaction to occur. The two DNA fragments used in the footprinting study (Portugal and Waring, 1987) contain few of the 16 distinct AT tetranucleotide sequences, hence the full sequence preference order for berenil is not known experimentally. The observed position of the berenil molecule has the hydrophobic phenyl rings contacting the hydrophobic regions of the backbone and the triazene group neighbouring the polar phosphodiester groups. This may be a further factor in the exact location of the drug within a general AT region. No thermodynamic data are available for berenil binding to polynucleotides; that for netropsin (Marky and Breslauer, 1987) shows that the process is enthalpy driven. The present results suggest a similar mode of binding for berenil although the incorporation of (redistributed) solvent at the drug binding sites for some sequences will enhance an entropic contribution.

The present structural result suggests that berenil should bind well to an AATT site in genomic DNA and is supported by the fact that the sole AATT site in the footprinted 102mer fragment examined is the region of the sequence most strongly protected from DNAase I. In general, these results suggest that berenil does not easily tolerate either C or G bases at the 3' side of the binding site, since we find that AAT(T) rather than AAT(C) is preferred for the drug. The nature of the flanking sequence at the 5' end may be of secondary

importance.

The widening of the minor groove of the AATT region compared to the native dodecamer reflects the interplay of two opposing factors: (i) the intrinsic narrowness of the AT sequence minor groove, and (ii) the widening essential for effective ligand binding. Recent hydroxyl radical footprinting studies (Laughton *et al.*, 1990) have suggested that the site size is in accord with the structural results presented here. The overall bending of the dodecamer is not significantly reduced compared to the native structure, although extrapolations to DNA in solution must be made with caution in view of the influence of crystal packing forces. The AT-selective DNA groove binder distamycin strongly diminishes the bending of kinetoplast DNA (Wu and Crothers, 1984) and is found to decrease the bending of the sequence d(CGCAAATTTGCG) in the crystal structure of its complex by 4° (Coll *et al.*, 1987). Berenil may therefore be considerably less perturbing in its effect on kinetoplast DNA structure, however the stabilisation of a cleavable topoisomerase II-DNA-berenil complex is likely to cause, the observed, complete loss of kinetoplast DNA in trypanosomes when treated with berenil (Riou and Bernard, 1980).

Site specific recognition of DNA by proteins is, in general, probably via the major groove, primarily for steric reasons. Non site-specific proteins, such as histone H1 (Turnell *et al.*, 1988) or DNAase I (Suck *et al.*, 1988), appear to interact with the minor groove. Crystal structures of protein-oligonucleotide complexes have shown a few instances where a water molecule mediates between an amino acid residue and a base. Especially relevant is the phage 434 operator-repressor complex (Aggarwal *et al.*, 1988) where Arg43 is bridged via three water molecules to base pairs in the minor groove. The coordination to N3 of adenine is strikingly similar to that in the present structure, implying that the role of water molecules in these circumstances is to fill the gaps not occupied by respective 'side-chains'. In the case of berenil, the triazene group

and the phenyl rings together impart selectivity towards AT, whereas the amidinium moieties select particular bases. The water molecule at the 5' end thus fulfils an essential role enabling an adenine to be sensed by the drug. The possibility of such interactions will have to be taken into account in the future design of sequence-specific agents. Bridging water molecules have been found to be important in stabilising the intercalative drug daunomycin when bound to the sequences d(CGTACG) and d(CGATCG) in the crystal (Wang *et al.*, 1987; Moore *et al.*, 1989).

Table 2.6. Final coordinates and temperature factors for the DNA dodecamer d(CGCGAATTCGCG)<sub>2</sub> complexed with berenil including solvent molecules.

	Atom	Residue		X	Y	Z	Occ	B(Iso)
ATOM	1	O5'	CYT	1	19.858	34.507	24.730	1.00 14.24
ATOM	2	O4'	CYT	1	19.204	31.120	25.022	1.00 15.96
ATOM	3	C5'	CYT	1	18.978	33.411	24.366	1.00 17.30
ATOM	4	C4'	CYT	1	19.846	32.164	24.320	1.00 16.71
ATOM	5	C3'	CYT	1	20.098	31.644	22.902	1.00 18.74
ATOM	6	C2'	CYT	1	19.061	30.537	22.757	1.00 18.57
ATOM	7	C1'	CYT	1	19.010	29.985	24.161	1.00 16.87
ATOM	8	N1	CYT	1	17.738	29.309	24.465	1.00 17.00
ATOM	9	C6	CYT	1	16.498	29.821	24.200	1.00 17.22
ATOM	10	C5	CYT	1	15.368	29.165	24.498	1.00 17.86
ATOM	11	N4	CYT	1	14.382	27.230	25.452	1.00 17.57
ATOM	12	C4	CYT	1	15.488	27.898	25.134	1.00 17.27
ATOM	13	N3	CYT	1	16.718	27.394	25.412	1.00 17.55
ATOM	14	O2	CYT	1	18.968	27.598	25.346	1.00 18.80
ATOM	15	C2	CYT	1	17.863	28.082	25.095	1.00 16.26
ATOM	16	O3'	CYT	1	21.407	31.060	22.803	1.00 23.06
ATOM	17	P	GUA	2	22.209	31.016	21.406	1.00 21.44
ATOM	18	O1P	GUA	2	21.206	31.280	20.346	1.00 16.79
ATOM	19	O2P	GUA	2	23.292	32.060	21.505	1.00 17.91
ATOM	20	O5'	GUA	2	22.767	29.553	21.432	1.00 19.04
ATOM	21	C5'	GUA	2	23.167	28.758	22.558	1.00 16.39
ATOM	22	O4'	GUA	2	22.054	26.643	22.518	1.00 16.47
ATOM	23	C4'	GUA	2	23.250	27.298	22.141	1.00 15.30
ATOM	24	C3'	GUA	2	23.414	27.066	20.624	1.00 12.02
ATOM	25	C2'	GUA	2	21.983	26.751	20.180	1.00 8.21
ATOM	26	C1'	GUA	2	21.422	26.019	21.386	1.00 7.58
ATOM	27	N9	GUA	2	19.963	26.151	21.505	1.00 9.63
ATOM	28	C8	GUA	2	19.157	27.182	21.068	1.00 5.68
ATOM	29	N7	GUA	2	17.900	27.010	21.313	1.00 3.83
ATOM	30	C5	GUA	2	17.856	25.791	21.975	1.00 4.67
ATOM	31	O6	GUA	2	15.579	25.439	22.472	1.00 7.47
ATOM	32	C6	GUA	2	16.750	25.075	22.505	1.00 5.05
ATOM	33	N1	GUA	2	17.074	23.888	23.088	1.00 4.69
ATOM	34	N2	GUA	2	18.476	22.237	23.783	1.00 4.63
ATOM	35	C2	GUA	2	18.360	23.428	23.167	1.00 6.34
ATOM	36	N3	GUA	2	19.434	24.068	22.690	1.00 6.42
ATOM	37	C4	GUA	2	19.113	25.247	22.108	1.00 7.38
ATOM	38	O3'	GUA	2	24.314	25.979	20.425	1.00 12.72
ATOM	39	P	CYT	3	24.924	25.435	19.061	1.00 12.23
ATOM	40	O1P	CYT	3	24.476	26.375	18.034	1.00 15.89
ATOM	41	O2P	CYT	3	26.410	25.395	19.266	1.00 15.07
ATOM	42	O5'	CYT	3	24.248	23.968	18.948	1.00 8.72
ATOM	43	C5'	CYT	3	24.167	23.228	20.180	1.00 10.18
ATOM	44	O4'	CYT	3	22.049	22.141	20.386	1.00 8.16
ATOM	45	C4'	CYT	3	23.410	21.933	20.054	1.00 11.36
ATOM	46	C3'	CYT	3	23.446	21.317	18.644	1.00 11.85
ATOM	47	C2'	CYT	3	22.069	21.729	18.094	1.00 10.01
ATOM	48	C1'	CYT	3	21.177	21.745	19.339	1.00 4.45

ATOM	49	N1	CYT	3	20.098	22.741	19.147	1.00	4.22
ATOM	50	C6	CYT	3	20.292	23.960	18.544	1.00	5.77
ATOM	51	C5	CYT	3	19.351	24.864	18.299	1.00	4.41
ATOM	52	N4	CYT	3	17.042	25.379	18.584	1.00	6.95
ATOM	53	C4	CYT	3	18.054	24.532	18.763	1.00	3.35
ATOM	54	N3	CYT	3	17.831	23.340	19.386	1.00	5.52
ATOM	55	O2	CYT	3	18.618	21.353	20.160	1.00	4.88
ATOM	56	C2	CYT	3	18.821	22.425	19.591	1.00	6.02
ATOM	57	O3'	CYT	3	23.657	19.902	18.677	1.00	14.33
ATOM	58	P	GUA	4	24.106	19.146	17.299	1.00	15.61
ATOM	59	O2P	GUA	4	25.478	18.627	17.465	1.00	13.09
ATOM	60	O1P	GUA	4	23.981	20.126	16.180	1.00	13.35
ATOM	61	O5'	GUA	4	22.936	18.091	17.233	1.00	10.11
ATOM	62	C5'	GUA	4	22.564	17.383	18.432	1.00	9.38
ATOM	63	O4'	GUA	4	20.164	17.563	18.226	1.00	10.12
ATOM	64	C4'	GUA	4	21.260	16.676	18.167	1.00	11.76
ATOM	65	C3'	GUA	4	21.167	15.976	16.789	1.00	13.48
ATOM	66	C2'	GUA	4	20.348	16.960	15.975	1.00	8.92
ATOM	67	C1'	GUA	4	19.397	17.535	17.014	1.00	9.91
ATOM	68	N9	GUA	4	18.963	18.899	16.670	1.00	9.68
ATOM	69	C8	GUA	4	19.689	19.874	16.021	1.00	9.73
ATOM	70	N7	GUA	4	19.061	21.001	15.855	1.00	10.45
ATOM	71	C5	GUA	4	17.831	20.770	16.465	1.00	6.01
ATOM	72	O6	GUA	4	16.640	22.797	16.226	1.00	9.35
ATOM	73	C6	GUA	4	16.718	21.633	16.624	1.00	7.89
ATOM	74	N1	GUA	4	15.662	21.061	17.273	1.00	8.63
ATOM	75	N2	GUA	4	14.571	19.390	18.346	1.00	8.22
ATOM	76	C2	GUA	4	15.696	19.778	17.736	1.00	6.41
ATOM	77	N3	GUA	4	16.716	18.923	17.617	1.00	7.32
ATOM	78	C4	GUA	4	17.757	19.490	16.968	1.00	8.16
ATOM	79	O3'	GUA	4	20.495	14.725	16.975	1.00	20.44
ATOM	80	P	ADE	5	20.201	13.469	16.101	1.00	24.01
ATOM	81	O1P	ADE	5	20.914	13.665	14.809	1.00	23.92
ATOM	82	O2P	ADE	5	20.583	12.234	16.869	1.00	26.06
ATOM	83	O5'	ADE	5	18.579	13.493	15.869	1.00	27.44
ATOM	84	C5'	ADE	5	17.743	13.321	17.034	1.00	20.19
ATOM	85	O4'	ADE	5	16.135	15.032	16.524	1.00	16.07
ATOM	86	C4'	ADE	5	16.302	13.641	16.697	1.00	16.20
ATOM	87	C3'	ADE	5	15.780	12.990	15.405	1.00	17.20
ATOM	88	C2'	ADE	5	15.868	14.129	14.385	1.00	13.45
ATOM	89	C1'	ADE	5	15.566	15.348	15.246	1.00	9.37
ATOM	90	N9	ADE	5	16.142	16.580	14.683	1.00	8.26
ATOM	91	C8	ADE	5	17.390	16.804	14.147	1.00	5.50
ATOM	92	N7	ADE	5	17.566	18.027	13.716	1.00	5.45
ATOM	93	C5	ADE	5	16.368	18.667	13.994	1.00	5.06
ATOM	94	C6	ADE	5	15.946	19.998	13.782	1.00	5.11
ATOM	95	N6	ADE	5	16.659	20.970	13.233	1.00	2.90
ATOM	96	N1	ADE	5	14.672	20.262	14.193	1.00	2.04
ATOM	97	C2	ADE	5	13.890	19.310	14.776	1.00	4.21
ATOM	98	N3	ADE	5	14.238	18.051	15.008	1.00	5.42
ATOM	99	C4	ADE	5	15.490	17.791	14.590	1.00	2.00
ATOM	100	O3'	ADE	5	14.436	12.558	15.610	1.00	16.05
ATOM	101	P	ADE	6	13.510	11.502	14.941	1.00	16.96
ATOM	102	O1P	ADE	6	14.272	10.907	13.816	1.00	18.99
ATOM	103	O2P	ADE	6	13.091	10.463	15.955	1.00	18.17
ATOM	104	O5'	ADE	6	12.243	12.338	14.458	1.00	15.56

ATOM	105	C5'	ADE	6	11.863	13.541	15.153	1.00	13.08
ATOM	106	O4'	ADE	6	11.674	15.556	13.902	1.00	17.19
ATOM	107	C4'	ADE	6	10.961	14.413	14.319	1.00	14.57
ATOM	108	C3'	ADE	6	10.395	13.773	13.034	1.00	14.21
ATOM	109	C2'	ADE	6	11.316	14.333	11.948	1.00	12.06
ATOM	110	C1'	ADE	6	11.620	15.716	12.471	1.00	11.37
ATOM	111	N9	ADE	6	12.868	16.316	11.968	1.00	7.92
ATOM	112	C8	ADE	6	14.115	15.792	11.782	1.00	9.28
ATOM	113	N7	ADE	6	14.993	16.652	11.325	1.00	10.73
ATOM	114	C5	ADE	6	14.272	17.835	11.206	1.00	8.87
ATOM	115	C6	ADE	6	14.652	19.126	10.776	1.00	8.40
ATOM	116	N6	ADE	6	15.865	19.478	10.372	1.00	3.48
ATOM	117	N1	ADE	6	13.650	20.054	10.802	1.00	7.66
ATOM	118	C2	ADE	6	12.387	19.758	11.213	1.00	8.10
ATOM	119	N3	ADE	6	11.976	18.567	11.630	1.00	12.43
ATOM	120	C4	ADE	6	12.968	17.651	11.610	1.00	8.02
ATOM	121	O3'	ADE	6	9.027	14.181	12.895	1.00	12.03
ATOM	122	P	THY	7	7.953	13.685	11.802	1.00	16.06
ATOM	123	O1P	THY	7	8.711	12.822	10.875	1.00	14.90
ATOM	124	O2P	THY	7	6.841	12.954	12.517	1.00	12.72
ATOM	125	O5'	THY	7	7.473	15.028	11.100	1.00	11.68
ATOM	126	C5'	THY	7	7.990	16.296	11.531	1.00	13.12
ATOM	127	O4'	THY	7	9.108	17.891	10.140	1.00	13.60
ATOM	128	C4'	THY	7	7.843	17.367	10.497	1.00	7.30
ATOM	129	C3'	THY	7	7.179	16.932	9.173	1.00	15.26
ATOM	130	C2'	THY	7	8.375	16.768	8.226	1.00	10.30
ATOM	131	C1'	THY	7	9.326	17.839	8.716	1.00	10.68
ATOM	132	N1	THY	7	10.748	17.575	8.418	1.00	9.71
ATOM	133	C6	THY	7	11.400	16.376	8.491	1.00	6.73
ATOM	134	C5	THY	7	12.701	16.232	8.199	1.00	8.19
ATOM	135	C4	THY	7	13.446	17.387	7.809	1.00	7.15
ATOM	136	O4	THY	7	14.674	17.379	7.530	1.00	7.37
ATOM	137	N3	THY	7	12.784	18.571	7.749	1.00	8.86
ATOM	138	C2	THY	7	11.461	18.719	8.047	1.00	10.76
ATOM	139	O2	THY	7	10.931	19.826	7.974	1.00	10.95
ATOM	140	M5	THY	7	13.407	14.901	8.285	1.00	6.07
ATOM	141	O3'	THY	7	6.235	17.935	8.789	1.00	17.73
ATOM	142	P	THY	8	5.032	17.703	7.756	1.00	19.00
ATOM	143	O1P	THY	8	5.306	16.428	7.060	1.00	19.62
ATOM	144	O2P	THY	8	3.674	17.771	8.358	1.00	22.69
ATOM	145	O5'	THY	8	5.265	18.990	6.795	1.00	22.69
ATOM	146	O4'	THY	8	7.878	20.590	6.073	1.00	18.15
ATOM	147	C5'	THY	8	5.831	20.206	7.325	1.00	19.10
ATOM	148	C4'	THY	8	6.532	20.982	6.239	1.00	18.04
ATOM	149	C3'	THY	8	5.868	20.850	4.848	1.00	22.07
ATOM	150	C2'	THY	8	6.814	19.886	4.113	1.00	20.76
ATOM	151	C1'	THY	8	8.150	20.254	4.689	1.00	17.16
ATOM	152	N1	THY	8	9.253	19.278	4.623	1.00	18.55
ATOM	153	C6	THY	8	9.184	17.947	4.868	1.00	15.99
ATOM	154	C5	THY	8	10.262	17.151	4.822	1.00	14.61
ATOM	155	C4	THY	8	11.522	17.743	4.517	1.00	16.15
ATOM	156	O4	THY	8	12.605	17.095	4.444	1.00	15.62
ATOM	157	N3	THY	8	11.598	19.082	4.285	1.00	14.16
ATOM	158	C2	THY	8	10.488	19.874	4.331	1.00	16.78
ATOM	159	O2	THY	8	10.593	21.085	4.113	1.00	17.58
ATOM	160	M5	THY	8	10.169	15.672	5.080	1.00	14.98

ATOM	161	O3'	THY	8	5.735	22.137	4.252	1.00	24.79
ATOM	162	P	CYT	9	4.900	22.457	2.907	1.00	24.41
ATOM	163	O1P	CYT	9	4.735	21.149	2.219	1.00	25.88
ATOM	164	O2P	CYT	9	3.603	23.120	3.245	1.00	29.83
ATOM	165	O5'	CYT	9	5.836	23.452	2.113	1.00	25.03
ATOM	166	C5'	CYT	9	7.037	23.940	2.735	1.00	25.00
ATOM	167	O4'	CYT	9	9.162	23.532	1.748	1.00	27.53
ATOM	168	C4'	CYT	9	8.042	24.400	1.729	1.00	24.78
ATOM	169	C3'	CYT	9	7.593	24.456	0.258	1.00	26.19
ATOM	170	C2'	CYT	9	8.137	23.168	-0.338	1.00	25.71
ATOM	171	C1'	CYT	9	9.434	22.996	0.444	1.00	22.09
ATOM	172	N1	CYT	9	9.775	21.569	0.583	1.00	16.64
ATOM	173	C6	CYT	9	8.855	20.594	0.881	1.00	18.29
ATOM	174	C5	CYT	9	9.152	19.302	1.020	1.00	17.41
ATOM	175	N4	CYT	9	10.880	17.671	1.027	1.00	17.32
ATOM	176	C4	CYT	9	10.520	18.947	0.887	1.00	15.49
ATOM	177	N3	CYT	9	11.446	19.906	0.629	1.00	15.38
ATOM	178	O2	CYT	9	11.954	22.077	0.225	1.00	19.97
ATOM	179	C2	CYT	9	11.108	21.221	0.470	1.00	16.49
ATOM	180	O3'	CYT	9	8.211	25.595	-0.364	1.00	28.34
ATOM	181	P	GUA	10	7.544	26.367	-1.596	1.00	33.00
ATOM	182	O1P	GUA	10	6.400	25.555	-2.093	1.00	31.75
ATOM	183	O2P	GUA	10	7.123	27.714	-1.060	1.00	32.03
ATOM	184	O5'	GUA	10	8.696	26.443	-2.689	1.00	32.49
ATOM	185	C5'	GUA	10	10.088	26.499	-2.325	1.00	32.36
ATOM	186	O4'	GUA	10	11.164	24.476	-3.053	1.00	27.82
ATOM	187	C4'	GUA	10	10.958	25.839	-3.371	1.00	30.31
ATOM	188	C3'	GUA	10	10.397	25.855	-4.802	1.00	29.13
ATOM	189	C2'	GUA	10	9.792	24.468	-4.967	1.00	25.75
ATOM	190	C1'	GUA	10	10.762	23.636	-4.159	1.00	21.65
ATOM	191	N9	GUA	10	10.233	22.345	-3.702	1.00	18.51
ATOM	192	C8	GUA	10	8.976	21.917	-3.384	1.00	18.43
ATOM	193	N7	GUA	10	8.900	20.666	-3.013	1.00	15.89
ATOM	194	C5	GUA	10	10.226	20.234	-3.080	1.00	16.33
ATOM	195	O6	GUA	10	10.191	17.979	-2.398	1.00	13.02
ATOM	196	C6	GUA	10	10.814	18.971	-2.788	1.00	15.51
ATOM	197	N1	GUA	10	12.172	18.927	-2.987	1.00	10.22
ATOM	198	N2	GUA	10	14.206	19.810	-3.550	1.00	10.59
ATOM	199	C2	GUA	10	12.887	19.994	-3.404	1.00	10.02
ATOM	200	N3	GUA	10	12.390	21.197	-3.682	1.00	15.29
ATOM	201	C4	GUA	10	11.054	21.245	-3.504	1.00	14.38
ATOM	202	O3'	GUA	10	11.485	25.987	-5.729	1.00	35.28
ATOM	203	P	CYT	11	11.858	27.142	-6.702	1.00	36.48
ATOM	204	O1P	CYT	11	11.030	26.855	-7.914	1.00	32.49
ATOM	205	O2P	CYT	11	11.601	28.502	-6.133	1.00	39.42
ATOM	206	O5'	CYT	11	13.441	26.923	-6.875	1.00	37.35
ATOM	207	C5'	CYT	11	14.287	26.951	-5.702	1.00	36.05
ATOM	208	O4'	CYT	11	14.704	24.600	-5.610	1.00	33.13
ATOM	209	C4'	CYT	11	15.333	25.851	-5.808	1.00	34.67
ATOM	210	C3'	CYT	11	16.044	25.771	-7.159	1.00	32.94
ATOM	211	C2'	CYT	11	15.280	24.656	-7.881	1.00	31.68
ATOM	212	C1'	CYT	11	14.941	23.736	-6.736	1.00	25.76
ATOM	213	N1	CYT	11	13.750	22.913	-6.981	1.00	23.48
ATOM	214	C6	CYT	11	12.537	23.396	-7.371	1.00	24.83
ATOM	215	C5	CYT	11	11.468	22.609	-7.570	1.00	25.74
ATOM	216	N4	CYT	11	10.613	20.374	-7.524	1.00	22.11

ATOM	217	C4	CYT	11	11.632	21.213	-7.338	1.00	22.11
ATOM	218	N3	CYT	11	12.838	20.746	-6.934	1.00	20.95
ATOM	219	O2	CYT	11	14.988	21.061	-6.391	1.00	21.74
ATOM	220	C2	CYT	11	13.919	21.561	-6.749	1.00	21.31
ATOM	221	O3'	CYT	11	17.427	25.399	-6.994	1.00	32.89
ATOM	222	P	GUA	12	18.530	25.671	-8.146	1.00	38.12
ATOM	223	O1P	GUA	12	17.976	26.751	-9.001	1.00	31.58
ATOM	224	O2P	GUA	12	19.833	25.931	-7.471	1.00	34.65
ATOM	225	O5'	GUA	12	18.522	24.272	-8.987	1.00	33.66
ATOM	226	C5'	GUA	12	18.985	23.132	-8.232	1.00	34.90
ATOM	227	O4'	GUA	12	17.907	21.217	-9.133	1.00	26.72
ATOM	228	C4'	GUA	12	19.142	21.913	-9.107	1.00	28.75
ATOM	229	C3'	GUA	12	19.534	22.165	-10.570	1.00	29.77
ATOM	230	C2'	GUA	12	18.189	22.061	-11.299	1.00	26.57
ATOM	231	C1'	GUA	12	17.463	21.009	-10.484	1.00	23.07
ATOM	232	N9	GUA	12	15.995	21.093	-10.544	1.00	18.29
ATOM	233	C8	GUA	12	15.174	22.157	-10.802	1.00	18.99
ATOM	234	N7	GUA	12	13.900	21.881	-10.776	1.00	16.87
ATOM	235	C5	GUA	12	13.870	20.526	-10.464	1.00	18.03
ATOM	236	O6	GUA	12	11.586	19.918	-10.365	1.00	17.46
ATOM	237	C6	GUA	12	12.780	19.638	-10.279	1.00	16.54
ATOM	238	N1	GUA	12	13.132	18.347	-9.974	1.00	16.11
ATOM	239	N2	GUA	12	14.608	16.668	-9.564	1.00	15.03
ATOM	240	C2	GUA	12	14.434	17.963	-9.862	1.00	15.20
ATOM	241	N3	GUA	12	15.495	18.755	-10.021	1.00	17.89
ATOM	242	C4	GUA	12	15.147	20.026	-10.312	1.00	17.67
ATOM	243	O3'	GUA	12	20.463	21.165	-11.034	1.00	27.97
ATOM	244	O5'	CYT	13	7.267	10.079	-9.027	1.00	24.88
ATOM	245	C5'	CYT	13	8.211	10.647	-9.961	1.00	22.75
ATOM	246	O4'	CYT	13	10.159	11.994	-9.842	1.00	18.13
ATOM	247	C4'	CYT	13	9.588	10.799	-9.352	1.00	17.82
ATOM	248	C3'	CYT	13	9.659	10.907	-7.822	1.00	22.13
ATOM	249	C2'	CYT	13	9.753	12.418	-7.577	1.00	17.43
ATOM	250	C1'	CYT	13	10.556	12.878	-8.769	1.00	15.61
ATOM	251	N1	CYT	13	10.228	14.253	-9.173	1.00	15.74
ATOM	252	C6	CYT	13	8.963	14.617	-9.557	1.00	16.21
ATOM	253	C5	CYT	13	8.642	15.848	-9.961	1.00	16.68
ATOM	254	N4	CYT	13	9.395	18.047	-10.425	1.00	15.77
ATOM	255	C4	CYT	13	9.686	16.812	-10.014	1.00	15.91
ATOM	256	N3	CYT	13	10.946	16.448	-9.656	1.00	17.10
ATOM	257	O2	CYT	13	12.400	14.881	-8.941	1.00	15.80
ATOM	258	C2	CYT	13	11.250	15.180	-9.246	1.00	15.26
ATOM	259	O3'	CYT	13	10.831	10.235	-7.371	1.00	25.39
ATOM	260	P	GUA	14	11.218	9.219	-6.239	1.00	29.53
ATOM	261	O1P	GUA	14	10.544	9.739	-5.014	1.00	23.15
ATOM	262	O2P	GUA	14	10.723	7.808	-6.497	1.00	28.79
ATOM	263	O5'	GUA	14	12.824	9.219	-6.259	1.00	22.22
ATOM	264	C5'	GUA	14	13.537	9.955	-7.272	1.00	21.34
ATOM	265	O4'	GUA	14	14.238	12.242	-6.934	1.00	21.03
ATOM	266	C4'	GUA	14	14.561	10.887	-6.676	1.00	21.38
ATOM	267	C3'	GUA	14	14.748	10.783	-5.153	1.00	20.63
ATOM	268	C2'	GUA	14	13.971	11.982	-4.623	1.00	20.50
ATOM	269	C1'	GUA	14	14.196	13.005	-5.716	1.00	20.42
ATOM	270	N9	GUA	14	13.145	14.037	-5.702	1.00	19.39
ATOM	271	C8	GUA	14	11.794	13.889	-5.517	1.00	20.59
ATOM	272	N7	GUA	14	11.118	15.000	-5.543	1.00	20.00

ATOM	273	C5	GUA	14	12.096	15.960	-5.782	1.00	19.57
ATOM	274	O6	GUA	14	10.968	18.063	-5.881	1.00	20.36
ATOM	275	C6	GUA	14	11.983	17.367	-5.934	1.00	19.50
ATOM	276	N1	GUA	14	13.177	17.987	-6.146	1.00	18.05
ATOM	277	N2	GUA	14	15.390	18.163	-6.457	1.00	16.45
ATOM	278	C2	GUA	14	14.365	17.339	-6.232	1.00	15.68
ATOM	279	N3	GUA	14	14.520	16.024	-6.100	1.00	18.30
ATOM	280	C4	GUA	14	13.346	15.392	-5.881	1.00	18.24
ATOM	281	O3'	GUA	14	16.135	10.879	-4.822	1.00	20.92
ATOM	282	P	CYT	15	16.789	10.207	-3.504	1.00	15.98
ATOM	283	O1P	CYT	15	15.677	9.795	-2.649	1.00	16.73
ATOM	284	O2P	CYT	15	17.694	9.115	-3.987	1.00	15.23
ATOM	285	O5'	CYT	15	17.642	11.446	-2.921	1.00	13.28
ATOM	286	C5'	CYT	15	18.490	12.094	-3.894	1.00	10.00
ATOM	287	O4'	CYT	15	17.436	14.317	-3.775	1.00	11.19
ATOM	288	C4'	CYT	15	18.613	13.561	-3.557	1.00	10.08
ATOM	289	C3'	CYT	15	19.005	13.797	-2.093	1.00	8.73
ATOM	290	C2'	CYT	15	17.686	14.273	-1.464	1.00	9.02
ATOM	291	C1'	CYT	15	17.027	15.036	-2.609	1.00	8.27
ATOM	292	N1	CYT	15	15.571	15.128	-2.378	1.00	11.04
ATOM	293	C6	CYT	15	14.799	14.105	-1.901	1.00	10.07
ATOM	294	C5	CYT	15	13.503	14.197	-1.643	1.00	10.19
ATOM	295	N4	CYT	15	11.576	15.588	-1.689	1.00	13.95
ATOM	296	C4	CYT	15	12.887	15.440	-1.914	1.00	11.45
ATOM	297	N3	CYT	15	13.625	16.476	-2.398	1.00	8.90
ATOM	298	O2	CYT	15	15.623	17.311	-3.066	1.00	9.78
ATOM	299	C2	CYT	15	14.963	16.360	-2.636	1.00	10.20
ATOM	300	O3'	CYT	15	20.025	14.781	-1.907	1.00	12.26
ATOM	301	P	GUA	16	20.917	14.625	-0.536	1.00	11.00
ATOM	302	O1P	GUA	16	19.976	14.121	0.497	1.00	13.72
ATOM	303	O2P	GUA	16	22.059	13.685	-0.748	1.00	14.58
ATOM	304	O5'	GUA	16	21.385	16.144	-0.351	1.00	14.12
ATOM	305	C5'	GUA	16	21.659	16.988	-1.503	1.00	14.11
ATOM	306	O4'	GUA	16	19.762	18.463	-1.192	1.00	11.86
ATOM	307	C4'	GUA	16	21.167	18.391	-1.205	1.00	14.18
ATOM	308	C3'	GUA	16	21.650	18.875	0.179	1.00	14.33
ATOM	309	C2'	GUA	16	20.412	18.643	1.046	1.00	12.53
ATOM	310	C1'	GUA	16	19.280	18.919	0.079	1.00	8.82
ATOM	311	N9	GUA	16	18.071	18.187	0.477	1.00	8.02
ATOM	312	C8	GUA	16	17.939	16.920	1.007	1.00	8.80
ATOM	313	N7	GUA	16	16.711	16.560	1.252	1.00	9.39
ATOM	314	C5	GUA	16	15.961	17.655	0.821	1.00	8.66
ATOM	315	O6	GUA	16	13.674	17.107	1.192	1.00	9.58
ATOM	316	C6	GUA	16	14.559	17.879	0.815	1.00	9.34
ATOM	317	N1	GUA	16	14.204	19.102	0.325	1.00	10.11
ATOM	318	N2	GUA	16	14.564	21.161	-0.576	1.00	10.32
ATOM	319	C2	GUA	16	15.101	20.018	-0.119	1.00	8.01
ATOM	320	N3	GUA	16	16.422	19.866	-0.132	1.00	9.12
ATOM	321	C4	GUA	16	16.782	18.659	0.351	1.00	11.23
ATOM	322	O3'	GUA	16	22.027	20.246	0.066	1.00	16.29
ATOM	323	P	ADE	17	22.537	21.221	1.219	1.00	17.26
ATOM	324	O1P	ADE	17	22.735	20.354	2.398	1.00	16.51
ATOM	325	O2P	ADE	17	23.679	22.029	0.729	1.00	17.00
ATOM	326	O5'	ADE	17	21.199	22.101	1.291	1.00	19.28
ATOM	327	C5'	ADE	17	20.873	22.849	0.079	1.00	20.04
ATOM	328	O4'	ADE	17	18.606	23.000	0.762	1.00	21.30

ATOM	329	C4'	ADE	17	19.745	23.776	0.470	1.00	21.47
ATOM	330	C3'	ADE	17	20.034	24.636	1.715	1.00	23.23
ATOM	331	C2'	ADE	17	19.331	23.832	2.815	1.00	18.58
ATOM	332	C1'	ADE	17	18.137	23.252	2.093	1.00	16.79
ATOM	333	N9	ADE	17	17.669	21.989	2.689	1.00	15.36
ATOM	334	C8	ADE	17	18.417	20.938	3.146	1.00	11.58
ATOM	335	N7	ADE	17	17.713	19.946	3.610	1.00	15.17
ATOM	336	C5	ADE	17	16.397	20.362	3.431	1.00	13.36
ATOM	337	C6	ADE	17	15.169	19.726	3.729	1.00	14.10
ATOM	338	N6	ADE	17	15.115	18.519	4.272	1.00	16.09
ATOM	339	N1	ADE	17	14.037	20.434	3.424	1.00	13.74
ATOM	340	C2	ADE	17	14.120	21.669	2.855	1.00	12.55
ATOM	341	N3	ADE	17	15.248	22.317	2.543	1.00	14.69
ATOM	342	C4	ADE	17	16.353	21.613	2.861	1.00	13.73
ATOM	343	O3'	ADE	17	19.510	25.951	1.490	1.00	28.21
ATOM	344	P	ADE	18	19.174	27.166	2.457	1.00	34.53
ATOM	345	O1P	ADE	18	20.211	27.202	3.523	1.00	32.12
ATOM	346	O2P	ADE	18	18.978	28.426	1.676	1.00	34.86
ATOM	347	O5'	ADE	18	17.780	26.611	3.033	1.00	33.51
ATOM	348	C5'	ADE	18	16.576	26.827	2.252	1.00	29.55
ATOM	349	O4'	ADE	18	15.157	25.579	3.696	1.00	23.67
ATOM	350	C4'	ADE	18	15.422	26.883	3.239	1.00	26.50
ATOM	351	C3'	ADE	18	15.704	27.758	4.471	1.00	26.39
ATOM	352	C2'	ADE	18	16.125	26.711	5.510	1.00	21.64
ATOM	353	C1'	ADE	18	15.270	25.523	5.133	1.00	18.22
ATOM	354	N9	ADE	18	15.868	24.236	5.543	1.00	14.97
ATOM	355	C8	ADE	18	17.164	23.812	5.610	1.00	12.45
ATOM	356	N7	ADE	18	17.319	22.585	6.027	1.00	14.43
ATOM	357	C5	ADE	18	16.015	22.149	6.252	1.00	13.58
ATOM	358	C6	ADE	18	15.505	20.918	6.709	1.00	10.52
ATOM	359	N6	ADE	18	16.182	19.826	7.040	1.00	8.26
ATOM	360	N1	ADE	18	14.145	20.870	6.802	1.00	11.72
ATOM	361	C2	ADE	18	13.358	21.917	6.484	1.00	15.89
ATOM	362	N3	ADE	18	13.765	23.104	6.047	1.00	14.31
ATOM	363	C4	ADE	18	15.115	23.152	5.961	1.00	12.79
ATOM	364	O3'	ADE	18	14.520	28.482	4.828	1.00	28.99
ATOM	365	P	THY	19	14.204	29.469	6.040	1.00	35.61
ATOM	366	O1P	THY	19	15.483	29.813	6.702	1.00	33.55
ATOM	367	O2P	THY	19	13.382	30.625	5.563	1.00	32.11
ATOM	368	O5'	THY	19	13.309	28.494	7.020	1.00	31.56
ATOM	369	C5'	THY	19	12.174	27.934	6.325	1.00	25.26
ATOM	370	O4'	THY	19	11.939	25.815	7.431	1.00	16.72
ATOM	371	C4'	THY	19	11.346	27.090	7.259	1.00	21.99
ATOM	372	C3'	THY	19	11.145	27.682	8.670	1.00	18.06
ATOM	373	C2'	THY	19	12.150	26.887	9.497	1.00	15.43
ATOM	374	C1'	THY	19	12.130	25.527	8.822	1.00	10.37
ATOM	375	N1	THY	19	13.348	24.748	9.093	1.00	7.06
ATOM	376	C6	THY	19	14.625	25.215	9.034	1.00	7.92
ATOM	377	C5	THY	19	15.696	24.472	9.325	1.00	6.98
ATOM	378	C4	THY	19	15.488	23.112	9.689	1.00	5.63
ATOM	379	O4	THY	19	16.417	22.309	9.968	1.00	4.29
ATOM	380	N3	THY	19	14.211	22.649	9.736	1.00	8.89
ATOM	381	C2	THY	19	13.120	23.420	9.458	1.00	6.29
ATOM	382	O2	THY	19	11.983	22.961	9.511	1.00	2.00
ATOM	383	M5	THY	19	17.081	25.059	9.239	1.00	8.21
ATOM	384	O3'	THY	19	9.806	27.482	9.146	1.00	20.36

ATOM	385	P	THY	20	9.184	28.338	10.378	1.00	22.82
ATOM	386	O1P	THY	20	10.120	29.413	10.756	1.00	24.22
ATOM	387	O2P	THY	20	7.833	28.794	9.948	1.00	19.91
ATOM	388	O5'	THY	20	9.140	27.206	11.564	1.00	17.45
ATOM	389	C5'	THY	20	8.265	26.151	11.107	1.00	11.58
ATOM	390	O4'	THY	20	9.456	24.240	12.027	1.00	5.76
ATOM	391	C4'	THY	20	8.260	24.995	12.054	1.00	9.25
ATOM	392	C3'	THY	20	8.032	25.383	13.531	1.00	8.59
ATOM	393	C2'	THY	20	9.424	25.259	14.140	1.00	10.94
ATOM	394	C1'	THY	20	10.005	24.108	13.352	1.00	7.98
ATOM	395	N1	THY	20	11.478	24.076	13.286	1.00	8.06
ATOM	396	C6	THY	20	12.262	25.147	12.968	1.00	5.93
ATOM	397	C5	THY	20	13.591	25.079	12.902	1.00	5.83
ATOM	398	C4	THY	20	14.213	23.816	13.140	1.00	6.98
ATOM	399	O4	THY	20	15.463	23.632	13.087	1.00	8.82
ATOM	400	N3	THY	20	13.419	22.753	13.431	1.00	8.80
ATOM	401	C2	THY	20	12.059	22.829	13.511	1.00	7.47
ATOM	402	O2	THY	20	11.382	21.837	13.776	1.00	2.58
ATOM	403	M5	THY	20	14.402	26.307	12.570	1.00	7.19
ATOM	404	O3'	THY	20	7.101	24.452	14.074	1.00	13.03
ATOM	405	P	CYT	21	6.221	24.600	15.359	1.00	14.76
ATOM	406	O1P	CYT	21	7.022	25.475	16.286	1.00	17.41
ATOM	407	O2P	CYT	21	4.870	25.131	15.047	1.00	16.34
ATOM	408	O5'	CYT	21	6.125	23.096	15.955	1.00	19.26
ATOM	409	C5'	CYT	21	7.069	22.069	15.617	1.00	19.16
ATOM	410	O4'	CYT	21	9.017	21.285	16.703	1.00	17.66
ATOM	411	C4'	CYT	21	7.603	21.329	16.816	1.00	17.76
ATOM	412	C3'	CYT	21	7.294	21.953	18.187	1.00	19.29
ATOM	413	C2'	CYT	21	8.583	22.721	18.511	1.00	18.53
ATOM	414	C1'	CYT	21	9.647	21.849	17.869	1.00	15.55
ATOM	415	N1	CYT	21	10.838	22.601	17.445	1.00	14.86
ATOM	416	C6	CYT	21	10.865	23.916	17.067	1.00	12.29
ATOM	417	C5	CYT	21	11.963	24.560	16.683	1.00	14.14
ATOM	418	C4	CYT	21	13.179	23.820	16.630	1.00	12.50
ATOM	419	N4	CYT	21	14.292	24.444	16.240	1.00	7.94
ATOM	420	N3	CYT	21	13.167	22.509	16.981	1.00	14.34
ATOM	421	O2	CYT	21	12.044	20.686	17.703	1.00	16.22
ATOM	422	C2	CYT	21	12.020	21.873	17.385	1.00	14.64
ATOM	423	O3'	CYT	21	7.022	20.930	19.154	1.00	19.41
ATOM	424	P	GUA	22	6.257	21.201	20.538	1.00	17.78
ATOM	425	O1P	GUA	22	5.953	22.649	20.564	1.00	16.91
ATOM	426	O2P	GUA	22	5.059	20.314	20.651	1.00	17.29
ATOM	427	O5'	GUA	22	7.409	20.746	21.558	1.00	18.51
ATOM	428	C5'	GUA	22	7.990	19.438	21.346	1.00	17.59
ATOM	429	O4'	GUA	22	10.341	19.902	21.512	1.00	16.06
ATOM	430	C4'	GUA	22	9.248	19.302	22.174	1.00	13.90
ATOM	431	C3'	GUA	22	9.191	19.974	23.538	1.00	14.80
ATOM	432	C2'	GUA	22	9.880	21.309	23.333	1.00	12.77
ATOM	433	C1'	GUA	22	10.954	20.910	22.333	1.00	10.13
ATOM	434	N9	GUA	22	11.456	22.077	21.598	1.00	8.16
ATOM	435	C8	GUA	22	10.868	23.256	21.227	1.00	8.36
ATOM	436	N7	GUA	22	11.650	24.080	20.591	1.00	8.26
ATOM	437	C5	GUA	22	12.858	23.400	20.525	1.00	6.38
ATOM	438	O6	GUA	22	14.358	24.824	19.392	1.00	9.20
ATOM	439	C6	GUA	22	14.110	23.760	19.968	1.00	7.21
ATOM	440	N1	GUA	22	15.081	22.809	20.114	1.00	5.29

ATOM	441	N2	GUA	22	15.895	20.782	20.803	1.00	6.91
ATOM	442	C2	GUA	22	14.865	21.625	20.743	1.00	4.39
ATOM	443	N3	GUA	22	13.721	21.241	21.293	1.00	8.03
ATOM	444	C4	GUA	22	12.757	22.173	21.141	1.00	8.73
ATOM	445	O3'	GUA	22	10.017	19.234	24.459	1.00	16.27
ATOM	446	P	CYT	23	9.456	18.347	25.618	1.00	17.82
ATOM	447	O1P	CYT	23	8.728	19.334	26.446	1.00	16.64
ATOM	448	O2P	CYT	23	8.525	17.343	24.922	1.00	23.34
ATOM	449	O5'	CYT	23	10.743	17.623	26.167	1.00	14.21
ATOM	450	C5'	CYT	23	11.581	16.920	25.207	1.00	20.94
ATOM	451	O4'	CYT	23	13.537	18.223	24.750	1.00	18.54
ATOM	452	C4'	CYT	23	13.052	17.143	25.525	1.00	20.43
ATOM	453	C3'	CYT	23	13.309	17.463	27.009	1.00	18.13
ATOM	454	C2'	CYT	23	13.449	18.990	26.962	1.00	17.88
ATOM	455	C1'	CYT	23	14.088	19.238	25.598	1.00	13.38
ATOM	456	N1	CYT	23	13.824	20.630	25.181	1.00	11.86
ATOM	457	C6	CYT	23	12.618	21.273	25.247	1.00	10.05
ATOM	458	C5	CYT	23	12.441	22.549	24.902	1.00	8.87
ATOM	459	N4	CYT	23	13.409	24.532	24.055	1.00	5.49
ATOM	460	C4	CYT	23	13.579	23.260	24.419	1.00	9.28
ATOM	461	N3	CYT	23	14.782	22.637	24.353	1.00	8.61
ATOM	462	O2	CYT	23	16.042	20.802	24.631	1.00	11.57
ATOM	463	C2	CYT	23	14.934	21.325	24.724	1.00	8.37
ATOM	464	O3'	CYT	23	14.476	16.840	27.558	1.00	20.22
ATOM	465	P	GUA	24	14.544	16.708	29.207	1.00	23.97
ATOM	466	O1P	GUA	24	13.461	17.515	29.804	1.00	22.69
ATOM	467	O2P	GUA	24	14.574	15.292	29.638	1.00	22.98
ATOM	468	O5'	GUA	24	15.985	17.379	29.479	1.00	20.53
ATOM	469	C5'	GUA	24	17.042	16.792	28.678	1.00	23.17
ATOM	470	O4'	GUA	24	17.934	18.871	27.823	1.00	22.92
ATOM	471	C4'	GUA	24	18.199	17.763	28.671	1.00	24.14
ATOM	472	C3'	GUA	24	18.554	18.307	30.075	1.00	21.50
ATOM	473	C2'	GUA	24	17.936	19.706	30.015	1.00	23.15
ATOM	474	C1'	GUA	24	18.076	20.102	28.532	1.00	17.58
ATOM	475	N9	GUA	24	17.059	21.133	28.254	1.00	15.69
ATOM	476	C8	GUA	24	15.723	21.093	28.605	1.00	14.52
ATOM	477	N7	GUA	24	15.047	22.157	28.300	1.00	15.10
ATOM	478	C5	GUA	24	16.000	22.969	27.691	1.00	13.41
ATOM	479	O6	GUA	24	14.841	24.948	27.088	1.00	14.93
ATOM	480	C6	GUA	24	15.865	24.268	27.134	1.00	12.75
ATOM	481	N1	GUA	24	17.020	24.780	26.618	1.00	12.71
ATOM	482	N2	GUA	24	19.213	24.764	26.075	1.00	9.95
ATOM	483	C2	GUA	24	18.196	24.092	26.631	1.00	10.56
ATOM	484	N3	GUA	24	18.368	22.869	27.141	1.00	12.90
ATOM	485	C4	GUA	24	17.235	22.361	27.658	1.00	12.44
ATOM	486	O3'	GUA	24	19.961	18.363	30.333	1.00	27.39
ATOM	487	N12	BER	25	6.711	22.101	10.206	1.00	34.97
ATOM	488	C1	BER	25	7.990	23.656	7.378	1.00	34.59
ATOM	489	N36	BER	25	7.654	17.879	14.849	1.00	38.51
ATOM	490	C62	BER	25	6.390	20.758	12.133	1.00	35.06
ATOM	491	C3	BER	25	8.343	25.039	5.464	1.00	35.90
ATOM	492	C4	BER	25	9.591	24.476	5.245	1.00	36.49
ATOM	493	C12	BER	25	7.282	21.213	11.193	1.00	34.70
ATOM	494	C19	BER	25	10.400	24.868	4.206	1.00	36.14
ATOM	495	C42	BER	25	8.140	19.502	13.186	1.00	36.55
ATOM	496	C2	BER	25	7.547	24.644	6.504	1.00	34.28

ATOM	497	N32	BER	25	10.159	26.095	3.610	1.00	38.17
ATOM	498	C5	BER	25	9.993	23.480	6.140	1.00	32.87
ATOM	499	C32	BER	25	9.034	19.974	12.219	1.00	36.42
ATOM	500	C21	BER	25	8.576	18.651	14.167	1.00	36.36
ATOM	501	N2	BER	25	7.495	22.237	9.226	1.00	34.84
ATOM	502	C22	BER	25	8.603	20.818	11.233	1.00	35.85
ATOM	503	N20	BER	25	9.922	18.599	14.471	1.00	35.84
ATOM	504	N1	BER	25	7.145	23.268	8.431	1.00	37.19
ATOM	505	C52	BER	25	6.821	19.894	13.140	1.00	36.08
ATOM	506	C6	BER	25	9.228	23.068	7.193	1.00	33.74
ATOM	507	N10	BER	25	11.664	24.312	4.093	1.00	36.42
ATOM	508	O	WAT	26	16.152	26.779	-11.305	1.00	47.60
ATOM	509	O	WAT	27	12.993	10.439	-1.709	1.00	19.51
ATOM	510	O	WAT	28	9.380	12.966	8.113	1.00	16.79
ATOM	511	O	WAT	29	16.336	9.643	16.697	1.00	27.92
ATOM	512	O	WAT	30	12.194	22.401	29.413	1.00	23.49
ATOM	513	O	WAT	31	5.527	15.364	-8.179	1.00	42.34
ATOM	514	O	WAT	32	8.941	14.473	-3.828	1.00	23.05
ATOM	515	O	WAT	33	22.941	14.849	-4.471	1.00	12.26
ATOM	516	O	WAT	34	16.708	30.629	0.722	1.00	29.13
ATOM	517	O	WAT	35	5.127	27.614	9.921	1.00	23.84
ATOM	518	O	WAT	36	28.111	23.460	17.935	1.00	24.36
ATOM	519	O	WAT	37	13.240	18.159	21.101	1.00	7.85
ATOM	520	O	WAT	38	7.625	16.900	-6.179	1.00	36.55
ATOM	521	O	WAT	39	2.468	19.210	19.624	1.00	19.87
ATOM	522	O	WAT	40	14.909	15.420	22.737	1.00	15.17
ATOM	523	O	WAT	41	22.302	18.091	5.073	1.00	21.92
ATOM	524	O	WAT	42	4.025	21.969	13.802	1.00	37.70
ATOM	525	O	WAT	43	11.520	17.155	16.240	1.00	32.01
ATOM	526	O	WAT	44	24.478	32.108	24.181	1.00	32.58
ATOM	527	O	WAT	45	13.059	15.036	32.042	1.00	27.10
ATOM	528	O	WAT	46	5.914	9.343	-6.550	1.00	41.50
ATOM	529	O	WAT	47	18.534	7.956	14.624	1.00	26.75
ATOM	530	O	WAT	48	19.008	20.318	7.524	1.00	39.60
ATOM	531	O	WAT	49	8.544	15.320	1.444	1.00	29.28
ATOM	532	O	WAT	50	19.860	18.095	12.392	1.00	30.54
ATOM	533	O	WAT	51	4.571	24.116	10.868	1.00	32.38
ATOM	534	O	WAT	52	17.515	27.606	16.352	1.00	32.05
ATOM	535	O	WAT	53	12.978	29.085	-9.133	1.00	41.57
ATOM	536	O	WAT	54	16.865	13.757	1.881	1.00	51.92
ATOM	537	O	WAT	55	19.780	13.841	3.444	1.00	39.03
ATOM	538	O	WAT	56	2.730	19.866	10.133	1.00	46.73
ATOM	539	O	WAT	57	15.235	6.797	16.094	1.00	43.89
ATOM	540	O	WAT	58	23.334	11.530	17.948	1.00	44.29
ATOM	541	O	WAT	59	21.294	29.505	16.981	1.00	51.01
ATOM	542	O	WAT	60	27.319	21.073	16.491	1.00	61.99
ATOM	543	O	WAT	61	27.199	16.136	17.597	1.00	52.19
ATOM	544	O	WAT	62	10.767	26.827	20.141	1.00	24.75
ATOM	545	O	WAT	63	7.909	28.378	-4.921	1.00	43.83
ATOM	546	O	WAT	64	13.270	29.629	11.014	1.00	43.87
ATOM	547	O	WAT	65	27.454	29.473	18.657	1.00	40.79
ATOM	548	O	WAT	66	16.064	8.624	13.173	1.00	31.73
ATOM	549	O	WAT	67	17.314	30.829	27.300	1.00	21.22
ATOM	550	O	WAT	68	13.990	14.661	4.782	1.00	20.68
ATOM	551	O	WAT	69	17.309	16.204	9.961	1.00	30.56
ATOM	552	O	WAT	70	16.907	29.289	20.200	1.00	15.22

ATOM	553	O	WAT	71	19.000	30.481	18.564	1.00	52.53
ATOM	554	O	WAT	72	13.733	27.378	21.392	1.00	39.77
ATOM	555	O	WAT	73	14.145	29.721	19.803	1.00	21.89

---

The coordinates listed are angstrom coordinates for one asymmetric unit.

The unit cell is an orthogonal space group  $P2_12_12_1$  with dimensions  $a = 24.51$ ,  $b = 39.98$ ,  $c = 66.23$  Å.

Occ is the occupancy which was not refined.

B(Iso) is the isotropic temperature factor. Å<sup>2</sup>

---

## **CHAPTER 3.**

The crystal structure of a complex between the dodecamer  $d(\text{CGCAAATTTGCG})_2$  and the minor groove-binding ligand berenil.

### 3.1 Introduction.

As discussed earlier in Chapter 2 the anti-trypanosomal drug berenil is known to bind to AT rich regions of double stranded DNA via the minor groove. In an effort to further understand the nature of specific interactions between ligand and DNA on binding, and to further explore the role of water in these interactions, this drug was co-crystallised with the DNA sequence d(CGCAAATTTGCG)<sub>2</sub>. This DNA sequence presents a much wider variety of AT binding sites than the previously reported d(CGCGAATTCGCG)<sub>2</sub> oligonucleotide-berenil complex (Brown *et al.*, 1990; chapter 2), and fortunately for this study the crystals that were grown diffracted to higher resolution than in the previous berenil-DNA complex. The sequence d(CGCAAATTTGCG)<sub>2</sub> was also chosen as its structure had previously been reported, both as native and complexed with the minor groove binding ligand distamycin, to exhibit three-centred hydrogen bonds in the A<sub>3</sub>T<sub>3</sub> core region. (Coll *et al.*, 1987) It has been proposed that three-centred hydrogen bonds stabilise long stretches of poly(dA).poly(dT) (Nelson *et al.*, 1987).

### Materials and Methods.

#### 3.2 Oligonucleotide synthesis.

The dodecanucleotide d(CGCAAATTTGCG) was prepared on an Applied Biosystems 391EP synthesiser using phosphoramidite chemistry, and purified by anion-exchange and reverse-phase (C8) HPLC procedures. It was synthesised using 6 x 1 µmole scale columns and then purified in three stages by H P L C. Firstly the oligonucleotide was purified by anion exchange chromatography on a Spherisorb SAX (5µm packing), a semi-preparative strong anion exchange column, using a gradient of 0.8 M potassium phosphate buffer (pH 6.4) against 20% acetonitrile. The detritylated sample

was taken up in a 50/50 volume of formamide and distilled water and left for 30 minutes at room temperature to prevent any secondary structure formation, which can be a problem with self-complementary sequences. The flow rate used was 2.5 ml per minute with the linear gradient of the 0.8 M potassium phosphate buffer applied from 0-100% over 35 minutes. The desired product was eluted from the column after 24 minutes of the gradient. The appropriate sample was collected and freeze dried (lyophilised) on a speedvac. The sample was then redissolved in ammonium acetate buffer and purified on a Brownlee semi-preparative octyl (C8) reverse phase column using a gradient of 80% acetonitrile (Buffer A) vs. 0.1 M ammonium acetate pH 7.0 (Buffer B) and a flow rate of 2 ml/minute going from 0 - 100% buffer B over 40 minutes. The product eluted as a single peak after 30 minutes. The desired fraction was freeze dried on a speedvac, then redissolved in a small volume of distilled water and desalted using a PD10 cartridge. The self complementary sequence was then annealed by slow cooling from 70°C and re-run down the reverse phase column to check purity and ensure no "unusual" loop or hairpin structures were present. The product was collected, concentrated, desalted and then freeze-dried to a fluffy white powder.

Berenil, was synthesised by Dr. T.C. Jenkins, as the di(N-acetylglycinate) salt, as described by Lane *et al.*, 1991 and was homogeneous (>98.5% w/w) by HPLC.

### 3.3 Crystallisation and data collection.

Crystals of the complex were grown using the hanging droplet vapour diffusion method under similar conditions to those of the complex of berenil and d(CGCGAATTCGCG)<sub>2</sub>. Optimum conditions for crystal growth were with 17µl droplets containing 3 µl of 15 mM berenil, 5 µl of 20 mg/ml DNA, 5 µl of 2-methylpentan-2,4-diol (MPD) at 35% vol/vol (MPD/sodium cacodylate buffer pH 7.0 containing 0.02% azide) and 4µl of 100 mM MgCl<sub>2</sub> against 35% vol/vol

MPD/sodium cacodylate buffer pH 7.0. All solutions were filtered through a 2  $\mu\text{m}$  filter. The droplets were set up at room temperature (18°C) and solutions added in the order buffer, MPD, DNA and finally berenil. The addition of berenil caused a small amount of precipitation which was easily redissolved by mixing the droplet with the aid of mixing mode of an EDP-PLUS automatic syringe. The same syringe was used for the addition of all the small volumes as it was found to be the quickest method for setting up of large amounts of screening trials and also gave the most reproducible results. The trays were then stored in an incubator (free from vibration) at 16°C. The crystals used for data collection grew in a period of 10-14 days. The crystals were long and needle-like, with typical dimensions of 0.1 mm x 0.1 mm x 0.6 mm and were yellow in colour.

Intensity data were collected on a Xentronics area detector, at the Laboratory of Molecular Biology, Oxford University with the kind assistance of Dr. Elspeth Garman and Dr. Mark Sanderson, using a Rigaku RU200 rotating anode source with the power at 50 mA / 50 KV. The crystal to detector distance was 8cm and the  $2\theta$  swing angle  $12^\circ$  with a frame collection time of 200 seconds. Two  $100^\circ \omega$  scans were carried out with  $\phi$  rotated  $60^\circ$  between them to obtain the unique data set (as the crystal had been identified by auto-indexing, using the Xengen software (Howard *et al.*, 1987), to be orthorhombic with space group  $P2_12_12_1$ ). This methodology scans the entire quarter of the reciprocal sphere needed for the unique data set in the space group  $P2_12_12_1$  with a degree of redundancy. A total of 12,415 of a possible 14,737 observations were collected to a resolution of  $2\text{\AA}$ . These were scaled and merged using the Xengen software package to give 4,067 unique reflections out of a possible 4,530 with a merging R factor of 6.24%.

Table 3.1. Resolution breakdown of data of collected.

Resolution limit Å	Number of Reflections		R-Factor <sup>a</sup>
	Possible	Collected	
3.69	814	812	1.95
2.93	767	767	4.04
2.56	749	749	11.02
2.33	745	743	20.08
2.16	729	690	30.00
2.03	726	306	34.47
	<u>4530</u>	<u>4067</u>	<u>6.24</u>

a. Unweighted absolute R-factor on  $F^*100$

### 3.4 Structure solution and refinement.

As the unit cell of this crystal  $a=24.64$ ,  $b=40.61$ ,  $c=65.07$  Å was highly isomorphous with that of the same sequence complexed with another minor groove binding drug distamycin ( $a=25.20$ ,  $b=41.07$ ,  $c=64.65$  Å); it was decided to use the coordinates of the DNA duplex from the co-crystal structure deposited in the Brookhaven Data Bank (Bernstein *et al.*, 1977) as a model to solve the berenil dodecamer complex by molecular replacement. The coordinates extracted from the data bank were already orthogonalised and simply had to be reformatted and converted into fractional coordinates for refinement.

It was decided to use the constrained/restrained refinement program CORELS (Sussman *et al.*, 1977) for the initial stages of refinement. CORELS allows one to perform rigid body refinement, thereby minimising the number of variable parameters in the refinement to eight ( $x,y,z$ , three angles, temperature factor and scale factor). Therefore one can use the medium to low resolution data to gain a large radius of convergence for the solution, whilst

keeping the ratio of the number of observations to refinable parameters high. Initially reflection data from 10-4 Å (542 unique reflections) were used together with the fractionalised, reformatted coordinates extracted from the Brookhaven Data Bank as a starting model which yielded an R factor of 39.2% after scaling and 5 cycles of positional refinement. At this stage, to ensure that the model was essentially correct, the first  $|2F_o - F_c|_{\alpha_{calc}}$  map was calculated. The PROTEIN (Steigemann, 1987) suite was used to calculate the maps which were viewed on a Silicon Graphics Iris 3130 using the TOM display package (Cambillau, 1988). Both  $|2F_o - F_c|_{\alpha_{calc}}$  and difference ( $|F_o - F_c|_{\alpha_{calc}}$ ) maps already showed density of comparable size and shape to that of a berenil molecule in the minor groove of the DNA symmetrically disposed about the helix 2-fold. Data was then used in the range 8 to 3 Å resolution, giving a total of 1366 reflections  $\geq 2\sigma$  and 5 more cycles of refinement carried out yielding an R factor of 36.3%. The next stage was to divide the structure into rigid groups of individual nucleotides, phosphates and sugars giving 48 rigid groups to allow greater flexibility in the structure during refinement. Five cycles of positional refinement followed by 5 cycles of temperature factor refinement reduced the R factor to 32.7%. The Afsig and Bfsig values and the scale were then updated and the resolution limit extended to 2.5 Å followed by a further 5 cycles of positional and temperature factor refinement which reduced the R factor to 32.4%. It was then decided to switch to the restrained least squares refinement program NUCLSQ (Westhof *et al.*, 1985). The weighting scheme used is shown in table 3.2. This weighting scheme consisted of heavier weights than in the refinement of the complex of berenil with d(CGCGAATTCGCG)<sub>2</sub> for most parameters. The reason for heavier weights was that initial refinement procedures, using the same weights as in the previous refinement, resulted in unusually high temperature factors and poor geometry for the phosphate groups. After 10 cycles of refinement using data

Table 3.2. Refinement parameters used and statistics from the NUCLSQ refinement of d(CGCAAATTTGCG)<sub>2</sub> complexed with berenil.

Resolution range	7-2.0 Å	
Number of reflections [I>2σ(I)]	3106	
Temperature of study	18°C	
Final R <sup>a</sup> factor	18.3%	
distances > 2σ	63	
	rms dev	σ value
sugar-base bond distances	0.011	0.020 Å
sugar base bond angle distances	0.026	0.030 Å
phosphate bond distances	0.031	0.030 Å
phosphate angle and H-bond distances	0.032	0.040 Å
planar groups	0.010	0.015 Å
chiral volumes	0.014	0.015 Å <sup>3</sup>
single torsion contacts	0.058	0.250 Å
multiple torsion contacts	0.149	0.250 Å
Isotropic temperature factors		
sugar-base bonds	2.654	3.5 Å <sup>2</sup>
sugar-base angles	3.446	3.5 Å <sup>2</sup>
phosphate bonds	4.713	5.0 Å <sup>2</sup>
phosphate angles, H bonds	5.591	10.0 Å <sup>2</sup>
weighting scheme applied to the structure factors	[1/(SIGAPP) <sup>2</sup> ] <sup>b</sup>	
AFSIG	2.5	
BFSIG	-40.0	

$$R^a = \frac{\sum |F_o - F_c|}{\sum F_o}$$

$$\text{SIGAPP}^b = \text{AFSIG} + \text{BFSIG} (\text{STHOL} - 0.166667)$$

from 7-2.5 Å the R factor was 25%. Maps of  $|2F_o - F_c| \alpha_{calc}$  and  $|F_o - F_c| \alpha_{calc}$  were calculated using data to 2.5 Å. Over the next 17 cycles the resolution limit was increased to first 2.2 Å and then 2.0 Å adjusting scale factor and Afsig and Bfsig accordingly resulting in an R factor of 24.9%.

Once the model started to converge to a solution it was decided to include solvent molecules. To this end the "peakpk" option of PROTEIN was used to list out peaks and holes of greater and less than  $2.5\sigma$  electron density respectively. The x,y,z positions of these peaks were then used as the basis for a positional search for possible water molecules using the program DRILL.

Each possible solvent molecule position was then studied carefully on the graphics, against  $|2F_o - F_c| \alpha_{calc}$  and  $|F_o - F_c| \alpha_{calc}$  maps, and adjusted so as to optimise hydration geometries. The criteria for hydration geometry were basically taken as proximity (2.2–3.5 Å) to a hydrogen bond donor/acceptor and, if the atom concerned was involved in more than one interaction, that the subtended angles (with the inclusion of idealised hydrogen positions) were  $180^\circ \pm 50^\circ$ . The addition of 36 solvent molecules over the next 20 cycles of refinement with the rejection of any waters that did not refine (by criteria of large increase in temperature factor or loss of hydration geometry) reduced the R factor to 22.3%. The addition of these waters, with care taken not to insert waters in the minor groove region, improved the density of the ligand slightly. At this stage the ligand was docked into the density. The quality of the density was such that it was possible to fit the curvature of the drug to it. The ligand subsequently refined satisfactorily to give an R factor of 20.6% and an average B factor for the drug of around  $35 \text{ \AA}^2$ . After further refinement and the addition of more water molecules to give a total of 65 solvent molecules the final R factor using  $2\sigma$  data from 7-2 Å was 18.3%.

Table 3.3. Final agreement statistics.

Dmin Å	Refs	W(Fo-Fc)**2	Shell R	Sphere R
6.00	73	3.53	0.291	0.291
5.00	135	3.14	0.203	0.229
4.00	301	2.51	0.146	0.178
3.00	803	3.45	0.161	0.169
2.50	879	3.04	0.208	0.177
2.20	707	3.21	0.222	0.181
2.00	208	2.68	0.266	0.183

Average weighted  $|F_o - F_c|$  discrepancy 0.15

Dmin. Minimum resolution limit of data.

Refs. Total number of reflections within that shell.

W(Fo-Fc)\*\*2 Weighted agreement of observed and calculated structure factors

R. Standard R-factor  $R = \frac{\sum(|F_o - F_c|)}{\sum(F_o)}$

## Results.

### 3.5 Overall Structural Features.

The drug is essentially symmetrically disposed about the DNA helix two-fold axis, spanning the four central AT base pairs in the minor groove (figure 3.1). It is closely isohelical with the floor and near parallel to the walls of the minor groove. The binding mode is distinctly different from the previously described structure with the berenil molecule displaced 0.5 Å based on an overlay of the phosphorus atoms. There is no evidence of a water molecule bridging between the ligand and the DNA, as was observed in the previous crystal structure of the complex with the dodecamer containing the *EcoRI* site. The berenil molecule exhibits an overall twist of approximately 35° with torsional rotations of 17° and 21° for the 5' and 3' amidinium moieties, respectively.

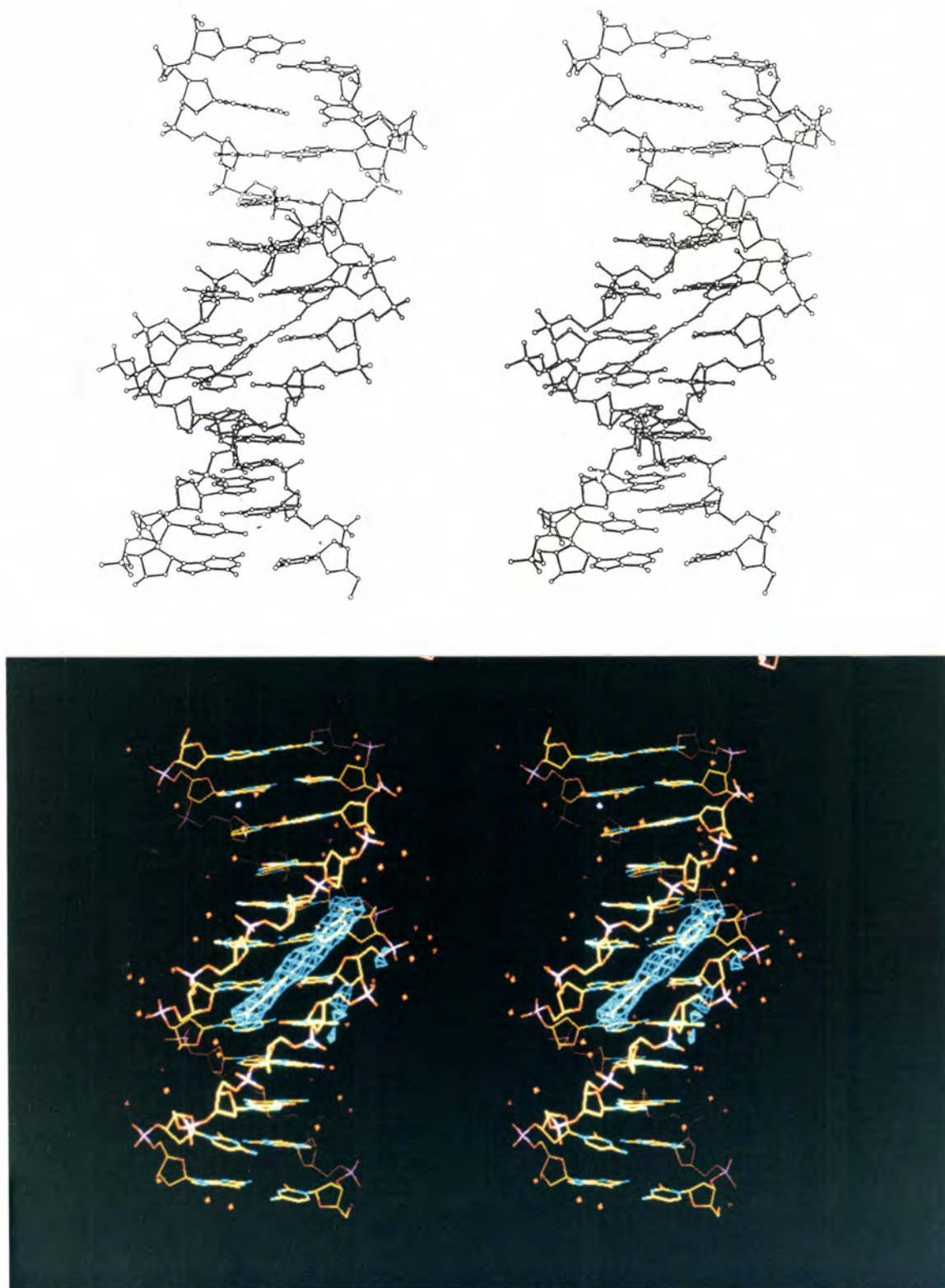


Figure. 3.1. Stereo views of complete structure showing berenil in the minor groove, the lower showing the final omit density at  $3\sigma$ .

In the present complex these amidinium moieties interact with O2 of thymine 8 (N10...O2 Thy(8) 3.01 Å) at the 3' end of the complex, and two weak interactions, O2 of thymine 20 and N3 of adenine 6 (N20...O2 Thy(20) 3.27 Å and N20...N3 Ade(6) 3.38 Å) (see figures 3.2 and 3.3). With the placement of idealised hydrogens for sp<sup>2</sup> nitrogens (H-N bond lengths 1.04 Å) the subtended angles for these hydrogen bonds at the 5' and 3' ends of the berenil would be 128.1° and 145.5° respectively. The two phenyl rings of berenil are perpendicularly aligned to the minor groove floor, and lie parallel to its walls. If hydrogens are generated in the central region of the structure the hydrogens attached to C1', C4' and C5' make numerous non-bonded contacts with the hydrogen atoms on the phenyl rings of the berenil, with distances which approximate the sum of the van der Waals radii of the atoms involved (3 Å or less). The hydrogen atom H4' of thymine 7 of strand 1 is in close contact and mid-way between the two distal carbon atoms of ring A. Atom H5', also from thymine 7 is in close contact with the carbon atom of the amidinium moiety at the 5' end of the structure. These contacts are balanced on the other side of the phenyl ring by the hydrogens of thymines 20 and 21. Atom H5' of thymine 21 makes three close contacts with the top side of the ring and H4' from this thymine also contacts the amidinium carbon atom. The H4' of thymine 20 also complements that from thymine 7 in contacting the ring carbon atom attached to the triazene group. All three nitrogens of the triazene are in close contact with the H4' and H5' atoms of thymines 8 and 20. Phenyl ring B again has a pattern of close contacts with H4' and H5' atoms, this time from thymines 9 and 19. As before, the majority of the contacts are with the top, outer-facing, carbon atoms of the ring. The sole exception is a contact between H1' of thymine 8 and one of the two inner-facing carbon atoms of ring B. The other inner-facing carbon atom of this ring has its hydrogen atom contacting the H2 of adenine 18. Its equivalent on phenyl ring A contacts the symmetry related

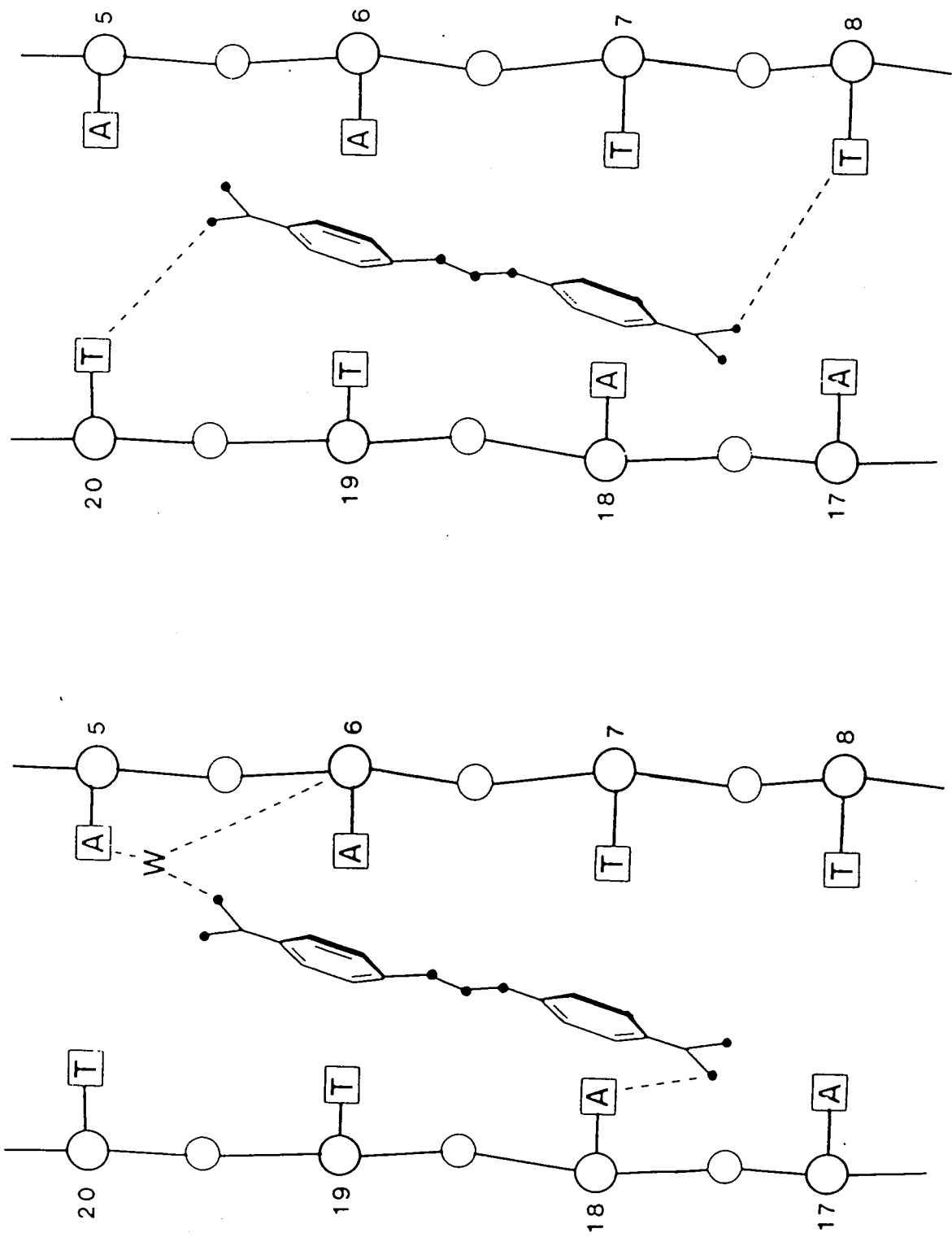
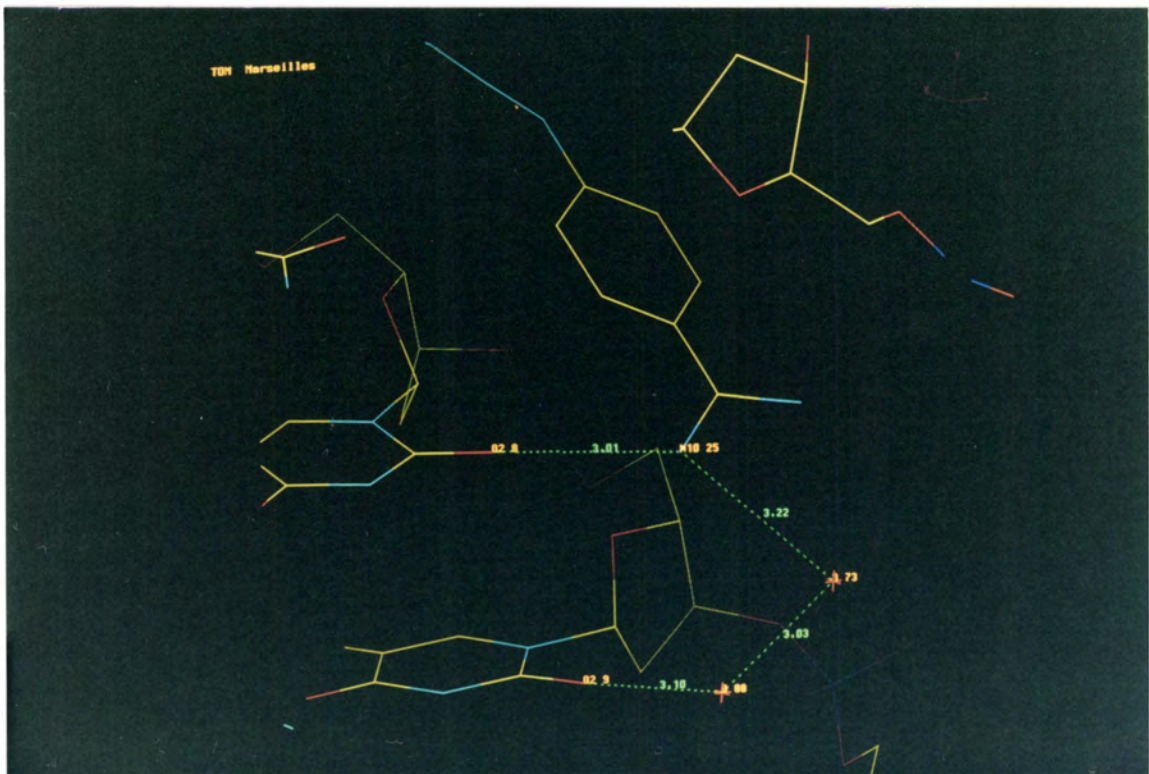
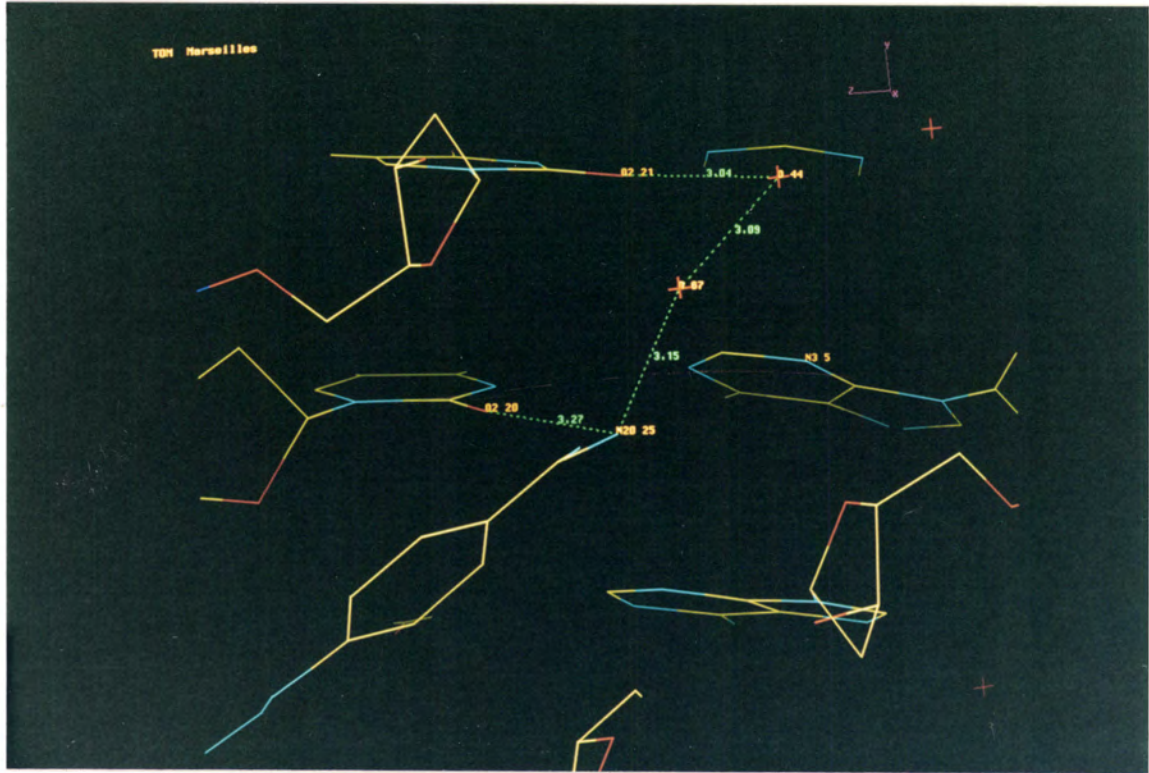


Figure. 3.2. Schematic of berenil binding modes.

a) with (CGCGAATTCGCG)<sub>2</sub>

b) with (CGCAAATTTGCG)<sub>2</sub>



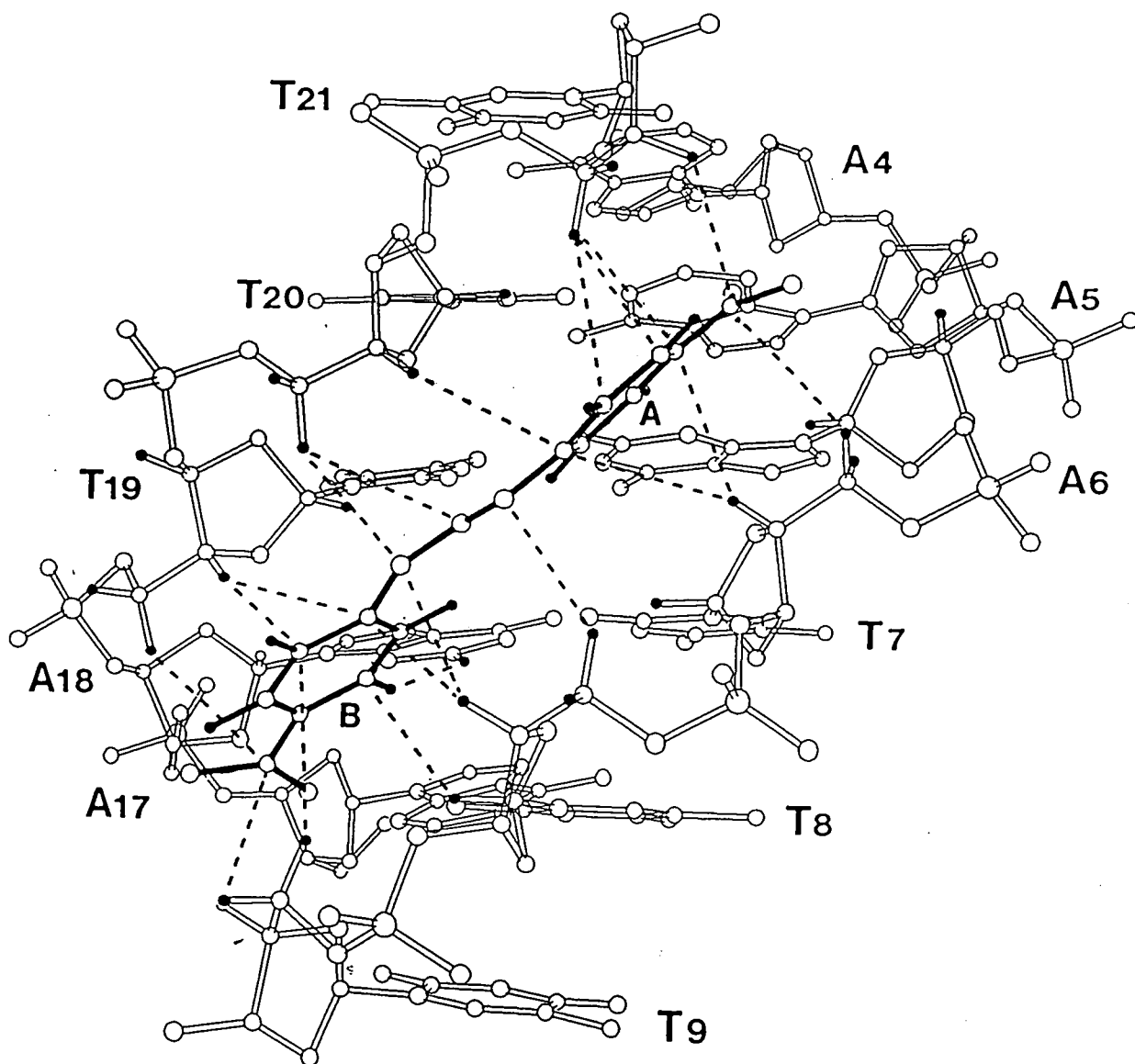


Figure. 3.3. Detailed interactions between berenil and the dodecamer.

H2 of adenine 6.

Similarities can be drawn between this complex and that observed for berenil with  $A_2T_2$  in that in native sequences for both  $A_2T_2$  and  $A_3T_3$  (i.e., the crystal structures of the DNA alone) the minor groove of the AT core regions is occupied by a spine of water molecules that runs down the middle of the minor groove linked together and to the DNA via an extensive hydrogen bond network.

### 3.6 Analysis of helical parameters.

Helix parameters were determined using NEWHELIX90 (Dickerson, 1990). The DNA structure is similar in overall appearance to the  $A_3T_3$ -distamycin structure. However the  $A_3T_3$ -distamycin backbone does exhibit some rather unusual torsion angles when examined in detail and thus the present structure shall only be compared with the native  $A_3T_3$  structure solved simultaneously in conjunction with Karen Edwards in this laboratory. The sugar-phosphate backbone is normal global B-DNA with standard  $B_I$  phosphates of average separation of one phosphorus atom to the adjacent phosphorus atom of 6.7 Å. The pseudorotation parameters of the sugars are not constant over the whole dodecamer, they are in fact very similar to those for the the *EcoRI* crystal structure varying between C3' *endo* and C2' *endo* (see tables 3.4 and 3.5). This is perhaps not surprising on the grounds that in the  $A_3T_3$  sequence a purine Gua(4) has been substituted by another purine Ade(4) and a pyrimidine Cyt(9) has been substituted by another pyrimidine Thy(9) thus keeping the nature of the helix identical when assessed on the basis of purines and pyrimidines which is believed to govern local helical parameters (Dickerson and Drew, 1981). There is one notable exception; the sugar pucker of Thy(7) being C1' *exo* instead of O4' *endo*, as observed in the crystal structure of the dodecamer containing the *EcoRI* site. The converse has occurred for the equivalent Thy(19) sugar pucker. This may be a significant factor in

Table 3.4.: Conformational parameters for the B-DNA dodecamer  
d(CGCAAATTTGCG)<sub>2</sub> complexed with berenil.

Residue	Angles, Degrees							Sugar Pucker
	$\chi$	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	
Cyt1	-103	—	—	172	153	-156	-117	C <sub>2'</sub> -endo
Gua2	-81	-53	168	27	140	-168	-128	C <sub>2'</sub> -endo
Cyt3	-139	-52	146	55	97	-166	-139	O <sub>4'</sub> -endo
Ade4	-84	-25	158	40	137	-177	-102	C <sub>2'</sub> -endo
Ade5	-95	-68	172	63	130	-182	-95	C <sub>1'</sub> -exo
Ade6	-93	-85	192	57	128	-169	-113	C <sub>1'</sub> -exo
Thy7	-118	-35	152	34	107	-172	-100	C <sub>1'</sub> -exo
Thy8	-124	-65	172	58	120	-174	-106	C <sub>1'</sub> -exo
Thy9	-119	-48	165	47	107	-161	-101	C <sub>1'</sub> -exo
Gua10	-73	-55	165	42	159	-104	-212	C <sub>2'</sub> -endo
Cyt11	-114	-75	144	56	134	-166	-87	C <sub>2'</sub> -endo
Gua12	-91	-85	184	52	122	—	—	C <sub>1'</sub> -exo
Cyt13	-110	—	—	38	159	-119	-179	C <sub>2'</sub> -endo
Gua14	-105	-43	132	31	123	-177	-103	C <sub>1'</sub> -exo
Cyt15	-125	-60	164	59	101	-165	-116	O <sub>4'</sub> -endo
Ade16	-92	-37	159	49	134	-206	-79	C <sub>2'</sub> -endo
Ade17	-86	-74	195	59	151	-174	-115	C <sub>2'</sub> -endo
Ade18	-80	-45	181	32	149	-176	-99	C <sub>2'</sub> -endo
Thy19	-133	-77	167	63	97	-189	-82	O <sub>4'</sub> -endo
Thy20	-117	-71	176	57	118	-172	-103	C <sub>1'</sub> -exo
Thy21	-118	-54	167	48	103	-174	-89	C <sub>1'</sub> -exo
Gua22	-79	-65	163	55	145	-126	-178	C <sub>2'</sub> -endo
Cyt23	-128	-51	132	38	110	-167	-110	C <sub>1'</sub> -exo
Gua24	-124	-61	160	50	86	—	—	C <sub>3'</sub> -endo
Mean	-105	-58	164	48 <sup>1</sup>	125	-165	-115	
±SD	19	16	17	11	21	22	33	

<sup>1</sup> Cyt1 value omitted because it represents end effects.

Table 3.4. (continued): Conformational parameters for the B-DNA dodecamer d(CGCGAATTCGCG)<sub>2</sub>.

Sugar residue	Angles, Degrees							Pucker
	$\chi$	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	
Cyt1	-105	—	—	174	157	-141	-144	C <sub>2'</sub> -endo
Gua2	-111	-66	170	40	128	-186	-98	C <sub>1'</sub> -exo
Cyt3	-135	-63	172	59	98	-177	-88	O <sub>4'</sub> -exo
Gua4	-93	-63	180	57	156	-155	-153	C <sub>2'</sub> -endo
Ade5	-126	-43	143	52	120	-180	-92	C <sub>1'</sub> -exo
Ade6	-122	-73	180	66	121	-186	-89	C <sub>1'</sub> -exo
Thy7	-127	-57	181	52	99	-186	-86	O <sub>4'</sub> -endo
Thy8	-126	-59	173	64	109	-189	-89	C <sub>1'</sub> -exo
Cyt9	-120	-58	180	60	129	-157	-94	C <sub>1'</sub> -exo
Gua10	-90	-67	169	47	143	-103	-210	C <sub>2'</sub> -endo
Cyt11	-125	-74	139	56	136	-162	-90	C <sub>2'</sub> -endo
Gua12	-112	-82	176	57	111	—	—	C <sub>1'</sub> -exo
Cyt13	-128	—	—	56	137	-159	-125	C <sub>2'</sub> -endo
Gua14	-116	-51	164	49	122	-182	-93	C <sub>1'</sub> -exo
Cyt15	-134	-63	169	60	86	-185	-86	O <sub>4'</sub> -endo
Gua16	-115	-69	171	73	136	-186	-98	C <sub>2'</sub> -endo
Ade17	-106	-57	190	54	147	-183	-97	C <sub>2'</sub> -endo
Ade18	-108	-57	186	48	130	-186	-101	C <sub>2'</sub> -endo
Thy19	-131	-58	174	60	109	-181	-88	C <sub>1'</sub> -exo
Thy20	-120	-59	179	55	122	-181	-94	C <sub>1'</sub> -exo
Cyt21	-114	-59	185	45	110	-177	-86	C <sub>1'</sub> -exo
Gua22	-88	-67	179	50	150	-100	-188	C <sub>2'</sub> -endo
Cyt23	-125	-72	139	45	113	-174	-97	C <sub>1'</sub> -exo
Gua24	-135	-65	171	47	79	—	—	C <sub>3'</sub> -endo
Mean	-117	-63	171	54 <sup>2</sup>	123	-169	-108	
±SD	14	8	14	8	21	25	34	

<sup>2</sup> Cyt1 value omitted because it represents end effect.

Table 3.5. Local helix parameters for the B-DNA dodecamer  
d(CGCGAATTCGCG)<sub>2</sub>.

Base-pairs	Propeller twist, $^{\circ}\theta_p$	Helix twist angle, $^{\circ}t^3$	Base-pairs per turn ( $n$ )	Rise per base pair, $\text{\AA}$ ( $h$ )
C1/G24	13.2			
		38.3	9.40	3.36
G2/C23	11.7			
		39.6	9.09	3.38
C3/G22	7.2			
		33.5	10.75	3.26
G4/C21	13.2			
		37.4	9.63	3.30
A5/T20	17.1			
		37.5	9.60	3.27
A6/T19	17.8			
		32.2	11.18	3.31
T7/A18	17.1			
		36.0	10.00	3.29
T8/A17	17.1			
		41.4	8.70	3.14
C9/G16	18.6			
		32.3	11.11	3.56
G10/C15	4.9			
		44.7	8.05	3.21
C11/G14	17.2			
		37.0	9.73	3.54
G12/C13	6.2			
Mean	13.4	37.3	9.75	3.33
SD	4.9	3.8	0.98	0.13

<sup>3</sup> Helical parameters obtained by using vectors between atoms C<sub>1</sub> and attached N of one base and the equivalent atoms of the next nucleotide along the chain.

Table 3.5.(continued ) Local helix parameters for the B-DNA dodecamer  
d(CGCAAATTTGCG)<sub>2</sub> complexed with berenil.

Base-pairs	Propeller twist, ° $\theta_p$	Helix twist angle, ° $t$	Base-pairs per turn ( $n$ )	Rise per base pair, Å ( $h$ )
C1/G24	12.3	38.3	9.40	3.28
G2/C23	12.8	38.0	9.47	2.88
C3/G22	7.0	30.0	12.0	3.97
A4/T21	12.9	35.1	10.26	3.21
A5/T20	22.4	37.7	9.55	3.10
A6/T19	18.4	31.0	11.61	3.29
T7/A18	18.0	38.5	9.35	3.18
T8/A17	18.1	35.5	10.14	3.08
T9/A16	12.1	32.8	10.98	4.01
G10/C15	4.9	40.0	9.0	2.87
C11/G14	21.2	40.0	9.0	3.54
G12/C13	11.2			
Mean	13.9	36.1	10.06	3.31
SD	5.3	3.3	0.99	0.37

<sup>4</sup> Helical parameters obtained by using vectors between atoms C<sub>1'</sub> and attached N of one base and the equivalent atoms of the next nucleotide along the chain.

accommodating the high propeller twist observed in the central A<sub>3</sub>T<sub>3</sub> region of this sequence (figure 3.4) or could be a result of berenil being bound.

Comparing minor groove width with sugar pucker reveals that the switch in sugar pucker for Thy19 from C4' *endo* in the A<sub>2</sub>T<sub>2</sub> to C1' *exo* in the A<sub>3</sub>T<sub>3</sub> sequence is accompanied by an increase of approximately 2 Å in the minor groove width, as measured from P10-P19. There is no evidence that the A<sub>3</sub>T<sub>3</sub> core induces any sequence specific bend in the DNA as has been observed for phased A-tracts in minicircle DNA from the kinetoplast body of *Leishmania tarentolae* (Marini *et al.*, 1982; Burkhoff and Tullius, 1987; Nadeau and Crothers 1989; Crothers *et al.*, 1990). The DNA exhibits an overall bend of 14.4°, towards the major groove, compared with a bend of 15.3° for the native A<sub>3</sub>T<sub>3</sub> and 14.8° for the A<sub>3</sub>T<sub>3</sub>-distamycin complex. The bend is consistent with the degree of bending normally found in dodecamer crystal structures that pack in the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. This supports the evidence for the crystallographically observed bend being due to crystal packing forces (DiGabriele *et al.*, 1989; Dickerson *et al.*, 1991).

It has been pointed out by Calladine (Calladine, 1982) that although propeller twist improves stacking interactions between bases along the helix it can induce steric clashes between adjacent purines on opposite strands. The clashes are due to interference between Gua O6 and Ade N6 in the major groove for pyrimidine -(5',3')- purine and in the case of a purine -(5',3')-pyrimidine clashes occur in the minor groove between Gua N3/N2 and Ade N3 atoms. These steric clashes can be relieved by a combination of four changes in helical parameters:-

- a) reducing propeller twists
- b) opening of roll angles between base pairs
- c) base-pair slide (Dy) to pull the purine out of the helical stack
- d) decrease in twist between base pairs

Analysis of the roll parameters (figure 3.5) for the sequence shows that it

Figure 3.4. Graph of propeller twist for the sequences listed.

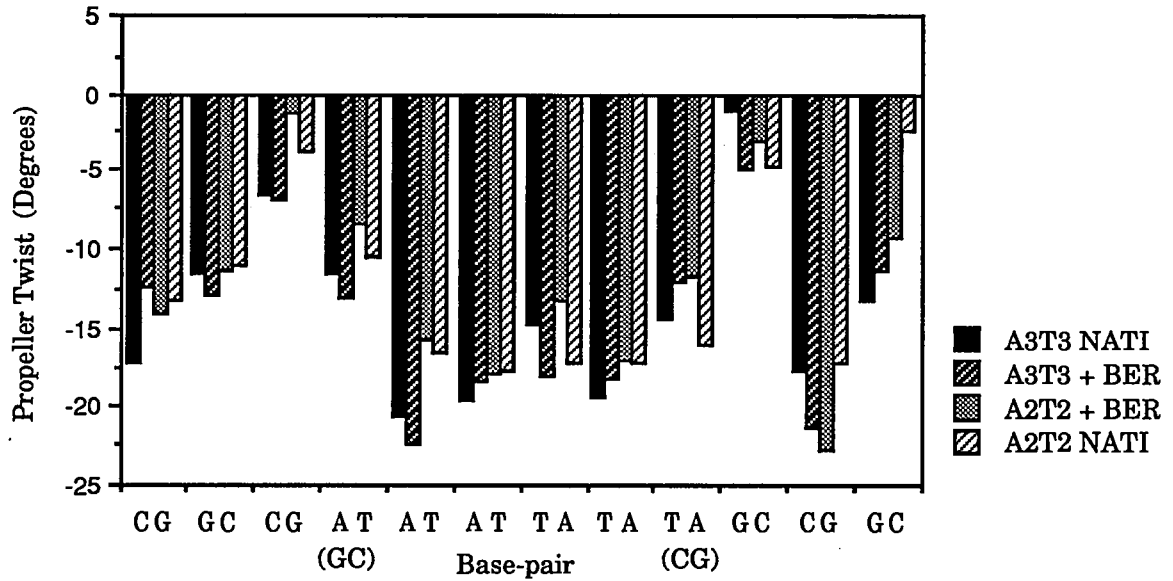
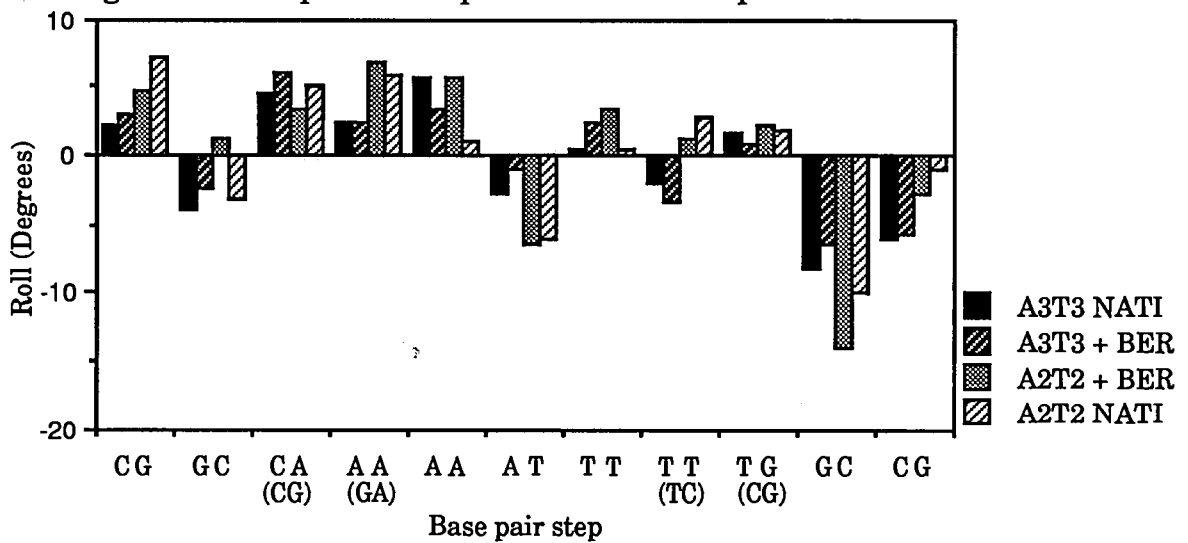


Figure 3.5. Graph of base-pair roll for the sequences listed.

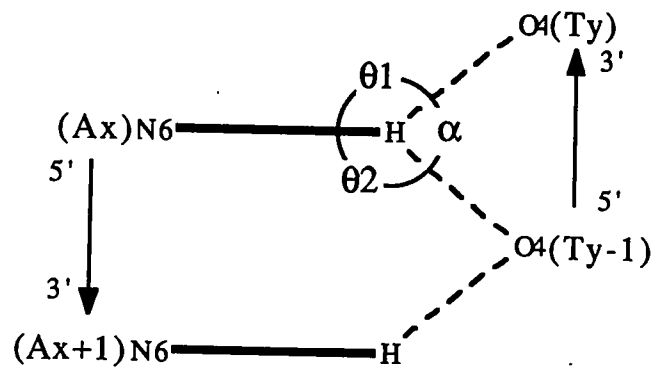


obeys the general principle of purine -(5'-3')- pyrimidine inducing a roll towards the minor groove (a positive roll, or an opening of the base pair edges in the minor groove) and pyrimidine -(5'-3')- purine inducing a roll towards the major groove (a negative roll). In the case of a purine-purine step or a pyrimidine-pyrimidine step the roll is usually positive. However in both the complex and the native  $A_3T_3$  structures the roll between base-pairs Thy8/Ade17 and Thy9/Ade16 (a homopyrimidine step) rather than being positive, as is the case for the corresponding Thy/Cyt step in the  $A_2T_2$  sequence, is negative, giving a roll towards the major groove. This would facilitate the necessary conformational change, on the minor groove side, to allow interaction between berenil and the O2 of Thy8. Also notable, when comparing the roll values with those of the  $A_2T_2$  sequence, is a lower positive roll value for base-pairs Ade4/Thy21 and Ade5/Thy20. This would result in a reduced roll towards the minor groove which would make the geometry more favourable for the berenil to interact with the O2 of Thy20. The fact that in both these cases the bases concerned in the sequence are those that have changed from those in the previously studied  $A_2T_2$  sequence is significant. Although the base changes have not altered the sequence in terms of purine/pyrimidine sequence, it appears that an Ade-Ade step can tolerate lower positive roll than a Gua-Ade step. Throughout the region spanned by the ligand in the complex the roll values are significantly lower than in the  $A_2T_2$  complex.

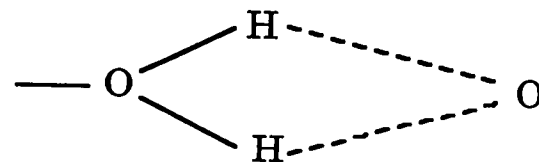
### 3.7 *Are three-centred hydrogen bonds present?*

The DNA exhibits large propeller twists in the  $A_3T_3$  core segment of the molecule as in the structure of the DNA-distamycin complex but the evidence for three centred hydrogen bonding is not striking. There is agreement with the geometrical conditions for three-centre hydrogen bonds as defined by Jeffrey (Jeffrey and Maluszynska, 1982; Jeffrey and Mitra, 1984; Jeffrey *et al.*, 1985, and figure 3.6) between the N6 amino group of Ade(5) and O4 of Thy(19) (if

Figure. 3.6. Three-centred vs. bifurcated hydrogen bond.



Scheme for a three-centre hydrogen bond.



A bifurcated hydrogen bond.

idealised hydrogen atoms are added to the N6 of adenine as determined in the crystal structure) for an unsymmetrical three-centre hydrogen bond. The same is true for the symmetry related Ade(19) N6 to O4 Thy(7) (see figure 3.7).

The distances Ade(5) N6...O4 Thy(19) 3.25 Å and Thy(7) O4...N6 Ade(17) 3.25 Å combined with the angular displacement of the atoms involved would result in a very weak interaction. There have been reports of three centred hydrogen bonds in polyA regions of crystal structures (Coll *et al.*, 1987; Nelson *et al.*, 1987; Drew *et al.*, 1981) and there is some evidence for their existence from NMR studies (Brown *et al.*, unpublished data). However both methods can only infer the presence of this feature as X-ray diffraction does not give hydrogen positions in macromolecular structures and  $^1\text{H}$  NMR gives only the hydrogen positions (it would be possible to do  $^{15}\text{N}$  and  $^{17}\text{O}$  NMR experiments as well to give the necessary information of the oxygen and nitrogen positions but this would be a difficult and costly experiment). Molecular dynamics studies have also been performed, both as part of NMR refinement and as purely modelling studies, which have indicated the presence of three centred hydrogen bonds. The behaviour of long AT stretches in simulations is a point of some debate. A molecular dynamics study of  $d(\text{CGCAAATTTGCG})_2$ , using XPLOR, with and without distamycin shows the DNA alone tries to transform from an initial standard B-DNA to A-DNA, only the inclusion of the ligand stabilises the central region as B-DNA (Boehncke *et al.*, 1991), yet our NMR studies and crystallographic data clearly show the DNA to be B-form. Calculations by Fritsch and Westhof (in press) on a polydA•polydT decamer using AMBER (Weiner *et al.*, 1984), without the inclusion of solvent molecules or counterions, held the end two base pairs rigid using the belly option thus only allowing movement in the central 6 bases. They report lifetimes for three centred hydrogen bonding but the criteria for their existence was as proposed by Jeffrey *et al.*, 1982, which seems generous in its distances of interactions.

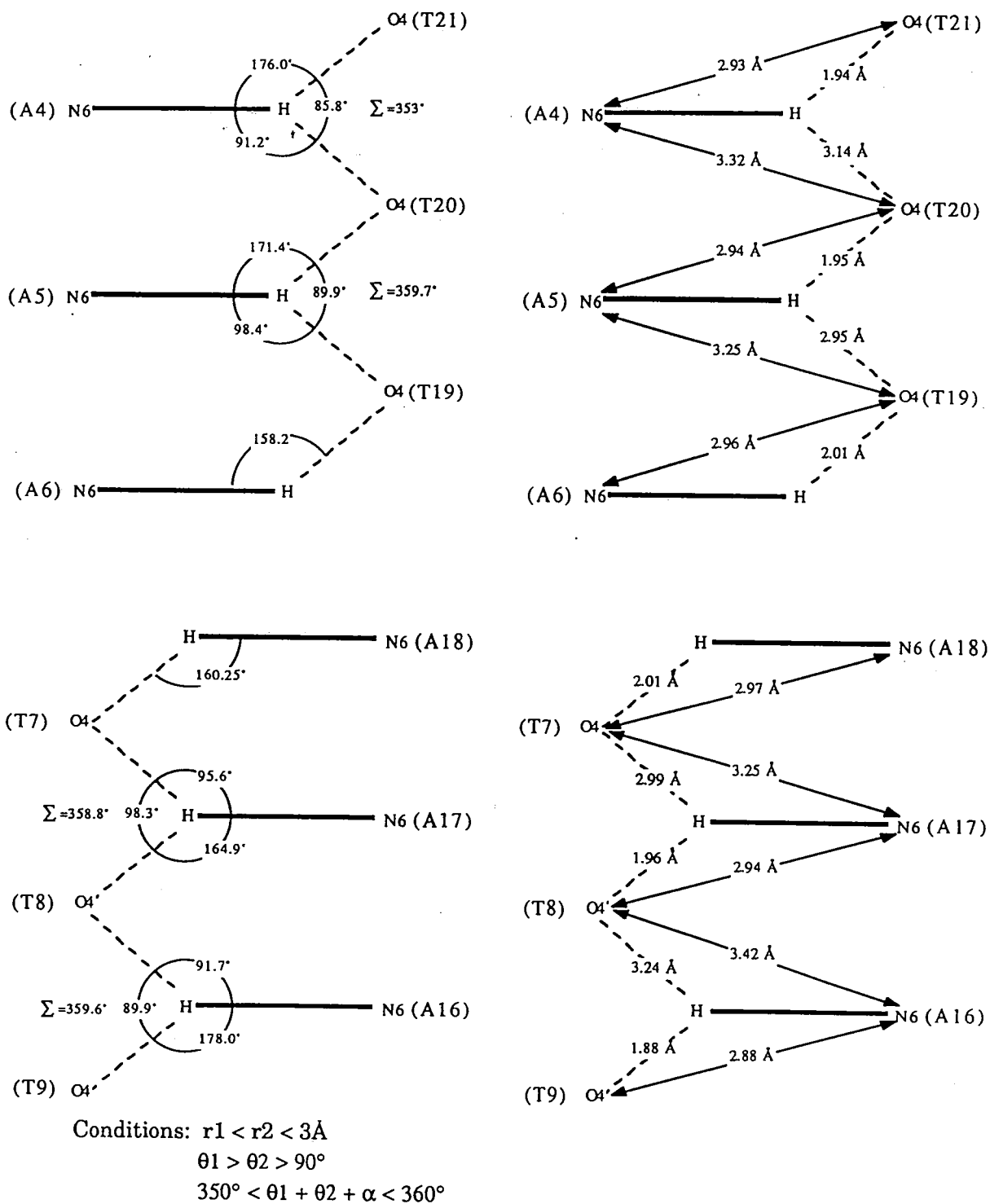


Figure. 3.7. Hydrogen bonding geometries for the major groove of the complex.

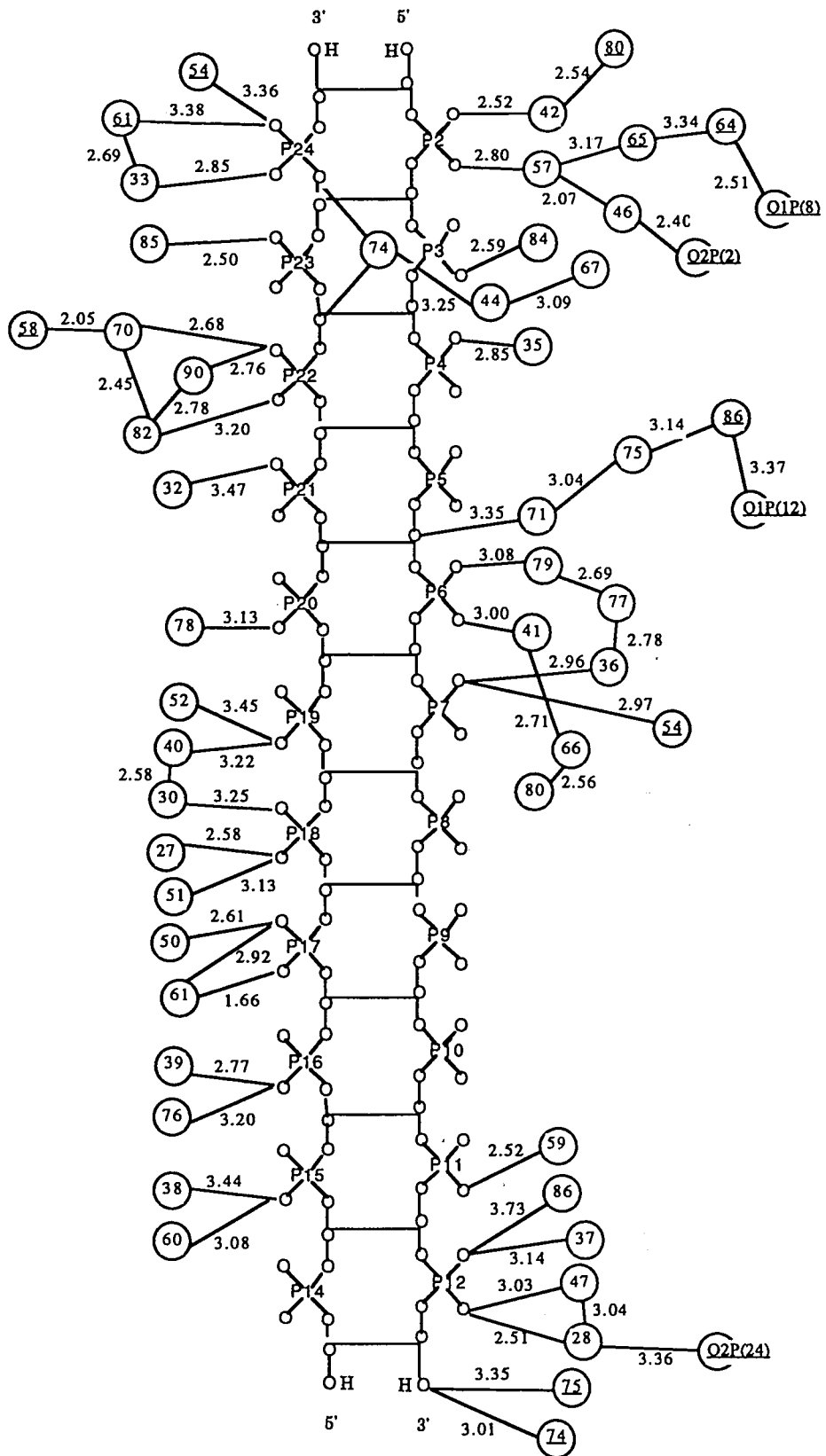


Figure. 3.8. Schematic representation of the phosphate backbone hydration.

### 3.8 Hydration of the crystal structure.

The oligonucleotide is well hydrated in the crystal environment. As has been observed for many oligonucleotides there is extensive hydration of the phosphate groups (hydrogen bonding to O1P and O2P ) (figure 3.8). There are numerous close contacts between the phosphorus atoms of one helix and symmetry related helices some of which are bridged by water molecules that form hydrogen bonds to the phosphate oxygens, O1P and O2P. The major groove also exhibits water networks particularly at the ends of the dodecamer, (figure 3.9). The central region, where the berenil is bound in the minor groove, is more sparsely populated with solvent molecules with little evidence to suggest that they might stabilise possible three-centred hydrogen bonding. This may be due to transmitted electrostatic effects of both the bound drug and associated solvent molecules in the minor groove. There is good evidence for the existence of a magnesium counterion in this structure at the 5' end of the helix in the major groove. A large peak in the electron density map was found situated close to the Gua2 base in the major groove, which refined well as an  $Mg^{2+}$  and had a final B factor of  $36.7 \text{ \AA}^2$ . The magnesium counterion coordinates a water network that provides a link to a symmetry related helix and probably plays a role in stabilising the crystal packing as the distances from W41 to O2P of Ade6 and W54 to O1P of Thy7, where Ade6 and Thy7 are of a symmetry related dodecamer, are both  $3.0 \text{ \AA}$ . A virtually identical magnesium hexahydrate cation has recently been reported in exactly the same position in a low temperature study of the interaction of Hoechst 33258 with the dodecamer  $d(CGCGAATTCGCG)_2$  (Quintana *et al.*, 1991). In the present structure the magnesium cation is surrounded by four water molecules in a plane with one water molecule above, but the water molecule expected below the plane for a full water complement, was not found (see figure 3.9a).

The minor groove exhibits the reverse situation to that of the major





groove, with solvent molecules being confined to the  $A_3T_3$  core region extending from either end of the bound berenil. There is an extensive network involving five water molecules at the 5' end and at the 3' end two water molecules link the amidinium to the O2 Thy(9) (figure 3.10). Beyond these water networks . . . contacts with symmetry related molecules occur for dodecamers in a  $P2_12_12_1$  space group (Dickerson *et al.*, 1987).

The minor groove edges of the top two base-pairs of one helix overlap the bottom two base-pairs of the helix related by a  $2_1$  screw axis. The base-pairs intersect obliquely with each guanine N2 donating a proton to a hydrogen bond with the N3 atom of the guanine of the symmetry related helix. Thus the end overlap is maintained by a network of four N2-H...N3 hydrogen bonds with an average N2...N3 distance of 3.12 Å and good hydrogen bond geometry. The helix overlap is further stabilised by hydrogen bonds of each of the 3'-OH groups which interact directly to the N2 of Gua (23) and to waters W71, W74 and W75 which in turn make contacts to the N3 of Ade(4) and N3 of Gua(22).

### 3.9 Discussion.

In summary then it appears that reduced roll and the associated large propeller twists within the  $A_3T_3$  core, coupled with the torsional rotations of the amidinium moieties, allow the berenil molecule to bind in a 1:4 mode to O2 of Thy8 and Thy20. The ends of the berenil molecule hydrogen bond to a thymine base on each strand with rest of the molecule in van der Waals contact with the backbone hydrogen atoms, mainly H4' and H5', belonging to the intervening thymines of the sequence. There is little contact between the most buried parts of the drug and the DNA, apart from those with H2 of adenines 6 and 18 in the centre of the sequence. The sugar conformations deviate slightly from the 'ideal' C2'-*endo* found in regular B-DNA to allow these close contacts to occur (and to accommodate the bend in the DNA that crystal packing of a dodecamer in this sequence dictates).

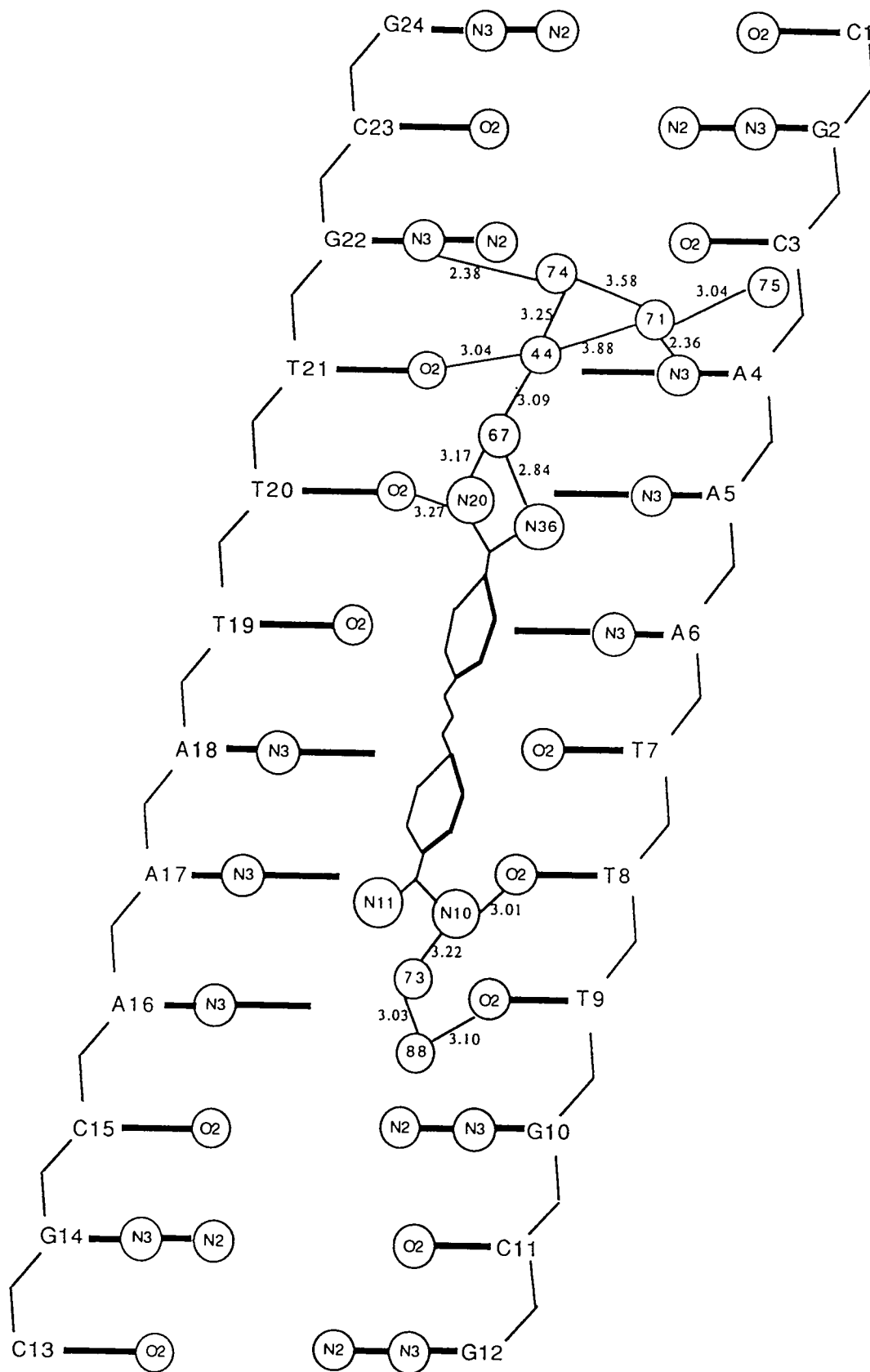


Figure. 3.10. Schematic representation of the minor groove hydration.

Table 3.6. Final coordinates and temperature factors for the DNA dodecamer d(CGCAAATTTGCG)<sub>2</sub> complexed with berenil including solvent molecules.

	Atom	Residue		X	Y	Z	Occ	B(Iso)
ATOM	1	O5'	CYT	1	18.707	34.226	90.428	1.00 44.60
ATOM	2	C5'	CYT	1	19.052	34.092	89.035	1.00 36.80
ATOM	3	C4'	CYT	1	20.089	33.004	88.944	1.00 36.46
ATOM	4	O3'	CYT	1	21.784	32.066	87.519	1.00 39.28
ATOM	5	C3'	CYT	1	20.459	32.577	87.532	1.00 34.45
ATOM	6	C2'	CYT	1	19.382	31.509	87.278	1.00 33.32
ATOM	7	C1'	CYT	1	19.478	30.746	88.567	1.00 32.69
ATOM	8	O4'	CYT	1	19.655	31.810	89.575	1.00 35.84
ATOM	9	N1	CYT	1	18.226	30.051	88.931	1.00 32.21
ATOM	10	C6	CYT	1	16.987	30.587	88.703	1.00 30.25
ATOM	11	C5	CYT	1	15.856	29.974	89.061	1.00 24.77
ATOM	12	C4	CYT	1	15.962	28.719	89.712	1.00 27.66
ATOM	13	N4	CYT	1	14.863	28.049	90.096	1.00 22.28
ATOM	14	N3	CYT	1	17.199	28.187	89.946	1.00 28.34
ATOM	15	C2	CYT	1	18.349	28.829	89.569	1.00 29.70
ATOM	16	O2	CYT	1	19.453	28.346	89.784	1.00 27.25
ATOM	17	P	GUA	2	22.748	32.009	86.270	1.00 50.02
ATOM	18	O1P	GUA	2	22.031	32.679	85.183	1.00 48.55
ATOM	19	O2P	GUA	2	24.029	32.728	86.699	1.00 53.76
ATOM	20	O5'	GUA	2	23.046	30.478	85.983	1.00 38.90
ATOM	21	C5'	GUA	2	23.526	29.670	87.083	1.00 34.46
ATOM	22	C4'	GUA	2	23.435	28.240	86.602	1.00 33.87
ATOM	23	O3'	GUA	2	24.315	26.900	84.741	1.00 33.51
ATOM	24	C3'	GUA	2	23.534	28.037	85.092	1.00 32.44
ATOM	25	C2'	GUA	2	22.060	27.883	84.702	1.00 31.47
ATOM	26	C1'	GUA	2	21.535	27.095	85.892	1.00 28.06
ATOM	27	O4'	GUA	2	22.203	27.684	87.005	1.00 31.91
ATOM	28	N9	GUA	2	20.067	27.249	85.977	1.00 26.12
ATOM	29	C8	GUA	2	19.308	28.309	85.561	1.00 21.92
ATOM	30	N7	GUA	2	18.022	28.155	85.808	1.00 25.32
ATOM	31	C5	GUA	2	17.948	26.892	86.406	1.00 22.25
ATOM	32	C6	GUA	2	16.839	26.137	86.875	1.00 24.17
ATOM	33	O6	GUA	2	15.644	26.453	86.842	1.00 23.54
ATOM	34	N1	GUA	2	17.179	24.926	87.415	1.00 20.28
ATOM	35	C2	GUA	2	18.455	24.468	87.467	1.00 19.95
ATOM	36	N2	GUA	2	18.623	23.241	88.007	1.00 17.54
ATOM	37	N3	GUA	2	19.520	25.138	87.044	1.00 20.48
ATOM	38	C4	GUA	2	19.197	26.332	86.517	1.00 21.89
ATOM	39	P	CYT	3	24.914	26.409	83.335	1.00 35.64
ATOM	40	O1P	CYT	3	24.379	27.306	82.340	1.00 33.05
ATOM	41	O2P	CYT	3	26.414	26.413	83.563	1.00 34.56
ATOM	42	O5'	CYT	3	24.445	24.898	83.159	1.00 31.83
ATOM	43	C5'	CYT	3	24.679	23.936	84.214	1.00 27.97
ATOM	44	C4'	CYT	3	23.516	22.969	84.220	1.00 29.09
ATOM	45	O3'	CYT	3	23.879	20.959	82.912	1.00 34.32
ATOM	46	C3'	CYT	3	23.245	22.283	82.906	1.00 29.56
ATOM	47	C2'	CYT	3	21.745	22.120	82.841	1.00 28.46
ATOM	48	C1'	CYT	3	21.213	22.689	84.142	1.00 24.29

ATOM	49	O4'	CYT	3	22.297	23.578	84.604	1.00	27.23
ATOM	50	N1	CYT	3	20.067	23.570	83.810	1.00	21.99
ATOM	51	C6	CYT	3	20.321	24.752	83.140	1.00	20.34
ATOM	52	C5	CYT	3	19.357	25.596	82.789	1.00	20.52
ATOM	53	C4	CYT	3	18.034	25.251	83.179	1.00	22.89
ATOM	54	N4	CYT	3	16.982	26.043	82.886	1.00	25.30
ATOM	55	N3	CYT	3	17.790	24.090	83.843	1.00	22.95
ATOM	56	C2	CYT	3	18.788	23.213	84.155	1.00	20.38
ATOM	57	O2	CYT	3	18.633	22.161	84.754	1.00	22.28
ATOM	58	P	ADE	4	24.012	20.212	81.552	1.00	33.15
ATOM	59	O1P	ADE	4	23.502	21.211	80.511	1.00	32.52
ATOM	60	O2P	ADE	4	25.392	19.769	81.253	1.00	43.60
ATOM	61	O5'	ADE	4	22.957	18.993	81.617	1.00	34.11
ATOM	62	C5'	ADE	4	22.615	18.518	82.945	1.00	29.43
ATOM	63	C4'	ADE	4	21.289	17.824	82.834	1.00	30.24
ATOM	64	O3'	ADE	4	20.498	15.728	81.936	1.00	32.53
ATOM	65	C3'	ADE	4	21.114	16.963	81.565	1.00	30.68
ATOM	66	C2'	ADE	4	20.338	17.901	80.661	1.00	28.51
ATOM	67	C1'	ADE	4	19.424	18.640	81.630	1.00	25.37
ATOM	68	O4'	ADE	4	20.205	18.717	82.834	1.00	29.04
ATOM	69	N9	ADE	4	19.010	19.984	81.188	1.00	25.38
ATOM	70	C8	ADE	4	19.710	20.926	80.492	1.00	23.92
ATOM	71	N7	ADE	4	19.059	22.047	80.264	1.00	25.84
ATOM	72	C5	ADE	4	17.825	21.832	80.882	1.00	22.97
ATOM	73	C6	ADE	4	16.684	22.669	81.006	1.00	21.87
ATOM	74	N6	ADE	4	16.565	23.891	80.498	1.00	20.69
ATOM	75	N1	ADE	4	15.661	22.100	81.676	1.00	21.76
ATOM	76	C2	ADE	4	15.718	20.845	82.209	1.00	19.36
ATOM	77	N3	ADE	4	16.760	20.029	82.138	1.00	24.96
ATOM	78	C4	ADE	4	17.788	20.585	81.455	1.00	24.60
ATOM	79	P	ADE	5	20.202	14.522	80.921	1.00	49.85
ATOM	80	O1P	ADE	5	21.020	14.697	79.737	1.00	48.16
ATOM	81	O2P	ADE	5	20.419	13.227	81.663	1.00	51.83
ATOM	82	O5'	ADE	5	18.623	14.681	80.641	1.00	43.35
ATOM	83	C5'	ADE	5	17.834	14.380	81.865	1.00	37.81
ATOM	84	C4'	ADE	5	16.440	14.798	81.435	1.00	34.88
ATOM	85	O3'	ADE	5	14.565	13.552	80.628	1.00	37.63
ATOM	86	C3'	ADE	5	15.866	13.994	80.251	1.00	31.31
ATOM	87	C2'	ADE	5	15.888	14.993	79.099	1.00	31.69
ATOM	88	C1'	ADE	5	15.676	16.301	79.815	1.00	27.25
ATOM	89	O4'	ADE	5	16.494	16.138	81.032	1.00	33.54
ATOM	90	N9	ADE	5	16.161	17.491	79.132	1.00	26.40
ATOM	91	C8	ADE	5	17.342	17.702	78.455	1.00	25.26
ATOM	92	N7	ADE	5	17.475	18.912	77.954	1.00	22.36
ATOM	93	C5	ADE	5	16.302	19.554	78.351	1.00	20.70
ATOM	94	C6	ADE	5	15.858	20.869	78.149	1.00	18.83
ATOM	95	N6	ADE	5	16.494	21.836	77.511	1.00	22.58
ATOM	96	N1	ADE	5	14.624	21.154	78.670	1.00	21.85
ATOM	97	C2	ADE	5	13.894	20.236	79.359	1.00	18.42
ATOM	98	N3	ADE	5	14.284	18.981	79.587	1.00	22.39
ATOM	99	C4	ADE	5	15.489	18.709	79.060	1.00	20.58
ATOM	100	P	ADE	6	13.458	12.715	79.834	1.00	38.76
ATOM	101	O1P	ADE	6	14.249	12.008	78.774	1.00	37.86
ATOM	102	O2P	ADE	6	12.685	11.797	80.700	1.00	36.11
ATOM	103	O5'	ADE	6	12.596	13.933	79.275	1.00	30.20
ATOM	104	C5'	ADE	6	11.559	14.502	80.114	1.00	26.87

ATOM	105	C4'	ADE	6	10.790	15.391	79.145	1.00	26.36
ATOM	106	O3'	ADE	6	8.816	15.014	77.798	1.00	29.03
ATOM	107	C3'	ADE	6	10.216	14.701	77.915	1.00	28.65
ATOM	108	C2'	ADE	6	11.061	15.273	76.757	1.00	23.29
ATOM	109	C1'	ADE	6	11.485	16.622	77.284	1.00	21.88
ATOM	110	O4'	ADE	6	11.657	16.394	78.689	1.00	26.08
ATOM	111	N9	ADE	6	12.729	17.154	76.698	1.00	22.53
ATOM	112	C8	ADE	6	13.944	16.553	76.490	1.00	20.16
ATOM	113	N7	ADE	6	14.841	17.345	75.930	1.00	20.38
ATOM	114	C5	ADE	6	14.190	18.559	75.807	1.00	17.30
ATOM	115	C6	ADE	6	14.612	19.793	75.279	1.00	19.06
ATOM	116	N6	ADE	6	15.802	20.122	74.791	1.00	19.47
ATOM	117	N1	ADE	6	13.646	20.756	75.266	1.00	18.04
ATOM	118	C2	ADE	6	12.391	20.557	75.768	1.00	21.28
ATOM	119	N3	ADE	6	11.945	19.416	76.282	1.00	20.29
ATOM	120	C4	ADE	6	12.894	18.457	76.256	1.00	18.80
ATOM	121	P	THY	7	7.702	14.457	76.919	1.00	29.58
ATOM	122	O1P	THY	7	8.318	13.353	76.125	1.00	27.73
ATOM	123	O2P	THY	7	6.466	14.027	77.661	1.00	30.79
ATOM	124	O5'	THY	7	7.252	15.647	75.865	1.00	30.76
ATOM	125	C5'	THY	7	7.284	17.007	76.353	1.00	29.28
ATOM	126	C4'	THY	7	7.540	17.909	75.169	1.00	28.50
ATOM	127	O3'	THY	7	5.827	18.250	73.503	1.00	29.90
ATOM	128	C3'	THY	7	6.926	17.442	73.841	1.00	29.34
ATOM	129	C2'	THY	7	8.047	17.523	72.807	1.00	27.99
ATOM	130	C1'	THY	7	9.048	18.392	73.484	1.00	28.73
ATOM	131	O4'	THY	7	8.922	18.084	74.928	1.00	28.91
ATOM	132	N1	THY	7	10.455	18.201	73.106	1.00	26.72
ATOM	133	C6	THY	7	11.152	17.044	73.249	1.00	22.60
ATOM	134	C5	THY	7	12.448	16.947	72.917	1.00	20.57
ATOM	135	C4	THY	7	13.128	18.096	72.442	1.00	21.42
ATOM	136	O4	THY	7	14.355	18.100	72.117	1.00	22.23
ATOM	137	N3	THY	7	12.411	19.245	72.319	1.00	21.80
ATOM	138	C2	THY	7	11.103	19.355	72.644	1.00	25.34
ATOM	139	O2	THY	7	10.474	20.415	72.521	1.00	27.85
ATOM	140	M5	THY	7	13.185	15.631	73.100	1.00	24.02
ATOM	141	P	THY	8	4.746	18.100	72.345	1.00	29.53
ATOM	142	O1P	THY	8	4.790	16.752	71.857	1.00	32.33
ATOM	143	O2P	THY	8	3.467	18.693	72.885	1.00	39.67
ATOM	144	O5'	THY	8	5.337	19.148	71.239	1.00	32.26
ATOM	145	C5'	THY	8	5.300	20.528	71.720	1.00	27.43
ATOM	146	C4'	THY	8	6.057	21.300	70.653	1.00	28.02
ATOM	147	O3'	THY	8	5.147	22.526	68.740	1.00	31.04
ATOM	148	C3'	THY	8	5.514	21.211	69.234	1.00	27.45
ATOM	149	C2'	THY	8	6.655	20.569	68.434	1.00	26.13
ATOM	150	C1'	THY	8	7.860	20.983	69.241	1.00	25.21
ATOM	151	O4'	THY	8	7.372	20.821	70.640	1.00	25.05
ATOM	152	N1	THY	8	9.016	20.086	69.150	1.00	23.26
ATOM	153	C6	THY	8	8.898	18.770	69.508	1.00	22.63
ATOM	154	C5	THY	8	9.920	17.913	69.462	1.00	21.21
ATOM	155	C4	THY	8	11.199	18.421	69.072	1.00	21.77
ATOM	156	O4	THY	8	12.221	17.678	69.000	1.00	20.24
ATOM	157	N3	THY	8	11.290	19.728	68.727	1.00	20.05
ATOM	158	C2	THY	8	10.240	20.585	68.759	1.00	21.55
ATOM	159	O2	THY	8	10.349	21.783	68.460	1.00	26.09
ATOM	160	M5	THY	8	9.720	16.475	69.859	1.00	18.92

ATOM	161	P	THY	9	4.408	22.681	67.315	1.00	32.44
ATOM	162	O1P	THY	9	4.078	21.263	66.924	1.00	38.23
ATOM	163	O2P	THY	9	3.245	23.562	67.419	1.00	43.23
ATOM	164	O5'	THY	9	5.537	23.261	66.378	1.00	33.13
ATOM	165	C5'	THY	9	6.313	24.398	66.872	1.00	33.89
ATOM	166	C4'	THY	9	7.495	24.407	65.909	1.00	34.40
ATOM	167	O3'	THY	9	7.261	25.544	63.821	1.00	38.86
ATOM	168	C3'	THY	9	7.079	24.285	64.439	1.00	34.36
ATOM	169	C2'	THY	9	7.951	23.180	63.873	1.00	34.81
ATOM	170	C1'	THY	9	8.986	22.953	64.927	1.00	30.45
ATOM	171	O4'	THY	9	8.323	23.310	66.176	1.00	32.87
ATOM	172	N1	THY	9	9.376	21.535	65.077	1.00	27.42
ATOM	173	C6	THY	9	8.486	20.541	65.356	1.00	24.11
ATOM	174	C5	THY	9	8.885	19.274	65.512	1.00	23.74
ATOM	175	C4	THY	9	10.275	18.989	65.421	1.00	20.13
ATOM	176	O4	THY	9	10.743	17.820	65.565	1.00	25.95
ATOM	177	N3	THY	9	11.147	19.988	65.155	1.00	22.18
ATOM	178	C2	THY	9	10.741	21.276	64.998	1.00	23.37
ATOM	179	O2	THY	9	11.524	22.185	64.777	1.00	25.36
ATOM	180	M5	THY	9	7.905	18.153	65.799	1.00	22.22
ATOM	181	P	GUA	10	6.670	26.141	62.506	1.00	45.48
ATOM	182	O1P	GUA	10	5.500	25.337	62.148	1.00	43.99
ATOM	183	O2P	GUA	10	6.414	27.619	62.695	1.00	46.20
ATOM	184	O5'	GUA	10	7.905	25.946	61.452	1.00	39.98
ATOM	185	C5'	GUA	10	9.171	26.543	61.882	1.00	37.81
ATOM	186	C4'	GUA	10	10.198	25.905	60.958	1.00	39.04
ATOM	187	O3'	GUA	10	10.864	25.657	58.602	1.00	45.10
ATOM	188	C3'	GUA	10	9.765	25.800	59.507	1.00	36.89
ATOM	189	C2'	GUA	10	9.016	24.447	59.546	1.00	36.75
ATOM	190	C1'	GUA	10	10.162	23.708	60.242	1.00	32.70
ATOM	191	O4'	GUA	10	10.433	24.557	61.381	1.00	36.19
ATOM	192	N9	GUA	10	9.748	22.376	60.678	1.00	29.88
ATOM	193	C8	GUA	10	8.486	21.873	60.821	1.00	28.33
ATOM	194	N7	GUA	10	8.466	20.646	61.257	1.00	31.48
ATOM	195	C5	GUA	10	9.812	20.301	61.407	1.00	26.89
ATOM	196	C6	GUA	10	10.440	19.115	61.843	1.00	25.13
ATOM	197	O6	GUA	10	9.881	18.076	62.200	1.00	28.74
ATOM	198	N1	GUA	10	11.815	19.152	61.875	1.00	24.26
ATOM	199	C2	GUA	10	12.495	20.268	61.491	1.00	25.47
ATOM	200	N2	GUA	10	13.828	20.191	61.543	1.00	24.12
ATOM	201	N3	GUA	10	11.958	21.426	61.088	1.00	26.28
ATOM	202	C4	GUA	10	10.610	21.369	61.068	1.00	28.33
ATOM	203	P	CYT	11	11.411	26.843	57.672	1.00	48.42
ATOM	204	O1P	CYT	11	10.627	26.669	56.435	1.00	53.65
ATOM	205	O2P	CYT	11	11.191	28.167	58.381	1.00	55.89
ATOM	206	O5'	CYT	11	12.975	26.583	57.522	1.00	43.89
ATOM	207	C5'	CYT	11	13.855	26.892	58.654	1.00	36.91
ATOM	208	C4'	CYT	11	14.924	25.804	58.615	1.00	35.70
ATOM	209	O3'	CYT	11	17.103	25.584	57.554	1.00	35.73
ATOM	210	C3'	CYT	11	15.693	25.698	57.294	1.00	35.00
ATOM	211	C2'	CYT	11	15.072	24.476	56.624	1.00	33.47
ATOM	212	C1'	CYT	11	14.735	23.598	57.789	1.00	31.78
ATOM	213	O4'	CYT	11	14.333	24.549	58.830	1.00	33.54
ATOM	214	N1	CYT	11	13.591	22.693	57.652	1.00	31.24
ATOM	215	C6	CYT	11	12.350	23.107	57.236	1.00	30.00
ATOM	216	C5	CYT	11	11.297	22.279	57.144	1.00	26.58

ATOM	217	C4	CYT	11	11.499	20.926	57.522	1.00	26.88
ATOM	218	N4	CYT	11	10.521	20.013	57.496	1.00	26.53
ATOM	219	N3	CYT	11	12.724	20.512	57.932	1.00	24.86
ATOM	220	C2	CYT	11	13.786	21.369	58.023	1.00	28.12
ATOM	221	O2	CYT	11	14.892	20.979	58.387	1.00	29.90
ATOM	222	P	GUA	12	18.268	25.763	56.487	1.00	41.28
ATOM	223	O1P	GUA	12	17.763	26.652	55.427	1.00	45.10
ATOM	224	O2P	GUA	12	19.515	26.230	57.118	1.00	48.85
ATOM	225	O5'	GUA	12	18.389	24.248	55.882	1.00	46.39
ATOM	226	C5'	GUA	12	19.197	23.253	56.546	1.00	38.74
ATOM	227	C4'	GUA	12	19.158	22.023	55.667	1.00	39.09
ATOM	228	O3'	GUA	12	20.624	21.361	53.800	1.00	39.44
ATOM	229	C3'	GUA	12	19.500	22.193	54.210	1.00	37.82
ATOM	230	C2'	GUA	12	18.236	21.755	53.455	1.00	36.31
ATOM	231	C1'	GUA	12	17.571	20.825	54.431	1.00	32.69
ATOM	232	O4'	GUA	12	17.874	21.442	55.726	1.00	37.01
ATOM	233	N9	GUA	12	16.105	20.841	54.353	1.00	30.06
ATOM	234	C8	GUA	12	15.314	21.933	54.021	1.00	31.99
ATOM	235	N7	GUA	12	14.037	21.690	54.054	1.00	29.06
ATOM	236	C5	GUA	12	13.968	20.362	54.451	1.00	28.78
ATOM	237	C6	GUA	12	12.860	19.513	54.685	1.00	27.48
ATOM	238	O6	GUA	12	11.672	19.801	54.555	1.00	29.67
ATOM	239	N1	GUA	12	13.200	18.234	55.049	1.00	27.90
ATOM	240	C2	GUA	12	14.483	17.828	55.205	1.00	25.22
ATOM	241	N2	GUA	12	14.634	16.561	55.576	1.00	26.21
ATOM	242	N3	GUA	12	15.560	18.583	55.010	1.00	28.94
ATOM	243	C4	GUA	12	15.237	19.834	54.633	1.00	27.51
ATOM	244	O5'	CYT	13	7.648	11.221	55.518	1.00	50.00
ATOM	245	C5'	CYT	13	8.671	10.437	54.834	1.00	43.65
ATOM	246	C4'	CYT	13	9.982	10.672	55.563	1.00	42.40
ATOM	247	O3'	CYT	13	11.219	10.522	57.665	1.00	46.15
ATOM	248	C3'	CYT	13	9.935	10.790	57.079	1.00	41.91
ATOM	249	C2'	CYT	13	9.558	12.280	57.223	1.00	38.91
ATOM	250	C1'	CYT	13	10.585	12.825	56.247	1.00	36.79
ATOM	251	O4'	CYT	13	10.553	11.895	55.153	1.00	37.94
ATOM	252	N1	CYT	13	10.250	14.169	55.745	1.00	30.27
ATOM	253	C6	CYT	13	9.008	14.494	55.283	1.00	31.91
ATOM	254	C5	CYT	13	8.727	15.724	54.802	1.00	34.65
ATOM	255	C4	CYT	13	9.782	16.683	54.795	1.00	32.86
ATOM	256	N4	CYT	13	9.548	17.913	54.333	1.00	36.75
ATOM	257	N3	CYT	13	11.017	16.350	55.251	1.00	31.66
ATOM	258	C2	CYT	13	11.275	15.078	55.726	1.00	30.24
ATOM	259	O2	CYT	13	12.391	14.758	56.123	1.00	31.39
ATOM	260	P	GUA	14	11.334	9.231	58.713	1.00	52.43
ATOM	261	O1P	GUA	14	10.398	9.608	59.754	1.00	53.09
ATOM	262	O2P	GUA	14	11.098	8.037	57.867	1.00	51.02
ATOM	263	O5'	GUA	14	12.887	9.320	59.149	1.00	43.04
ATOM	264	C5'	GUA	14	13.796	9.604	58.036	1.00	37.32
ATOM	265	C4'	GUA	14	14.690	10.725	58.543	1.00	35.27
ATOM	266	O3'	GUA	14	16.334	10.636	60.326	1.00	33.22
ATOM	267	C3'	GUA	14	14.944	10.701	60.040	1.00	33.80
ATOM	268	C2'	GUA	14	14.281	11.980	60.541	1.00	33.65
ATOM	269	C1'	GUA	14	14.449	12.890	59.324	1.00	32.17
ATOM	270	O4'	GUA	14	14.170	11.976	58.218	1.00	32.31
ATOM	271	N9	GUA	14	13.347	13.868	59.331	1.00	31.48
ATOM	272	C8	GUA	14	12.017	13.568	59.546	1.00	28.45

ATOM	273	N7	GUA	14	11.236	14.595	59.487	1.00	28.01
ATOM	274	C5	GUA	14	12.103	15.651	59.214	1.00	28.82
ATOM	275	C6	GUA	14	11.857	17.036	59.038	1.00	29.56
ATOM	276	O6	GUA	14	10.760	17.588	59.090	1.00	29.20
ATOM	277	N1	GUA	14	12.998	17.767	58.784	1.00	28.07
ATOM	278	C2	GUA	14	14.235	17.223	58.713	1.00	25.60
ATOM	279	N2	GUA	14	15.210	18.108	58.485	1.00	26.93
ATOM	280	N3	GUA	14	14.530	15.931	58.869	1.00	25.65
ATOM	281	C4	GUA	14	13.412	15.213	59.123	1.00	28.94
ATOM	282	P	CYT	15	16.984	10.534	61.784	1.00	34.69
ATOM	283	O1P	CYT	15	15.910	10.051	62.649	1.00	39.70
ATOM	284	O2P	CYT	15	18.197	9.645	61.719	1.00	45.82
ATOM	285	O5'	CYT	15	17.512	12.012	62.070	1.00	34.12
ATOM	286	C5'	CYT	15	18.480	12.642	61.218	1.00	26.61
ATOM	287	C4'	CYT	15	18.426	14.116	61.556	1.00	29.83
ATOM	288	O3'	CYT	15	20.064	14.957	63.137	1.00	32.14
ATOM	289	C3'	CYT	15	18.717	14.482	63.001	1.00	28.65
ATOM	290	C2'	CYT	15	17.711	15.541	63.372	1.00	27.35
ATOM	291	C1'	CYT	15	16.861	15.708	62.174	1.00	24.70
ATOM	292	O4'	CYT	15	17.152	14.628	61.263	1.00	28.09
ATOM	293	N1	CYT	15	15.402	15.615	62.454	1.00	25.25
ATOM	294	C6	CYT	15	14.784	14.465	62.877	1.00	25.08
ATOM	295	C5	CYT	15	13.436	14.429	63.066	1.00	27.54
ATOM	296	C4	CYT	15	12.699	15.627	62.812	1.00	22.06
ATOM	297	N4	CYT	15	11.386	15.696	62.968	1.00	23.51
ATOM	298	N3	CYT	15	13.347	16.739	62.415	1.00	21.51
ATOM	299	C2	CYT	15	14.695	16.760	62.200	1.00	20.97
ATOM	300	O2	CYT	15	15.284	17.779	61.836	1.00	22.05
ATOM	301	P	ADE	16	20.717	15.054	64.582	1.00	31.99
ATOM	302	O1P	ADE	16	19.663	14.559	65.499	1.00	36.91
ATOM	303	O2P	ADE	16	22.006	14.339	64.686	1.00	44.68
ATOM	304	O5'	ADE	16	20.951	16.618	64.712	1.00	27.87
ATOM	305	C5'	ADE	16	21.353	17.263	63.469	1.00	25.53
ATOM	306	C4'	ADE	16	21.013	18.717	63.678	1.00	29.54
ATOM	307	O3'	ADE	16	22.117	20.597	64.725	1.00	37.43
ATOM	308	C3'	ADE	16	21.511	19.318	64.992	1.00	32.16
ATOM	309	C2'	ADE	16	20.259	19.391	65.851	1.00	33.15
ATOM	310	C1'	ADE	16	19.121	19.489	64.868	1.00	28.31
ATOM	311	O4'	ADE	16	19.636	18.908	63.638	1.00	32.15
ATOM	312	N9	ADE	16	17.938	18.693	65.252	1.00	26.46
ATOM	313	C8	ADE	16	17.876	17.450	65.799	1.00	26.36
ATOM	314	N7	ADE	16	16.659	17.003	66.007	1.00	27.52
ATOM	315	C5	ADE	16	15.846	18.035	65.565	1.00	25.65
ATOM	316	C6	ADE	16	14.446	18.173	65.499	1.00	24.91
ATOM	317	N6	ADE	16	13.554	17.284	65.877	1.00	23.51
ATOM	318	N1	ADE	16	13.986	19.367	64.972	1.00	24.60
ATOM	319	C2	ADE	16	14.851	20.313	64.543	1.00	24.65
ATOM	320	N3	ADE	16	16.181	20.240	64.569	1.00	27.17
ATOM	321	C4	ADE	16	16.620	19.087	65.090	1.00	26.63
ATOM	322	P	ADE	17	22.228	21.771	65.747	1.00	42.35
ATOM	323	O1P	ADE	17	22.644	21.138	67.003	1.00	52.37
ATOM	324	O2P	ADE	17	23.107	22.900	65.239	1.00	47.85
ATOM	325	O5'	ADE	17	20.754	22.470	65.773	1.00	41.62
ATOM	326	C5'	ADE	17	20.528	23.245	64.543	1.00	37.72
ATOM	327	C4'	ADE	17	19.337	24.110	64.875	1.00	36.08
ATOM	328	O3'	ADE	17	18.768	26.222	65.909	1.00	35.19

ATOM	329	C3'	ADE	17	19.564	25.048	66.079	1.00	34.13
ATOM	330	C2'	ADE	17	19.121	24.151	67.243	1.00	33.78
ATOM	331	C1'	ADE	17	17.886	23.529	66.612	1.00	31.34
ATOM	332	O4'	ADE	17	18.268	23.286	65.246	1.00	35.60
ATOM	333	N9	ADE	17	17.492	22.275	67.276	1.00	27.10
ATOM	334	C8	ADE	17	18.248	21.284	67.829	1.00	27.08
ATOM	335	N7	ADE	17	17.576	20.301	68.343	1.00	25.36
ATOM	336	C5	ADE	17	16.248	20.666	68.089	1.00	25.90
ATOM	337	C6	ADE	17	15.033	20.021	68.389	1.00	23.81
ATOM	338	N6	ADE	17	14.917	18.847	69.000	1.00	23.50
ATOM	339	N1	ADE	17	13.914	20.683	67.985	1.00	23.52
ATOM	340	C2	ADE	17	13.968	21.901	67.360	1.00	24.89
ATOM	341	N3	ADE	17	15.072	22.559	67.042	1.00	25.88
ATOM	342	C4	ADE	17	16.186	21.877	67.439	1.00	26.48
ATOM	343	P	ADE	18	18.786	27.513	66.846	1.00	43.12
ATOM	344	O1P	ADE	18	19.850	27.237	67.835	1.00	39.66
ATOM	345	O2P	ADE	18	18.916	28.772	66.072	1.00	50.19
ATOM	346	O5'	ADE	18	17.285	27.505	67.491	1.00	37.31
ATOM	347	C5'	ADE	18	16.225	27.233	66.502	1.00	32.58
ATOM	348	C4'	ADE	18	14.952	27.290	67.308	1.00	31.61
ATOM	349	O3'	ADE	18	13.663	28.768	68.694	1.00	37.59
ATOM	350	C3'	ADE	18	14.984	28.273	68.499	1.00	32.69
ATOM	351	C2'	ADE	18	15.439	27.363	69.644	1.00	27.79
ATOM	352	C1'	ADE	18	14.710	26.104	69.306	1.00	25.61
ATOM	353	O4'	ADE	18	14.621	26.047	67.862	1.00	27.62
ATOM	354	N9	ADE	18	15.358	24.898	69.853	1.00	26.08
ATOM	355	C8	ADE	18	16.669	24.565	70.048	1.00	19.48
ATOM	356	N7	ADE	18	16.859	23.404	70.575	1.00	20.91
ATOM	357	C5	ADE	18	15.570	22.904	70.764	1.00	22.56
ATOM	358	C6	ADE	18	15.097	21.682	71.291	1.00	19.89
ATOM	359	N6	ADE	18	15.809	20.666	71.740	1.00	22.42
ATOM	360	N1	ADE	18	13.732	21.580	71.310	1.00	20.47
ATOM	361	C2	ADE	18	12.906	22.539	70.848	1.00	20.52
ATOM	362	N3	ADE	18	13.293	23.700	70.321	1.00	21.20
ATOM	363	C4	ADE	18	14.636	23.810	70.328	1.00	19.42
ATOM	364	P	THY	19	13.175	29.869	69.716	1.00	42.10
ATOM	365	O1P	THY	19	14.345	30.685	70.093	1.00	35.35
ATOM	366	O2P	THY	19	12.071	30.567	68.974	1.00	36.47
ATOM	367	O5'	THY	19	12.564	29.061	70.959	1.00	36.46
ATOM	368	C5'	THY	19	11.219	28.553	70.588	1.00	34.52
ATOM	369	C4'	THY	19	10.943	27.554	71.694	1.00	34.17
ATOM	370	O3'	THY	19	9.506	28.447	73.418	1.00	39.26
ATOM	371	C3'	THY	19	10.866	28.208	73.087	1.00	34.27
ATOM	372	C2'	THY	19	11.588	27.233	73.998	1.00	30.91
ATOM	373	C1'	THY	19	11.896	26.055	73.126	1.00	28.56
ATOM	374	O4'	THY	19	11.990	26.624	71.779	1.00	29.17
ATOM	375	N1	THY	19	13.138	25.349	73.464	1.00	25.61
ATOM	376	C6	THY	19	14.407	25.812	73.243	1.00	22.28
ATOM	377	C5	THY	19	15.496	25.105	73.588	1.00	24.71
ATOM	378	C4	THY	19	15.324	23.802	74.141	1.00	22.27
ATOM	379	O4	THY	19	16.253	23.034	74.499	1.00	21.94
ATOM	380	N3	THY	19	14.055	23.371	74.329	1.00	24.34
ATOM	381	C2	THY	19	12.943	24.082	74.011	1.00	26.88
ATOM	382	O2	THY	19	11.820	23.607	74.199	1.00	27.39
ATOM	383	M5	THY	19	16.906	25.621	73.386	1.00	26.01
ATOM	384	P	THY	20	8.836	28.906	74.746	1.00	41.37

ATOM	385	O1P	THY	20	9.745	29.889	75.371	1.00	44.11
ATOM	386	O2P	THY	20	7.483	29.495	74.414	1.00	52.41
ATOM	387	O5'	THY	20	8.673	27.611	75.657	1.00	40.71
ATOM	388	C5'	THY	20	7.705	26.579	75.312	1.00	36.24
ATOM	389	C4'	THY	20	7.909	25.475	76.340	1.00	34.45
ATOM	390	O3'	THY	20	6.682	25.259	78.370	1.00	38.79
ATOM	391	C3'	THY	20	7.764	25.954	77.785	1.00	32.87
ATOM	392	C2'	THY	20	9.100	25.694	78.429	1.00	32.70
ATOM	393	C1'	THY	20	9.681	24.606	77.570	1.00	29.83
ATOM	394	O4'	THY	20	9.173	24.890	76.243	1.00	32.93
ATOM	395	N1	THY	20	11.155	24.662	77.550	1.00	27.85
ATOM	396	C6	THY	20	11.874	25.779	77.232	1.00	24.17
ATOM	397	C5	THY	20	13.219	25.775	77.212	1.00	25.49
ATOM	398	C4	THY	20	13.909	24.573	77.518	1.00	23.86
ATOM	399	O4	THY	20	15.166	24.459	77.544	1.00	26.01
ATOM	400	N3	THY	20	13.170	23.485	77.837	1.00	26.45
ATOM	401	C2	THY	20	11.803	23.477	77.856	1.00	26.30
ATOM	402	O2	THY	20	11.167	22.478	78.143	1.00	27.18
ATOM	403	M5	THY	20	13.971	27.030	76.848	1.00	26.98
ATOM	404	P	THY	21	6.000	25.499	79.802	1.00	36.76
ATOM	405	O1P	THY	21	6.365	26.807	80.264	1.00	36.79
ATOM	406	O2P	THY	21	4.526	25.284	79.496	1.00	46.69
ATOM	407	O5'	THY	21	6.623	24.317	80.635	1.00	33.19
ATOM	408	C5'	THY	21	6.567	22.928	80.251	1.00	31.09
ATOM	409	C4'	THY	21	7.582	22.291	81.207	1.00	30.92
ATOM	410	O3'	THY	21	6.719	21.661	83.355	1.00	33.94
ATOM	411	C3'	THY	21	7.377	22.729	82.658	1.00	30.07
ATOM	412	C2'	THY	21	8.782	23.010	83.185	1.00	30.23
ATOM	413	C1'	THY	21	9.701	22.551	82.099	1.00	26.10
ATOM	414	O4'	THY	21	8.898	22.636	80.862	1.00	26.57
ATOM	415	N1	THY	21	10.856	23.395	81.786	1.00	24.40
ATOM	416	C6	THY	21	10.723	24.719	81.448	1.00	23.57
ATOM	417	C5	THY	21	11.778	25.475	81.116	1.00	22.35
ATOM	418	C4	THY	21	13.057	24.870	81.090	1.00	22.23
ATOM	419	O4	THY	21	14.121	25.487	80.797	1.00	24.14
ATOM	420	N3	THY	21	13.180	23.546	81.416	1.00	23.62
ATOM	421	C2	THY	21	12.101	22.786	81.767	1.00	23.44
ATOM	422	O2	THY	21	12.226	21.605	82.060	1.00	25.80
ATOM	423	M5	THY	21	11.630	26.924	80.726	1.00	27.19
ATOM	424	P	GUA	22	6.187	21.718	84.851	1.00	37.23
ATOM	425	O1P	GUA	22	5.886	23.172	85.053	1.00	37.17
ATOM	426	O2P	GUA	22	5.022	20.800	84.897	1.00	38.85
ATOM	427	O5'	GUA	22	7.382	21.247	85.801	1.00	36.46
ATOM	428	C5'	GUA	22	7.900	19.883	85.723	1.00	31.77
ATOM	429	C4'	GUA	22	9.252	19.927	86.406	1.00	31.97
ATOM	430	O3'	GUA	22	10.250	19.728	88.567	1.00	37.62
ATOM	431	C3'	GUA	22	9.252	20.467	87.825	1.00	32.64
ATOM	432	C2'	GUA	22	9.669	21.933	87.604	1.00	30.35
ATOM	433	C1'	GUA	22	10.824	21.645	86.660	1.00	26.33
ATOM	434	O4'	GUA	22	10.216	20.699	85.736	1.00	31.15
ATOM	435	N9	GUA	22	11.364	22.843	86.003	1.00	21.72
ATOM	436	C8	GUA	22	10.741	24.033	85.710	1.00	21.11
ATOM	437	N7	GUA	22	11.519	24.930	85.125	1.00	21.77
ATOM	438	C5	GUA	22	12.766	24.281	85.079	1.00	18.04
ATOM	439	C6	GUA	22	14.008	24.699	84.565	1.00	18.81
ATOM	440	O6	GUA	22	14.279	25.779	84.012	1.00	20.19

ATOM	441	N1	GUA	22	15.021	23.749	84.676	1.00	19.13
ATOM	442	C2	GUA	22	14.806	22.547	85.248	1.00	19.70
ATOM	443	N2	GUA	22	15.858	21.734	85.287	1.00	21.03
ATOM	444	N3	GUA	22	13.646	22.080	85.723	1.00	19.51
ATOM	445	C4	GUA	22	12.667	23.010	85.600	1.00	18.86
ATOM	446	P	CYT	23	9.785	19.018	89.946	1.00	47.42
ATOM	447	O1P	CYT	23	8.870	17.986	89.419	1.00	49.16
ATOM	448	O2P	CYT	23	9.311	20.126	90.818	1.00	40.17
ATOM	449	O5'	CYT	23	11.125	18.356	90.532	1.00	46.58
ATOM	450	C5'	CYT	23	11.913	17.511	89.634	1.00	40.00
ATOM	451	C4'	CYT	23	13.338	18.023	89.823	1.00	39.62
ATOM	452	O3'	CYT	23	14.631	17.519	91.859	1.00	40.63
ATOM	453	C3'	CYT	23	13.692	18.409	91.248	1.00	37.91
ATOM	454	C2'	CYT	23	14.264	19.805	91.163	1.00	36.87
ATOM	455	C1'	CYT	23	14.424	20.041	89.692	1.00	33.65
ATOM	456	O4'	CYT	23	13.456	19.209	89.048	1.00	38.19
ATOM	457	N1	CYT	23	14.124	21.450	89.380	1.00	33.12
ATOM	458	C6	CYT	23	12.894	22.023	89.562	1.00	31.66
ATOM	459	C5	CYT	23	12.658	23.306	89.257	1.00	26.89
ATOM	460	C4	CYT	23	13.737	24.065	88.723	1.00	23.91
ATOM	461	N4	CYT	23	13.562	25.353	88.417	1.00	20.58
ATOM	462	N3	CYT	23	14.949	23.489	88.547	1.00	25.88
ATOM	463	C2	CYT	23	15.178	22.181	88.879	1.00	27.79
ATOM	464	O2	CYT	23	16.280	21.682	88.703	1.00	27.14
ATOM	465	P	GUA	24	14.833	17.629	93.454	1.00	39.77
ATOM	466	O1P	GUA	24	13.781	18.628	93.844	1.00	45.78
ATOM	467	O2P	GUA	24	14.688	16.321	94.091	1.00	34.40
ATOM	468	O5'	GUA	24	16.326	18.246	93.623	1.00	34.50
ATOM	469	C5'	GUA	24	17.457	17.479	93.089	1.00	28.26
ATOM	470	C4'	GUA	24	18.574	18.502	92.946	1.00	29.50
ATOM	471	O3'	GUA	24	19.690	18.737	95.152	1.00	27.71
ATOM	472	C3'	GUA	24	18.800	19.343	94.202	1.00	24.97
ATOM	473	C2'	GUA	24	19.345	20.646	93.655	1.00	27.24
ATOM	474	C1'	GUA	24	18.736	20.735	92.282	1.00	26.71
ATOM	475	O4'	GUA	24	18.251	19.424	91.931	1.00	28.38
ATOM	476	N9	GUA	24	17.645	21.714	92.230	1.00	27.93
ATOM	477	C8	GUA	24	16.331	21.564	92.634	1.00	28.60
ATOM	478	N7	GUA	24	15.590	22.624	92.451	1.00	26.14
ATOM	479	C5	GUA	24	16.472	23.542	91.879	1.00	28.28
ATOM	480	C6	GUA	24	16.290	24.878	91.423	1.00	27.40
ATOM	481	O6	GUA	24	15.264	25.556	91.462	1.00	30.06
ATOM	482	N1	GUA	24	17.418	25.471	90.948	1.00	26.06
ATOM	483	C2	GUA	24	18.613	24.837	90.870	1.00	25.36
ATOM	484	N2	GUA	24	19.621	25.548	90.350	1.00	24.40
ATOM	485	N3	GUA	24	18.842	23.586	91.254	1.00	28.93
ATOM	486	C4	GUA	24	17.728	23.002	91.755	1.00	27.56
ATOM	487	N12	BER	25	6.722	22.693	74.401	1.00	33.76
ATOM	488	C1	BER	25	7.929	24.390	71.662	1.00	35.75
ATOM	489	N36	BER	25	7.658	18.920	79.398	1.00	29.32
ATOM	490	C62	BER	25	6.389	21.531	76.431	1.00	34.69
ATOM	491	C3	BER	25	8.190	25.714	69.683	1.00	36.03
ATOM	492	C4	BER	25	9.390	25.101	69.404	1.00	37.00
ATOM	493	C12	BER	25	7.298	21.929	75.468	1.00	34.11
ATOM	494	C19	BER	25	10.125	25.458	68.291	1.00	39.56
ATOM	495	C42	BER	25	8.161	20.467	77.648	1.00	31.92
ATOM	496	C2	BER	25	7.449	25.373	70.790	1.00	37.56

ATOM	497	N32	BER	25	9.959	26.725	67.757	1.00	39.64
ATOM	498	C5	BER	25	9.846	24.122	70.282	1.00	36.70
ATOM	499	C32	BER	25	9.063	20.894	76.685	1.00	31.58
ATOM	500	C21	BER	25	8.570	19.732	78.735	1.00	33.06
ATOM	501	N2	BER	25	7.631	23.326	73.770	1.00	35.50
ATOM	502	C22	BER	25	8.644	21.605	75.592	1.00	33.25
ATOM	503	N20	BER	25	9.923	19.578	78.988	1.00	32.17
ATOM	504	N1	BER	25	7.138	24.078	72.774	1.00	35.33
ATOM	505	C52	BER	25	6.830	20.792	77.524	1.00	32.90
ATOM	506	C6	BER	25	9.129	23.765	71.414	1.00	35.68
ATOM	507	N10	BER	25	11.118	24.626	67.822	1.00	38.86
ATOM	508	MG	BER	26	14.131	29.909	84.793	1.00	36.32
ATOM	509	O	WAT	27	19.505	27.038	70.386	1.00	57.38
ATOM	510	O	WAT	28	21.365	25.978	58.797	1.00	65.69
ATOM	511	O	WAT	29	7.707	16.760	60.801	1.00	53.06
ATOM	512	O	WAT	30	16.849	31.274	65.961	1.00	61.49
ATOM	513	O	WAT	31	17.041	16.207	74.837	1.00	45.40
ATOM	514	O	WAT	32	3.965	25.816	76.112	1.00	52.20
ATOM	515	O	WAT	33	12.574	21.211	93.811	1.00	44.09
ATOM	516	O	WAT	34	13.020	13.893	66.287	1.00	43.74
ATOM	517	O	WAT	35	23.147	23.712	79.184	1.00	49.66
ATOM	518	O	WAT	36	9.955	13.669	73.685	1.00	37.12
ATOM	519	O	WAT	37	17.635	29.556	54.229	1.00	49.10
ATOM	520	O	WAT	38	16.053	12.354	65.200	1.00	45.04
ATOM	521	O	WAT	39	18.002	14.027	67.647	1.00	59.13
ATOM	522	O	WAT	40	16.790	31.972	68.447	1.00	51.69
ATOM	523	O	WAT	41	10.159	10.810	79.418	1.00	32.11
ATOM	524	O	WAT	42	20.429	31.379	83.732	1.00	45.25
ATOM	525	O	WAT	43	5.655	20.313	89.784	1.00	52.61
ATOM	526	O	WAT	44	11.879	18.770	83.088	1.00	63.29
ATOM	527	O	WAT	45	9.035	22.047	52.525	1.00	60.74
ATOM	528	O	WAT	46	12.781	6.981	56.351	1.00	75.42
ATOM	529	O	WAT	47	21.136	28.756	57.535	1.00	58.68
ATOM	530	O	WAT	48	10.366	13.279	63.983	1.00	49.24
ATOM	531	O	WAT	49	10.026	15.347	66.176	1.00	55.32
ATOM	532	O	WAT	50	24.810	24.005	66.879	1.00	53.49
ATOM	533	O	WAT	51	21.289	29.816	68.870	1.00	51.77
ATOM	534	O	WAT	52	15.309	28.991	72.943	1.00	44.50
ATOM	535	O	WAT	53	19.165	17.690	75.052	1.00	40.20
ATOM	536	O	WAT	54	14.417	31.696	85.391	1.00	34.12
ATOM	537	O	WAT	55	13.983	28.496	86.179	1.00	31.58
ATOM	538	O	WAT	56	10.647	26.271	88.664	1.00	45.57
ATOM	539	O	WAT	57	24.428	35.319	87.675	1.00	80.81
ATOM	540	O	WAT	58	11.443	22.807	52.642	1.00	55.78
ATOM	541	O	WAT	59	12.835	29.077	60.066	1.00	38.84
ATOM	542	O	WAT	60	12.968	9.848	63.534	1.00	76.17
ATOM	543	O	WAT	61	23.898	21.584	62.754	1.00	44.34
ATOM	544	O	WAT	62	18.608	17.905	69.137	1.00	48.54
ATOM	545	O	WAT	63	13.550	15.282	69.039	1.00	48.33
ATOM	546	O	WAT	64	3.997	14.758	73.158	1.00	50.67
ATOM	547	O	WAT	65	3.023	15.834	76.164	1.00	66.19
ATOM	548	O	WAT	66	8.220	9.616	77.947	1.00	29.42
ATOM	549	O	WAT	67	9.063	18.559	81.845	1.00	54.83
ATOM	550	O	WAT	68	11.487	10.071	84.044	1.00	68.98
ATOM	551	O	WAT	69	10.674	27.339	84.532	1.00	42.42
ATOM	552	O	WAT	70	2.420	20.325	84.500	1.00	57.45

ATOM	553	O	WAT	71	15.720	18.348	83.420	1.00	45.77
ATOM	554	O	WAT	72	16.053	16.370	72.267	1.00	64.58
ATOM	555	O	WAT	73	11.982	26.445	65.311	1.00	36.71
ATOM	556	O	WAT	74	13.412	19.708	85.801	1.00	36.78
ATOM	557	O	WAT	75	17.664	16.532	84.890	1.00	46.93
ATOM	558	O	WAT	76	19.498	11.363	65.317	1.00	44.10
ATOM	559	O	WAT	77	12.115	12.045	74.323	1.00	45.29
ATOM	560	O	WAT	78	12.813	30.413	75.019	1.00	47.97
ATOM	561	O	WAT	79	14.089	13.011	75.872	1.00	37.52
ATOM	562	O	WAT	80	6.653	11.724	78.722	1.00	41.62
ATOM	563	O	WAT	81	14.952	27.992	80.602	1.00	55.26
ATOM	564	O	WAT	82	2.937	21.990	82.776	1.00	56.60
ATOM	565	O	WAT	83	14.491	28.301	83.146	1.00	36.35
ATOM	566	O	WAT	84	27.656	28.053	85.131	1.00	54.22
ATOM	567	O	WAT	85	9.831	22.392	91.749	1.00	47.62
ATOM	568	O	WAT	86	16.267	24.512	53.292	1.00	65.48
ATOM	569	O	WAT	87	5.214	23.127	61.237	1.00	56.44
ATOM	570	O	WAT	88	13.867	24.204	64.549	1.00	59.77
ATOM	571	O	WAT	89	11.928	13.633	70.295	1.00	53.09
ATOM	572	O	WAT	90	4.063	24.451	83.426	1.00	54.08
ATOM	573	O	WAT	91	21.555	14.652	84.812	1.00	68.28

---

The coordinates listed are angstrom coordinates for one asymmetric unit.

The unit cell is an orthogonal space group  $P2_12_12_1$  with dimensions  $a = 24.64$ ,  $b = 40.61$ ,  $c = 65.07 \text{ \AA}$ .

Occ is the occupancy which was not refined.

B(Iso) is the isotropic temperature factor.

---

## CHAPTER 4.

Nuclear magnetic resonance spectroscopic and molecular modelling  
studies of the interaction of berenil with the dodecamer  
 $d(\text{CGCAAATTTGCG})_2$ .

#### 4.1 Introduction.

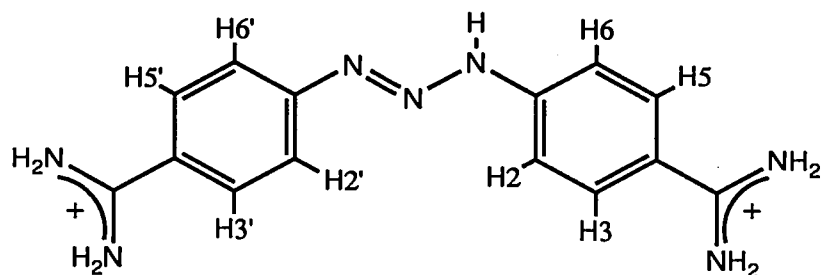
Whilst single crystal X-ray studies remain the main biophysical tool for the three-dimensional structure determination of large macromolecules, it is now possible to investigate the structure of small proteins (<40 kD) and oligonucleotides (<30 base-pairs) using nuclear magnetic resonance spectroscopy (NMR). X-ray crystal structures of biological macromolecules are based on the positions of the elements C, N, O, S and P. The deduced spatial arrangement of these atoms in a crystal environment is basically a time-averaged snapshot of the "frozen" structure. High-frequency thermal vibrations are averaged over the data collection period, and are reflected in the temperature factors (and/or occupancies) of individual atoms in the refined structure. NMR techniques complement X-ray crystallography in that the sample is in a solution state, rather than a crystalline state, which can yield direct, quantitative stoichiometry and dynamic information such as the rate of exchange of protons and association/dissociation rates for interacting ligands within the system.

Molecular dynamics is now often employed to exploit the experimental information from NMR to determine and refine the three-dimensional structure of macromolecules. Full resonance assignments using sequential resonance assignment strategies based on through-bond and through-space connectivities obtained by one-dimensional, two-dimensional and heteronuclear experiments are now possible. These assignments can then, in turn, be used to derive a large set of approximate interproton distances by means of pre-steady-state nuclear Overhauser effect (NOE) measurements (NOE build up). In general, as is the case with macromolecular crystallography, the number of experimental observations is exceeded by the number of variable parameters for the structure and similarly the experimental data have to be augmented with information about the geometry and non-bonded interactions of the macromolecule.

For this research the X-PLOR version 2.1 package (Brünger, 1990) was used for three-dimensional structure determination of the oligonucleotide ligand complex using NMR data.

The interaction of the bis(benzamidinium) compound berenil (figure 4.1) with the AT-rich dodecamer  $d(\text{CGCAAATTTGCG})_2$  has been examined by high-resolution  $^1\text{H}$  and  $^{31}\text{P}$  NMR, optical spectroscopy, and molecular mechanics/dynamics calculations. Extensive assignments for the free DNA and the complex formed with the ligand were obtained using nuclear Overhauser enhancement (NOESY) and total correlation (TOCSY) NMR spectroscopy. Complexation induces large changes in chemical shift for protons situated in the DNA minor groove region associated with the Ade5-Thy9 base segment. In addition, NOEs are observed between Ade5/6(H2) and Thy7/8(H1') protons with the ligand protons, confirming the extent of the binding site within the minor groove. The binding site suggested by the NMR data is too long. However, rapid exchange on the chemical shift timescale indicates ready dissociation of the ligand from the DNA. Calculations of the relative binding enthalpies for the possible A/T binding sites indicate that this dodecamer contains at least two overlapping sites for which the binding affinity for the ligand is similar. The conformation of the DNA is not altered substantially, and the binding of berenil is probably similar to that observed in previous NMR and crystallographic studies reported for this ligand bound to dodecamers (Lane *et al.*, 1991; Brown *et al.*, 1990; chapters 2 and 3).

The high propeller twists in the core region of the dodecamer sequence are maintained, implying that the binding site may be preformed.



Berenil

Figure 4.1. The chemical structure of berenil with the numbering scheme used for this chapter.

## Materials and Methods.

### 4.2 Materials.

The dodecanucleotide d(CGCAAATTTGCG) was prepared on an Applied Biosystems 391-EP synthesiser using phosphoramidite chemistry, and purified by anion-exchange and reverse-phase (C8) HPLC procedures (see chapter 3 for details). The DNA was annealed by heating to 363 K for 5 minutes in 1 mL of a buffer containing 0.1M KCl, 10 mM sodium phosphate, and 0.2 mM EDTA, pH 7.0, and allowing to cool slowly. The solution was then lyophilised, and the duplex redissolved in 0.5 mL of 100% D<sub>2</sub>O, pD\* 7.0, containing 0.1mM 4,4-dimethylsilapentane-1-sulphonate for internal chemical shift referencing. Berenil, as the *N*-acetylglycinate salt, was synthesised as described previously (Lane *et al.*, 1991). Berenil was dissolved in D<sub>2</sub>O containing 0.2M KCl, pD\* 7.0, and the concentration determined by optical absorbance.

## Methods

### 4.3 Optical spectroscopy of DNA-berenil complex.

The ultraviolet absorbance difference spectrum was recorded with 2.5 μM duplex and 123 μM berenil in 5 mM sodium phosphate and 50 mM KCl,

using tandem cuvettes ( $d=0.44$  cm) on a Pye Unicam SP8-100 spectrophotometer. Circular dichroism (CD) spectra were recorded in a 1 cm path length cuvettes using a Jasco Model J-600 spectropolarimeter. The duplex was titrated by addition of small aliquots of buffered berenil solution (5 mM sodium phosphate and 100 mM KCl, pH 7.0 at 298 K), recording the dichroism changes from 500 to 240 nm. Spectra were also obtained for other DNA sequences containing AT tracts (see table 4.1).

Concentrations were determined from absorbances, using extinction coefficients of  $23.1 \text{ mM}^{-1} \text{ cm}^{-1}$  at 380 nm for berenil and  $3.83 \text{ mM}^{-1} \text{ cm}^{-1}$  at 258 nm for the dodecamer duplex.

#### 4.4 NMR spectroscopy.

The optical and NMR spectroscopy were carried out at <sup>the</sup> National Institute for Medical Research, Mill Hill with Dr. Andrew Lane.

$^1\text{H}$  NMR spectra were recorded at 303 K on a Bruker AM400 NMR spectrometer (9.4 T) equipped with a 16-bit digitiser and an ASPECT3000 computer. One-dimensional (1D) spectra were obtained by using a  $90^\circ$  excitation pulse, a 1.87-s acquisition time and a relaxation delay of 1.5 s. The residual HOD signal was suppressed by low-power irradiation from the decoupler, which was phase-coherent with the receiver.

One-dimensional truncated NOEs were obtained using the method of Wagner & Wüthrich (1979) and analysed as described previously (Lane, 1990; Lane *et al.*, 1991a). Phase-sensitive NOESY spectra were recorded by using TPPI (Marion & Wüthrich, 1984) with a 0.21-s acquisition time, a  $t_1(\text{max})$  of 50 ms, and a recycle time of 2 s. A flat base-plane was achieved by adjusting the receiver phase and the delay between the last pulse and the start of sampling (Frenkiel *et al.*, 1990). TOCSY spectra were recorded with a 0.21-s acquisition time, a  $t_1(\text{max})$  value of 50 ms, and a 2-s recycle time. Isotropic mixing was

Table 4.1: Binding properties for various DNA-berenil complexes<sup>a</sup>

DNA	sequence	difference spectrum		$K_d / \mu\text{M}$
		$\lambda_{\text{max}} / \text{nm}$ ( $\Delta\epsilon / \text{mM}^{-1} \text{cm}^{-1}$ )		
–	d(CGTACG) <sub>2</sub>	355(-1), 408(2)		>5 <sup>b</sup>
<i>EcoRI</i>	d(CGCGAATTCGCG) <sub>2</sub>	350(-8), 405(4)		1.0±0.1
–	d(CGCAAATTTGCG) <sub>2</sub>	355(-11), 410(5)		1.3±0.1
<i>trpO</i> <sup>c</sup>	d(CGTACTGATTA ATCAGTACG) <sub>2</sub>	355(-30), 415(11)		2.0±1.0
<i>trpO</i>	d(CGTACTAGTTAACTAGTACG) <sub>2</sub>	355(-2), 415(7)		3.5±1.5
<i>trpRO</i>	d(CGTACTCTTTAGCGAGTACA)· d(TGTACTCGCTAAAGAGTACG)	355(-28), 412(11)		4.5±1.5

<sup>a</sup> Visible difference spectra and  $K_d$  values determined at 298 K in aqueous buffer containing 100 mM KCl and 5 mM sodium phosphate, pH 7.00.

<sup>b</sup> Saturation cannot be achieved.

<sup>c</sup> Consensus tryptophan operator sequence.

obtained by using MLEV-17 (Bax & Davis, 1985) for a period of 50 ms. Data matrices were zero-filled to 4096 by 2048 complex points prior to Fourier transformation.

<sup>31</sup>P NMR spectra were recorded at 4.7 T on a Bruker 200 spectrometer with composite pulse decoupling of the protons. Acquisition times for the one-dimensional spectra were typically 2.5 s. Heteronuclear correlation experiments were performed as described by Lane *et al.* 1991b.

#### 4.5 Molecular Modelling.

Initial coordinates for the d(HO·CGCAAATTTGCG·OH) dodecamer duplex were generated using the program GENHELIX2 (Jenkins, unpublished), with an idealised global B-DNA conformation selected on the basis of the derived nucleotide parameters. Coordinates for the ligand were taken from the crystal structure (Pearl *et al.*, 1987), with the berenil molecule asymmetric about the triazene core. Interactive graphics molecular modelling was performed using the GEMINI program (Beveridge, 1990), with visualisation on a Silicon Graphics IRIS4D-20G workstation. All calculations were carried out on an Alliant FX40/3 computer.

Initial models for the DNA-ligand complexes were generated by docking each ligand at locations within the minor groove with the inner concave surface of the molecule facing the groove floor to give a close isohelical fit (Goodsell & Dickerson, 1986), and with the amidinium moieties close to hydrogen bond acceptors disposed on opposite strands, with H(donor)...N3(adenine) or H(donor)...O2(pyrimidine) separations of 2.2 -2.5 Å. Models with 1:1 stoichiometry were generated for binding sites involving only the d(.CAAATTT.) region since earlier molecular modelling and solution studies have shown conclusively that binding is less favourable at G/C-rich sites (Luck *et al.*, 1988; Brown *et al.*, 1990; Laughton *et al.*, 1990; Sansom *et al.*,

1990; Fox *et al.*, 1990; Lane *et al.*, 1991). In the case of berenil, this arrangement is achieved for model complexes where the central nitrogen atom of the ligand is positioned either in the plane of a DNA base-pair or between adjacent base-pair planes, described as class-1 (1,3-bp) or class-2 (1,2-bp) binding modes, respectively (Laughton *et al.*, 1990, figure 4.2). A total of eleven discrete models (table 4.2) for the DNA-berenil complex, representing the possible class-1 or -2 binding modes, were generated for subsequent energy minimisation.

#### 4.6 Molecular mechanics protocol.

Each model, together with the component DNA and ligand molecules, was energy-minimised at the all-atom level by using CHARMM force-field parameters (Brooks *et al.*, 1983) with the X-PLOR 2.1 program (Brünger, 1990a), adapted (J. Jian-Sheng & R.E. Hubbard, University of York, U.K., 1990) for double-precision application with the Alliant array processor system. Atom-centred MNDO charges for the ligands were derived using the AMPAC molecular orbital package (QCPE Program No. 506, Dept. of Chemistry, University of Indiana), with corrected parameter values for nitrogen (Stewart, 1989). Additional CHARMM parameters were interpolated from earlier studies from this laboratory by Dr. Terry Jenkins.

The X-PLOR protocol used for energy minimisation involved conjugate-gradient (Powell-type molecular mechanics) refinement to an ultimate rms gradient of  $\leq 5 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  ( $\leq 0.12 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ ). Glycosidic torsion angle ( $\chi$ ) restraints were applied throughout the refinements, such that  $\chi(\text{pyrimidine}) = -115 \pm 10^\circ$  and  $\chi(\text{purine}) = -110 \pm 10^\circ$ , where  $\chi$  is defined by O4'-C1'-N1-C2 for pyrimidines and O4'-C1'-N9-C4 for purines. Additional torsional constraints were applied to the linking moieties and termini of the ligands to ensure that torsion angles were held within  $\pm 20^\circ$  of the crystallographic values.

Table 4.2: Calculated energies for the dodecamer-berenil complexes<sup>a</sup>

sequence <sup>b</sup>	$E_{\text{bind}}$	component energies			perturbation <sup>c</sup>	
		$\Delta E_{\text{vdW}}$	$\Delta E_{\text{q}}$	$\Delta E_{\text{HB}}$	DNA	ligand
<i>berenil:</i>						
<u>ATT</u>	-323	-223	-78	-41	2	7
<u>AAT</u>	-317	-220	-87	-29	28	4
<u>ATT</u>	-316	-208	-70	-31	1	7
<u>TTT</u>	-311	-226	-79	-26	14	9
<u>AAA</u>	-311	-224	-77	-40	20	5
<u>TTT</u>	-301	-196	-72	-36	27	4
<u>AAA</u>	-281	-179	-77	-29	26	6
<u>AAA</u>	-276	-181	-67	-38	31	7
<u>CAA</u>	-274	-183	-74	-36	32	10
<u>TTT</u>	-267	-175	-65	-34	41	5
<u>CAA</u>	-263	-148	-65	-31	9	6

<sup>a</sup> Energies in kJ mol<sup>-1</sup>, where

$$E_{\text{bind}} = E_{\text{complex}} - [E_{\text{DNA}} + E_{\text{ligand}}]$$

and the decomposed van der Waals ( $\Delta E_{\text{vdW}}$ ), electrostatic ( $\Delta E_{\text{q}}$ ) and H-bonded ( $\Delta E_{\text{HB}}$ ) binding terms are given by

$$\Delta E = E_{\text{term}(\text{complex})} - [E_{\text{term}(\text{DNA})} + E_{\text{term}(\text{ligand})}]$$

<sup>b</sup> Base sequence for binding site in d(CGCAAATTTGCG)<sub>2</sub>, where the indicated bases represent the possible class-1 or class-2 complexes, respectively (see text).

<sup>c</sup> Perturbation energies for complexed versus free molecules, where

$$\Delta E_{\text{DNA}} = E_{\text{DNA}(\text{bound})} - E_{\text{DNA}} ; \Delta E_{\text{ligand}} = E_{\text{ligand}(\text{bound})} - E_{\text{ligand}}$$

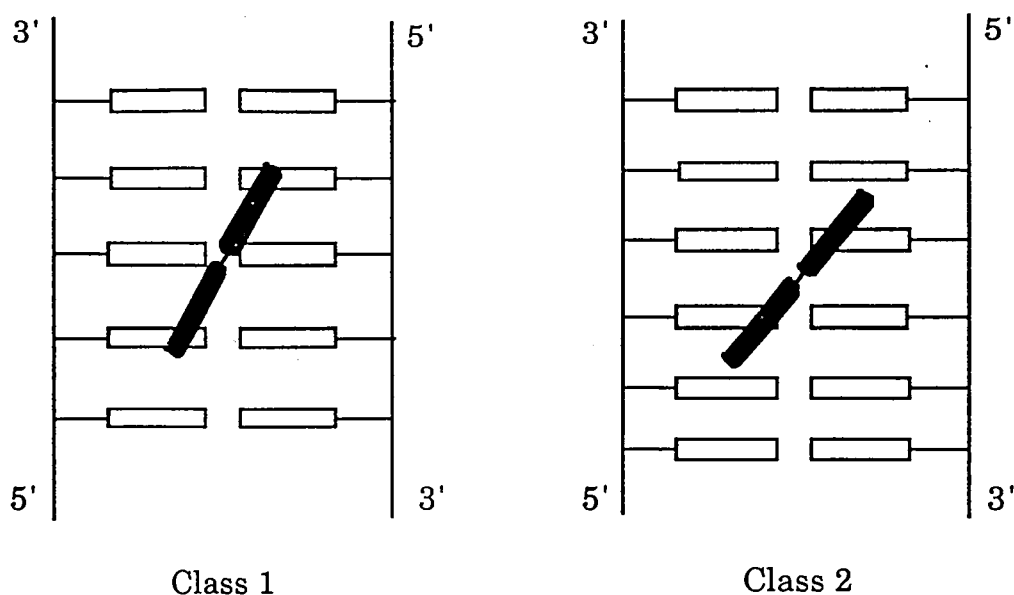


Figure 4.2. Schematic representations of "class I" and "class II" binding.

The effects of solvent and counterions were simulated by using a distance-dependent dielectric constant with  $\epsilon=4/r_{ij}$  (Orozco *et al.*, 1990). Non-bonded energy terms were included up to 9.5 Å, with switching between 7.5-8.0 Å, and a factor of 0.4 was used to damp 1,4 electrostatic interactions. The H-bonded interactions were switched on for donor to acceptor distances of 5.5-6.5 Å, and terms up to 7.5 Å were included. Explicit distance restraints corresponding to Watson-Crick base-pair geometry were not employed. Typically, each minimisation required 900-1400 cycles of refinement ( $\leq 2000$  cpu s) for satisfactory convergence. Perturbation energies were calculated for the ligands and DNA separated from each minimised complex. This involves taking the minimised complex and separating the DNA and ligand followed by one cycle of minimisation performed on each to give the various energy components. These energies were then compared to those found for the DNA and ligand minimised separately (see table 4.2).

#### 4.7 Molecular dynamics protocol for DNA-berenil complexes.

Since this oligonucleotide is reported to exhibit unusually high propeller

twists within the A3T3 core (Coll *et al.*, 1987), the DNA was first minimised, by conjugate gradient method, in the absence of ligand using the NOE distances and  $\chi$ -angles determined for the uncomplexed DNA (table 4.3). This procedure yielded a global B-type DNA structure with satisfactory propeller twists for the central AT base stretch. Examination of the refined structures from the molecular mechanics study revealed that two low-energy models would satisfy both the observed NMR behaviour and all observed interproton NOEs, with the ligand disposed either centrosymmetrically about the diad axis ("class-1" AT 1,2-bp model) or displaced in the 3'-direction of the Watson strand, thereby spanning Ade6-Thy8 (ATT 1,3-bp model) in a class-2 fashion (Laughton *et al.*, 1990; figure 4.2). On this basis, the AT and ATT models were selected for X-PLOR molecular dynamics refinement.

The dynamic annealing (Brünger, 1990) protocol for structure refinement used, stepwise: (i) 400 cycles of conjugate-gradient minimisation using  $\chi$ -angle dihedral constraints (Table 4.3) for the DNA bases, and values of  $-115/-110(\pm 10^\circ)$  for undetermined pyrimidine and purine bases respectively, (ii) 0.21 ps dynamic heating to 300 K (0.15 fs timestep) with incremental increase in weighting on NOE distance restraints, (iii) restrained Verlet molecular dynamics for 10 ps (0.5-fs timestep) using  $\chi$  constraints and NOE distance restraints (Tables 4.3 and 4.4), with coupling to a heatbath at 300 K (100 ps<sup>-1</sup> friction coefficient) and sampling at 0.5 ps intervals, and (iv) 400 cycles of unrestrained conjugate-gradient minimisation of the rms-averaged snapshots to a final rms gradient of  $\leq 6$  kJ mol<sup>-1</sup> n m<sup>-1</sup>. A dielectric constant of form  $\epsilon = cr_{ij}$  was used throughout, with  $c=1$  during steps (ii) and (iii), and  $c=4$  for steps (i) and (iv). No attempt was made to restrain either terminal or internal base-pairs. Typically, each three-stage minimisation required ~28000 cpu seconds.

Table 4.3: NMR properties of d(CGCAAATTTTCGC)<sub>2</sub> in the absence and presence of berenil<sup>a</sup>

base	property	value		dimension
		-berenil	+berenil	
C3	$\tau_{\text{H5-H6}}$	3.1±0.3	nd	ns
C3	$\chi$	-112±2 <sup>b</sup>	nd	deg
A4	$\chi$	-106±2 <sup>c</sup>	-103±2 <sup>c</sup>	deg
A5	$r_{\text{H2-A6(H2)}}$	0.345±0.015 <sup>d</sup>	0.36±0.02 <sup>d</sup>	nm
T7	$\chi$	nd	-116±2 <sup>e</sup>	deg
C11	$\tau_{\text{H5-H6}}$	3.3±0.3	3.4±0.4	ns
C11	$\chi$	-117±5 <sup>f</sup>	-120±2 <sup>f</sup>	deg
	$\Delta\delta(\text{P}_i)$	0.61 <sup>g</sup>	0.59 <sup>g</sup>	ppm

<sup>a</sup> Property at 303 K, pD\* 7.0; nd = not determined

<sup>b</sup> Value = -114±2° in d(CGCGAATTCGCG)<sub>2</sub>(*EcoRI*)

<sup>c</sup> Value = -112° (G4) in *EcoRI*

<sup>d</sup> Value = 0.35±0.015 nm in *EcoRI*

<sup>e</sup> Value = -115° in *EcoRI*

<sup>f</sup> Value = -117° in *EcoRI*

<sup>g</sup>  $\Delta\delta(\text{P}_i)$  = 0.51 ppm in *EcoRI*

Data for *EcoRI* taken from Lane et al. (1991b).

Acknowledgement to Dr Andrew Lane for data acquisition, interpretation and analysis by NUCFIT (Lane, 1990)

## RESULTS

### 4.8 Optical spectroscopy.

The absorbance difference spectrum for berenil with this sequence has extrema at 410 and 355 nm, with  $\Delta\epsilon$  values of 5 and  $-11 \text{ mM}^{-1} \text{ cm}^{-1}$ , respectively. This behaviour is closely similar to that reported for the *EcoRI*-berenil complex (Lane *et al.*, 1991b). The circular dichroism also shows similar behaviour, with maxima at 380 and 335 nm and a diminution of the DNA band at 280 nm (Figure 4.3). The dissociation constant determined at 303 K by titration was  $1.3 \pm 0.1 \text{ }\mu\text{M}$ , and this value compares with  $1.0 \pm 0.1 \text{ }\mu\text{M}$  for the *EcoRI*-berenil complex (Lane *et al.*, 1991b). Under these conditions, the binding of the drug at low concentrations of the ligand is quasi-stoichiometric, and a limit of  $\leq 1.2:1$  for the ligand:DNA molar ratio in the complex was estimated.

The characteristics of berenil binding to the two DNA sequences are compared in table 4.5. Clearly, the binding characteristics are very similar, suggesting that similar binding sites are involved in the DNA sequences.

The optical absorption changes determined for berenil complexes with the other DNA sequences of different composition are collected in table 4.1, together with the estimated  $K_d$  values. There is a clear progression of binding affinity, with  $A_n T_n > T_n A_n \geq \text{TpA}$ . It is noteworthy that saturation could not be achieved for the  $d(\text{CGTACG})_2$  duplex containing a TpA class-1 binding site, suggesting that this represents a very unfavourable binding interaction.

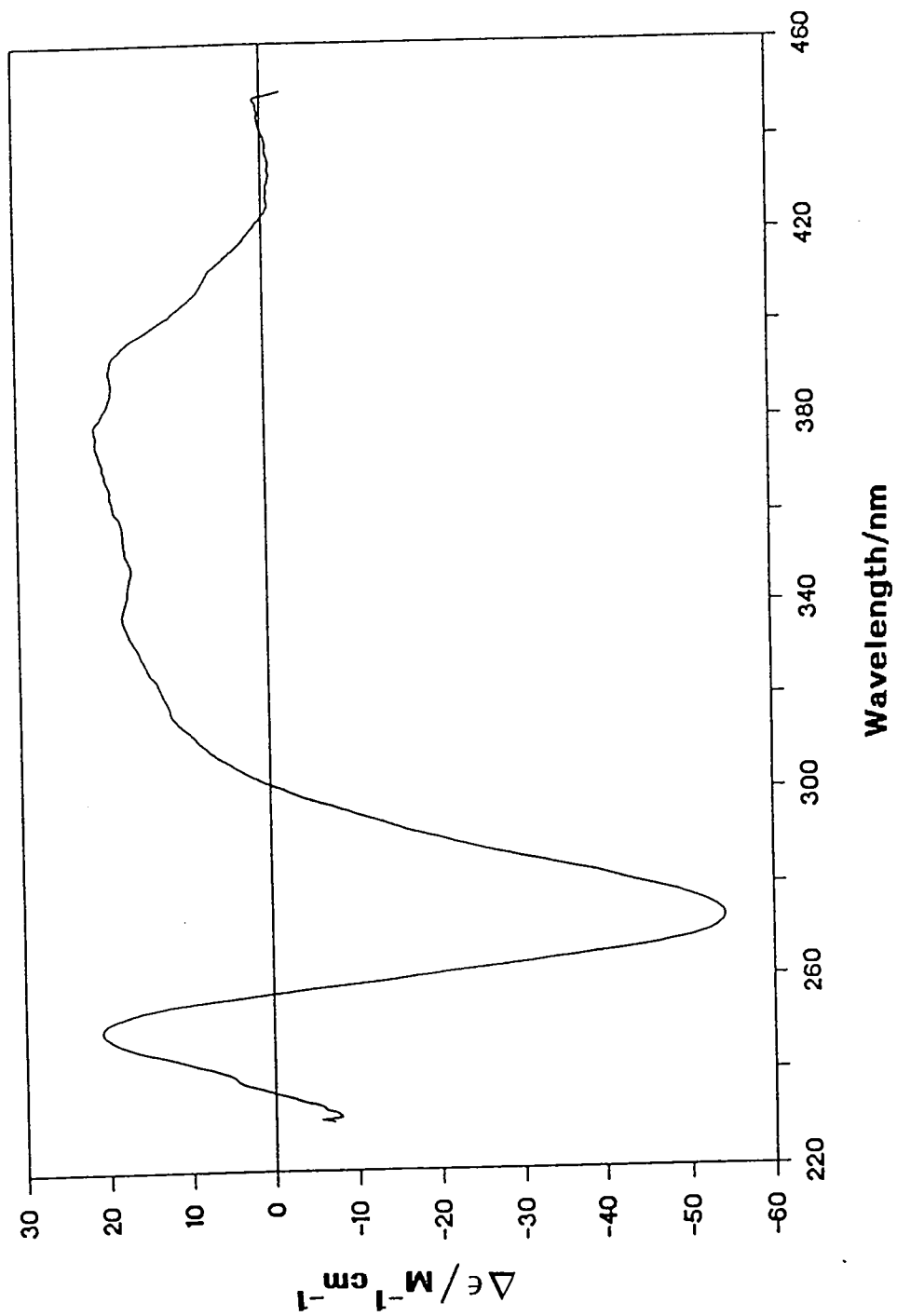


Figure 4.3. Difference circular dichroism spectra for A3T3 in the presence and absence of berenil.

Table 4.4: Interproton NOEs in the dodecamer-berenil complex.<sup>a</sup>

H <sub>a</sub>	H <sub>b</sub>	$\sigma$ s <sup>-1</sup>	$r_{app}$ nm	$r_{corr}$ nm	$r_{lower}$ nm	$r_{upper}$ nm
A5(H2)	A6(H2)	-0.1	0.35	0.37	0.35	0.40
A5(H2)	L(H3)	-0.25	0.30	0.32	0.30	0.35
L(H3)	A5(H2)	-0.3	0.29	0.31	0.29	0.34
A5(H2)	L(H2)	-0.08	0.36	0.38	0.33	0.45
A6(H2)	L(H3)	-0.4	0.28	0.30	0.27	0.35
T8(H1')	A6(H2)	-0.08	0.36	0.38	0.33	0.45
T8(H1')	L(H2)	-0.1	0.35	0.37	0.33	0.45
T8(H1')	L(H3)	-0.1	0.35	0.37	0.33	0.45

<sup>a</sup> See footnote of table 4.3

Table 4.5: Optical changes on binding of berenil to A<sub>n</sub>T<sub>n</sub> dodecamer duplexes<sup>a</sup>

parameter <sup>b</sup>	dimension	value with	
		n=2	n=3
$\lambda_{max}(A)$	nm	355,405	355,405
$\Delta\epsilon(A)$	mM <sup>-1</sup> cm <sup>-1</sup>	-8.1,3.9	-10.5,4.8
$\lambda_{max}(cd)$	nm	272,335,380	275,335,380
$\Delta\epsilon(cd)$	M <sup>-1</sup> cm <sup>-1</sup>	-62,16,20	-55,16,20
$K_d$	$\mu$ M	1.0±0.1	1.3±0.1

<sup>a</sup> i.e. d(CGCGAATTCGCG)<sub>2</sub> and d(CGCAAATTTGCG)<sub>2</sub>

<sup>b</sup> A refers to the optical absorption change; cd refers to the circular dichroism effect (see text for conditions).

#### 4.9 NMR spectroscopy of the uncomplexed DNA.

Nonexchangeable protons of the free DNA duplex (table 4.6) were assigned by using NOESY and TOCSY methods. Figure 4.4 shows NOE and TOCSY connectivities. The TOCSY experiment clearly shows the presence of twelve correlations for H1'-H3' and for H3'-H4', allowing unambiguous assignments of all sugar protons other than H5'/H5'' to be obtained. Further, the NOE intensities are consistent with right-handed B-like DNA. Comparison of the spectra of this dodecamer with those of the *EcoRI* sequence obtained under similar conditions shows that the induced changes in chemical shift are significant (i.e.  $|\Delta\delta| > 0.05$  ppm) only for bases immediately adjacent to the substitution sites, with the changes being larger for the H1' than for the base protons (except Ade5(H2) and Ade6(H2), which are also found in the minor groove).

The apparent correlation time of  $3.2 \pm 0.2$  ns at 303 K, determined from the cross-relaxation rate constant for the H5-H6 vectors of Cyt3 and Cyt11, is identical, within experimental error, with the *EcoRI* dodecamer (Lane *et al.*, 1991a). Further, glycosidic torsion angles ( $\chi$ ) for certain residues were determined from 1D NOE timecourses using NUCFIT (Lane, 1990).

The  $\chi$ -angles determined in the  $d(\text{CGCAAATTTGCG})_2$  sequence for Cyt3 and Cyt11 are within  $\pm 4^\circ$  of those found in the  $d(\text{CGCGAATTCGCG})_2$  dodecamer. These data strongly suggest that the local conformations at the ends of the duplexes are identical to those in the native *EcoRI* dodecamer.

Table 4.6: NMR assignments for nonexchangeable protons of  
 $d(\text{CGCAAATTTTCGC})_2$  in the absence of ligand<sup>a</sup>

base	H6/8	Me/ H2/H5	H1'	H2'	H2''	H3'	H4'
C1	7.62	5.89	5.76	1.95	2.38	4.66	4.03
G2	7.92	-	5.86	2.61	2.70	4.95	4.32
C3	7.29	5.42	5.38	1.83	2.20	4.77	4.08
A4	8.18	7.14	5.78	2.72	2.87	5.02	4.36
A5	8.09	7.01	5.88	2.63	2.89	5.04	4.43
A6	8.08	7.51	6.12	2.52	2.92	4.99	4.46
T7	7.09	1.24	5.92	2.02	2.58	4.85	4.21
T8	7.41	1.53	6.13	2.17	2.61	4.90	4.26
T9	7.27	1.65	5.86	2.08	2.45	4.91	4.15
G10	7.90	-	5.79	2.59	2.63	4.98	4.37
C11	7.35	5.42	5.80	1.89	2.32	4.78	4.16
G12	7.93	-	6.14	2.61	2.38	4.65	4.15

<sup>a</sup> Proton assignments at 303 K, pD\* 7.0

The sequence numbering is as follows:

5'– C1 G2 C3 A4 A5 A6 T7 T8 T9 G10 C11 G12  
 G24 C23 G22 T21 T20 T19 A18 A17 A16 C15 G14 C13 –5'

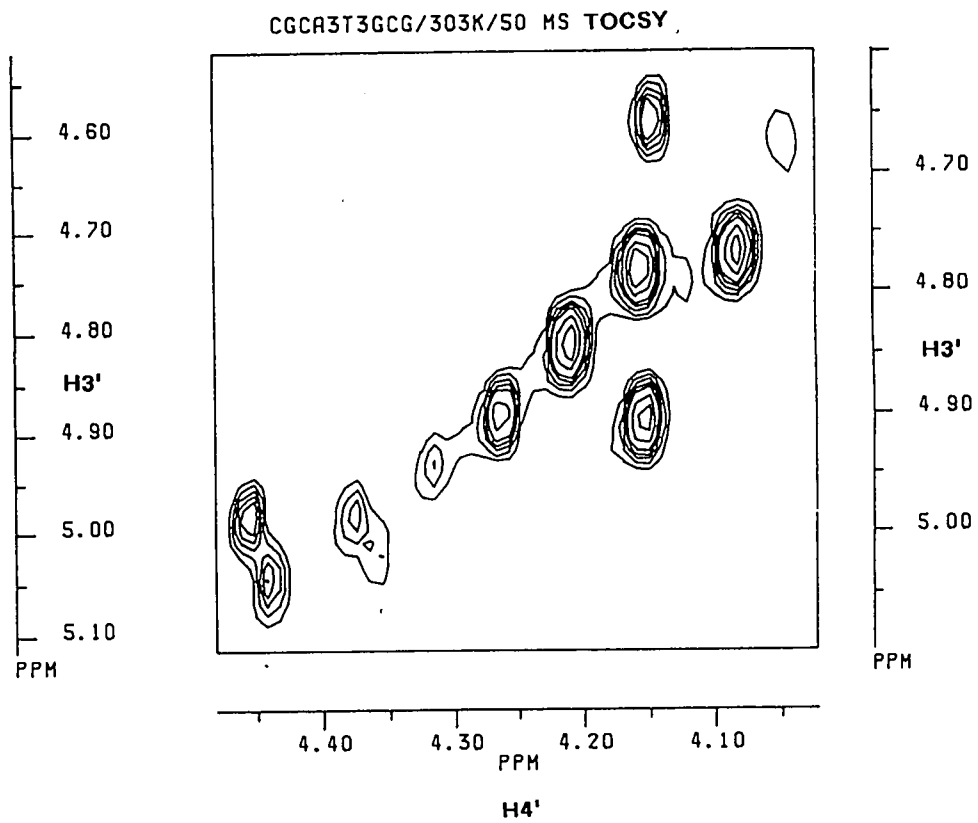
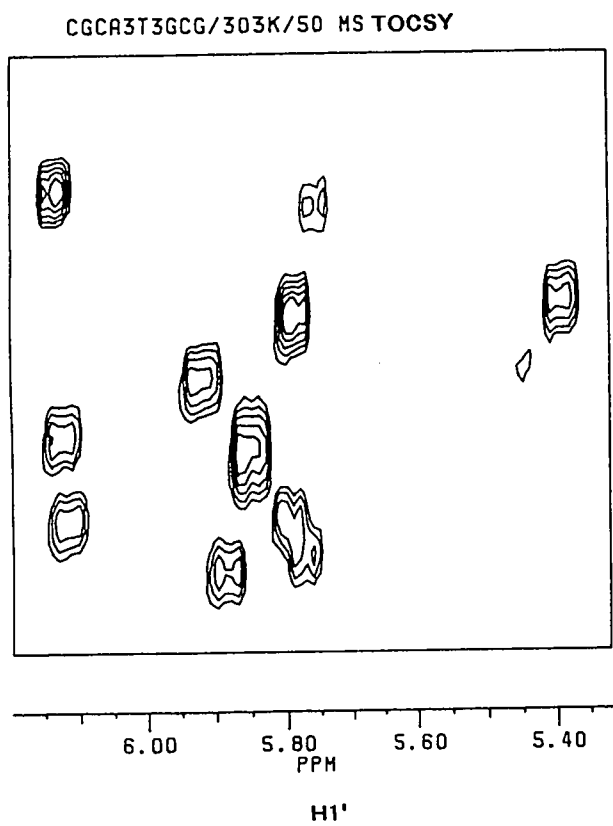


Figure 4.4. Partial 2D Spectra for A3T3 in the absence of ligand.

#### 4.10 NMR spectroscopy of the DNA-berenil complex.

Resonances for the nonexchangeable protons for the dodecamer in the presence of berenil (table 4.7) were assigned by TOCSY and NOESY experiments. Titration of the dodecamer duplex with berenil produces gradual changes in chemical shift as the concentration of ligand is increased, typical of fast exchange on the chemical shift timescale. The largest shifts are observed for Thy7/8 proton resonances (figure 4.5), with the greatest changes for those protons found in the minor groove. The chemical shift of Ade6(H2) also changes dramatically. The most significant shift differences (Figure 4.6) are centred around Thy7/8, but extend significantly to Ade5 and Thy8. The pattern of shifts, in which the greatest perturbations are for protons located in the minor groove, indicates that the ligand binds in the minor groove. However, the perturbation of chemical shifts extends over more bases than was observed in the native d(CGCGAATTCGCG)<sub>2</sub> dodecamer [*EcoRI* dodecamer] (Lane *et al.*, 1991a), suggesting either that the binding site is more extensive, or that the ligand actually binds to two or more overlapping AT-rich regions within the Ade4-Thy9 base segment. Of the phosphodiester, only the chemical shifts of TpT (and possibly ApA) are shifted significantly (table 4.8). However, the overall chemical shift dispersion of the phosphate resonances remains essentially the same, suggesting that binding does not induce substantial rearrangement in the phosphate backbone.

In the complex, strong NOEs are observed between the L(H2/2') and L(H3/3') ligand aromatic protons, with no symmetry-breaking in either the ligand or the DNA. Further NOEs were observed between L(H2)-Ade5(H2) (weak) and L(H3)-Ade5/6(H2) (strong). These NOEs confirm that the ligand binds via the minor groove, and in the AT-rich region of the DNA duplex. Relative NOESY cross-peak intensities for the DNA protons are barely altered in the complex, indicating that the conformation of the DNA is little affected by binding.

Table 4.7: NMR assignments for nonexchangeable protons of  $d(\text{CGCAAATTTTCGC})_2$  in the presence of berenil<sup>a</sup>

base	H6/8	Me/ H2/H5	H1'	H2'	H2''	H3'	H4'
C1	7.60	5.88	5.75	1.92	2.38	4.68	4.05
G2	7.91	-	5.86	2.63	2.68	4.95	4.30
C3	7.29	5.41	5.41	1.83	2.23	4.79	4.11
A4	8.19	7.15	5.83	2.74	2.89	5.03	4.37
A5	8.07	7.25	5.91	2.63	2.85	5.04	4.43
A6	8.02	7.93	5.98	2.42	2.74	4.91	3.97
T7	6.88	1.21	5.35	1.78	2.28	4.68	nd
T8	7.20	1.46	5.55	1.92	2.39	4.79	nd
T9	7.18	1.58	5.45	1.93	2.30	4.82	3.83
G10	7.85	-	5.83	2.61	2.66	4.99	4.37
C11	7.34	5.41	5.79	1.89	2.32	4.80	4.16
G12	7.93	-	6.14	2.60	2.38	4.68	4.16

<sup>a</sup> Proton assignments at 303 K, pD\* 7.0; - = not determined.

Table 4.8:  $^{31}\text{P}$  NMR assignments for  $\text{d}(\text{CGCAAATTTTCGC})_2$  in the absence (-B) and presence (+B) of Berenil<sup>a</sup>.

resonance	$\delta(-\text{B})$	$R_2$	$R_1$	$\Delta R$	$\Delta\sigma^b$	assignment	$\delta(+\text{B})$
	ppm				ppm		ppm
		s <sup>-1</sup>					
1(1)	-17.2	4.35	0.74	3.98	176		17.24
2(2)	-17.27	3.8	0.54	3.53	161	C1/C3/C11	17.27,17.29
3(1)	-17.40	3.55	0.49	3.3	154	C1/C3/C11	17.39
4(2)	-17.52	4.2	0.67	3.9	173	A6	17.51,17.57
5(1)	-17.65	4.5	0.85	4.1	179	A5	17.60
6(3)	-17.72	5.2	0.89	4.75	197	T7/T8/T9	17.66,17.81
7(1)	-17.8	5.2	0.69	4.9	201	T7/T8/T9	17.83

<sup>a</sup>see footnote table 4.3

$R_1$  and  $R_2$  are the spin-lattice and spin-spin relaxation rate constants respectively

$\Delta R = R_2 - 0.5R_1 = 88.88\omega_p^2\Delta\sigma^2\tau + 1$  at 81 MHz. The correlation time for the cytosine vectors is  $3.2\pm 0.1\text{ns}$ , i.e.  $\sim 4.3\text{ ns}$  for end-over-end tumbling, or  $\sim 4.0\text{ ns}$  for  $S^2 = 0.8$ . Using  $\tau = 4.2\text{ ns}$ , the value can be calculated for each peak.

<sup>b</sup> Value for the *EcoRI* dodecamer was  $146\pm 7\text{ ppm}$  (Lane *et al.*, 1991b).

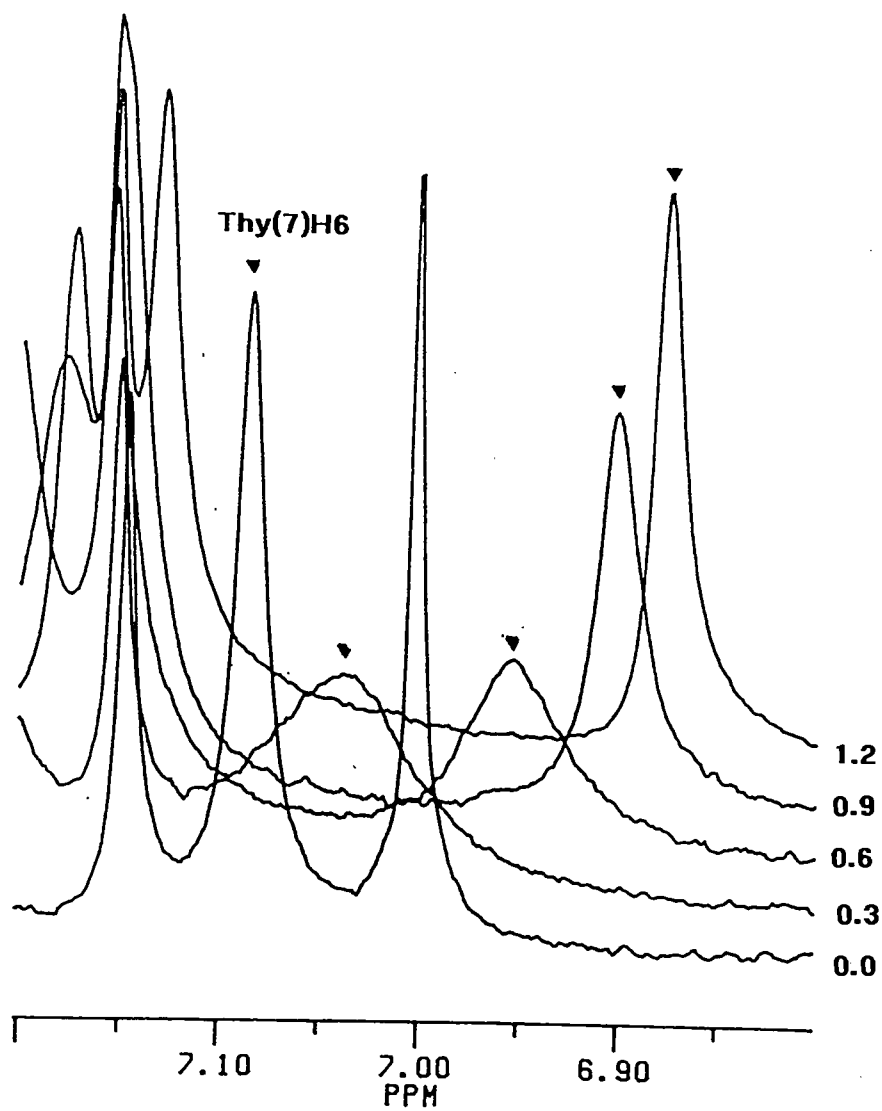


Figure 4.5. Change in chemical shift and spectral linewidth of H6 (Thy7) with increasing concentration of berenil.

a) 0.0 b) 0.3 c) 0.6 d) 0.9 e) 1.2 mM berenil.

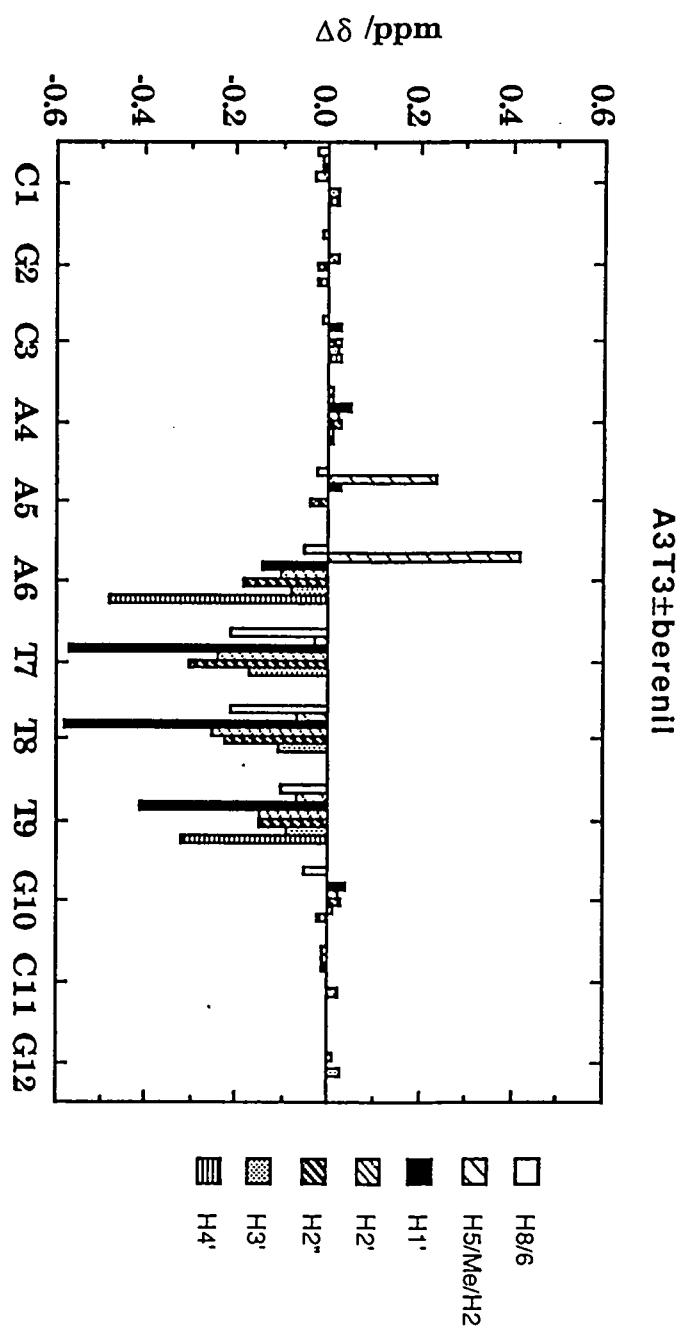


Figure 4.6. Plot of chemical shift differences for A3T3 in the presence of berenil.

The Thy7(H6) resonance not only shifts 0.25 ppm upfield during titration with berenil, but its linewidth also changes. Specifically, the linewidth first increases from 6 Hz in the absence of berenil to >35 Hz at 50% saturation, after which the linewidth again decreases as the fractional saturation increases, and approaches that of the free DNA (figure 4.7). This is expected for moderately fast exchange behaviour. As the lineshapes are determined for both 0% and 100% saturation, the "excess" linewidth can be used to estimate the rate constants for the binding and dissociation processes, since the dissociation constant is known independently. At 50% saturation, where the linewidth is ~40 Hz, the pseudofirst-order forward ( $k_1$ ) and reverse ( $k_{-1}$ ) rate constants are equal, from which a value of ~420 s<sup>-1</sup> can be calculated. According to Sandstrom (1982), the lineshape depends on the fractional saturation and the rate constants as

$$L = 4\pi p_b p_a^2 \delta\nu^2 / k_b + p_a L_a + p_b L_b$$

where  $L$  is the linewidth and  $p_a$  and  $p_b$  are the relative populations. Plotting  $L$  versus  $p_b p_a^2$  gives a slope of  $4\pi \delta\nu^2 / k_b$ , from which  $k_b$  can be calculated, where  $\delta\nu$ , the chemical shift difference, is 84 Hz. Fitting the observed linewidths gave values of  $510 \pm 30$  s<sup>-1</sup> and  $3 \times 10^8$  M<sup>-1</sup> s<sup>-1</sup> for  $k_{-1}$  and  $k_1$ , respectively. This is consistent with a binding reaction in which there is little rearrangement.

NOEs were observed between ligand aromatic ring protons and both Ade5/6(H2) and Thy8(H1'), identifying the actual binding focus as AATT/ATTT (see Table 4.4). The significant Ade6(H2)-Thy8(H1') NOE is consistent with a large propeller twist for one or both of these base-pairs, as expected for poly(dA)-poly(dT) (Aymami *et al.*, 1989; Yanagi *et al.*, 1991).

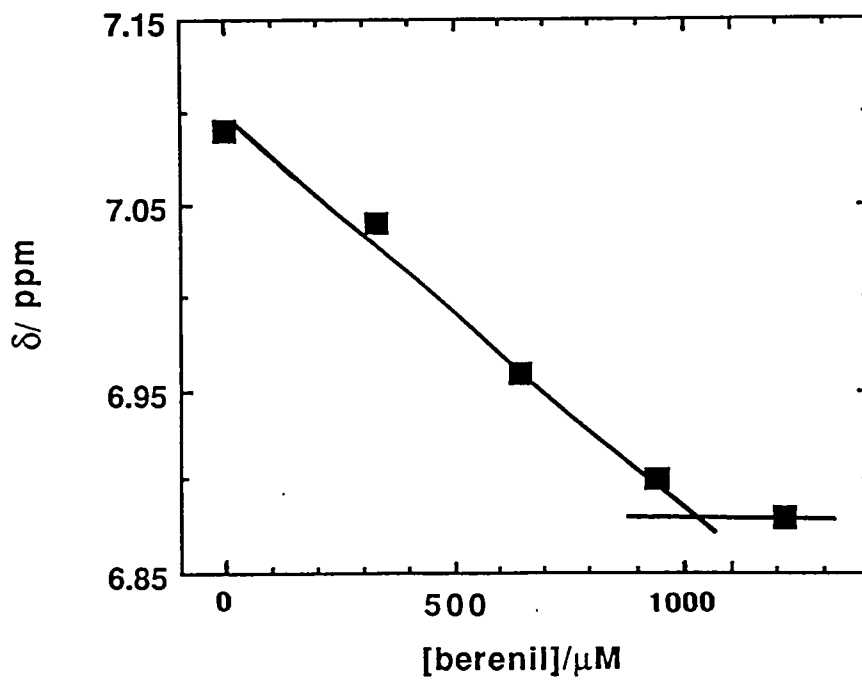
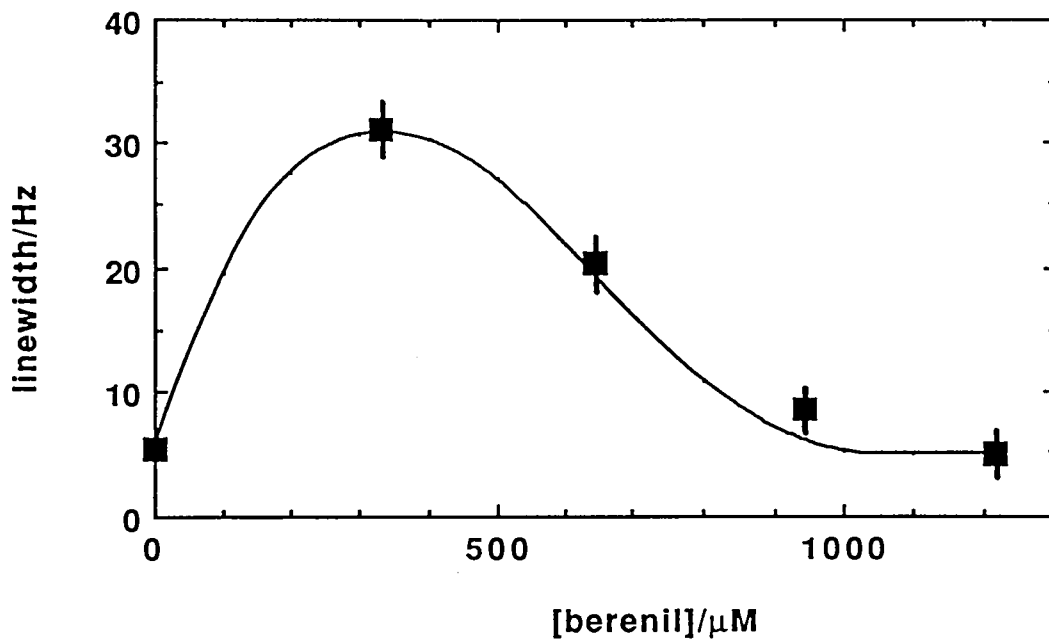


Figure 4.7 Plots of linewidth and chemical shift for A3T3 in the presence of varying concentration of berenil.

#### 4.11 Molecular modelling and energy minimisation of the DNA-berenil complex.

Initial models for the DNA-berenil complex were generated as described in Materials and Methods for each of the 11 possible class-1 (1,2- bp) or class-2 (1,3-bp) A/T(C)-binding sites present in this sequence (Laughton *et al.*, 1990) with (i) equimolar duplex:ligand stoichiometry, (ii) the inner concave surface of the ligand facing the convex floor of the B-DNA minor groove, and (iii) with hydrogen-bonded contacts between the amidinium protons and either N3(adenine) or O2(thymine/cytosine). The global B-like DNA conformation was selected on the basis of the derived NMR parameters (see above). Preliminary studies established that, in the absence of ligand, the global B-type integrity of the dodecamer is retained during subsequent energy refinement.

The structural models were then refined by conjugate-gradient energy minimisation using the X-PLOR program (Brünger, 1990) to an ultimate rms gradient of  $\leq 5 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . Glycosidic torsional constraints for a typical B-DNA structure were applied throughout the refinement (see Methods). The binding enthalpies, decomposed energies and perturbation terms calculated for the DNA-berenil complexes are collected in Table 4.2.

Table 4.2 reveals that there is little difference ( $\leq 17 \text{ kJ mol}^{-1}$ ) in computed binding energy for the five highest-ranked models where, in all cases, the ligand spans a distance of 3 base-pairs. Further, sites involving spanned cytosine residues are clearly not favoured, in accord with results from footprinting studies (Laughton *et al.*, 1990). It is evident that H-bonded interactions to the amidinium protons of the ligand are in the order A-A > A-T  $\approx$  T-A, where the indicated acceptor bases are on opposite strands of the DNA. Unexpectedly, the most favourable binding site computed for this sequence is the central AT site, a "class-1" model (Laughton *et al.*, 1990), with hydrogen-bonded contacts to N3 acceptor atoms of adenosine residues disposed on

opposite strands in a 1,2-fashion, rather than the more usual nonadjacent or 1,3-bp mode (Brown *et al.*, 1990; Lane *et al.*, 1991b). These results indicate that the AT-rich core of this dodecamer presents a broad binding region for berenil that contains a number of overlapping and energetically-similar binding sites.

It is interesting that in our molecular mechanics study the AT 1:2 pseudo-centrosymmetric model is more favoured than the equivalent model in a similar study using the 8 base-pair core of the *EcoRI* dodecamer (Brown *et al.*, 1990). This feature probably reflects greater conformational flexibility in the core regions of the two sequences as AT base-pairs can propeller twist, buckle and shear more readily than G-C base-pairs. Indeed the favoured AT 1,2-bp complex in this study exhibits higher propeller twists for the central base-pair, which is tolerated by the flanking AT bases, but which would not have been accommodated so readily in the *EcoRI* sequence (Aymami *et al.*, 1990). However, it should be noted that we have employed more extensive force-field parametrisation than was used in the earlier study and with longer non-bonded cutoff distances and explicit hydrogen bond terms introduced rather than the primitive 1-4 electrostatic terms used in the earlier study; hence energies in table 4.2 are considerably greater. The ranking order for binding seems to be dominated by van der Waals ( $\Delta E_{vdW}$ ) and hydrogen-bonded ( $\Delta E_{HB}$ ) interactions, rather than electrostatic ( $\Delta E_q$ ) terms. In all models, induced perturbation of the drug is low with the berenil molecule retaining its general planarity. Overall, these data indicate that the AT-rich core region presents a broader binding region for berenil than the *EcoRI* dodecamer study.

#### 4.12 Dynamic simulation of the DNA-berenil complex.

Preliminary examination of the possible DNA-berenil complexes established that no single model would satisfy all the observed interproton NOEs (table 4.4). On the basis of the induced NMR chemical shift changes

observed for the complex (figure 4.6), two starting models for MD simulations, each representing the extrema of the Ade6-Thy9 binding site, were selected for more extended molecular dynamics calculations. Thus, the centrosymmetric AT 1,2-bp and non-centrosymmetric ATT 1,3-bp models, shown to be favoured by molecular mechanics calculations (see above), were selected. Each model satisfied a high proportion of the observed NOEs, but together accounted for all observed drug-DNA interproton contacts.

The two drug-DNA complexes were investigated by using a simulated isothermal dynamic annealing procedure. The molecular dynamics protocol used (see Methods) is similar to that adopted for the berenil-*EcoRI* complex Lane *et al.* (1991). The protocol differed in that (i) the DNA was previously minimised with NOE interproton constraints observed for the isolated DNA, and (ii) dynamic averaging of samples accumulated at 0.5 ps intervals during 10 ps of heatbath-coupled annealing at 300 K. Thermal equilibration of the system is achieved within our 1 ps heating period, as judged from component energy terms and velocity distribution. Following equilibration and rms averaging of the 20 accumulated samples the structure was subjected to 400 cycles of Powell conjugate gradient minimisation.

Despite numerous attempts, employing a variety of heating, annealing, dielectric distance cutoffs, etc. simulations of the isolated DNA duplex resulted in a tendency of transition from B to A-type DNA. During this research a similar study of the same dodecamer sequence was published which also observed a tendency for B to A-type DNA transition during prolonged dynamics simulations, in both the presence and absence of counterions and solvent molecules (Boehncke *et al.*, 1991). Improved structures are achieved following only very protracted dynamics simulations, particularly if time-averaged distance restraints are employed (Swaminathan *et al.*, 1991; Pearlman & Kollman, 1991). We find that annealing is incomplete at times <5 ps, largely as a consequence of the limited number of NOE distance restraints employed

(table 4.4), but satisfactory system convergence is achieved after 10 ps of dynamic annealing, as judged by temperature stability and low rms for the system.

For each of the models NOE's were added and removed in a stepwise manner to establish which of the distance restraints employed between the ligand and each strand of the DNA were degenerate and equivalent.

#### 4.13 Centrosymmetric DNA-berenil model.

In the case of the centrosymmetric AT 1,2 bp model interproton NOEs were gradually introduced from table 4.4, with the assumption of two-fold degeneracy for all NOEs and equivalence for the two DNA strands. The final model used a total of 18 interproton distance restraints and 6 measured glycosidic constraints, with the remaining  $\chi$ -angles targeted to values typical for B-DNA (Saenger, 1984), and had a final  $E_{\text{complex}} = -585.8 \text{ kJ mol}^{-1}$ . NOE violations were  $< \pm 0.1 \text{ \AA}$  for 11 of 18 NOEs prior to final conjugate-gradient relaxation and  $< \pm 0.3 \text{ \AA}$  in the fully refined model. As reported in previous studies (Lane *et al.*, 1991b) ligand alignment with triazene NH-N=N core disposed in the 5'→3' direction gave more favourable energy of interaction.

Detailed analysis of the final model shows that the ligand is closely isohelical with the minor groove floor, and is essentially parallel with the hydrophobic minor groove walls. The minor groove of the DNA is narrower in the region of the ligand, as has been noted for other minor groove-binding ligands, somewhat clamping the ligand in place. The terminal amidinium groups are little perturbed with inter-plane torsional rotations of  $< 6^\circ$  for both groups, which agree closely with crystallographic observations (Pearl *et al.*, 1987). The amidinium moieties provide close hydrogen-bonded contacts with Ade6 and Ade18, with N3...H(amidine) separations of 2.33 and 2.07  $\text{\AA}$ , respectively, and subtended N...H-N(amidine) angles of  $169^\circ$  and  $174^\circ$ . Further

weak contacts at the 5'-end include two Ade6(O4')...H(amidine) contacts, at separations of 2.58 and 2.83 Å, and two interactions with Thy21 on the opposite strand, with O2...H(amidine) and O4'...H(amidine) separations of 2.63 and 2.49 Å, respectively. At the 3'-end of the complex additional contacts to the sugar residues are provided by two H(amidine)...Ade18(O4') and H(amidine)...Thy9(O4') separations of 2.30/2.86 and 2.61 Å, respectively. Thus, there is a potential three-centre hydrogen bond at the both ends of the binding site, with an amidinium proton effectively located between the N3 and O4' acceptors of both Ade6 and Ade18.

High-value propeller twist angles (average 19°) are retained for the AT-rich stretch of the DNA. The backbone is normal B-DNA with standard B<sub>I</sub> phosphates and C2'-*endo* sugars, with the exception of C3 which has a sugar pucker of O4'-*endo* rather than the C3'-*endo* or C2'-*endo* puckers associated with A- or B-DNA, respectively (Saenger, 1984). The unusual pucker adopted by C3 has recently been confirmed in the crystal structure of the berenil-DNA complex formed with this sequence (see chapter 3). Helix analysis of the duplex shows an average helical rise of 3.2±0.7 Å and a mean helical rotation of 36±1.2° (10.0 bp/turn).

#### 4.14 *Non-centrosymmetric DNA-berenil model.*

For the ATT 1,3-bp model, which is the predominant binding mode suggested by the induced chemical shift changes (figure 4.5), it was apparent that the observed NOEs could no longer be treated with two-fold degeneracy and consequently they were applied or removed in a stepwise manner until structural convergence was achieved. In the final model, 19 interproton NOE restraints were used together with the same dihedral constraints as in the previous model. The distance restraints used, together with those applied in the above model, account for all except one of the observed NOEs (table 4.4). NOE violations for 11 of 19 NOEs were ≤±0.1 Å prior to final conjugate-gradient

refinement and  $\leq \pm 0.7$  Å in the final model. The final  $E_{\text{complex}}$  ( $-618.4$  kJ mol $^{-1}$ ) is of lower energy than the refined centrosymmetric AT model by  $32.6$  kJ mol $^{-1}$ , largely due to the non-bonded contribution being more favourable.

The details of the complex display the close isohelical fit with the floor, together with the parallel alignment of the ligand with the walls, of the minor groove. The closest contacts are hydrogen-bonded Ade6/17(N3)···H(amidine) of  $2.22$  and  $2.09$  Å (subtended angles  $148^\circ$  and  $172^\circ$ , respectively) at the 5'- and 3'-ends, respectively. The only other close distance interaction is at the 3'-end with A17(O4')···H(amidine)  $2.16$  Å (subtended angle  $122^\circ$ ). The perturbation on the amidinium moieties is slightly greater than in the previous model with torsional rotations of  $9^\circ$  and  $13^\circ$  at the 5'- and 3'-ends, respectively. High propeller twists are retained by base-pairs in the central Ade4-Thy9 region, other than for bases Ade6 and Thy7 (figure 4.8). Globally, the backbone is normal B-DNA, with standard B<sub>I</sub> phosphates and C2'-*endo* sugar puckers, although that adopted by C3 is again O4'-*endo*.

Helix analysis of the duplex shows an average helical rise of  $3.2 \pm 0.7$  Å and a mean helical rotation of  $36 \pm 9^\circ$  ( $10.0$  bp/turn), and is thus slightly overwound compared with  $3.5 \pm 0.5$  Å and  $37 \pm 5^\circ$  ( $9.7$  bp/turn) for the *EcoRI* dodecamer (Lane *et al.*, 1991a). The duplex thus adopts a general B-like DNA conformation.

#### 4.15 Discussion.

The thermodynamics, stoichiometry and optical properties of berenil binding to  $d(\text{CGCAAATTTGCG})_2$  are remarkably similar to those for binding to the *EcoRI* dodecamer, suggesting a similar mode of interaction with the two DNA duplexes. The pattern of chemical shift perturbations due to ligand binding is also similar for the two DNA sequences, and is consistent with asymmetric binding in the A/T-rich segment of the minor groove.

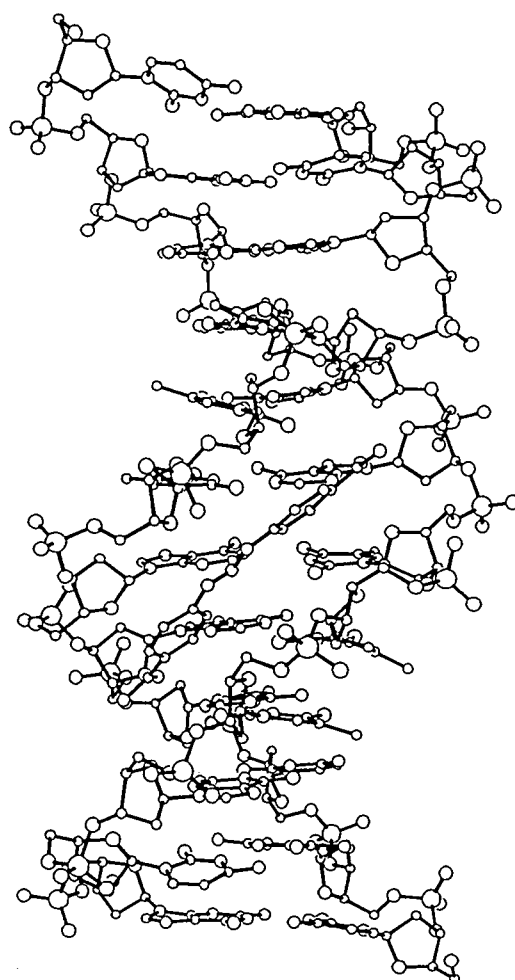
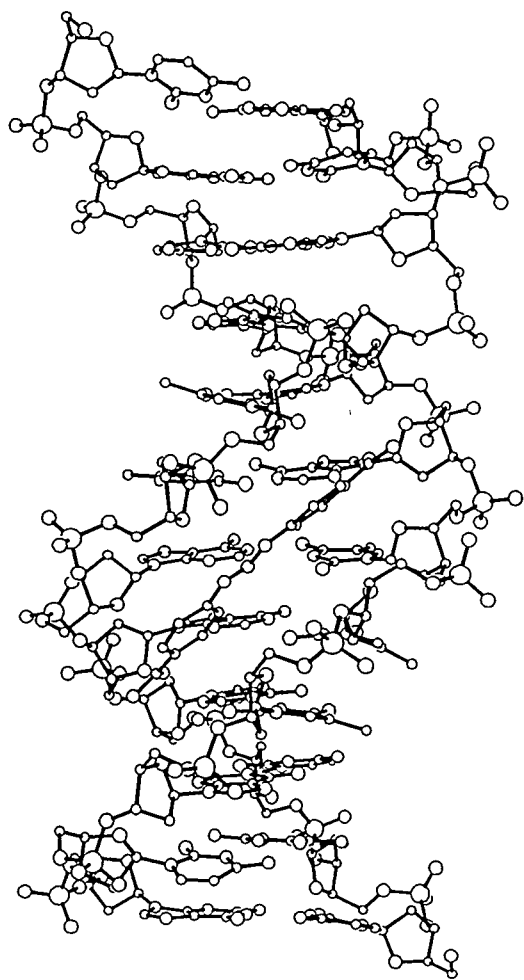


Figure 4.8 Stereo view of non-centrosymmetric berenil model.

This location is confirmed by the observation of NOEs from the ligand aromatic protons to Ade(H2) and Thy(H1') associated with bases Ade5-Thy9. Further, magnetic equivalence is preserved for the two strands of the DNA and for the berenil H2 and H6, H3 and H5 aromatic protons.

The number of nucleotides that respond to berenil binding is greater in this dodecamer than in the *EcoRI* dodecamer, suggesting that the binding site may be more extended in this sequence. Previous studies have indicated a 3 base pair spanning model for berenil in standard B-DNA although, in the presence of high propeller-twists associated with runs of A·T base-pairs, it is possible for up to 4 bases to be involved (see chapter 3). In this study the NOEs observed between the ligand and the DNA, and the changes in chemical shifts for the DNA resonances upon complexation indicate an extended binding site which spans 4-5 base pairs. These data cannot be satisfied by a single site model and imply that berenil binds two or more overlapping sites in the AT tract with similar fractional occupancies. The observed NOE's and chemical shift perturbations suggest that the binding sites are asymmetrically disposed with respect to the diad axis of the DNA duplex.

The influence of berenil on the structure of the DNA is small, and negligible outside of the immediate binding site. This behaviour is similar to that observed for the binding to *EcoRI*, and is consistent with the binding site(s) being preformed.

Induced perturbations of the DNA duplex by the ligand are small and are not propagated beyond the immediate binding site. This behaviour is similar to that observed for berenil binding to the *EcoRI* sequence (which has high propeller twists in the crystal structure), and suggests that high propeller twist values associated with extended oligo(A)<sub>n</sub> tracts may well influence the recognition processes.

## **CHAPTER 5.**

The Crystal structure of a complex between  
the dodecamer d(CGCAAATTTGCG)<sub>2</sub>  
and the minor groove-binding ligand Hoechst 33258.

## 5.1 Introduction.

The bisbenzimidazole derivative Hoechst 33258 (figure 5.1) is a fluorochrome widely used in chromosome staining that also possesses antihelminthic properties (Lämmle *et al.*, 1971). Hoechst 33258 has been determined to have both a low-level tight binding mode and a high-level low-affinity binding that are differentiated by their fluorescent characteristics which has been supported by viscosity and electro-optical data (Comings, 1975). Extensive UV, CD, fluorescence studies (Latt and Wohlleb, 1975; Bontemps *et al.*, 1975; Latt and Stetten, 1976; Jorgensen *et al.*, 1988), melting temperature studies (Comings, 1975) and footprinting (Harshmann and Dervan, 1985; Portugal and Waring, 1988; Murray and Martin, 1988), have shown that Hoechst 33258 binds preferentially to AT-rich DNA.

The length of the Hoechst 33258 specific binding site was first proposed to be at least 3 base pairs (Muller and Gautier, 1975). After initial major groove models were proposed (Bontemps *et al.*, 1975) based on fluorescence data, CD studies (Zasedatelev *et al.*, 1980; Mikhailov, 1981) implied a minor groove interaction spanning four AT base pairs which compares with remarkable similarity to the studies of netropsin and distamycin (Zasedatelev, 1982). The site was also confirmed to be at least four consecutive AT base pairs in length as determined using  $^{125}\text{I}$  labelled Hoechst 33258 (Martin and Holmes, 1983).

The more recent footprinting studies (Harshmann and Dervan, 1985; Portugal and Waring, 1988) confirm the preferential binding to AT rich regions over GC rich regions, although they provide evidence that GC base pairs are not completely excluded at the ends of the binding site.

The different binding modes for Hoechst 33258-DNA has been further investigated by Loontjens *et al.*, 1990 who have shown that at low drug/DNA ratios a sequence-specific tight binding is observed which is not salt dependent. At higher drug/DNA ratios several other binding modes become possible,

which include a non-specific salt-dependent binding that is mediated by electrostatic attraction between the positively charged drug and the negatively charged DNA.

In attempts to deduce the specific binding mode of Hoechst 33258 several NMR (Embrey *et al.*, 1991; Fede *et al.*, 1991; Kumar *et al.*, 1991; Parkinson *et al.*, 1990, Searle and Embrey, 1990) and X-ray crystallographic studies (Pjura *et al.*, 1987; Teng *et al.*, 1988; Carrondo *et al.*, 1989, Quintana *et al.*, 1991) have been carried out.

The first two crystal structures (Pjura *et al.*, 1987; Teng *et al.*, 1988) both investigated the interaction of Hoechst 33258 with the same dodecamer sequence, d(CGCGAATTCGCG)<sub>2</sub>. The structures reported differed somewhat in that the structure proposed by Pjura *et al.*, had the protonated piperazine ring protruding into the adjacent CG region i.e. an ATTC “pip-down” model. The model of Teng *et al.*, on the other hand, which also has the piperazine ring pointing downwards (towards the 3' of the Watson strand), has the drug disposed centrally in the AATT core region i.e. an AATT “pip-down” model. A study by Carrondo *et al.*, of the binding of Hoechst 33258 with a different dodecamer d(CGCGATATCGCG)<sub>2</sub> showed the drug to be bound with the piperazine moiety up (towards the 5' end of the Watson strand) and lying roughly in the plane of the fourth GC base pair of the sequence with the remainder of the drug extending over the the next three AT base pairs, an GATA “pip-up” model.

A reexamination of the binding of Hoechst 33258 to the d(CGCGAATTCGCG)<sub>2</sub> dodecamer has been very recently reported (Quintana *et al.*, 1991). This study examined binding at different temperatures and concluded that at low temperature the drug occupied an AATT binding site with the preferred drug orientation being “pip-down”. However they also noted that even at 0°C they could not distinguish between the two possible orientations.

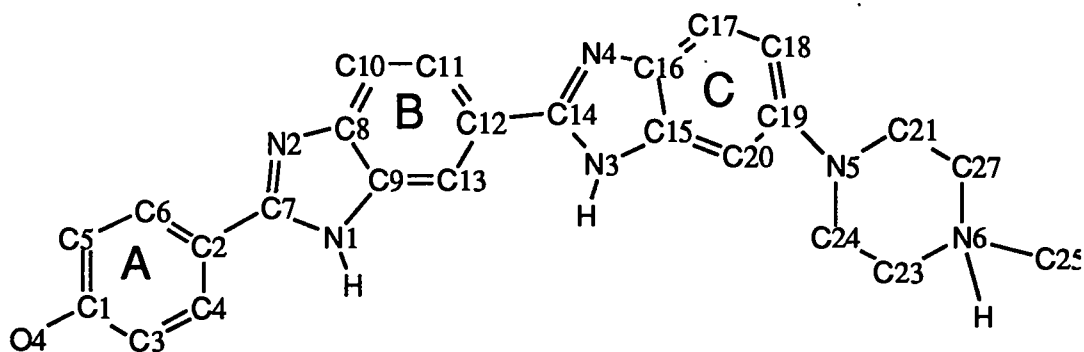


Figure 5.1. Chemical structure of Hoechst 33258 with crystallographic numbering used

## Materials and Methods.

### 5.2 Oligonucleotide synthesis.

The dodecanucleotide d(CGCAAATTTGCG) was prepared on an Applied Biosystems 391EP synthesiser using phosphoramidite chemistry, and purified by anion-exchange and reverse-phase (C8) HPLC procedures, as described in chapter 3.

Hoechst 33258 (2'-[4-hydroxyphenyl]-5-[4-methyl-1-piperazinyl]-2,5'-bi-1H-benzimidazole) was purchased from the Sigma Chemical Co. Ltd.

### 5.3 Crystallisation and data collection.

Crystals of the complex were grown using the hanging droplet vapour diffusion method by Dr. Jane Skelly. Optimum conditions for crystal growth were with 22  $\mu$ l droplets containing 5  $\mu$ l of 5 mM Hoechst 33258, 5  $\mu$ l of 20 mg/ml DNA [ d(CGCAAATTTGCG)<sub>2</sub> ], 10  $\mu$ l of 2-methylpentan-2,4-diol (MPD) at 23% vol/vol (MPD/sodium cacodylate buffer pH 7.0 containing 0.02% azide) and 2  $\mu$ l of 100 mM MgCl<sub>2</sub> against 23% vol/vol MPD/sodium cacodylate buffer pH 7.0. All solutions were filtered through a 2  $\mu$ m filter. The trays were then stored in an incubator (free from vibration) at 15°C. The crystals used for data collection grew in a period of 18-24 days. The crystals were long and needle-

like, with typical dimensions of 0.15 mm x 0.15 mm x 0.5 mm.

Intensity data were collected on a Xentronics area detector, using an Enraf-Nonius FR571 rotating anode source with the power at 40 mA / 60 KV. The crystal to detector distance was 8cm and the  $2\theta$  swing angle  $10^\circ$  with a frame collection time of 180 seconds. Two  $100^\circ$   $\omega$  scans were carried out with  $\phi$  rotated  $60^\circ$  between them to obtain the unique data set (as the crystal had been identified by auto-indexing, using the Xengen software (Howard *et al.*, 1987), as orthorhombic, space group  $P2_12_12_1$ ). A total of 11,799 of a possible 13,852 observations were collected to a resolution of  $2.25\text{\AA}$ . These were scaled and merged using the Xengen software package to give 3,114 unique reflections out of a possible 3,437 with a merging R factor of 6.07%.

---

Table 5.1. Resolution breakdown of data of collected.

---

Resolution limit $\text{\AA}$	Number of Reflections		R-Factor <sup>a</sup>
	Possible	Collected	
4.12	620	619	2.68
3.27	581	581	4.00
2.86	566	566	9.96
2.60	565	565	19.97
2.41	562	540	33.02
2.27	543	243	39.66
	<hr/> 3437	<hr/> 3314	<hr/> 6.07

---

a. Unweighted absolute R-factor on  $F^*100$

---

#### 5.4 Structure Solution and Refinement.

The structure crystallises in the orthorhombic space group  $P2_12_12_1$  with unit cell dimensions of  $a=25.27$ ,  $b=41.32$ ,  $c=65.11$   $\text{\AA}$ . The crystals are isomorphous with those of the same sequence complexed with the previously

studied minor groove binding drug, berenil ( $a=24.63$ ,  $b=40.60$ ,  $c=65.06$  Å) (see chapter 3). Therefore it was decided to use the same strategy to solve the structure and the final coordinates of this structure to solve the complex by molecular replacement. As previously described the coordinates simply had to be reformatted and converted into fractional coordinates for refinement.

For the same reasons as in chapters 2 and 3, it was decided to use the constrained/restrained refinement program CORELS (Sussman *et al.*, 1977) for the initial stages of refinement. Initially reflection data from 10-4 Å (662 unique reflections) were used together with the fractionalised, reformatted coordinates as a starting model which yielded an R factor of 36.0% after scaling and five cycles of positional refinement. At this stage the first  $|2F_o - F_c|_{\alpha calc}$  maps were calculated, to ensure that the model was essentially correct. The PROTEIN (Steigemann, 1987) suite was used to calculate the maps which were viewed on a Silicon Graphics Iris 3130 using the TOM display package (Cambillau, 1988). Both  $|2F_o - F_c|_{\alpha calc}$  and difference ( $|F_o - F_c|_{\alpha calc}$ ) maps already showed density in the minor groove of the DNA in the central AT region. Data was then used in the range 8 to 3 Å resolution, giving a total of 1375 reflections and 5 more cycles of refinement did not lower the R factor significantly, indicating that the molecule was correctly oriented. The structure was then divided into rigid groups of individual nucleotides, phosphates and sugars giving 48 rigid groups and allowing greater flexibility in the structure during refinement. Ten cycles of positional refinement followed by 5 cycles of temperature factor refinement reduced the R factor to 25.6%. The Afsig and Bfsig values and the scale were then updated and the resolution limit extended to 2.5 Å followed by a further 5 cycles of positional and temperature factor refinement which gave an R factor to 28.3%.

It was then decided to switch to the restrained least squares refinement program NUCLSQ (Westhof *et al.*, 1985). The weighting scheme consisted of

heavy weights, similar to those used in the refinement of the complex of berenil with  $d(\text{CGCAAATTTGCG})_2$ , for most parameters. Heavier weights were employed for the initial refinement procedure, in order to maintain good geometries and bond distances. After 10 cycles of refinement using the 2213 reflections greater than  $2\sigma$  from in the resolution range 7-2.2 Å the R factor was 27.6%. Maps of  $|2F_o - F_c|_{\alpha_{\text{calc}}}$  and  $|F_o - F_c|_{\alpha_{\text{calc}}}$  showed density in the minor groove. It was decided then to add possible solvent molecules (excluding the minor groove region), to enhance the quality of the minor groove density to enable a better fit of the ligand. Over the next forty cycles of refinement 50 putative solvent peaks were assigned as waters which resulted in an R factor of 21.3%. Examination of the minor groove density clearly indicated the presence of the ligand, however only the central portion of the drug, the two benzimidazole moieties could be fitted to the density. The quality of the density in the central region was such that one could see the angle of rotation between the two benzimidazole moieties (figure 5.2). Beyond this well defined central region the density was poor and it was not possible to distinguish which density could be assigned to the piperazine moiety and which to the phenol group. Since the density in the AT region of the minor groove, which was thought to correspond to the benzimidazole moieties, could be fitted with the Hoechst 33258 with the piperazine located towards the 5' end of the Watson strand and in the opposite orientation equally well, it was decided to refine both proposed complexes in tandem. The two models for the orientations for the Hoechst 33258 are related by a  $180^\circ$  rotation of the whole drug about the mid-point of the bond linking the two benzimidazole moieties.

The refinements both used the same weighting scheme and yielded near identical results in terms of R factor, temperature factors of the ligand, close contacts with the DNA and omit difference fourier maps. A few extra solvent peaks were assigned to each model but it was impossible to unequivocally

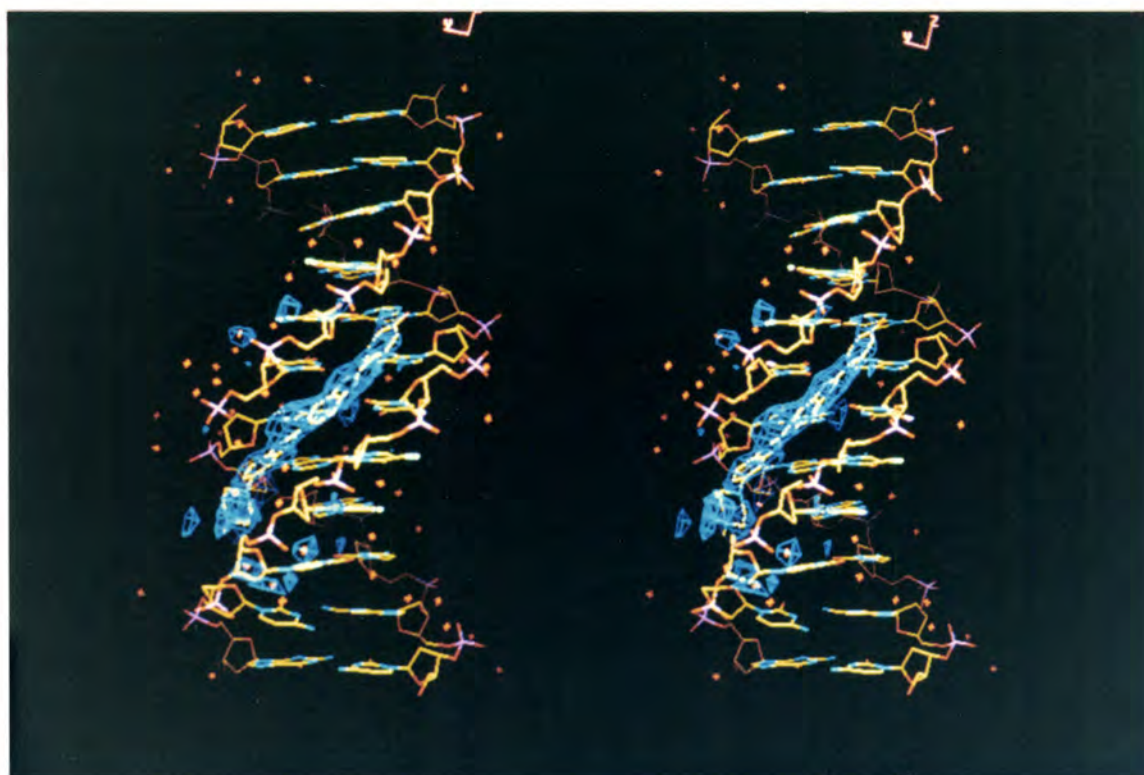


Figure 5.2. Difference fourier map showing  $2\sigma$  density in the minor groove.

determine a unique model. The refinements were therefore terminated with a reasonable R factor of 19.7% for the pip-up model and 19.6% for the pip-down model, without the addition of further water molecules.

### 5.5 Analysis of models detailing their various interactions with the DNA.

As both the orientations of the Hoechst 33258 refined to virtually the same R factor and the temperature factors associated with the ligand in each case were near identical, no definitive model could be assigned as the structural solution of the complex with the  $d(\text{CGCAAATTTGCG})_2$ . Both models occupy the same position with respect to the DNA duplex, being bound in the minor groove spanning the same five base pairs from Ade5 to Thy9 in AATTT pip-up and AATTT pip-down orientations figures 5.3 and 5.4). Each model forms what could be described as “specific” hydrogen bonds to the AT region of the DNA.

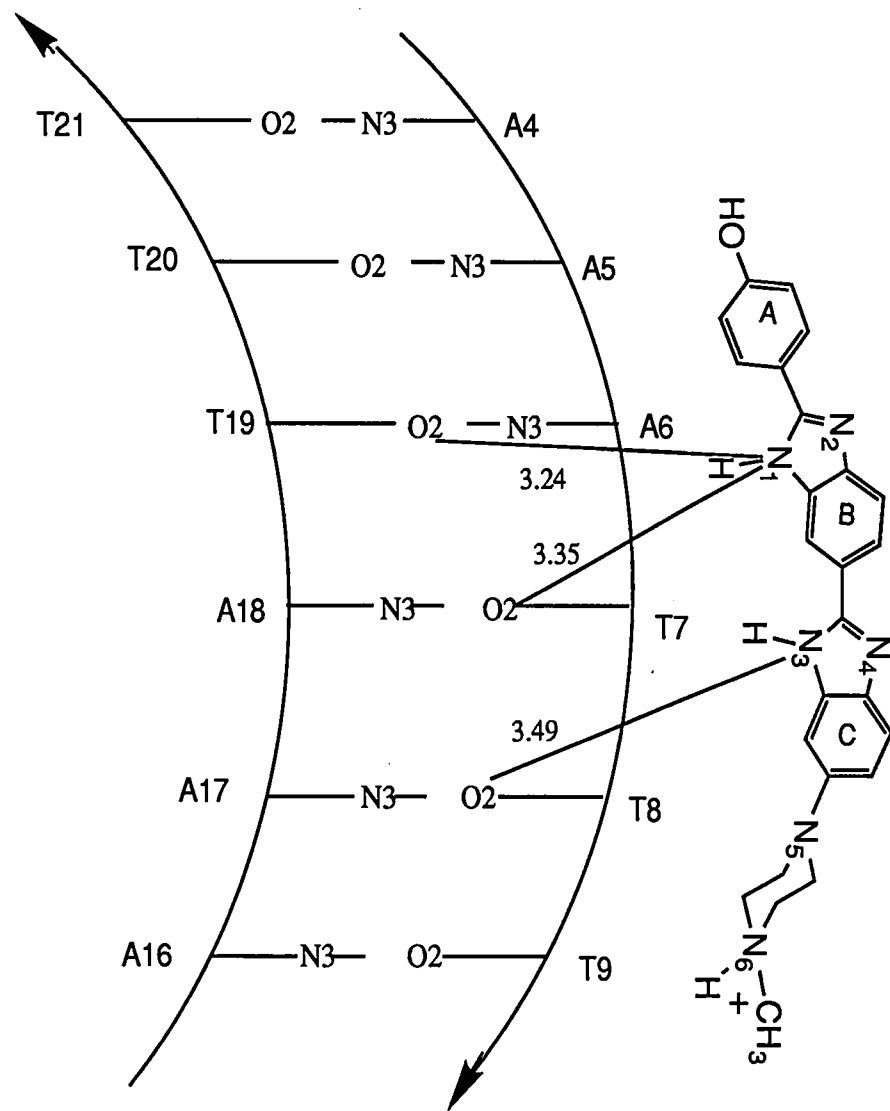


Figure 5.3. CGCAAATTTGCG-Hoechst (Pip-down)

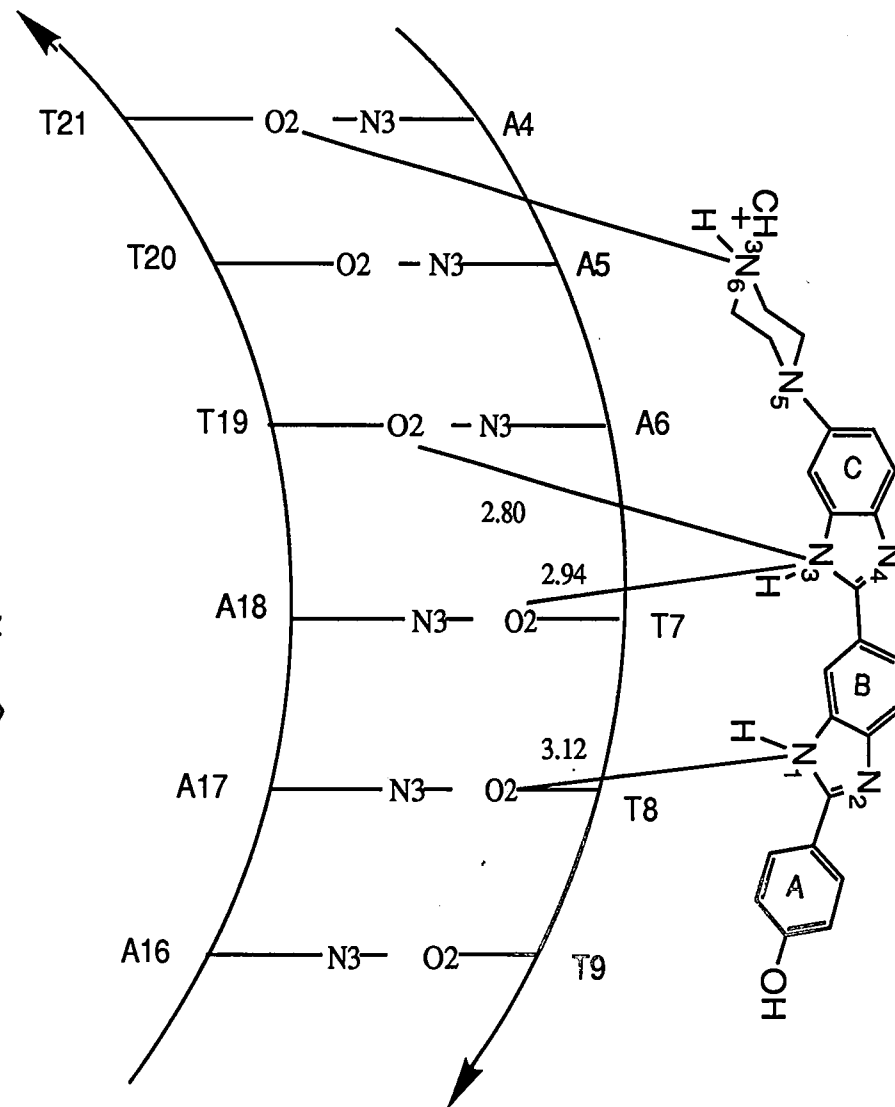
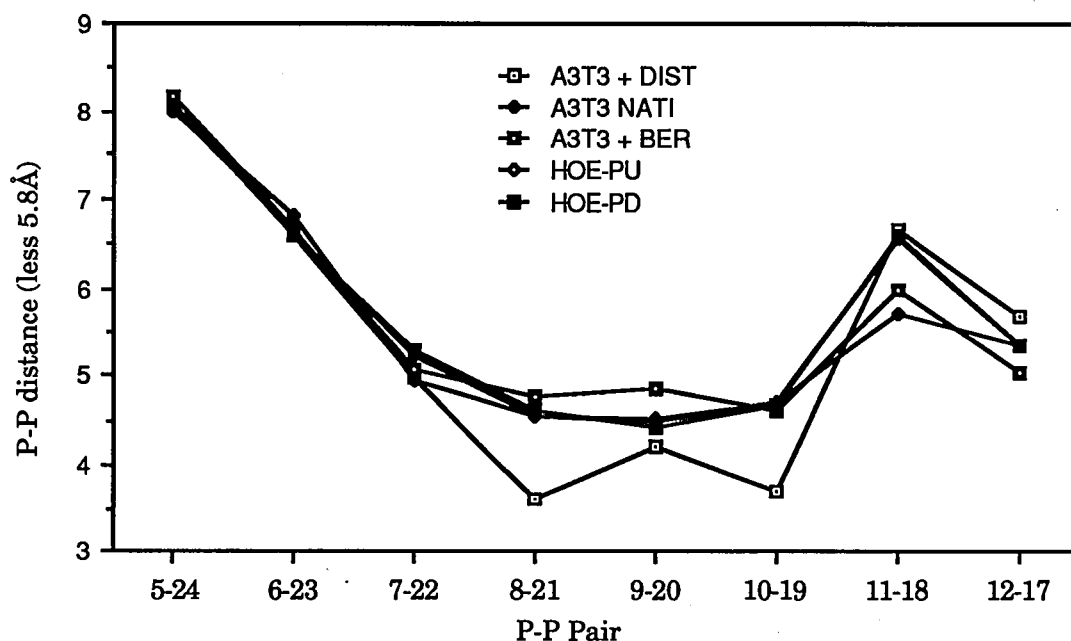


Figure 5.4. CGCAAATTTGCG-Hoechst (Pip-up)

### 5.6 Analysis of the "pip-down" model.

As mentioned above the ligand spans five base pairs from Ade5 to Thy9 with the methyl group piperazine moiety lying over the Thy9-Ade16 base pair and the hydroxyl group of the phenol moiety lying in the same plane as the Ade5-Thy20 base pair. There are several potential hydrogen bonds formed between Hoechst 33258 (Hoe) and the DNA, not only to the base pairs but also to the O4' of the backbone sugars. Thy21 O4'...O4 Hoe 3.06 Å, Thy20 O4'...N1 Hoe 3.15 Å, Thy19 O2...N1 Hoe 3.24 Å, Thy7 O2...N1 Hoe 3.35 Å, Thy8 O2...N3 Hoe 3.49 Å, Thy9 O4'...N3 Hoe 3.34 Å. It is interesting to note that the high propeller twists that are characteristic of the AT core of this sequence, facilitate the formation of the possible hydrogen bonds bringing both the O2 atoms of Thy19, Thy7 and Thy8 into favourable positions adjacent to N1 and N3 of the ligand respectively. As has been observed for other minor groove binding ligands the narrow minor groove width (figure 5.5) seems to favour drug binding, with the walls of the groove forming good van der Waals interactions clamping the ligand in place.

Figure 5.5 Graph of Minor groove Width based on P-P distances.



### 5.7 Analysis of the "pip-up" model.

The piperazine-up model also spans five base pairs, with the methyl group of the piperazine moiety lying in the same plane as Ade5-Thy20. The hydroxyl group of the phenol moiety lies in the plane of the Thy9-Ade16 base pair, and may form a hydrogen bond to the sugar of Gua10 with Gua10 O4'...O4 Hoe 3.54 Å. Possible hydrogen bonds to the bases that may add to the AT specificity of Hoechst 33258 are Thy7 O2...N3 Hoe 2.94 Å, Thy19 O2...N3 Hoe 2.80 Å, Thy8 O2...N1 Hoe 3.12 Å, Ade18 N3...N1 Hoe 3.52 Å and Thy9 O4'...N1 Hoe 2.98 Å.

### 5.8 Discussion.

Comparison of each of the previously solved crystallographic structures (figures 5.6-5.9) shows that the Hoechst 33258 is found in the AT region of the minor groove, with at least three possible hydrogen bonds to the base pairs. Both the "pip-up" and "pip-down" models presented in this study are consistent with these observations.

Using the information provided in this study it is clear that the high propeller twists inherent in the AT core region of this dodecamer facilitate possible hydrogen bond interactions between the two N-H hydrogen bond donors of the benzimidazole rings and the O2 oxygens of the thymine residues, which seem to be the more favoured targets as hydrogen bond acceptors over N3 of adenines. Also the density present in the minor groove is confined to the AATTT stretch of bases which indicates that although two possible orientations of the drug may be present there is no evidence for shuffling between adjacent sites which supports the findings of Searle *et al.*, 1991.

Previous studies have noted (Pjura *et al.*, 1987; Fede *et al.*, 1991), at least two possible orientations for the piperazine moiety. The data were not of sufficient quality to determine whether the piperazine ring has more than one conformation in this study, although the poor quality of the electron density of

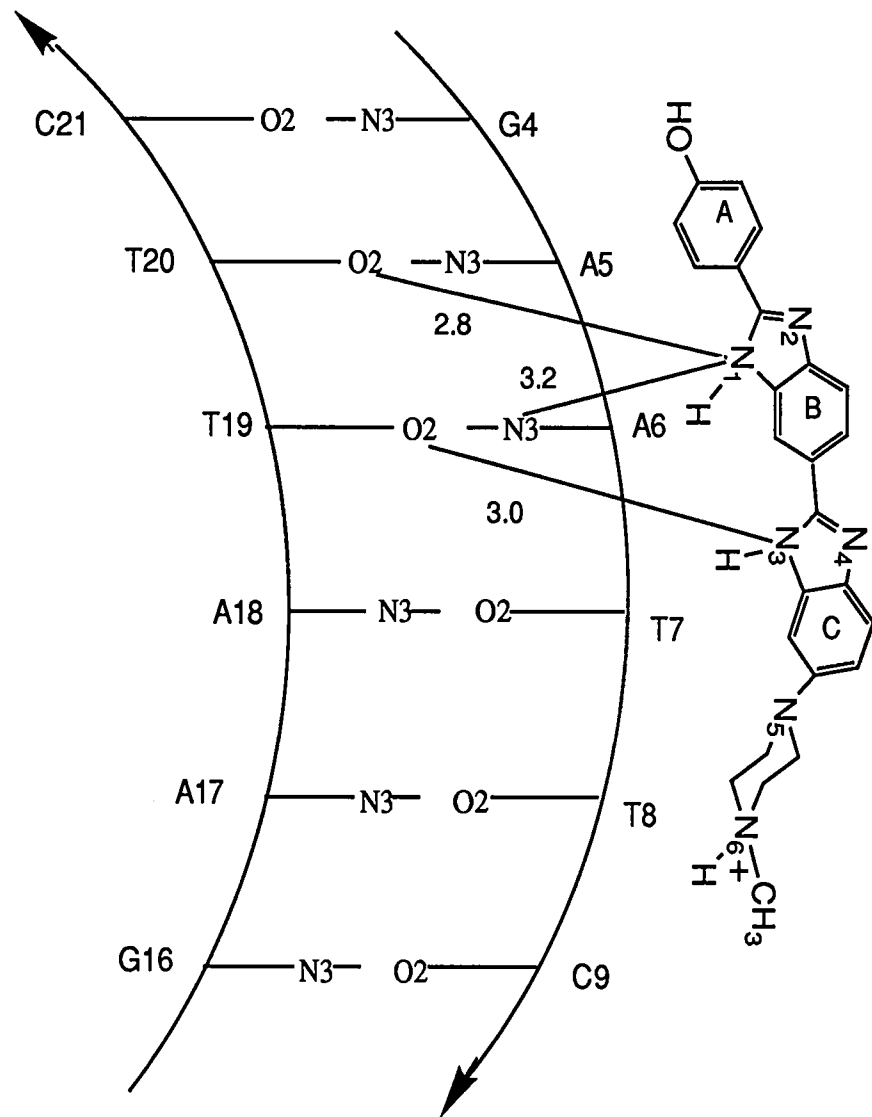


Figure 5.6. CGCGAATTCGCG-Hoechst 33258 (Teng et al., 1988)

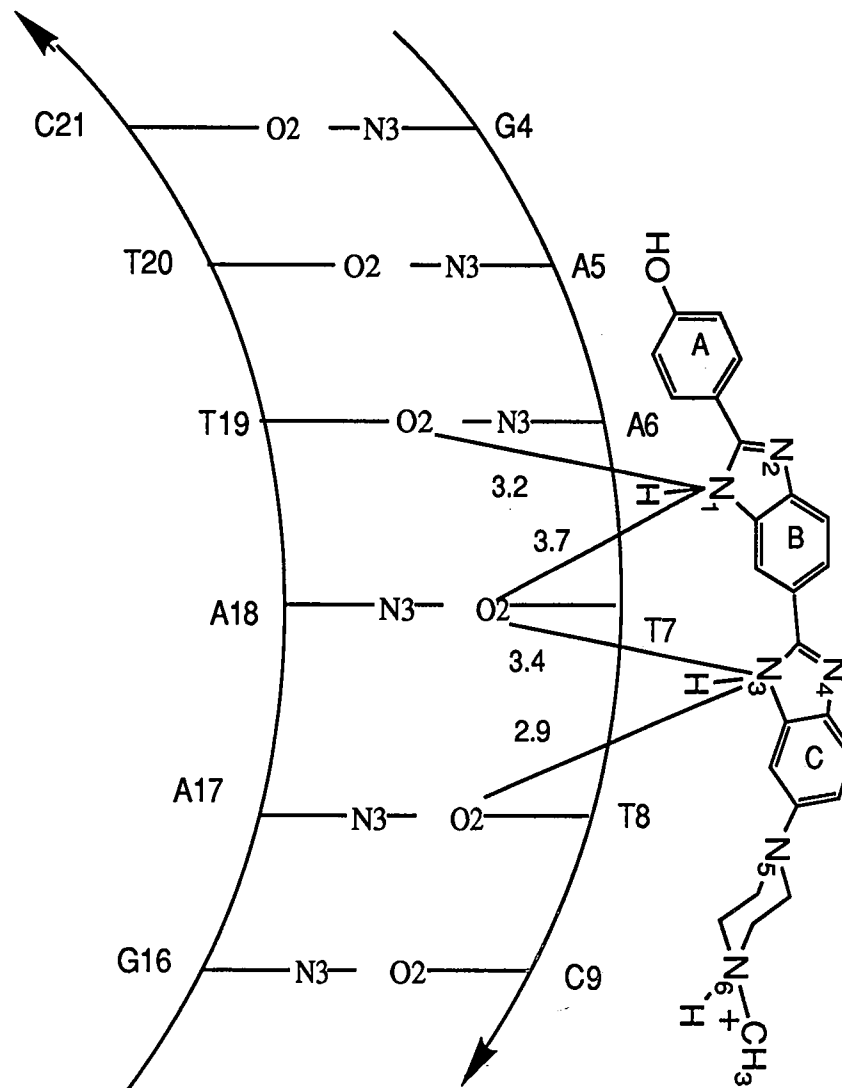


Figure 5.7. CGCGAATTCGCG-Hoechst 33258 (Pjura et al., 1987)

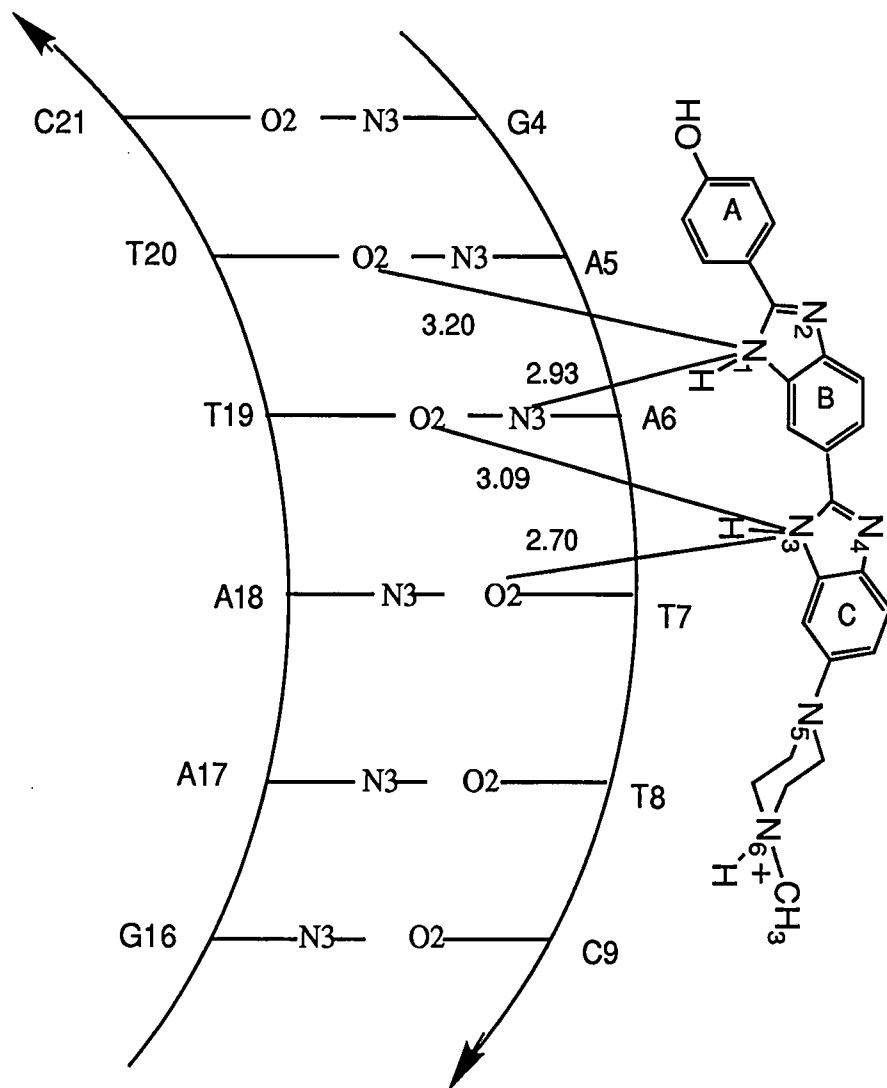


Figure 5.8. CGCGAATTCGCG-Hoechst -25°C (Quintana et al.,1991)

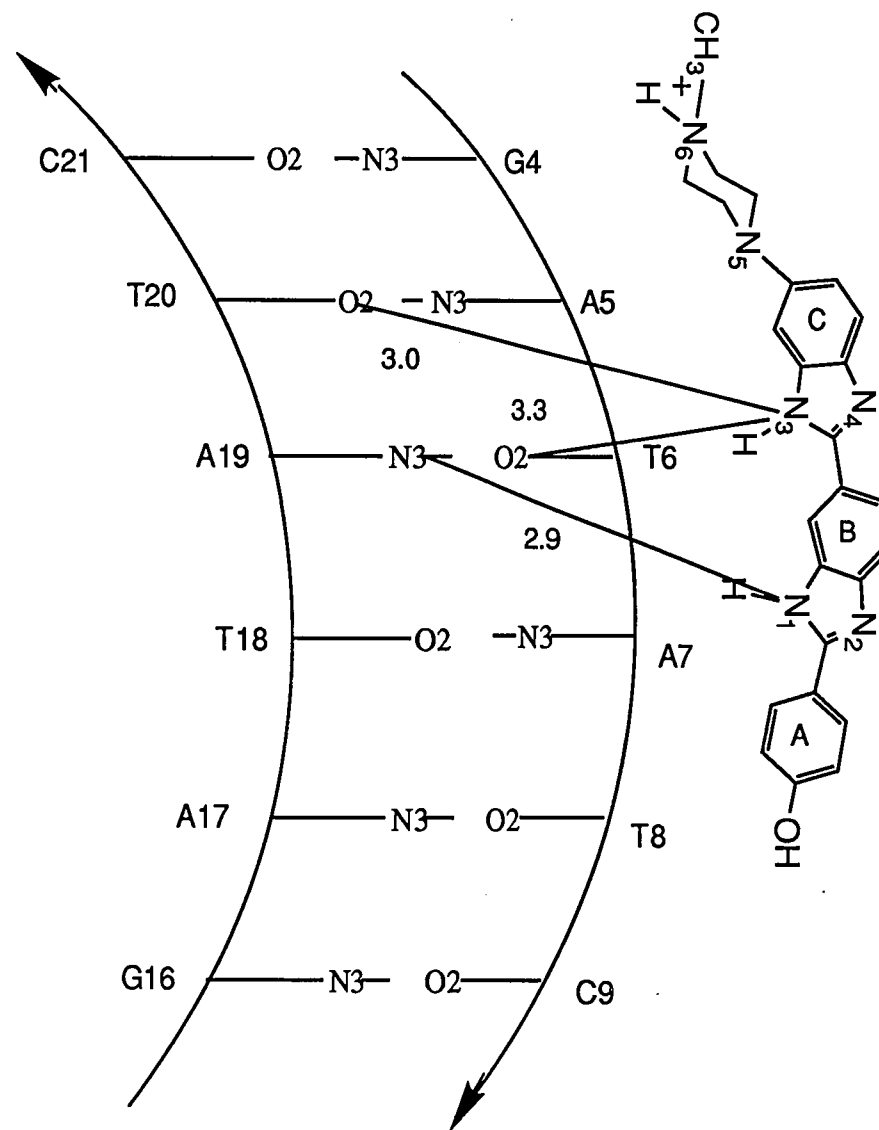


Figure 5.9. CGCGATATCGCG-Hoechst (Carrondo et al., 1989)

this moiety could be indicative more than one conformation.

The dodecamer used in this study also offers a more favourable sequence of bases for narrowing of the minor groove, than found in any of the other crystallographic or NMR spectroscopic studies. The consistent narrowness of the minor groove of this sequence over a larger range than for instance the d(CGCGAATTCGCG)<sub>2</sub> dodecamer may indicate that the drug has to perturb the DNA less on binding to this sequence. This may have the effect that the main factor in stabilising each of the two proposed models is close van der Waals rather than specific recognition of hydrogen bond acceptors on the minor groove floor. The above argument could also be applied to the observations of Quintana *et al.*, (1991) to explain why more than one orientation is possible at molar ratios greater than 1:1, only the most highly favoured model is selected out by low drug:DNA ratios.

The observation of two equally possible drug orientations at 16°C with a high drug:DNA ratio in the crystallisation conditions is completely in accord with those found by Quintana *et al.*, (1991), published during this study. It would seem that in order to obtain a unique solution to the crystal structure one would have to reduce the initial drug:DNA ratio in the crystallisation conditions to a 1:1 ratio, grow the crystals at 4°C and collect the data at low temperature, perhaps -100°C, based upon the observations of Quintana *et al.*

## **CHAPTER 6.**

General conclusions on minor groove-binding ligands.

This thesis presents crystallographic and NMR spectroscopic data on the minor groove-binding ligand berenil. The results show two different binding modes dependent on the nucleotide sequence of the DNA duplex, and also suggest that the crystal structure and the NMR derived solution structure differ only very slightly in the case of  $d(\text{CGCGAATTCGCG})_2$  but are distinctly different in the study with the  $d(\text{CGCAAATTTGCG})_2$ . The thesis also presents crystallographic data for the binding of Hoechst 33258 to the latter sequence, which also shows sequence selectivity from within a range of "specific" binding sites. The questions arising from these results are:-

How does the nucleotide sequence determine the mode of binding of a minor groove-binding ligand and what structural features of the DNA and the drug determine the sequence specificity within a favoured binding region?

### *6.1. The factors influencing binding of positively-charged minor groove-binding ligands.*

#### *Charge interactions.*

Initial long range attraction is to the negativity of the DNA as a whole, with initial selectivity for the more electronegative minor groove region. The crystallographic and spectroscopic experiments are carried out in high-salt conditions, which increases the dielectric constant of the solution (crystal), and therefore increases the cooperativity of binding. Selectivity for AT regions is further enhanced by virtue of the fact that they have been calculated to be more electronegative than GC regions.

#### *Steric considerations.*

Ligands that bind via the minor groove tend not to have bulky substituents, they are relatively planar. Where more than four base pairs are spanned the ligands often consist of planar groups, linked by a rotatable bond so that the drug can follow the curvature of the minor groove of the DNA. The inner binding surface of minor groove binders is normally crescent-shaped to give a

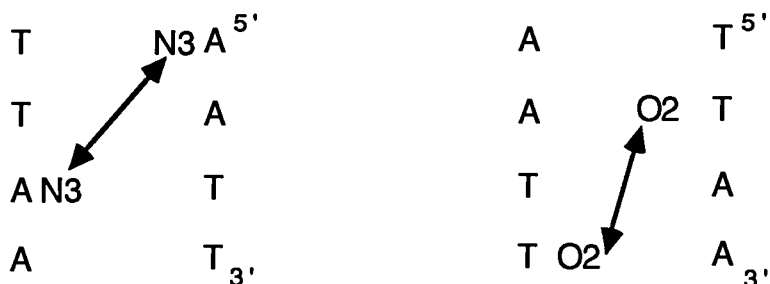
close isohelical fit with the minor groove floor. The contours of the minor groove floor impart a great deal to the selectivity of these ligands in that the NH<sub>2</sub> groups present on guanines in the minor groove sterically inhibit their binding. Therefore by excluding guanines as possible binding sites one has increased AT selectivity.

The width of the minor groove is also a sequence-dependent feature of the DNA which has been shown to play an important role in the interaction between the drug and the DNA by providing good van der Waals interactions. A run of either adenines, or adenines followed by thymines (reading from 5' to 3') will cause a gradual decrease in minor groove width coupled with a bend in the DNA duplex. Interruption of this adenine tract with guanine or cytosine or a thymine tract, does not produce this narrowing.

*Specific patterns of hydrogen bond donors and acceptors.*

Within an AT tract there is probably a broad energy well from which a range of particular models can be isolated. The formation of hydrogen bonds to the acceptors/donors on the minor groove floor further stabilises the complex. It is important to note the pattern of possible hydrogen bond acceptors within the region that act to stabilise the non-covalent complex. When trying to establish a hierarchy of sequence preference and model selection within AT tracts, it is the displacement of the N3 and O2 donors for adenine and thymine respectively that is important, and the angle with for which optimal hydrogen bond geometry would be satisfied. There is a subtle difference in positions of these acceptors within the minor groove with O2 of thymine positioned slightly further from the middle of the minor groove compared with the N3 of adenine. This may be a key factor in the selection of possible 1:2, 1:3 or 1:4 models. Also one must take into account the flanking regions which provide steric influence in the minor groove region adjacent to the potential binding site. Flanking GC base pairs also impose, by virtue of the greater rigidity in terms of propeller twist, due to possession of three base pair hydrogen bonds,

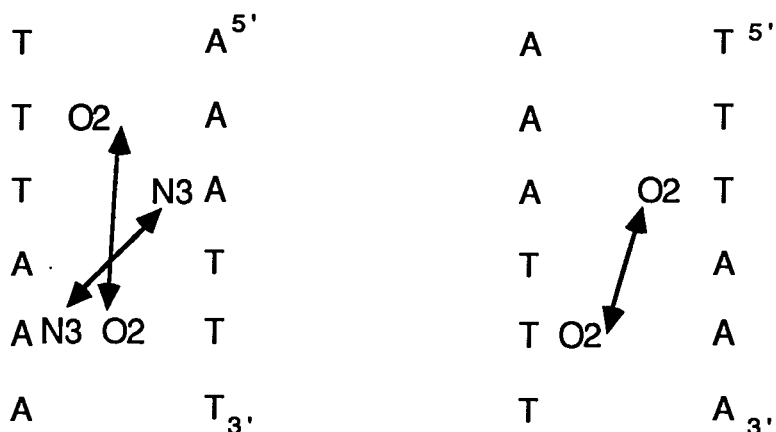
limitations on the degree of propeller twisting in the AT site. Further the influence of flanking bases on the AT site must be considered in terms of base pair stacking interactions and the freedom to propeller twist and roll, particularly for short AT regions where the drug might overlap into these flanking regions. Various possible binding sites for berenil are examined below.



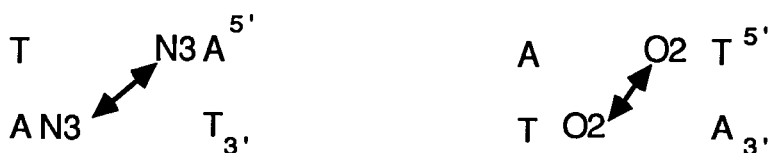
Of the two forms of a 1:3 interaction, the N3-N3 interaction has been found in both NMR and crystallographic studies. The bend in the DNA and high propeller twists of the AT base pairs may be important structural features which are "sensed" by the drug to give this preferred mode of binding. Although there is no quantitative evidence for bending in the NMR data, chemical shift anisotropy measurements cannot be fitted by a precise cylindrical model which is indicative of some degree of bend. The left of the two sequences above could possibly bind in a 1:4, O2-O2 mode but perhaps the propeller twists of the base pairs at the end of the AT tract are not sufficiently high due to the "stiffness" of the adjacent CG or GC base pair. The flanking sequence is also important in the drug selecting AAT in preference to the ATT site due to clashes with the N2 group of a guanine, although an adjacent cytosine would be better tolerated.

With the longer AT tract there are several possible 1:3 binding modes (see below). It now also becomes possible for the berenil to span four base pairs from thymine O2 to thymine O2, due to the high propeller twist associated with these base pairs (this is dependent of the flanking sequence). There is also evidence for the existence of the 1:4 O2-O2 model in the complex of berenil with

d(GCAATTGC)<sub>2</sub> (Shafer *et al.*, 1990). This 1:4 model may form more favourable hydrophobic contacts than the 1:3 model in the crystal structure where the inherent bend (partially or totally due to crystal packing forces) selects this model out. This may not be true in the solution studies on the same sequence where at least two binding modes, probably 1:3, are in dynamic equilibrium.



Another 1:3 model spanning O2-O2 is also possible given the appropriate sequence. However there is no structural evidence for this mode of binding.



With only a two base pair model the 1:2 N3-N3 model is just possible if the previous 5' base is not a guanine or the next 3' base is not cytosine as the NH2 group of the guanine in the GC base pair would not be tolerated. The 1:2 O2-O2 model would not be at all favourable as it would not be possible to form any stabilising hydrogen bonds to the O2 acceptors.

The above explanations of the the possible 1:2, 1:3 and 1:4 binding modes is consistent with crystallographic, NMR spectroscopic and footprinting data if the order of ranking for berenil binding is:

$$1:3 \text{ N3-N3} \approx 1:4 \text{ O2:O2} > 1:3 \text{ O2-O2} > 1:2 \text{ N3-N3}$$

This hierarchy for preferred binding is not only dependent on the sequence of

the spanned bases but also the bases flanking this sequence. In AT tracts consisting of six base pairs, there appears to be a broad binding region in which several possible binding modes exist in dynamic equilibrium.

Modelling using mechanics can give an idea of the sequence preference due to basic steric and electrostatic interactions, but clearly such simple calculations are not sufficiently elaborate to take into account sequence-dependent structural variability in DNA. DNA sequence induces structural variations in DNA, such as bending, high propeller twists or large changes in roll parameters, all of which effect the patterns of both hydrogen bond donors/acceptors and groove width, to which minor groove-binding drugs are sensitive. The extension of "simple" *in vacuo* studies by the use of molecular dynamics has shown there is a tendency for the minor groove to expand for the d(CGCAAATTTGCG)<sub>2</sub> dodecamer (Boehncke *et al.*, 1991), a conformational feature of the DNA to which a minor groove binding ligand is highly sensitive.

The study involving the Hoechst 33258 compound, although inconclusive, would seem to infer that minor groove width is of key importance in perhaps determining the sequence to which the ligand binds. The Hoechst ligand is displaced to the 3' end of the duplex where the groove is narrower. Hydrogen bonds will still be an important factor in stabilising the complex.

## **APPENDIX I.**

Crystallographic and refinement theory.

X-rays are scattered by the electron clouds surrounding atomic nuclei. If the atomic scattering sites form a periodic arrangement then an incident beam of X-rays will only be scattered in certain defined directions. If the basic repeating unit of a periodic arrangement, the crystal, is a parallelepiped described by three vectors  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  (the unit cell) and  $\mathbf{s}$  is the vector ratio of the vector difference between the incident and scattered beam directions, then three conditions limiting reflections hold simultaneously:

$$\mathbf{a} \cdot \mathbf{s} = h \quad \mathbf{b} \cdot \mathbf{s} = k \quad \mathbf{c} \cdot \mathbf{s} = l \quad \text{The Laue Equations} \quad \text{I.1}$$

where  $h$ ,  $k$  and  $l$  are integers which are equivalent to the Miller indices which define planes through a crystal.

The total scattered amplitude of a diffracted beam from a crystal is defined with respect to that from an array of electrons, one at each unit cell origin, and is written as

$$F_{hkl} = \sum_{j=1}^N f_j \exp^{2\pi i \mathbf{r}_j \cdot \mathbf{s}} \quad \text{I.2}$$

where  $f_j$  is the atomic scattering factor of the  $j^{\text{th}}$  atom (defined as the amplitude of the coherent scattered radiation from the atom relative to that from a single electron situated at the atomic centre). The summation is over the  $N$  atoms in the unit cell. The position of the  $j^{\text{th}}$  atom in the unit cell is  $\mathbf{r}_j$  which in terms of fractional atomic coordinates is

$$\mathbf{r}_j = x_j \mathbf{a} + y_j \mathbf{b} + z_j \mathbf{c} \quad \text{I.3}$$

The scattering vector  $\mathbf{s}$  can be described in terms of the reciprocal lattice vector  $G_{hkl}$ :

$$G_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad \text{I.4}$$

and

$$\mathbf{s} = 2\pi G_{hkl} \quad \text{I.5}$$

where  $\mathbf{a}^*$ ,  $\mathbf{b}^*$  and  $\mathbf{c}^*$  are the base vectors in reciprocal space which are related to the real vectors  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  for a right handed system as described below:

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \quad \text{I.6}$$

Combining equations I.3, I.4 and I.5 an expression for  $\mathbf{r}_j \cdot \mathbf{s}$  may be written as:

$$\begin{aligned} \mathbf{r}_j \cdot \mathbf{s} &= 2\pi(x_j \mathbf{a} + y_j \mathbf{b} + z_j \mathbf{c}) \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \\ &= 2\pi(hx_j + ky_j + lz_j) \end{aligned} \quad \text{I.7}$$

Thus an alternative form of equation I.2 can be written in terms of the Miller indices, thus the relative amplitude of a diffraction spot associated with the  $hkl$  reciprocal lattice point is given by the structure factor  $F_{hkl}$ :

$$F_{hkl} = \sum_{j=1}^N f_j \exp 2\pi i (hx_j + ky_j + lz_j) \quad \text{I.8}$$

Equation I.8 is complex and can be resolved into its real and imaginary components:

$$\begin{aligned} F_{hkl} &= A_{hkl} + iB_{hkl} \\ A_{hkl} &= \sum_{j=1}^N f_j \cos 2\pi(hx_j + ky_j + lz_j) \\ B_{hkl} &= \sum_{j=1}^N f_j \sin 2\pi(hx_j + ky_j + lz_j) \end{aligned} \quad \text{I.9}$$

The phase angle  $\phi_{hkl}$  associated with  $F_{hkl}$  is given by:

$$\phi_{hkl} = \tan^{-1} \frac{B_{hkl}}{A_{hkl}} \quad \text{I.10}$$

From equations I.9 Friedel's Law can be derived:

$$|\overline{F_{hkl}}| = |F_{hkl}| \quad \text{I.11}$$

and 
$$\phi_{\overline{hkl}} = -\phi_{hkl} \quad \text{I.12}$$

Due to the intrinsic vibration of the atoms within the crystal and the fall off in scattered amplitude with increasing  $\theta$  (measured between the incident and reflected beams and the reflecting plane, the atomic scattering factor  $f_j$  should be written:

$$f_j = f_0 \exp -B_j \frac{\sin^2 \theta}{\lambda} \quad \text{I.13}$$

where  $\lambda$  is the wavelength of the incident radiation,  $B_j$  is the temperature factor which is a measure of the root mean square amplitude of thermal motion.  $B$  is expressed in  $\text{\AA}^2$ , and  $B = 8\pi\bar{u}^2$  where  $\bar{u}$  is the mean square displacement of atoms from their mean positions.

The above approach the scattering amplitudes have been derived for point scatterers. However, the more correct approach is to treat the crystal as having continuously varying electron density. If the crystal is divided into small volumes  $dv$  with a point charge  $\rho(\underline{r}) dv$  where  $\rho(\underline{r})$  is the electron density distribution the expression for the total scattering amplitude becomes:

$$F(\underline{s}) = \int_v \rho(\underline{r}) \exp 2\pi i \underline{s} \cdot \underline{r} dv \quad \text{I.14}$$

The inverse transform of I.14 is:

$$\rho(\underline{r}) = \int_v F(\underline{s}) \exp -2\pi i \underline{s} \cdot \underline{r} dv^* \quad \text{I.15}$$

As  $F(s)$  only has magnitude at reciprocal lattice points I.15 may be expressed:

$$\rho(\underline{r}) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} \exp -2\pi i(hy_j + kx_j + lz_j) \quad \text{I.16}$$

If Friedel's law holds then:

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hkl}| \cos (2\pi(hy_j + kx_j + lz_j) - \phi_{hkl}) \quad \text{I.17}$$

The magnitudes of the  $F_{hkl}$ 's can be measured from the intensity of the diffracted beams. However, the phases  $\phi_{hkl}$  have to be derived. Once the phases are derived the electron density can be calculated with the peaks occurring at the atoms sites.

The  $|F_{hkl}|$  is related to the measured diffracted intensities by:

$$|F_{hkl}| = \sqrt{\frac{KI_{hkl}}{Lp}} \quad \text{I.18}$$

where  $p$  is the polarisation factor,  $L$  is the Lorentz factor and  $K$  is scaling constant.

$$K = \left( \frac{e^2}{4\pi\epsilon_0 mc^2} \right)^2 N^2 \lambda^3 \quad \text{I.19}$$

where  $N$  is the total number of unit cells within a volume  $dv$  and  $\lambda$  is the wavelength of the incident radiation. The other constants are the atomic constants.

$K$  is only evaluated when the absolute magnitudes of the structure amplitudes are required. It is usual to work with  $|F_{rel}|$

$$|F_{rel}| = R|F_d| = \sqrt{\frac{I_{hkl}}{Lp}} \quad \text{I.20}$$

During refinement a scaling factor is used to scale  $|F_{\text{rel}}|$  to  $|F_o|$  so that comparison can be made with  $|F_{\text{calc}}|$ .

The polarisation factor arises to compensate for the partial polarisation of the scattered beam (as the incident radiation is usually unpolarised) in the direction parallel to the crystal plane. The polarisation factor is given by:

$$p = \frac{1 + \cos^2 2\theta}{2} \quad \text{I.21}$$

The Lorentz factor arises due to the finite time a reciprocal lattice point takes to intercept the Ewald sphere, which is a function of their scattering angles.  $L$  is dependent on the diffraction geometry used. For equi-inclination geometry:

$$L = \frac{\sin\theta}{\sin 2\theta \sqrt{\sin^2\theta - \sin^2\mu}} \quad \text{I.22}$$

where  $\mu$  is the equi-inclination angle.

When collecting X-ray diffraction data the phase of the scattered beam is unknown, only the intensity of the beam is recorded. The method of molecular replacement assigns the phases calculated from the starting model as a first estimate. Any method used to satisfactorily determine the three-dimensional structure must adjust the parameters of the model; the positions  $x_i, y_i, z_i$  and the thermal parameters  $B_i$  (temperature factors) of the  $i$  constituent atoms within the unit cell to optimise the statistical agreement between the scattered amplitudes observed and those calculated from the model.

To reconstruct the electron density map of the unit cell (to which one can fit the optimal atomic model) the magnitude and phase of each of the diffracted spots are needed.

In the process of refinement as the model parameters change from one iteration to the next so the calculated phases also change and these are monitored by the R factor summed over all measured reflections hkl, defined as:-

$$R = \frac{\sum_{hkl} |F_{o,hkl}| - |F_{c,hkl}|}{\sum_{hkl} |F_{o,hkl}|} \quad \text{I.23}$$

hence the lower the R factor the closer the agreement between the observed and calculated amplitude.

The problems faced in macromolecular crystallography are compounded by the limit of resolution. By Bragg's law

$$d_{\min} = \frac{n\lambda}{2\sin \theta_{\max}} \quad \text{I.24}$$

$d_{\min}$  is the limit of resolution (i.e. minimum separation of the Bragg planes)  $\theta_{\max}$  is the maximal diffraction angle. Macromolecules are do not pack in a perfect lattice, there are small differences between one unit cell and the next (known as static disorder) which gives rise to diffuse scattered intensity. The atoms also undergo thermal motion (dynamic disorder) so over a data collection of hours to days we see the time-averaged position, which again reduces the limit of resolution. A further factor affecting data collection, reducing the resolution, is that of radiation damage, this occurs in two forms. i) the production of free radicals,  $H\cdot$ ,  $OH\cdot$  and solvated electrons all of which are highly reactive. The radiation-induced yield of such free radicals produced is proportional to the amount of solvent (water) present in the crystal, typically 40-50% w/w for oligonucleotide crystals. ii) Heating effects causing greater

thermal motion. These can be reduced by cooling the crystal.

When solving any crystallographic structure the number of observations must be equal to or greater than the number of parameters to be evaluated. The greater the ratio of observations to parameters, the more reliable the derived parameters tend to be. For this reason one tries to maximise the number of observations or minimise the number of parameters used to describe the model. With macromolecules the number of parameters is often greater than the number of observations (unique reflections) it is possible to collect, due to the rapid fall off of intensities of diffracted beams with the diffraction angle  $\theta$ .

Due to the limitations of a macromolecular data set, additional information such as bond lengths, bond angles, torsion angles and van der Waals radii can be incorporated into the refinement procedure. These reduce the number of parameters by fixing certain geometries, or include the additional information as extra observations.

Two refinement programs were used in the each of the crystallographic structure refinements, Joel Sussman's constrained-restrained program CORELS and Eric Westhof's restrained refinement program for nucleic acids NUCLSQ. Both programs use the least-squares method of refinement. Restraints involve restricting model features, such as bond lengths, bond angles, van der Waals radii and torsion angles, to a realistic range conformations thereby increasing the number of observations. Constraints confine certain features to specific values, such as planar groups in nucleic acids, thereby reducing the number of independent variables. Restraints can be thought of as the addition of ideal configurations with springs pulling the real atoms towards them; constraints can be thought of as the imposition of rigid rods between atoms, reducing their degrees of freedom.

CORELS was developed for proteins and nucleic acids to take advantage

of the intrinsic rigid groups found in these molecules to overcome the limitations of an underdetermined problem.

The quantity  $Q$  to be minimised in the least squares procedure consists of the sum of four separate observational functions.

$$Q = w_F DF + w_D DD + w_V DV + w_T DT \quad \text{I.25}$$

where  $w_F$ ,  $w_D$ ,  $w_V$  and  $w_T$  are the overall weights for each term.

The first term,  $DF$  comprises of the sum of the structure factor differences summed over all reflections  $hkl$ :

$$DF = \sum_{hkl} \frac{1}{\sigma_F} (|F_{o,hkl}| - k|F_{c,hkl}|)^2 \quad \text{I.26}$$

where  $\sigma_F$  is the standard deviation of  $F_o$  at the end of the refinement. Different weighting schemes with  $\sigma_F$  equal to half the residuals ( $\sigma_F = 1/2 |F_o - F_c|$ ) are often used during refinement and  $k$  is the scale factor.

The second term,  $DD$  restrains the stereochemistry of the structure and represents the sum of all the distance restraints  $d$ :

$$DD = \sum_d \frac{1}{\sigma_d} (D_{o,d} - D_{c,d})^2 \quad \text{I.27}$$

where  $D_{o,d}$  is the "ideal" distance between a specified pair of atoms. This distance could correspond to a bond length, or be due to a bond angle or torsion angle restraint.  $D_{c,d}$  is the distance calculated from the model. The weighting for these stereochemical restraints  $\sigma_D$  corresponds to the standard deviation in distances of type  $d$  observed in small structures (from which the distance restraints arise).

The non-bonded close contact term  $DV$  is the sum over all non-bonded distance restraints  $v$ . The term refers to the repulsive van der Waals potential and is therefore only considered when  $D_{c,v} < D_{o,v}$ , where  $D_{o,v}$  is the "close contact" distance between a specified pair of atoms and  $D_{c,v}$  is the distance

calculated from the model.

$$DV = \sum_v \frac{1}{\sigma_v} (D_{o,v} - D_{c,v})^2 \quad \text{I.28}$$

$\sigma_v$  corresponds to the standard deviation of values for non-distances of a particular type.

The final term, DT, restrains the structure from deviating grossly from the input coordinates. The sum is over all atoms  $i$  with three positional coordinates  $j$  for each atom:

$$DT = \sum_i w_i \sum_{j=1}^3 (X_{T,i,j} - X_{i,j})^2 \quad \text{I.29}$$

where  $X_{T,i,j}$  are the coordinates of the input atoms and  $X_{i,j}$  are the corresponding coordinates of the working model.

Each constrained group is fully described relative to an arbitrary origin by means of a translation and rotation vector, with in addition any internal dihedral angles and temperature factors. Then the quantity  $Q$  (equation I.25) is minimised, which will proceed dependent upon the group parameters as well as the scale factor relating  $F_o$  and  $F_c$ .

$$Q = Q(X_i, R_i, \psi_{i1}, \psi_{i2}, \dots, B_{i1}, B_{i2}, \dots, k) \quad \text{I.30}$$

where  $X_i$  and  $R_i$  are the translation and rotation vectors of group  $i$ .  $\psi_{in}$  are the  $n$  internal dihedrals of the group,  $B_{im}$  are the  $m$  temperature factors and  $k$  is the scale factor of the structure.

The minimisation of  $Q$  with respect to  $X_i$  and  $R_i$  is performed using the appropriate derivatives of  $Q$  from (equation I.25). The least squares normal equations are constructed from the sets of orientational and thermal derivatives and the resulting matrix solved by the conjugate gradient iterative method.

CORELS described above was developed from the Konnert-Hendrickson

program PROLSQ, which was subsequently developed for nucleic acids by Eric Westhof resulting in NUCLSQ. NUCLSQ as mentioned before is a restrained least squares refinement program which works on the same principle as all least squares programs which is to minimise the weighted sum over all observations of the squared residuals:

$$\phi(\mathbf{X}) = \sum_{hkl} w_{hkl} [g_{hkl}^{\text{obs}} - g_{hkl}^{\text{calc}}(\mathbf{X})]^2 \quad \text{I.31}$$

The appropriate weights  $w_{hkl}$  are the inverse of the variances of the observations. In NUCLSQ each piece of information is treated as an observational equation of the form:

$$g^{\text{obs}} = g^{\text{calc}}(\mathbf{X}) + \varepsilon \quad \text{I.32}$$

where  $\varepsilon$  is the residual describing the discrepancy between  $g^{\text{obs}}$  and  $g^{\text{calc}}$  computed for the parameters  $\mathbf{X}$ .

NUCLSQ breaks down the observations into different classes which are in turn treated as separate observational functions, which are then summed together:

$$\Phi(\mathbf{X}) = \sum_i \phi_i \quad \text{I.33}$$

The observational function related to the structure factor is given below:

$$\phi_1 = \sum_{hkl}^{\text{reflections}} \frac{1}{\sigma_F^2(hkl)} (|F_{hkl}^{\text{obs}}| - |F_{hkl}^{\text{calc}}|)^2 \quad \text{I.34}$$

where the weighting factor is determined by  $\sigma_F$  and  $h$  is  $hkl$ .

There are three types of interatomic distances that can be used to restrict the stereochemistry of the structure. Actual bond distances, the next-nearest-neighbour distances from triples of atoms to define bond angles and the first to fourth distances that describe dihedral angles and possible hydrogen bonds. The resulting observational function is:

$$\phi_2 = \sum_j^{\text{distances}} \frac{1}{\sigma_D^2(j)} (d_j^{\text{ideal}} - d_j^{\text{model}})^2 \quad \text{I.35}$$

where  $\sigma_D$  is the standard deviation for distances of a particular type.

Planarity can be used to impose further restraints on the system. NUCLSQ restrains the deviations of atoms from the least squares plane of the group to give:

$$\phi_3 = \sum_k^{\text{planes}} \sum_i^{\text{coplanar atoms}} \frac{1}{\sigma_P^2(i,k)} (\mathbf{m}_k \cdot \mathbf{r}_{i,k} - d_k)^2 \quad \text{I.36}$$

where  $\mathbf{m}_k$  and  $d_k$  are parameters of the least squares plane and  $\sigma_P$  is the standard deviation permitted.

Another stereochemical restraint that can be employed is the stereoconfiguration at chiral centres (e.g. asymmetric carbon atoms). The chiral volume is the triple scalar product from a central atom to three attached atoms. This give rise to the function:

$$\phi_4 = \sum_1^{\text{chiral centres}} \frac{1}{\sigma_C^2(1)} (V_1^{\text{ideal}} - V_1^{\text{model}})^2 \quad \text{I.37}$$

Contacts between non-bonded atoms are also important determinants of macromolecular conformation. These contacts are characterised by potential energy functions  $U(d)$  such as the Lennard-Jones function that have a steep repulsive barrier for close approach and a shallow weak attractive well. The repulsive part of the potential can be approximated by:

$$U(d) - U(d^{\text{min}}) \approx \frac{1}{\sigma^{2n}} (d - d^{\text{min}})^{2n} \quad d < d^{\text{min}} \quad \text{I.38}$$

The best fit in the range  $d^{\text{min}} - 2\text{\AA}$  and  $d^{\text{min}}$  have  $n=2$  for contacts among H, C,

N, O, S and P thus the observational function taken for the repulsive contacts where  $d^{\text{model}} < d^{\text{min}}$  is:

$$\phi_5 = \sum_m^{\text{nonbonded contacts}} \frac{1}{\sigma_N^4(m)} (d_m^{\text{min}} - d_m^{\text{model}})^4 \quad \text{I.39}$$

$d^{\text{min}}$  depends on the atoms in contact and whether they are single torsion separated pairs of atoms or multiple torsion separated or in a special case hydrogen-bonded atom pairs.  $\sigma_N$  varies with atom types but in practice the fitted values are relatively uniform.

Torsion angles are among the least restricted stereochemical features, although certain restrictions, related to non-bonded contacts, apply.

$$\phi_6 = \sum_t^{\text{conformational torsion angles}} \frac{1}{\sigma_T^2(t)} (\chi_t^{\text{ideal}} - \chi_t^{\text{model}})^2 \quad \text{I.40}$$

where  $\chi^{\text{ideal}}$  is the angle of the torsional potential minimum to which  $\chi^{\text{model}}$  is nearest.  $\sigma_T$  is the standard deviation of the expected distribution.

Thermal parameters within biological macromolecules reflect the high rms displacements from mean atomic positions. Deviations from these positions are normally in the order of tenths of an angstrom whilst for covalent bonded distances and variations associated with bond angles rms deviations are limited to hundredths of an angstrom. The relationship of displacement of one atom and connected atoms must be maintained by the stereochemistry. NUCLSQ imposes the stereochemistry by restraining the variances of time-dependant interatomic distance distribution to suitably small values.

The variance for isotropic thermal motion is the sum of the magnitude

of the differences in the mean square displacements  $\overline{u^2}$  of two atoms joined by a bonding distance. Since these displacements are directly related to the thermal parameters by  $B=8\pi\overline{u^2}$  the appropriate observation function can be written:

$$\phi_7 = \sum_j^{\text{distances}} \frac{1}{\sigma_B^2(j)} (B_j^{\text{origin}} - B_j^{\text{target}})^2 \quad \text{I.41}$$

where  $B^{\text{origin}}$  and  $B^{\text{target}}$  denote the thermal parameters of the atom pair related by a certain bonding distance.

Restraints for anisotropic thermal parameters are separately imposed along selected, orthogonal directions rather than in the aggregate as in the isotropic case. The observational function is as follows:

$$\phi_8 = \frac{1}{\sigma_U^2(j)} \left[ \Delta_U^2 \cos^2 \theta + \frac{\Delta_U^4 (\sin^4 \theta - 3 \cos^2 \theta \sin^2 \theta)}{d_j^2} \right]^2 \quad \text{I.42}$$

where

$$\Delta_U^2 = \pm \left( \overline{U_{U,j}^2}^{\text{origin}} - \overline{U_{U,j}^2}^{\text{target}} \right) \quad \text{I.43}$$

The angle between a given bond direction and specific direction of motion  $\nu$  is  $\theta(\nu,j)$  and the mean square displacement along the direction specified by  $\nu$  is  $\overline{\theta_\nu^2}$ . Standard deviations for anisotropic temperature factors relate to mean square displacements rather than to B values as is the case for isotropic thermal motion.

Restraints can also be utilised when dealing with non-crystallographic symmetry, relating near identical copies of the macromolecule within the asymmetric unit. The restraints imposed are a compromise between imposing copies of the molecule to be identical, whilst allowing for real deviations in structure. Both positional and temperature factor restraints

utilising the similarity in atomic positions and temperature factors of the related molecule can be expressed:

$$\phi_9 = \sum_i^{\text{unique atoms}} \frac{1}{\sigma_{\text{SP}}^2(i)} \sum_k^{\text{copies}} (\mathbf{r}'_{i,k} - \overline{\mathbf{r}'_{i,k}})^2 \quad \text{I.44}$$

where  $\mathbf{r}'_{i,k} = \mathbf{R}_k \mathbf{r}_{i,k} + \mathbf{t}_k \quad \text{I.45}$

and  $\mathbf{R}$  is a rotation matrix and  $\mathbf{t}$  a translation vector that when applied to each object  $k$  brings them to a common coincidence.

For the isotropic case of non-crystallographic symmetry restraints:

$$\phi_{10} = \sum_i^{\text{unique atoms}} \frac{1}{\sigma_{\text{SB}}^2(i)} \sum_k^{\text{copies}} (B_{i,k} - \overline{B_{i,k}})^2 \quad \text{I.46}$$

Another form of restraint that can be imposed is on excessive shifts of atoms or groups of atoms. If a region is poorly determined by the crystallographic data instability may occur during minimisation. The application of permissible breadths of distribution  $\sigma_{\text{EP}}$  and  $\sigma_{\text{EB}}$  from current values  $\mathbf{r}_i^0$  and  $\mathbf{B}_i^0$  for positional and temperature factors respectively.

$$\phi_{11} = \sum_i^{\text{atoms}} \frac{1}{\sigma_{\text{EP}}^2} (\mathbf{r}_i - \mathbf{r}_i^0)^2 + \frac{1}{\sigma_{\text{EB}}^2} (B_i - B_i^0)^2 \quad \text{I.47}$$

The final form of restraint that can be imposed is that on occupancy factors. It may be that a group of atoms which have variable occupancy factors can all be expected to have the same occupancy, such as may arise for a non-covalently bound ligand.

The observational function is of the form:

$$\phi_{12} = \sum_k^{\text{groups}} \frac{1}{\sigma_Q^2(k)} \sum_i^{\text{atoms in groups}} (Q_{i,k} - \overline{Q_{i,k}})^2 \quad \text{I.48}$$

where  $Q_{i,k}$  is the occupancy factor of atom  $i$  in group  $k$  and  $\sigma_Q$  is the expected standard deviation for the distribution.

## **APPENDIX II.**

**Nuclear magnetic resonance and X-PLOR refinement.**

The main source of geometrical information employed in the determination of three-dimensional structures of macromolecules by NMR spectroscopy is derived from nuclear Overhauser enhancement (NOE) measurements, that enable one to deduce the identification and separation of protons in space.

Considering the simplest system of two protons, *i* and *j*, each possessing magnetisation, separated by a distance  $r_{ij}$ , such that magnetisation can be exchanged by the process known as cross-relaxation (provided  $r_{ij} < 5\text{\AA}$ ). In the equilibrium state the cross-relaxation rates in both directions are equal, so the magnetisations associated with each of the protons at equilibrium are also equal.

The cross-relaxation rate is proportional to two variables:  $r^{-6}$ , where *r* is the distance vector between the two protons, and  $\tau_{\text{eff}}$ , the effective correlation time of the interproton vector.

If the magnetisation of one of the spins is perturbed, the magnetisation of the second will change. The change in magnetisation of proton *i* upon perturbation of the magnetisation of proton *j* is known as the nuclear Overhauser effect (NOE). The perturbation is induced by selective irradiation of one of the resonances by a radiofrequency pulse. If proton *j* is close to proton *i* in space, its magnetisation will be reduced. The initial build-up rate of the NOE is equal to the cross-relaxation rate and hence proportional to  $r^{-6}$ .

$$N_{ij}(t) \sim \sigma_{ij} t \quad \text{II.1}$$

where *N* is the NOE, *t* is the irradiation period and

$$\sigma_{ij} = \frac{\gamma^4 (\hbar/2\pi)^2}{10r_{ij}^6} \left( \tau_{\text{app}} - \frac{6\tau_{\text{app}}}{1 + 4\omega^2\tau_{\text{app}}^2} \right) \quad \text{II.2}$$

where  $r_{ij}$  is the distance between the two protons;  $\sigma$  is the cross-relaxation rate and  $\tau_{\text{app}}$  is the correlation time (the tumbling time for the protons). The

Larmor precession frequency  $\omega$  (Hz) is the rate at which the bulk magnetisation precesses about the applied magnetic field  $B_0$ .

$$\omega = \gamma B_0 / 2\pi \quad \text{II.3}$$

where  $\gamma$  is gyromagnetic ratio for the proton.

In the experiments outlined in chapter 4 the two-dimensional (2D) pulse sequences used were NOESY (nuclear Overhauser spectroscopy), and TOCSY (total correlation spectroscopy). The 2D experiment has the obvious advantage that peak overlap is reduced by spreading the spectrum in two dimensions. All 2D experiments consist of a preparation time, an evolution period, a mixing time and a detection time, as outlined in figure II.1.

The NOESY pulse sequence is shown in figure II.2. In the preparation period the magnetisation is allowed to reach equilibrium before the application of the first  $90^\circ$  pulse. After the application of the first  $90^\circ$  pulse the spins are in the  $xy$  plane and are labelled by their appropriate chemical shifts during the evolution period  $t_1$ . The application of the second  $90^\circ$  pulse produces transfer of magnetisation between spins during the mixing time  $\tau$ . This transfer of magnetisation is a through space dipolar transfer (in contrast to a through bond scalar transfer). The third  $90^\circ$  pulse is the detection pulse, after which the free induction decay is acquired for a time  $t_2$ . The 2D spectrum is obtained by Fourier transformation in the  $t_1$  and  $t_2$  time domains (figure II.1).

The TOCSY pulse sequence is outlined in figure II.2, and consists of an initial  $90^\circ$  pulse followed by an evolution time  $t_1$ . The subsequent spin lock pulse effectively reduces the line broadening often associated with signals from macromolecules. The spin lock pulse is a long irradiation time (ms) compared to the short ( $\mu\text{s}$ ) pulses used for the  $90^\circ$  pulses.

The standard mixing time is replaced by the MLEV17 pulse sequence, which is a 50 ms series of pulses to ensure isotropic mixing to enhance the resolution and sensitivity.

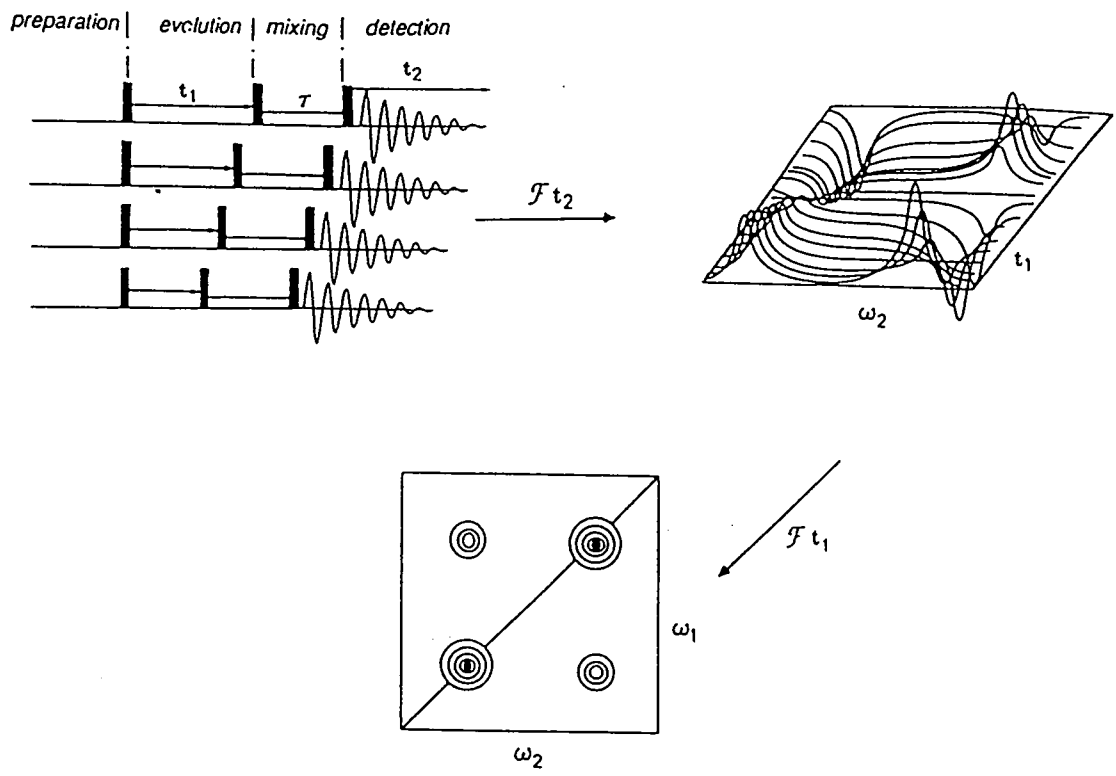


Figure II.1. A generalised representation of the 2D-NMR experiment.

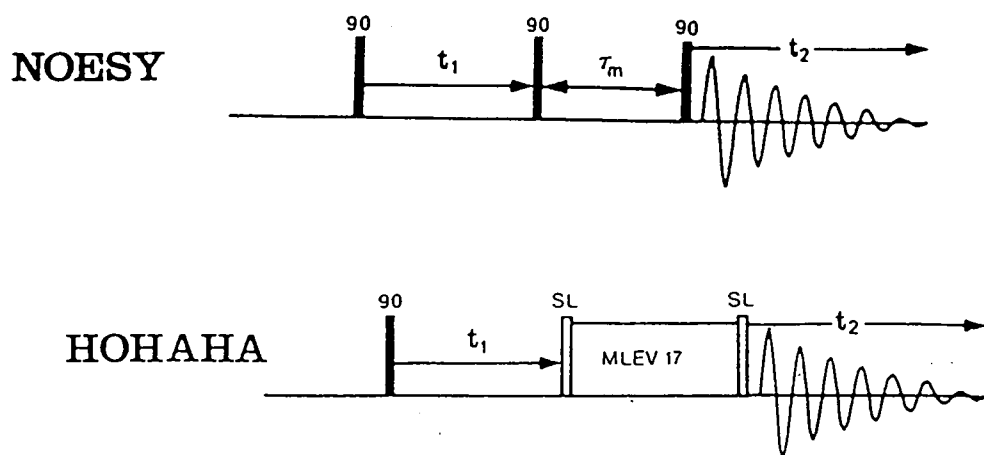


Figure II.2 Pulse schemes for NOESY and TOCSY experiments.

The energy function used in this thesis with X-PLOR takes the form of:

$$E_{\text{TOTAL}} = E_{\text{EMPIRICAL}} + E_{\text{EFFECTIVE}} \quad \text{Eq. II.4}$$

The standard unrestrained molecular mechanics calculations consider only  $E_{\text{EMPIRICAL}}$ . The  $E_{\text{EFFECTIVE}}$  term comprises of the restraining energy terms derived from experimental observations, such as NOE interproton distance restraints or dihedral angles, and is thus only considered in the case of restrained molecular mechanics and dynamics.

The empirical energy function which is minimised comprises of the terms described:

$$E_{\text{EMPIRICAL}} = \sum_{p=1}^N [w_{\text{BOND}}^p E_{\text{BOND}} + w_{\text{ANGL}}^p E_{\text{ANGL}} + w_{\text{DIHE}}^p E_{\text{DIHE}} + w_{\text{IMPR}}^p E_{\text{IMPR}} + w_{\text{HBON}}^p E_{\text{HBON}} + w_{\text{VDW}}^p E_{\text{VDW}} + w_{\text{ELEC}}^p E_{\text{ELEC}}] \quad \text{Eq. II.5}$$

where each energy term has a corresponding weighting element  $w_N^p$ .

The first four terms in Eq. II.5 are conformational energy terms.

The first term describes the covalent bond energy,

$$E_{\text{BOND}} = \sum_{\text{bonds}} k_b (r - r_0)^2 \quad \text{Eq. II.6}$$

where  $r$  = actual bond length,  $r_0$  = equilibrium bond length and  $k_b$  = stretching force constant.

The second term describes the bond angle energy,

$$E_{\text{ANGL}} = \sum_{\text{angles}} k_a (\theta - \theta_0)^2 \quad \text{Eq. II.7}$$

where  $\theta$  = actual bond angle,  $\theta_0$  = equilibrium bond angle and  $k_a$  = bending force constant.

The third and fourth terms describe the energy involving dihedral angles and planarity, respectively

$$E_{\text{IMPR}} + E_{\text{DIHE}} = \sum_{\text{dihedrals,impropers}} \begin{cases} k_d(1 + \cos(n\phi + \delta))^2 & \text{if } n > 0 \\ k_d(\phi - \delta)^2 & \text{if } n=0 \end{cases} \quad \text{Eq. II.8}$$

where  $n$  is the number of rotational minima whose maximum value is  $k_d$  and minimum value at  $\phi$  is 0, and  $\delta$  is the angular deviation from this minimum.

The three remaining terms of Eq. II.5 describe non-bonded interactions. The explicit hydrogen bond energy term is used particularly for nucleic acids and takes the rather complex form:

$$E_{\text{HBON}} = \sum_{i < j} \left( \frac{A}{r_{\text{AD}}^i} - \frac{B}{r_{\text{AD}}^j} \right) \cos^m(\theta_{\text{A-H-D}}) \quad \text{Eq. II.9}$$

$$\begin{aligned} & * \cos^n(\theta_{\text{AA-A-H}}) \text{SW}(r_{\text{AD}}^2, r_{\text{hon}}^2, r_{\text{hoff}}^2) \\ & * \text{SW}[\cos^2(\theta_{\text{A-H-D}}), \cos^2(\theta_{\text{hon}}), \cos^2(\theta_{\text{hoff}})] \end{aligned}$$

where AA, A, D and H are acceptor antecedent, acceptor and donor heavy atoms and hydrogen atom respectively.  $\theta_{\text{A-H-D}}$  is the angular cutoff (acceptor-hydrogen-donor);  $0^\circ$  corresponds to a linear hydrogen bond. The terms  $i, j, m$  and  $n$  are positive integers where  $i$  is the repulsive exponential (default=6),  $j$  is the attractive exponential (default=4),  $m$  is the (A-H-D) angle exponential (default=4) and  $n$  is the (AA-A-H) angle exponential (default=2). The switching function, SW, maintains a continuous smooth function to the cutoff point, and is employed so as to avoid the ill-conditioning of a sudden cutoff on the minimisation; this function takes the form:

$$\text{SW}(R, R_{\text{on}}, R_{\text{off}}) = \begin{cases} 0 & \text{if } R > R_{\text{off}} \\ \frac{(R^2 - R_{\text{off}}^2) * (R^2 - R_{\text{off}}^2 - 3(R^2 - R_{\text{on}}^2))}{R_{\text{off}}^2 - R_{\text{on}}^2} & \text{if } R_{\text{off}} > R > R_{\text{on}} \\ 1 & \text{if } R < R_{\text{on}} \end{cases}$$

Eq. II.10

where the angles  $\theta_{\text{hon}}$  and  $\theta_{\text{hoff}}$  (Eq. II.6) are switching function parameters for hydrogen bond-angle cutoffs. The distances  $r_{\text{on}}$  and  $r_{\text{off}}$  are the switching function parameters for hydrogen bond distance cutoffs.

For van der Waals and electrostatic calculations there are four possible options. The first is truncation which uses the standard Lennard-Jones and Coulomb functions. The second involves a switched van der Waals and a shifted electrostatic function. The third, which was used in this thesis, uses a switched van der Waals with a  $1/R$  dielectric electrostatic function. The fourth uses only a purely repulsive function.

The van der Waals term is given by:

$$f_{VDW}(R) = \begin{cases} \frac{A}{R^{12}} - \frac{B}{R^6} = 4\epsilon \left( \left( \frac{\sigma}{R} \right)^{12} - \left( \frac{\sigma}{R} \right)^6 \right) \text{heavy}(R - R_{\text{cut}}) & \text{truncation} \\ \left( \frac{A}{R^{12}} - \frac{B}{R^6} \right) \text{SW}(R, R_{\text{on}}, R_{\text{off}}) & \text{switched} \\ C_{\text{rep}}(\max(0, (k^{\text{rep}} R_{\text{min}})^{\text{irexp}} - R^{\text{irexp}}))^{\text{rexp}} & \text{repel} \end{cases} \quad \text{Eq. II.11}$$

In the first two cases the van der Waals energy is approximated by the Lennard-Jones potential energy function, with the attractive force proportional to  $R^{-6}$  and the repulsive force varying as  $R^{-12}$ . The A and B terms are the van der Waals parameters for the atomic dispersion and repulsion and  $r$  is their separation. The distance  $r_{\text{on}}$  is that at which the switching function becomes effective, whereas  $r_{\text{off}}$  is the distance at which the switching function forces the energy to zero and  $r_{\text{cut}}$  is the non-bonded interaction cutoff. The A, B,  $\epsilon$  and  $\sigma$  parameters are related to the well depth  $E_{\text{min}}$  and minimum distance  $R_{\text{min}}$  by

$$R_{\text{min}} = \sigma \sqrt[6]{2} \quad \text{Eq. II.12}$$

$$E_{\text{min}} = -\epsilon \quad \text{Eq. II.13}$$

$$A = 4\sigma^{12}\epsilon \quad \text{Eq. II.14}$$

$$B = 4\sigma^6\epsilon \quad \text{Eq. II.15}$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad \text{Eq. II.16}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad \text{Eq. II.17}$$

The electrostatic term is driven by:

$$f_{\text{ELEC}}(R) = \begin{cases} Q_i Q_j \frac{C}{\epsilon_0 R} \text{heavy}(R - R_{\text{cut}}) & \text{for pure truncation} \\ Q_i Q_j \frac{C}{\epsilon_0 R} \left(1 - \frac{R^2}{R_{\text{off}}^2}\right)^2 & \text{for shifted option} \\ Q_i Q_j \frac{C}{\epsilon_0 R^2} \text{SW}(R, R_{\text{on}}, R_{\text{off}}) & \text{for } \frac{1}{R} \text{ option} \\ 0 & \text{for repel option} \end{cases}$$

Eq. II.18

where  $\epsilon_0$  is the dielectric constant and  $R_{\text{cut}}$ ,  $R_{\text{on}}$  and  $R_{\text{off}}$  are as in equation II.7, and  $Q_i$  and  $Q_j$  are the atomic charges of the atoms  $i$  and  $j$ , respectively. As with the van der Waals term there are several schemes for truncating the electrostatic potential. In this thesis the  $1/R$  option was used together with a damping factor of 0.4 for the 1-4 electrostatic terms.

The intramolecular interaction energy is the summation of the individual non-bonded interaction energies.

$$E_{\text{ELEC}} = \sum_{i < j} f_{\text{ELEC}}(R_{ij}) + e_{14} \sum_{(ij) \in \{1-4\}} f_{\text{ELEC}}(R_{ij}) \quad \text{Eq. II.19}$$

$$E_{\text{VDW}} = \sum_{i < j} f_{\text{VDW}}(R_{ij}) + \sum_{(ij) \in \{1-4\}} f_{\text{VDW}}(R_{ij}) \quad \text{Eq. II.20}$$

The choice of dielectric constant must account for the screening effect of counterions and solvent molecules that are not explicitly included in the model. A value of 4.0 is typically used for gas phase minimisations of biological macromolecules.

### Conjugate Gradient Minimisation.

Molecular mechanics minimisation within X-PLOR is by the conjugate gradient method, using the Powell algorithm.

A diagrammatical representation of a two-dimensional conjugate gradient method search is depicted below (figure II.3).

At the starting point A the gradient is  $g_1$ , the first search direction  $s_1$  is taken along the negative gradient ( i.e.  $s_1 = -g_1$ ) along the line ABD.

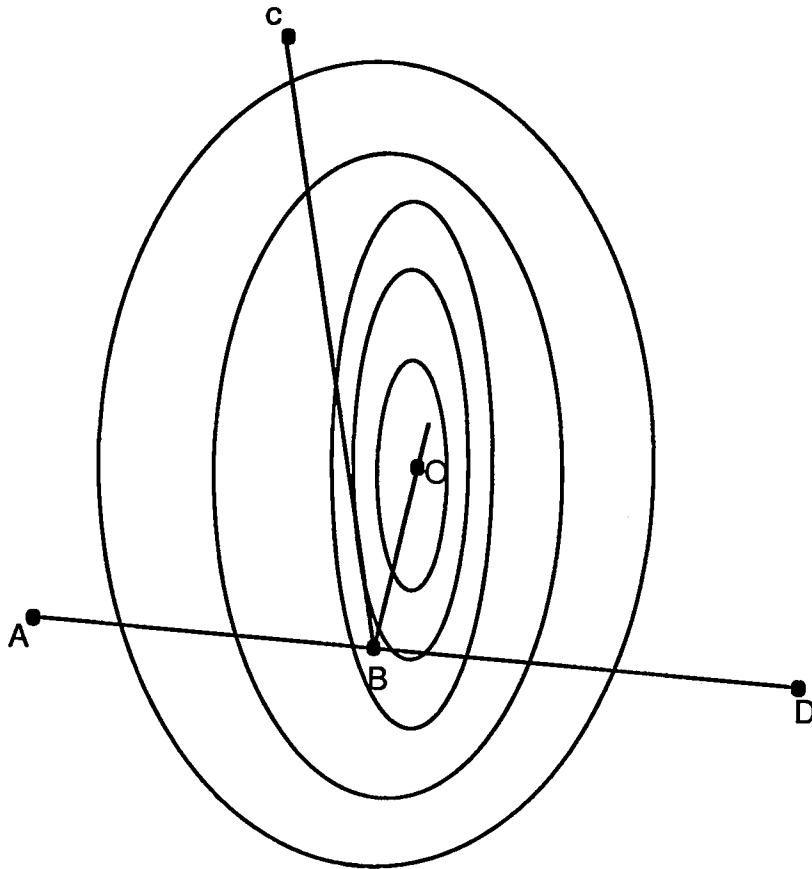


Figure II.3 A simple two-dimensional minimisation.

The minimum energy along line ABD lies at point B and the new negative gradient lies along the line BC. Rather than using the line of steepest descent (along BC), the conjugate gradient method uses a linear combination of the current gradient  $-g_2$  and the previous search direction  $s_1$  for the next search direction.

$$s_k = -g_k + b_k s_{k-1} \quad \text{Eq. II.21}$$

$$b_k = \frac{|g_k|^2}{|g_{k-1}|^2} \quad \text{Eq. II.22}$$

where

It can be shown that for an  $n$ -dimensional, quadratic surface, the search

direction will pass through the minimum on the  $n^{\text{th}}$  step provided that the minimum along each successive search is found. For the simple two-dimensional system in figure II.1 it can be seen that the search will pass through the minimum at O on the second step. The step size at each successive step is increased or decreased in tandem with the increase or decrease of the energy of the system.

In general for molecular mechanics the potential surface is not quadratic and numerical errors can accumulate in the determination of new search directions. The accumulation of errors is due to the dependency on previous search directions rather than simply the previous gradients. This problem is overcome by interrupting the search algorithm every  $m$  steps and re-orienting the search direction based on the current gradient alone (i.e.  $b_k=0$ ).

For a non-quadratic potential surface, even with a less than optimal step size, the conjugate gradient method will generally yield a superior search direction than the steepest descent method, regardless of the dimensionality of the surface. Although the conjugate gradient method improves the search method for the global minimum, it is impossible to overcome the so called 'multiple minima problem'.

### Molecular Dynamics.

Molecular dynamics consists of solving Newton's equations of motion.

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\nabla_{x_i} E_{\text{TOTAL}} \quad \text{Eq. II.23}$$

where the index  $i$  is for all free atoms and the gradient  $\nabla_{x_i} E_{\text{TOTAL}}$  is derived from the X-PLOR energy function (Eq. II.1). The atomic mass and position of each atom are  $m_i$  and  $x_i$  respectively.

The initial velocities for each atom within the system are assigned by a Maxwellian distribution which takes the form

$$\left(\frac{m_i}{2\pi k_B T_I}\right)^{3/2} \exp\left(\frac{-m_i \delta^2}{2k_B T_I}\right) \quad \text{Eq. II.24}$$

where  $\delta$  is a random number between 0 and 1, and  $T_I$  is the initial temperature specified by the user.

Temperature control of the system during these studies was maintained by coupling to a temperature bath. During the employment of this methodology coordinates are restrained by use of a friction coefficient which is derived from the equation

$$b_i = b_i^I \left(\frac{T_0}{T - 1}\right) \quad \text{Eq. II.25}$$

where  $b_i^I$  is friction coefficient property of the atom  $i$ .  $T_0$  is the target temperature of the heatbath and  $T$  is the temperature of the system at the particular step

$$T = \sum_i m_i \left(\frac{dx_i}{dt}\right)^2 / (n_{\text{deg}} k_B) \quad \text{Eq. II.26}$$

where  $n_{\text{deg}}$  is the number of degrees of freedom of the system.

The  $E_{\text{EFFECTIVE}}$  term (Eq. II.1) is comprised of additional restraining energy terms which are derived from experimental observations.

Additional restraints applied to the molecular dynamics in the case of the NMR refinement described in this thesis are in the form of a square-well function

$$E_{\text{NOE}} = \min(\text{ceil}, \text{SC}) \Delta^{\text{exp}} \quad \text{Eq. II.27}$$

where  $S$  (incremented during the heating stage) is the scale factor applied to

the restraints, and C (1) is the additional scale factor for the square-well function. The ceil (100) term is the ceiling value for the force constants and exp (2) is the exponent term for the square-well function. The figures in brackets are the values used for this study. The deviation  $\Delta$  is defined as

$$\Delta = \begin{cases} R - (d + d_{\text{plus}} - d_{\text{off}}) & d + d_{\text{plus}} - d_{\text{off}} < R \\ 0 & d - d_{\text{minus}} + d_{\text{off}} < R < d + d_{\text{plus}} - d_{\text{off}} \\ (d - d_{\text{minus}} + d_{\text{off}}) - R & R < d - d_{\text{minus}} + d_{\text{off}} \end{cases} \quad \text{Eq.II.28}$$

The values  $d$ ,  $d_{\text{plus}}$ ,  $d_{\text{minus}}$  are the distances with plus and minus errors of the NOE interproton distances deduced from the NMR experiments and  $d_{\text{off}}$  is the square-well function offset (0).

The functional form of the effective energy term  $E_{\text{CDIH}}$  is given by

$$E_{\text{CDIH}} = S \sum C \text{well}(2\pi(\phi - \phi_0), \Delta\phi)^{\text{ed}} \quad \text{Eq.II.29}$$

where S is an overall weighting factor. The constant C, the angle range  $\Delta\phi$ , the angle centroid  $\phi_0$  and the exponent (ed) for each dihedral are specified by the user in the input parameter file. The sum extends over all the restrained dihedral angles and the square-well potential,  $\text{well}(a,b)$

$$\text{well}(a,b) = \begin{cases} a - b & \text{if } a < b \\ 0 & \text{if } -b < a < b \\ a + b & \text{if } a < -b \end{cases} \quad \text{Eq.II.30}$$

## **REFERENCES.**

Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C. (1988) *Science* **242** p899-907

Recognition of a DNA operator by the *cro* repressor of phage 434: a view at high resolution.

Anderson, J.E., Ptashane, M. and Harrison S.C. (1987) *Nature* **326** p846-852

Structure of the repressor-operator complex of bacteriophage 434.

Anderson, W.F., Ohlandorf, D.H., Takeda, Y. and Matthews B.W.

(1981) *Nature* **290** p754-758

Structure of the *cro* repressor from bacteriophage  $\lambda$  and its interaction with DNA.

Arents, G., Burlingame, R.W., Wang, B.-C., Love, W.E.

and Moudrianakis, E.M. (1991) *Proc. Natl. Acad. Sci. USA* **88** p10148-10152

The nucleosomal core histone octamer at 3.1Å resolution: A tripartite protein assembly and a left-handed superhelix.

Arnott, S., Chandrasekaran, R., Hall, I. and Puigjaner, L.C.

(1983) *Nucleic Acids Research* **11** p4141-4155

Heteronomous DNA.

Aymami, J., Coll, M., Frederick, C.A., Wang, A.H.-J. and Rich, A.

(1989) *Nucleic Acids Research* p3229-3245

The propeller DNA conformation of poly(dA)·poly(dT).

Baguley, B.C (1982) *Mol. Cell. Biochem.* **43** p167-181

Nonintercalative DNA-binding antitumour compounds.

Benne, R. (1989) *Biochim. Biophys. Acta.* **1007** p131-139

RNA editing in trypanosome mitochondria.

Bernard, J. and Riou, G.F. (1980) *Biochemistry* **19** p4197-4201

In vivo effects of intercalating and nonintercalating drugs on the tertiary structure of kinetoplast deoxyribonucleic acid.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D.,

Rodgers, J.R., Kennard, O., Shimanouchi, T and Tasumi, M.

(1977) *Eur. J. Biochem.* **80** p319-324

The protein data bank: a computer based archival file for macromolecular structures.

Beveridge, A. (1990) *GEMINI An in-house molecular modelling package* .

Bitonti, A.J., Dumont, J.A. and McCann, P.P. (1986) *Biochem. J.* **237** p685-689

Characterisation of *Trypanosoma brucei brucei* S-adenosyl-L-methionine decarboxylase and its inhibition by berenil, pentamidine and methylglyoxal bis(guanylhydrazone).

Blum, B., Bakalara, N. and Simpson, L. (1990) *Cell* **60** p189-198

A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information.

Boehncke, K., Nonella, M., Schulten, K. and Wang, A.H.-J.

(1991) *Biochemistry* **30** p5465-5475

Molecular dynamics investigation of the interaction between DNA and distamycin.

Bontemps, J., Houssier, C., and Fredericq, E. (1975) *Nucleic Acids Research* **2** p975-984

Physico-chemical study of the complexes of 'Hoechst 33258' with DNA and nucleohistone.

Braithwaite, A.W. and Baguley, B.C. (1980) *Biochemistry* **19** p1101-1106

Existence of an extended series of antitumour compounds which bind to deoxyribonucleic acid by nonintercalative means.

Brennan, R.G., Roderick, S.L., Takeda, Y. and Matthews, B.W.

(1990) *Proc. Natl. Acad. Sci. USA* **87** p8165-8169

Protein-DNA conformational changes in the crystal structure of a  $\lambda$  cro-operator complex.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) *J. Comp. Chem.* **4** p187-217

CHARMM: A program for macromolecular energy, minimisation, and

dynamic calculations.

Brown, D.G., Sanderson, M.R., Skelly, J.V., Jenkins, T.C., Brown, T., Garman, E., Stuart, D.I. and Neidle, S.

(1990) *EMBO J.* **9** p1329-1334

The crystal structure of a berenil-dodecanucleotide complex: the role of water in sequence-specific ligand binding.

Burkhoff, A.M. and Tullius, T.D. (1987) *Cell* **48** p935-943

The unusual conformation adopted by the adenine tracts in kinetoplast DNA.

Burlingame, R.W., Love, W.E., Wang, B.C. Hamlin, R., Xuong, N.H., and Moudrianakis, E.N. (1985) *Science* **228** p546-553

Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3Å.

Calladine, C.R. (1982) *J. Mol. Biol.* **161** p343-352

Mechanism of sequence-dependant stacking of bases in B-DNA.

Cambillau, C.M. (1988) *TOM, a version of FRODO amended to run on a Silicon Graphics IRIS* .

Carrondo, M., Coll, M., Aymami, J., Wang, A.H.-J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1989) *Biochemistry* **28** p7849-7859

Binding of a Hoechst dye to d(CGCGATATCGCG) and its influence on the conformation of a DNA fragment.

Chandra, P., Steel, L.K., Ebner, U., Woltersdorf, M., Laube, H., Kornhuber, B., Mildner, B. and Götz, A. (1977) *Pharmac. Ther. A.* **1** p231-287

Chemical inhibitors of oncoviral DNA polymerases: biological implications and their mode of action.

Chen, G.L., Yang, L., Rowe, T.C., Halligan, B.D., Tewey, K.M. and Liu, L.F. (1984) *J. Biol. Chem.* **259** p13560-13566

Nonintercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian topoisomerase II.

Clarke, N.D., Beamer, L.J., Goldberg, H.R., Berkower, C. and Pabo, C.O.

(1991) *Science* **254** p267-270

The DNA binding arm of  $\lambda$  repressor : Critical contacts from a flexible region.  
N terminal specific contacts in the major groove

Clore, G.M. and Gronenborn, A.M. (1989) *CRC Critical Reviews in Biochemistry and Molecular Biology* **24** p479-564

Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy.

Coll, M., Aymami, J., van der Marel, G.A., van Boom, J.H., Rich, A. and Wang, A.H.-J. (1989) *Biochemistry* **28** p310-320

Molecular structure of the netropsin-d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment.

Coll, M., Frederick, C.A., Wang, A. H.-J. and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA* **84** p8385-8389

A bifurcated hydrogen-bonded conformation in the d(A•T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin.

Comings, D.E. (1975) *Chromosoma* **52** p229-243

Mechanism of chromosome banding. VIII. Hoechst 33258-DNA interaction.

Crothers, D.M., Haran, T.E. and Nadeau, J.G. (1990) *J. Biol. Chem.* **265** p7093-7096

Intrinsically bent DNA.

De Clercq, E. and Dann, O. (1980) *J. Med. Chem.* **23** p787-795

Diaryl amidine derivatives as oncornaviral DNA polymerase inhibitors.

Denny, W.A. (1989) *Anti-Cancer Drug Design* **4** p241-263

DNA-intercalating ligands as anti-cancer drugs: prospects for future design.

Dervan, P.B. (1986) *Science* **232** p461-471

Design of sequence-specific DNA-binding molecules

Dickerson, R.E. (1990) NEWHEL90 available from bitnet address RED@UCLAUE.

Dickerson, R.E. and Drew, H.R. (1981) *Proc. Natl. Acad. Sci. USA* **78**

p7318-7322

Kinematic model for B-DNA.

Dickerson, R.E., Goodsell, D.S., Kopka, M.L. and Pjura, P.E. (1987)

*J. Biomol. Str. Dynam.* **5** p557-579

The effect of crystal packing on oligonucleotide double helix structure.

Dickerson, R.E., Kopka, M.L. and Pjura, P. (1983) *Biochemistry* **80** p7099-7103

A random-walk model for helix bending in B-DNA.

Drew, H.R. and Dickerson, R.E. (1981) *J. Mol. Biol.* **151** p535-556

Structure of a B-DNA dodecamer. III. Geometry of hydration.

Drew, H.R., Wing, R.M., Tanako, T., Broka, C., Tanaka, S., Itakura, K., and

Dickerson, R.E. (1981) *Proc. Natl. Acad. Sci. USA* **78** p2179-2183

Structure of a B-DNA dodecamer: Conformation and dynamics.

EMBO Workshop on DNA Curvature and Bending (1989) *EMBO J.* **8** p1-4

Definition and nomenclature of nucleic acid structure parameters.

Embrey, K.J., Searle, M.S, and Craik, D.J. (1991) *J. Chem. Soc. Chem.*

*Commun.* p1770-1771

Probing the interaction of Hoechst 33258 with an A-T rich oligonucleotide duplex using <sup>1</sup>H NMR spectroscopy.

Fairfield, F.R., Bauer, W.R. and Simpson, M.V. (1985)

*Biochemica et Biophysica Acta* **824** p45-47

Studies on mitochondrial type I topoisomerase and on its function.

Fede, A., Labhardt, A., Bannwarth, W. and Leupin, W. (1991) *Biochemistry* **30**

p11377-11388

Dynamics and binding mode of Hoechst 33258 to d(GTGGAATTCCAC)<sub>2</sub> in the 1:1 solution complex as determined by two-dimensional <sup>1</sup>H NMR.

Feigon, J., Denny, W.A., Leupin, W. and Kearns, D.R. (1984) *J. Med. Chem.* **23**

p450-455

Interactions of antitumour drugs with natural DNA: <sup>1</sup>H NMR study of binding mode and kinetics.

- Felsenfield, G., Davies, D.R. and Rich, A. (1957) *J. Am. Chem. Soc.* **79**  
p2023-2027
- Fox, K.R., Sansom, C.E. and Stevens, M.F.G. (1990) *FEBS* **266** p150-154  
Footprinting studies on the sequence-selective binding of pentamidine to DNA.
- Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R.E. (1982) *J. Biol. Chem.* **257** p14686-14707  
Reversible bending and helix geometry in a B-DNA dodecamer:  
CGCGAATT<sup>Br</sup>CGCG.
- Frederick, C.A., Williams, L.D., Ughetto, G.A., van der Marel, G., van Boom, J.H. and Wang, A.H.-J. (1990) *Biochemistry* **28** p2538-2549  
Structural comparisons of anti-cancer drug-DNA complexes: Adriamycin and daunomycin.
- Freemont, P.S., Lane, A.L. and Sanderson, M.R. (1991) *Biochem. J.* **278** p1-23  
Structural aspects of protein-DNA recognition.
- Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. and Feeney, J. (1990) *J. Magn. Reson.* **90** p420-425  
HMQC-NOESY-HMQC, a three-dimensional NMR experiment which allows detection of nuclear overhauser effects between protons with overlapping signals.
- Fritsch, V. and Westhof, E. (1991) *J. Am. Chem. Soc.* **113** p8271-8277  
Three-center hydrogen bonds in DNA: Molecular dynamics of poly(dA)•poly(dT).
- Gago, F., Reynolds, C.A. and Richards, W.G. (1989) *Molecular Pharmacology* **35** p232-241  
The binding of nonintercalative drugs to alternating DNA sequences.
- Gasser, S.M. and Laemmli, U.K. (1986) *Cell* **46** p521-530  
Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*.

- Gasser, S.M. and Laemmli, U.K. (1987) *Trends Genet.* **3** p16-22  
A glimpse at chromosomal order.
- Gasser, S.M. Laroche, T., Falquet, J., Boy de la Tour, E. and Laemmli, U.K. (1986) *J. Mol. Biol.* **188** p613-629  
Metaphase chromosome structure. Involvement of topoisomerase II.
- Goodsell, D. and Dickerson, R.E. (1986) *J. Med. Chem.* **29** p727-733  
Isohelical analysis of DNA groove-binding drugs.
- Gorenstein, D.G., Schroeder, S.A., Fu, J.M., Metz, J.T., Roongta, V. and Jones, C.R. (1988) *Biochemistry* **27** p7223-7237  
Assignments of  $^{31}\text{P}$  NMR resonances in oligodeoxyribonucleotides: Origin of sequence-specific variations in the deoxyribose phosphate backbone conformation and the  $^{31}\text{P}$  chemical shifts of double-helical nucleic acids.
- Gresh, N., and Pullman, B. (1986) *Mol. Pharmacol.* **25** p452-458  
A theoretical study of the nonintercalative binding of berenil and stilbamidine to double-stranded (dA-dT)<sub>n</sub> oligomers.
- Griffin, L.C. and Dervan, P.B. (1989) *Science* **245** p967-971  
Recognition of thymine-adenine base pairs by guanine in a pyrimidine triple-helix motif.
- Griffith, J., Bleyman, M., Rauch, C.A., Kitchin, P.A. and Englund, P.T. (1986) *Cell* **46** p717-724  
Visualisation of the bent helix in kinetoplast DNA by electron microscopy.
- Gronenborn, A.M. and Clore, G.M. (1989) *Biochemistry* **28** p5978-5984  
Analysis of relative contributions to the Nuclear Overhauser interproton distance restraints and empirical energy function in the calculation of oligonucleotide structures using restrained molecular dynamics.
- Hare, D.R., Wemmer, D.E., Chou, S.-H., Dronby, G. and Reid, B.R. (1983) *J. Mol. Biol.* **171** p319-336  
Assignment of the non-exchangable proton resonances of

d(C-G-C-G-A-A-T-T-C-G-C-G) using two dimensional nuclear magnetic resonance methods.

Harrison, S.C. and Aggarwal, A.K. (1990) *Ann. Rev. Biochem.* **59** p933-969

DNA recognition by proteins with the helix-turn-helix motif.

Hunter W.N., Brown, T. and Kennard, O. (1986) *J. Biomol. Str. Dynam.* **4**  
p173-191

Structural features and hydration of d(CGCGAATTAGCG); a double helix containing two G-A mispairs.

Izaurralde, E., Käs, E. and Laemmli, U.K. (1989) *J. Mol. Biol.* **210** p573-585

High preferential nucleation of histone H1 assembly on scaffold associated regions.

Jeffrey, G.A. and Maluszynska, H. (1982) *Int. J. Biol. Macromol.* **4** p173-185

A survey of hydrogen bond geometries in the crystal structures of amino acids.

Jeffrey, G.A. and Mitra, J. (1984) *J. Am. Chem. Soc.* **106** p5546-5553

Three-center (bifurcated) hydrogen bonding in the crystal structures of amino acids.

Jordan, S.R. and Pabo, C.O. (1988) *Science* **242** p893-899

Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Korberg, T.B. and Pabo, C.O. (1990) *Cell* **63** p579-590

Crystal structure of engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions.

Klevit, R.E., Wemmer, D.E. and Reid, B.R. (1986) *Biochemistry* **25** p3296-3303

<sup>1</sup>H NMR studies of the interaction between Distamycin A and a symmetrical DNA dodecamer.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985a)

*J. Mol. Biol.* **183** p553-563

Binding of an antitumor drug to DNA. Netropsin and CGCGAATTBrCGCG.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985b) *Proc. Natl. Acad. Sci. USA* **82** p1376-1380

The molecular origin of DNA-drug specificity in netropsin and distamycin.

Kumar, S., Yadagiri, B., Zimmermann, J., Pon., R.T. and Lown, J.W. (1991) *J. Biomol. Str. Dynam.* **8** p331-357

Sequence specific molecular recognition and binding by a GC recognizing Hoechst 33258 analogue to the decadeoxyribonucleotide d(CATGGCCATG)<sub>2</sub>: Structural and dynamic aspects deduced from high field <sup>1</sup>H NMR studies.

Lämmler, G., Herzog, H., Saupe, E. and Schutze, H.R. (1971) *WHO Bull.* **44** p751-756

Chemotherapeutic studies on *Litomosoides carinii* infection of *Mastomys natalensis*. 1. The filaricidal action of 2,6-bis-benzimidazoles.

Lane, A.N. (1990) *Biochim. Biophys. Acta.* **1049** p189-204

The determination of the conformational properties of nucleic acids in solution from NMR data.

Lane, A.N., Jenkins, T.C., Brown, D.J.S., and Brown, T. (1991)

*Biochem. J.* **279** p269-281

NMR determination of the solution conformation and dynamics of the A•G mismatch in the d(CGCAAATTCGCG)<sub>2</sub> dodecamer.

Lane, A.N., Jenkins, T.C., Brown, T. and Neidle, S. (1991) *Biochemistry* **30** p1372-1385

Interaction of berenil with the Eco RI dodecamer d(CGCGAATTCGCG)<sub>2</sub> in solution studied by NMR.

Lane, M.J., Dabrowiak, J.C. and Vournakis, J.N. (1983)

*Proc. Natl. Acad. Sci. USA* **80** p3260-3264

Sequence specificity of actinomycin D and netropsin binding to pBR322 analysed by protection from DNase I.

Larsen, T.A., Goodsell, D.S., Cascio, D., Grzeskowiak, K. and Dickerson, R.E. (1989) *J. Biomol. Str. Dynam.* **7** p477-491

The structure of DAPI bound to DNA.

Latt, S.A. and Stetten, G. (1976) *J. Histochem. Cytochem.* **24** p24-33

Spectral studies on 33258 Hoechst and related bisbenzimidazole dyes useful for fluorescent detection of DNA synthesis.

Latt, S.A. and Wohlleb, J.C. (1975) *Chromosoma* **52** p297-316

Optical studies of the interaction of 33258 Hoechst with DNA, chromatin, and metaphase chromosomes.

Latt, S.A., Stetten, G., Juergens, L.A., Willard, H.F. and Scher, C.D. (1975) *J. Histochem. Cytochem.* **23** p493-505

Recent developments in the detection of DNA synthesis by 33258 Hoechst fluorescence

Laughton, C.A., Jenkins, T.C., Fox, K.R. and Neidle, S. (1990) *Nucleic Acids Research* **18**p4479-4488

Interaction of berenil with the tyrT sequence studied by footprinting and molecular modelling. Implications for design of sequence-specific DNA recognition agents.

Lavery, R. and Pullman, B. (1981) *Int. J. Quantum Chem.* **20** p259-272

Molecular electrostatic potential on the surface envelopes of macromolecules: B-DNA

Lefèvre, J.-F., Lane, A.N. and Jardetzky, O. (1987) *Biochemistry* **26** p5076-5090

Solution structure of the Trp operator of *Escherichia coli* determined by NMR.

Leupin, W., Chazin, W.J., Hyberts, S., Denny, W.A. and Wütrich, K. (1986) *Biochemistry* **25** p5902-5908

NMR studies of the complex between the decadeoxynucleotide d(GCATTAATGC) and a minor groove binding drug SN 6999.

Liaw, Y.-C., Gao, Y.-G., Robinson, H., van der Marel, G., van Boom, J.H., Rich, A. and Wang A.H.-J. (1989) *Biochemistry* **28** p9913-9918

Antitumour drug nogalamycin binds DNA in both grooves simultaneously: Molecular structure of nogalamycin-DNA complex.

Liu, L.F. (1984) *CRC Critical Reviews in Biochemistry* **15** p1-24

DNA topoisomerases: enzymes that catalise the breaking and rejoining of DNA.

Loontjens, F.G., Regenfuss, P., Zechel, A., Dumortier, L. and Clegg, R.M. (1990) *Biochemistry* **29** p9029-9039

Binding characteristics of Hoechst33258 with calf thymus DNA, poly[d(A-T)], and d(CCGGAATTCCGG): Multiple stoichiometries and determination of tight binding with a wide spectrum of site affinities.

Lowe, P.R., Sansom, C.E., Schwalbe, C.H. and Stevens, M.F.G. (1989)

*J. Chem. Soc. Chem. Commun.* **24** p1164-1165

Crystal structure and molecular modelling of the antimicrobial drug pentamidine.

Lown, J.W. (1988) *Anti-Cancer Drug Design* **3** p25-40

Lexitropsins: rational design of DNA sequence reading agents as novel anti-cancer agents and potential cellular probes.

Luck, G., Zimmer, CH., and Schweizer, D. (1988) *Studia Biophysica* **125** p107-119

DNA binding studies of the nonintercalative ligand pentamidine: dA·dT base pair preference.

Maher III, L.J., Wold, B. and Dervan, P.B. (1989) *Science* **238** p725-730

Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation.

Marini, J.C., Effron, P.N., Goodman, T.C., Singleton, C.K., Wells R.D., Wartell, R.M. and Englund, P.T. (1988) *J. Biol. Chem.* **259** p8974-8979

Physical characterisation of a kinetoplast DNA fragment with unusual properties.

Marini, J.C., Levens, S.D., Crothers, D.M. and Englund P.T. (1982)

*Proc. Natl. Acad. Sci. USA* **79** p7664-7668

Bend in helical structure in kinetoplast DNA

Marky, L.A., and Breslauer, K.J. (1987) *Proc. Natl. Acad. Sci. USA* **84**  
p4363-4359

Origins of netropsin binding affinity and specificity: Correlations of thermodynamic and structural data.

Martin, R. and Holmes, N. (1983) *Nature* **302** p452-454

Use of an  $^{125}\text{I}$ -labelled DNA ligand to probe DNA structure.

McClarín, J.A., Frederick, C.A., Wang, B.C., Greene, P., Boyer, H.W., Grable, J. and Rosenberg, J.M. (1986) *Science* **234** p1526-1541

Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å resolution.

McKay, D.B. and Steitz, T.A. (1981) *Nature* **290** p744-749

Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA

Mikhailov, M.V., Zasedatelev, A.S., Krylov, A.S., and Gursky, G.V. (1981) *Molec. Biol. SSSR.* **15** p541-554

Mechanism of the recognition of AT base pairs in DNA by molecules of the the dye Hoechst 33258.

Mildner, B. and Chandra, P. (1979) *Cell Molec. Biol.* **25** p399-407

Molecular mechanism of action of diaminephenylindole (DAPI). II. Effect of DAPI on the template activity of DNA and polydeoxynucleotides an the DNA-polymerase system for bacteria, eukaryotic cells and RNA tumour viruses.

Mondragón, A., Subbiah, S., Almo, S.C., Drottar, M. and Harrison, S.C. (1989a) *J. Mol. Biol.* **205** p189-200

Structure of the amino-terminal domain of phage 434 repressor at 2.0Å resolution.

Mondragón, A., Wolberger, C. and Harrison, S.C. (1989b) *J. Mol. Biol.* **205**  
p179-188

Structure of phage 434 cro protein at 2,35Å resolution.

- Moore, M.H., Hunter W.N., d'Estaintot, B.L. and Kennard, O. (1989)  
*J. Mol. Biol.* **206** p693-705  
 Drug-DNA interactions in the crystal structure of d(CGATCG) complexed with daunomycin.
- Moser, H.E. and Dervan, P.B. (1987) *Science* **238** p645-650  
 Sequence specific cleavage of double helical DNA by triple helix formation.
- Müller, W and Gautier, F. (1975) *Eur. J. Biochem.* **54** p385-394  
 Interactions of heterochromatic compounds with nucleic acids. A-T specific non-intercalating ligands.
- Nadeau, J.G. and Crothers, D.M. (1989) *Proc. Natl. Acad. Sci. USA* **86**  
 p2622-2626  
 Structural basis for DNA bending.
- Neidle, S., Brown, D.G., Jenkins, T.C., Laughton C.A., Sanderson, M.R. and Skelly, J.V. (1990) *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions* Kluwer Academic Publishers p43-57  
 Drug-DNA Recognition: Sequence-specificity of the DNA minor groove binder berenil.
- Nelson, E.M., Tewey, K.M. and Liu, L.F. (1984) *Proc. Natl. Acad. Sci. USA* **81**  
 p1361-  
 Mechanism of antitumour drug action: poisoning of mammalian DNA topoisomerase II on DNA by 4'-(9-acridinylamino)methanesulfon-m-anisidide.
- Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987) *Nature* **330**  
 p221-226  
 The structure of an oligo(dA)•oligo(dT) tract and its biological implications.
- Nerdal, W., Hare, D.R. and Reid, B.R. (1988) *J. Mol. Biol.* **201** p717-739  
 Three-dimensional structure of wild-type lac Pringbow promoter DNA in solution.
- Newton, B.A., (1975) *In Antibiotics III, mechanism of action of antimicrobial and antitumour agents.* Corcoran, I.W., and Hahn, F.G. (eds) Publ. Sringer-

*Verlag, Berlin.* 3 p34-47

Berenil: A trypanocide with selective activity against extranuclear DNA.

Ott, J. and Eckstein, F. (1985) *Biochemistry* 24 p2530-2535

<sup>31</sup>P NMR spectral analysis of the dodecamer d(CGCGAATTCGCG)

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J.* 9 p3085-3092

Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution.

Antennapedia homeodomain DNA complex

Pabo, C.O. and Lewis, M. (1982) *Nature* 298 p443-447

The operator-binding domain of  $\lambda$  repressor: structure and DNA recognition

Parkinson, J.A., Barber, J., Douglas, K.T., Rosamond, J. and Sharples, D. (1990) *Biochemistry* 29 p10181-10190

Minor groove recognition of the self-complementary duplex d(CGCGAATTCGCGC)<sub>2</sub> by Hoechst 33258: a high field NMR study.

Patel, D.J. (1978) *Eur. J. Biochem.* 99 p369-379

Netropsin · dG-dG-dA-dA-dT-dT-dC-dC complex. Antibiotic binding at Adenine-Thymine base pairs in the minor groove of the self complementary octanucleotide complex.

Patel, D.J. and Shapiro, L. (1985) *Biochimie* 67 p887-915

Molecular recognition in noncovalent antitumor agent-DNA complexes: NMR studies of the base and sequence dependent recognition of the DNA minor groove by netropsin.

Pearl, L.H., Skelly, J.V., Hudson, B.D. and Neidle, S. (1987)

*Nucleic Acids Research* 15 p3469-3477

The crystal structure of the DNA-binding drug berenil. Molecular modelling studies of berenil-DNA complexes.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.* 220 p457-479

Are time-averaged restraints necessary for nuclear magnetic resonance refinement?

Pelton, J.G. and Wemmer, D.E. (1988) *Biochemistry* **27** p8088-8096

Structural modeling of the Distamycin A-d(CGCGAATCGCG) complex using 2D NMR and molecular mechanics.

Pelton, J.G. and Wemmer, D.E. (1990) *J. Am. Chem. Soc.* **112** p1393-1399

Binding modes of distamycin A with d(CGCAAATTTGCG)<sub>2</sub> determined by two dimensional NMR.

Phillips, S.E.V. (1991) *Curr. Opinions Struct. Biol.* **1** p89-98

Specific  $\beta$ -sheet interactions.

Pjura, P.E., Grzeskowiak, G. and Dickerson, R.E. (1987) *J. Mol. Biol.* **197**

p257-271

Binding of Hoechst 33258 to the minor groove of B-DNA.

Portugal, J. and Waring, M.J. (1986) *Nucleic Acids Research* **14** p8735-8754

Antibiotics which can alter the rotational orientation of nucleosome core DNA.

Portugal, J. and Waring, M.J. (1987) *Biochemie* **69** p825-840

Analysis of the effects of antibiotics on the structure of nucleosome core particles determined by DNAase I cleavage.

Portugal, J. and Waring, M.J. (1987) *Eur. J. Biochem.* **167** p281-289

Comparison of binding sites in DNA for berenil, netropsin and distamycin. A footprinting study.

Postel, E.H., Flint, S.J., Kessler, D.J. and Hogan, M.E. (1991) *Proc. Natl. Acad. Sci. USA* **88** p8227-8231

Evidence that a triple helix-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels.

Powell, M.J.D. (1977) *Mathematical Programming* **12** p241-254

Quigley, G.J., Ughetto, G., van der Marel, G., van Boom, J.H., Wang, A.H.-J. and Rich, A. (1986) *Science* **232** p1255-1258

Non-Watson-Crick G-C and A-T base pairs in a DNA-antibiotic complex

Quigley, G.J., Wang, A.H.-J., Ughetto, G., van der Marel, G., van Boom, J.H., and Rich, A. (1980) *Proc. Natl. Acad. Sci. USA* **77** p7204-7208

Molecular structure of an anticancer drug-DNA complex: Daunomycin plus d(CGTAACG)

Rafferty, J.B., Somers, W.S., Saint-Girons, I. and Phillips, S.E.V. (1989) *Nature* **341** p705-710

Three-dimensional crystal structures of *Escherichia coli* metJ repressor with and without corepressor.

Rajagopal, P., Gilbert, D.E., van der Marel, G.A., van Boom, J.H. and Feigon, J. (1988) *J. Magn. Reson.* **78** p526-537

Observation of exchangeable proton resonances of DNA in two-dimensional NOE spectra using a presaturation pulse; Application to d(CGCGAATTCGCG) and d(CGCGA<sup>m6</sup>ATTCGCG)

Ray, D. (1989) *Mol. Cell. Biol.* **9** p1365-1367

Conserved sequence blocks in kinetoplast DNA minicircles from diverse species of trypanosomes.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) *Nature* **311** p532-537

Structure of the nucleosome core particle at 7.0Å resolution.

Rinkel, L.J. and Altona, C. (1987) *J. Biomol. Str. Dynam.* **4** p621-649

Conformational analysis of the deoxyribofuranose ring in DNA by means of the proton-proton coupling constants: A graphical method.

Riou, G. and Bernard, J. (1980) *Biochem. and Biophys. Res. Comm.* **96** p350-354

Berenil induces the complete loss of kinetoplast DNA sequences.

Saenger, W. (1988) *Principles of Nucleic Acid Structure.* Springer-Verlag, Berlin.

Sands, M., Kron, M.A., and Brown, R.B. (1985) *Rev. Infect. Diseases* **7**  
p625-634

Pentamidine: A review.

Sandstrom, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press,  
New York

Sansom, C.E., Laughton, C.A., Neidle, S., Schwalbe, C.H. and Stevens,  
M.F.G. (1990) *Anti-Cancer Drug Design* **5** p243-248

Structural studies on bio-active compounds. Part XIV. Molecular modelling of  
the interactions between pentamidine and DNA.

Schultz, S.C., Shields, G.C. and Steitz, T.A. (1990) *J. Mol. Biol.* **213** p159-166

Craystallisation of *Escherichia coli* catabolite gene activator protein with its  
DNA binding site. The use of modular DNA.

Scott, M.P., Tamkun, J.W. and Hartzell, G.W. (1989) *BBA Rev. Cancer.* **989**  
p25-48

Searle, M.S, and Embrey, K.J. (1990) *Nucleic Acids Research* **18** p3753-3762

Sequence-specific interaction of Hoechst 33258 with the minor groove of an  
adenine-tract DNA duplex studied in solution by <sup>1</sup>H NMR spectroscopy.

Shapiro, T.A. and Englund, P.T. (1990) *Proc. Natl. Acad. Sci. USA* **87** p950-954

Selective cleavage of kinetoplast DNA minicircles promoted by  
antitrypanosomal drugs.

Sklenár, V. and Feigon, J. (1990) *Nature* **345** p836-838

Formation of a stable triplex from a single DNA strand.

Steigemann, W. (1974) *Dissertation, Technische Universität, München.*

Steitz, T.A. (1990) *Q. Rev. Biophys.* **23** p205-280

Structural studies of protein-nucleic acid interaction: the sources of sequence-  
specific binding.

Strobel, S.A. and Dervan, P.B. (1990) *Science* **249** p73-75

Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-  
helix formation.

Strum, N.R, and Simpson, L. (1991) *Nucleic Acids Research* **19** p6277-6281

*Leishmania tarentolae* minicircles of different sequence classes encode single guide RNAs located in the variable region approximately 150 bp from the conserved region.

Suck, D., Lahm, A. and Oefner, C. (1988a) *Nature* **322** p464-468

Structure refined to 2.0Å of a nicked DNA octanucleotide complex with DNase I

Sussman, J.L., Holbrook, S.R., Church, G.M. and Kim, S.-H. (1977)

*Acta Crystallogr.* **33** p800-804

A structure-factor least-squares refinement procedure for macromolecule structures using constrained and restrained parameters.

Swaminathan, S., Ravishanker, G. and Beveridge, D.L. (1991)

*J. Am. Chem. Soc.* **113** p5027-5040

Molecular dynamics of B-DNA including water and counterions: A 140-ps trajectory for d(CGCGAATTCGCG) based on the GROMOS force field.

Taylor, J.S., Schulz, P.G. and Dervan, P.B. (1984) *Tetrahedron* **40** p457-465

DNA affinity cleaving. Sequence specific cleavage of DNA by distamycin-EDTA·Fe(II) and EDTA-distamycin·Fe(II)

Teng, M., Usmann, N., Frederick, C.A. and Wang, A.H.-J. (1988) *Nucleic Acids Research* **16** p2671-2690

The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG)

Tidor, B., Irikura, K.K., Brooks, B.R. and Karplus, M. (1983)

*J. Biomol. Str. Dynam.* **1** p231-252

Dynamics of DNA oligomers.

Turnell, W.G., Satchwell, S.C., and Travers, A.A. (1988) *FEBS Lett.* **232** p263-268

A decapeptide motif for binding to the minor groove of DNA.

Ughetto, G., Wang, A.H.-J., Quigley, G.J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1985) *Nucleic Acids Research* **13** p2305-2323

A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment.

van de Ven, F.J.M. and Hilbers, C.W. (1988) *Eur. J. Biochem.* **178**p1-38

Nucleic acids and nuclear magnetic resonance.

Van Dyke, M.W., Hertzberg, R.P. and Dervan, P.B. (1982)

*Proc. Natl. Acad. Sci. USA* **79** p5470-5474

Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA: DNA cleavage inhibition patterns with methidiumpropyl-EDTA·Fe(II).

Wagner, G. and Wütrich, K. (1979) *J. Magn. Reson.* **33** p675-680

Truncated driven nuclear overhauser effect (TOE). A new technique for studies of selected 1H-1H overhauser effects in the presence of spin diffusion.

Walzer, P.D., Kim, C.K., Foy, J., Linke, M.J. and Cushion, M.T.

(1986) *Antimicrobial agents and Chemotherapy* **32** p896-905

Cationic antitrypanosomal and other antimicrobial agents in the therapy of experimental *Pneumocystis carinii* pneumonia.

Wang, A.H.-J. and Teng, M.-K. (1990) in *Crystallographic and Modeling Methods in Molecular Design Springer-Verlag Publishers* p123-150

Molecular recognition of DNA minor groove binding drugs.

Wang, A.H.-J., Ughetto, G., Quigley, G.J. and Rich, A. (1986) *J. Biomol. Str. Dynam.* **4** p319-342

Interactions of a quinoxaline antibiotic and DNA: The molecular structure of triostin A-d(GCGTACGC) complex.

Wang, A.H.-J., Ughetto, G., Quigley, G.J. and Rich, A. (1987) *Biochemistry* **26** p1152-1163

The interaction between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to d(CGTCACG) at 1.2Å resolution.

Wang, A.H.-J., Ughetto, G., Quigley, G.J., Hakoshima, T., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) *Science* **225** p1125-1121

The molecular structure of a DNA- triostin A complex.

- Ward, B., Rehfuss, R., Goodisman, J. and Dabrowiak, J.C. (1988)  
*Biochemistry* **27** p1198-1204  
 Determination of netropsin-DNA binding constants from footprinting data.
- Weber, I.T. and Steitz, T.A. (1987) *J. Mol. Biol.* **198** p311-326  
 Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5Å resolution.
- Weiner, P.K. and Kollman, P.A. (1981) *J. Comput. Chem.* **2** p287-295  
 "AMBER" assisted model building with energy refinement. A general program for modeling molecules and their interactions.
- Weisblum, B. and Haensler, E. (1974) *Chromosoma* **46** p255-260  
 Fluorimetric properties of the bisbenzimidazole derivative Hoechst 33258, a fluorescent probe specific for AT concentration in chromosomal DNA.
- Westhof, E., Dumas, P. and Moras, D. (1985) *J. Mol. Biol.* **184** p119-145  
 Crystallographic refinement of yeast aspartic transfer RNA.
- Wilbur, D.J., DeFries, T. and Jonas, J. (1976) *J. Chem. Phys.* **65** p1783-1788  
 Self-diffusion in compressed liquid heavy water.
- Williams, L.D, Elgi, M., Gao, Q., van der Marel, G van Boom, J.H., Rich, A. and Frederick, C.A. (1990) *Proc. Natl. Acad. Sci. USA* **87** p2225-2229  
 Structure of nogalamycin bound to a DNA hexamer d(<sup>m5</sup>CGTsA<sup>m5</sup>CG)
- Wilson W.D., Tanious, F.A., Barton, H.J., Jones, R.L., Fox, K., Wydra, R.L. and Streckowski, L. (1990) *Biochemistry* **29** p8452-8461  
 Binding of 4',6'-diamidino-2-phenylindole (DAPI) to GC and mixed sequences in DNA: Intercalation of a classical groove binding molecule.
- Wing, R.M., Pjura, P., Drew, H.R., and Dickerson, R.E. (1984) *EMBO J.* **3** p1201-1206  
 The primary mode of binding of cisplatin to a B-DNA dodecamer: CGCGAATTCGCG

Wispelwey, B and Pearson, R.D. (1991) *Infecti.Control Hosp. Epidemiol.* **12** p375-382

Pentamidine: A review.

Wolberger, C., Dong, Y., Ptashne, M. and Harrison S.C. (1988) *Nature* **335** p789-795

Structure of a phage 434 *cro*/DNA complex.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) *Cell* **67** p517-528

Crystal structure of a MAT $\alpha$ 2 homeodomain-operator complex suggests a general model for homeodomain DNA interactions.

Wolffe, A.P. (1989) *EMBO J.* **8** p527-537

Dominant and specific repression of *Xenopus* oocyte 5S RNA genes and stellite I DNA by histone H1

Woynarowski, J.M., McHugh, M., Sigmund, R.D. and Beerman, T.A. (1989) *Molecular Pharmacology* **35** p177-182

Modulation of topoisomerase II catalytic activity by minor groove binding agents distamycin, Hoechst 33258 and 4',6-diamidino-2-phenylindole (DAPI).

Wu, H.-M. and Crothers, D.M. (1984) *Nature* **308** p509-513

The locus of sequence directed-directed and protein induced DNA bending

Yanagi, K., Prive', G.G. and Dickerson, R.E. (1991) *J. Mol. Biol.* **217** p201-214

Analysis of local helix geometry in three B-DNA decamers and eight dodecamers.

Yoshida, M., Banville, D.L. and Shafer, R.H. (1990) *Biochemistry* **29** p6585-6592

Structural analysis of d(GCAATTGC)<sub>2</sub> and its complex with berenil by Nuclear Magnetic Resonance Spectroscopy.

Zarkezevska, K., Lavery, R. and Pullman, B. (1983) *Nucleic Acids Research* **11** p8825-8839

Theoretical studies of the selective binding to DNA of two non-intercalating ligands: netropsin and SN 18071

Zasedatelev, A.S., Mikhailov, M.V., Krylov, A.S., and Gursky, G.V. (1980)  
*Dokl. Akad. Nauk. SSSR.* **255** p756-760

Mechanism of recognition of A·T pairs in DNA by Hoechst 33258.

Zasedatelev, A.S., Mikhailov, M.V., Krylov, A.S., and Gursky, G.V. (1982)  
*Studia Biophys.* **87** p197-198

Ligand-DNA recognition: Molecular mechanism of AT-specific binding of dye  
"Hoechst 33258" to DNA.

Zimmer, CH. and Wähnert, U. (1986) *Prog. Biophys. Mol. Biol.* **47** p31-112

Nonintercalating DNA-binding ligands: Specificity of the interaction and their  
use as tools in biophysical, biochemical and biological investigations of the  
genetic material.

## **REFERENCES.**

Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C. (1988) *Science* **242** p899-907

Recognition of a DNA operator by the *cro* repressor of phage 434: a view at high resolution.

Anderson, J.E., Ptashane, M. and Harrison S.C. (1987) *Nature* **326** p846-852

Structure of the repressor-operator complex of bacteriophage 434.

Anderson, W.F., Ohlandorf, D.H., Takeda, Y. and Matthews B.W.

(1981) *Nature* **290** p754-758

Structure of the *cro* repressor from bacteriophage  $\lambda$  and its interaction with DNA.

Arents, G., Burlingame, R.W., Wang, B.-C., Love, W.E.

and Moudrianakis, E.M. (1991) *Proc. Natl. Acad. Sci. USA* **88** p10148-10152

The nucleosomal core histone octamer at 3.1Å resolution: A tripartite protein assembly and a left-handed superhelix.

Arnott, S., Chandrasekaran, R., Hall, I. and Puigjaner, L.C.

(1983) *Nucleic Acids Research* **11** p4141-4155

Heteronomous DNA.

Aymami, J., Coll, M., Frederick, C.A., Wang, A.H.-J. and Rich, A.

(1989) *Nucleic Acids Research* p3229-3245

The propeller DNA conformation of poly(dA)·poly(dT).

Baguley, B.C (1982) *Mol. Cell. Biochem.* **43** p167-181

Nonintercalative DNA-binding antitumour compounds.

Benne, R. (1989) *Biochim. Biophys. Acta.* **1007** p131-139

RNA editing in trypanosome mitochondria.

Bernard, J. and Riou, G.F. (1980) *Biochemistry* **19** p4197-4201

In vivo effects of intercalating and nonintercalating drugs on the tertiary structure of kinetoplast deoxyribonucleic acid.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D.,

Rodgers, J.R., Kennard, O., Shimanouchi, T and Tasumi, M.

(1977) *Eur. J. Biochem.* **80** p319-324

The protein data bank: a computer based archival file for macromolecular structures.

Beveridge, A. (1990) *GEMINI An in-house molecular modelling package* .

Bitonti, A.J., Dumont, J.A. and McCann, P.P. (1986) *Biochem. J.* **237** p685-689

Characterisation of *Trypanosoma brucei brucei* S-adenosyl-L-methionine decarboxylase and its inhibition by berenil, pentamidine and methylglyoxal bis(guanylhydrazone).

Blum, B., Bakalara, N. and Simpson, L. (1990) *Cell* **60** p189-198

A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information.

Boehncke, K., Nonella, M., Schulten, K. and Wang, A.H.-J.

(1991) *Biochemistry* **30** p5465-5475

Molecular dynamics investigation of the interaction between DNA and distamycin.

Bontemps, J., Houssier, C., and Fredericq, E. (1975) *Nucleic Acids Research* **2** p975-984

Physico-chemical study of the complexes of 'Hoechst 33258' with DNA and nucleohistone.

Braithwaite, A.W. and Baguley, B.C. (1980) *Biochemistry* **19** p1101-1106

Existence of an extended series of antitumour compounds which bind to deoxyribonucleic acid by nonintercalative means.

Brennan, R.G., Roderick, S.L., Takeda, Y. and Matthews, B.W.

(1990) *Proc. Natl. Acad. Sci. USA* **87** p8165-8169

Protein-DNA conformational changes in the crystal structure of a  $\lambda$  cro-operator complex.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) *J. Comp. Chem.*, **4** p187-217

CHARMM: A program for macromolecular energy, minimisation, and

dynamic calculations.

Brown, D.G., Sanderson, M.R., Skelly, J.V., Jenkins, T.C., Brown, T., Garman, E., Stuart, D.I. and Neidle, S.

(1990) *EMBO J.* **9** p1329-1334

The crystal structure of a berenil-dodecanucleotide complex: the role of water in sequence-specific ligand binding.

Burkhoff, A.M. and Tullius, T.D. (1987) *Cell* **48** p935-943

The unusual conformation adopted by the adenine tracts in kinetoplast DNA.

Burlingame, R.W., Love, W.E., Wang, B.C. Hamlin, R., Xuong, N.H., and Moudrianakis, E.N. (1985) *Science* **228** p546-553

Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3Å.

Calladine, C.R. (1982) *J. Mol. Biol.* **161** p343-352

Mechanism of sequence-dependant stacking of bases in B-DNA.

Cambillau, C.M. (1988) *TOM, a version of FRODO amended to run on a Silicon Graphics IRIS* .

Carrondo, M., Coll, M., Aymami, J., Wang, A.H.-J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1989) *Biochemistry* **28** p7849-7859

Binding of a Hoechst dye to d(CGCGATATCGCG) and its influence on the conformation of a DNA fragment.

Chandra, P., Steel, L.K., Ebner, U., Woltersdorf, M., Laube, H., Kornhuber, B., Mildner, B. and Götz, A. (1977) *Pharmac. Ther. A.* **1** p231-287

Chemical inhibitors of oncoviral DNA polymerases: biological implications and their mode of action.

Chen, G.L., Yang, L., Rowe, T.C., Halligan, B.D., Tewey, K.M. and Liu, L.F. (1984) *J. Biol. Chem.* **259** p13560-13566

Nonintercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian topoisomerase II.

Clarke, N.D., Beamer, L.J., Goldberg, H.R., Berkower, C. and Pabo, C.O.

(1991) *Science* **254** p267-270

The DNA binding arm of  $\lambda$  repressor : Critical contacts from a flexible region.  
N terminal specific contacts in the major groove

Clore, G.M. and Gronenborn, A.M. (1989) *CRC Critical Reviews in Biochemistry and Molecular Biology* **24** p479-564

Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy.

Coll, M., Aymami, J., van der Marel, G.A., van Boom, J.H., Rich, A. and Wang, A.H.-J. (1989) *Biochemistry* **28** p310-320

Molecular structure of the netropsin-d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment.

Coll, M., Frederick, C.A., Wang, A. H.-J. and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA* **84** p8385-8389

A bifurcated hydrogen-bonded conformation in the d(A•T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin.

Comings, D.E. (1975) *Chromosoma* **52** p229-243

Mechanism of chromosome banding. VIII. Hoechst 33258-DNA interaction.

Crothers, D.M., Haran, T.E. and Nadeau, J.G. (1990) *J. Biol. Chem.* **265** p7093-7096

Intrinsically bent DNA.

De Clercq, E. and Dann, O. (1980) *J. Med. Chem.* **23** p787-795

Diaryl amidine derivatives as oncornaviral DNA polymerase inhibitors.

Denny, W.A. (1989) *Anti-Cancer Drug Design* **4** p241-263

DNA-intercalating ligands as anti-cancer drugs: prospects for future design.

Dervan, P.B. (1986) *Science* **232** p461-471

Design of sequence-specific DNA-binding molecules

Dickerson, R.E. (1990) NEWHEL90 available from bitnet address RED@UCLAUE.

Dickerson, R.E. and Drew, H.R. (1981) *Proc. Natl. Acad. Sci. USA* **78**

p7318-7322

Kinematic model for B-DNA.

Dickerson, R.E., Goodsell, D.S., Kopka, M.L. and Pjura, P.E. (1987)

*J. Biomol. Str. Dynam.* **5** p557-579

The effect of crystal packing on oligonucleotide double helix structure.

Dickerson, R.E., Kopka, M.L. and Pjura, P. (1983) *Biochemistry* **80** p7099-7103

A random-walk model for helix bending in B-DNA.

Drew, H.R. and Dickerson, R.E. (1981) *J. Mol. Biol.* **151** p535-556

Structure of a B-DNA dodecamer. III. Geometry of hydration.

Drew, H.R., Wing, R.M., Tanako, T., Broka, C., Tanaka, S., Itakura, K., and

Dickerson, R.E. (1981) *Proc. Natl. Acad. Sci. USA* **78** p2179-2183

Structure of a B-DNA dodecamer: Conformation and dynamics.

EMBO Workshop on DNA Curvature and Bending (1989) *EMBO J.* **8** p1-4

Definition and nomenclature of nucleic acid structure parameters.

Embrey, K.J., Searle, M.S, and Craik, D.J. (1991) *J. Chem. Soc. Chem.*

*Commun.* p1770-1771

Probing the interaction of Hoechst 33258 with an A-T rich oligonucleotide duplex using <sup>1</sup>H NMR spectroscopy.

Fairfield, F.R., Bauer, W.R. and Simpson, M.V. (1985)

*Biochemica et Biophysica Acta* **824** p45-47

Studies on mitochondrial type I topoisomerase and on its function.

Fede, A., Labhardt, A., Bannwarth, W. and Leupin, W. (1991) *Biochemistry* **30**

p11377-11388

Dynamics and binding mode of Hoechst 33258 to d(GTGGAATTCCAC)<sub>2</sub> in the 1:1 solution complex as determined by two-dimensional <sup>1</sup>H NMR.

Feigon, J., Denny, W.A., Leupin, W. and Kearns, D.R. (1984) *J. Med. Chem.* **23**

p450-455

Interactions of antitumour drugs with natural DNA: <sup>1</sup>H NMR study of binding mode and kinetics.

- Felsenfield, G., Davies, D.R. and Rich, A. (1957) *J. Am. Chem. Soc.* **79**  
p2023-2027
- Fox, K.R., Sansom, C.E. and Stevens, M.F.G. (1990) *FEBS* **266** p150-154  
Footprinting studies on the sequence-selective binding of pentamidine to DNA.
- Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R.E. (1982) *J. Biol. Chem.* **257** p14686-14707  
Reversible bending and helix geometry in a B-DNA dodecamer:  
CGCGAATT<sup>Br</sup>CGCG.
- Frederick, C.A., Williams, L.D., Ughetto, G.A., van der Marel, G., van Boom, J.H. and Wang, A.H.-J. (1990) *Biochemistry* **28** p2538-2549  
Structural comparisons of anti-cancer drug-DNA complexes: Adriamycin and daunomycin.
- Freemont, P.S., Lane, A.L. and Sanderson, M.R. (1991) *Biochem. J.* **278** p1-23  
Structural aspects of protein-DNA recognition.
- Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. and Feeney, J. (1990)  
*J. Magn. Reson.* **90** p420-425  
HMQC-NOESY-HMQC, a three-dimensional NMR experiment which allows detection of nuclear overhauser effects between protons with overlapping signals.
- Fritsch, V. and Westhof, E. (1991) *J. Am. Chem. Soc.* **113** p8271-8277  
Three-center hydrogen bonds in DNA: Molecular dynamics of poly(dA)•poly(dT).
- Gago, F., Reynolds, C.A. and Richards, W.G. (1989) *Molecular Pharmacology* **35** p232-241  
The binding of nonintercalative drugs to alternating DNA sequences.
- Gasser, S.M. and Laemmli, U.K. (1986) *Cell* **46** p521-530  
Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*.

- Gasser, S.M. and Laemmli, U.K. (1987) *Trends Genet.* **3** p16-22  
A glimpse at chromosomal order.
- Gasser, S.M. Laroche, T., Falquet, J., Boy de la Tour, E. and Laemmli, U.K. (1986) *J. Mol. Biol.* **188** p613-629  
Metaphase chromosome structure. Involvement of topoisomerase II.
- Goodsell, D. and Dickerson, R.E. (1986) *J. Med. Chem.* **29** p727-733  
Isohelical analysis of DNA groove-binding drugs.
- Gorenstein, D.G., Schroeder, S.A., Fu, J.M., Metz, J.T., Roongta, V. and Jones, C.R. (1988) *Biochemistry* **27** p7223-7237  
Assignments of  $^{31}\text{P}$  NMR resonances in oligodeoxyribonucleotides: Origin of sequence-specific variations in the deoxyribose phosphate backbone conformation and the  $^{31}\text{P}$  chemical shifts of double-helical nucleic acids.
- Gresh, N., and Pullman, B. (1986) *Mol. Pharmacol.* **25** p452-458  
A theoretical study of the nonintercalative binding of berenil and stilbamidine to double-stranded (dA-dT) $_n$  oligomers.
- Griffin, L.C. and Dervan, P.B. (1989) *Science* **245** p967-971  
Recognition of thymine-adenine base pairs by guanine in a pyrimidine triple-helix motif.
- Griffith, J., Bleyman, M., Rauch, C.A., Kitchin, P.A. and Englund, P.T. (1986) *Cell* **46** p717-724  
Visualisation of the bent helix in kinetoplast DNA by electron microscopy.
- Gronenborn, A.M. and Clore, G.M. (1989) *Biochemistry* **28** p5978-5984  
Analysis of relative contributions to the Nuclear Overhauser interproton distance restraints and empirical energy function in the calculation of oligonucleotide structures using restrained molecular dynamics.
- Hare, D.R., Wemmer, D.E., Chou, S.-H., Dronby, G. and Reid, B.R. (1983) *J. Mol. Biol.* **171** p319-336  
Assignment of the non-exchangable proton resonances of

d(C-G-C-G-A-A-T-T-C-G-C-G) using two dimensional nuclear magnetic resonance methods.

Harrison, S.C. and Aggarwal, A.K. (1990) *Ann. Rev. Biochem.* **59** p933-969

DNA recognition by proteins with the helix-turn-helix motif.

Hunter W.N., Brown, T. and Kennard, O. (1986) *J. Biomol. Str. Dynam.* **4** p173-191

Structural features and hydration of d(CGCGAATTAGCG); a double helix containing two G-A mispairs.

Izaurralde, E., Käs, E. and Laemmli, U.K. (1989) *J. Mol. Biol.* **210** p573-585

High preferential nucleation of histone H1 assembly on scaffold associated regions.

Jeffrey, G.A. and Maluszynska, H. (1982) *Int. J. Biol. Macromol.* **4** p173-185

A survey of hydrogen bond geometries in the crystal structures of amino acids.

Jeffrey, G.A. and Mitra, J. (1984) *J. Am. Chem. Soc.* **106** p5546-5553

Three-center (bifurcated) hydrogen bonding in the crystal structures of amino acids.

Jordan, S.R. and Pabo, C.O. (1988) *Science* **242** p893-899

Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Korberg, T.B. and Pabo, C.O. (1990) *Cell* **63** p579-590

Crystal structure of engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions.

Klevit, R.E., Wemmer, D.E. and Reid, B.R. (1986) *Biochemistry* **25** p3296-3303

<sup>1</sup>H NMR studies of the interaction between Distamycin A and a symmetrical DNA dodecamer.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985a)

*J. Mol. Biol.* **183** p553-563

Binding of an antitumor drug to DNA. Netropsin and CGCGAATTBrCGCG.

Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985b)  
*Proc. Natl. Acad. Sci. USA* **82** p1376-1380

The molecular origin of DNA-drug specificity in netropsin and distamycin.

Kumar, S., Yadagiri, B., Zimmermann, J., Pon., R.T. and Lown, J.W. (1991)  
*J. Biomol. Str. Dynam.* **8** p331-357

Sequence specific molecular recognition and binding by a GC recognizing  
Hoechst 33258 analogue to the decadeoxyribonucleotide d(CATGGCCATG)<sub>2</sub>:  
Structural and dynamic aspects deduced from high field <sup>1</sup>H NMR studies.

Lämmler, G., Herzog, H., Saupe, E. and Schutze, H.R. (1971) *WHO Bull.* **44**  
p751-756

Chemotherapeutic studies on *Litomosoides carni* infection of *Mastomys natalensis*. 1. The filaricidal action of 2,6-bis-benzimidazoles.

Lane, A.N. (1990) *Biochim. Biophys. Acta.* **1049** p189-204

The determination of the conformational properties of nucleic acids in solution  
from NMR data.

Lane, A.N., Jenkins, T.C., Brown, D.J.S., and Brown, T. (1991)

*Biochem. J.* **279** p269-281

NMR determination of the solution conformation and dynamics of the A•G  
mismatch in the d(CGCAAATTCGCG)<sub>2</sub> dodecamer.

Lane, A.N., Jenkins, T.C., Brown, T. and Neidle, S. (1991) *Biochemistry* **30**  
p1372-1385

Interaction of berenil with the Eco RI dodecamer d(CGCGAATTCGCG)<sub>2</sub> in  
solution studied by NMR.

Lane, M.J., Dabrowiak, J.C. and Vournakis, J.N. (1983)

*Proc. Natl. Acad. Sci. USA* **80** p3260-3264

Sequence specificity of actinomycin D and netropsin binding to pBR322  
analysed by protection from DNase I.

Larsen, T.A., Goodsell, D.S., Cascio, D., Grzeskowiak, K. and Dickerson, R.E.  
(1989) *J. Biomol. Str. Dynam.* **7** p477-491

The structure of DAPI bound to DNA.

Latt, S.A. and Stetten, G. (1976) *J. Histochem. Cytochem.* **24** p24-33

Spectral studies on 33258 Hoechst and related bisbenzimidazole dyes useful for fluorescent detection of DNA synthesis.

Latt, S.A. and Wohlleb, J.C. (1975) *Chromosoma* **52** p297-316

Optical studies of the interaction of 33258 Hoechst with DNA, chromatin, and metaphase chromosomes.

Latt, S.A., Stetten, G., Juergens, L.A., Willard, H.F. and Scher, C.D. (1975) *J. Histochem. Cytochem.* **23** p493-505

Recent developments in the detection of DNA synthesis by 33258 Hoechst fluorescence

Laughton, C.A., Jenkins, T.C., Fox, K.R. and Neidle, S. (1990) *Nucleic Acids Research* **18** p4479-4488

Interaction of berenil with the tyrT sequence studied by footprinting and molecular modelling. Implications for design of sequence-specific DNA recognition agents.

Lavery, R. and Pullman, B. (1981) *Int. J. Quantum Chem.* **20** p259-272

Molecular electrostatic potential on the surface envelopes of macromolecules: B-DNA

Lefèvre, J.-F., Lane, A.N. and Jardetzky, O. (1987) *Biochemistry* **26** p5076-5090

Solution structure of the Trp operator of *Escherichia coli* determined by NMR.

Leupin, W., Chazin, W.J., Hyberts, S., Denny, W.A. and Wütrich, K. (1986) *Biochemistry* **25** p5902-5908

NMR studies of the complex between the decadeoxynucleotide d(GCATTA ATGC) and a minor groove binding drug SN 6999.

Liaw, Y.-C., Gao, Y.-G., Robinson, H., van der Marel, G., van Boom, J.H., Rich, A. and Wang A.H.-J. (1989) *Biochemistry* **28** p9913-9918

Antitumour drug nogalamycin binds DNA in both grooves simultaneously: Molecular structure of nogalamycin-DNA complex.

Liu, L.F. (1984) *CRC Critical Reviews in Biochemistry* **15**p1-24

DNA topoisomerases: enzymes that catalise the breaking and rejoining of DNA.

Loontjens, F.G., Regenfuss, P., Zechel, A., Dumortier, L. and Clegg, R.M. (1990) *Biochemistry* **29** p9029-9039

Binding characteristics of Hoechst33258 with calf thymus DNA, poly[d(A-T)], and d(CCGGAATTCCGG): Multiple stoichiometries and determination of tight binding with a wide spectrum of site affinities.

Lowe, P.R., Sansom, C.E., Schwalbe, C.H. and Stevens, M.F.G. (1989)

*J. Chem. Soc. Chem. Commun.* **24** p1164-1165

Crystal structure and molecular modelling of the antimicrobial drug pentamidine.

Lown, J.W. (1988) *Anti-Cancer Drug Design* **3**p25-40

Lexitropsins: rational design of DNA sequence reading agents as novel anti-cancer agents and potential cellular probes.

Luck, G., Zimmer, CH., and Schweizer, D. (1988) *Studia Biophysica* **125** p107-119

DNA binding studies of the nonintercalative ligand pentamidine: dA-dT base pair preference.

Maher III, L.J., Wold, B. and Dervan, P.B. (1989) *Science* **238** p725-730

Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation.

Marini, J.C., Effron, P.N., Goodman, T.C., Singleton, C.K., Wells R.D., Wartell, R.M. and Englund, P.T. (1988) *J. Biol. Chem.* **259** p8974-8979

Physical characterisation of a kinetoplast DNA fragment with unusual properties.

Marini, J.C., Levens, S.D., Crothers, D.M. and Englund P.T. (1982)

*Proc. Natl. Acad. Sci. USA* **79** p7664-7668

Bend in helical structure in kinetoplast DNA

- Marky, L.A., and Breslauer, K.J. (1987) *Proc. Natl. Acad. Sci. USA* **84**  
p4363-4359  
Origins of netropsin binding affinity and specificity: Correlations of  
thermodynamic and structural data.
- Martin, R. and Holmes, N. (1983) *Nature* **302** p452-454  
Use of an <sup>125</sup>I-labelled DNA ligand to probe DNA structure.
- McClarín, J.A., Frederick, C.A., Wang, B.C., Greene, P., Boyer, H.W., Grable,  
J. and Rosenberg, J.M. (1986) *Science* **234** p1526-1541  
Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å  
resolution.
- McKay, D.B. and Steitz, T.A. (1981) *Nature* **290** p744-749  
Structure of catabolite gene activator protein at 2.9 Å resolution suggests  
binding to left-handed B-DNA
- Mikhailov, M.V., Zasedatelev, A.S., Krylov, A.S., and Gursky, G.V. (1981)  
*Molec. Biol. SSSR.* **15** p541-554  
Mechanism of the recognition of AT base pairs in DNA by molecules of the the  
dye Hoechst 33258.
- Mildner, B. and Chandra, P. (1979) *Cell Molec. Biol.* **25** p399-407  
Molecular mechanism of action of diaminephenylindole (DAPI). II. Effect of  
DAPI on the template activity of DNA and polydeoxynucleotides an the DNA-  
polymerase system for bacteria, eukaryotic cells and RNA tumour viruses.
- Mondragón, A., Subbiah, S., Almo, S.C., Drottar, M. and Harrison, S.C.  
(1989a) *J. Mol. Biol.* **205** p189-200  
Structure of the amino-terminal domain of phage 434 repressor at 2.0Å  
resolution.
- Mondragón, A., Wolberger, C. and Harrison, S.C. (1989b) *J. Mol. Biol.* **205**  
p179-188  
Structure of phage 434 cro protein at 2,35Å resolution.

- Moore, M.H., Hunter W.N., d'Estaintot, B.L. and Kennard, O. (1989)  
*J. Mol. Biol.* **206** p693-705  
 Drug-DNA interactions in the crystal structure of d(CGATCG) complexed with daunomycin.
- Moser, H.E. and Dervan, P.B. (1987) *Science* **238** p645-650  
 Sequence specific cleavage of double helical DNA by triple helix formation.
- Müller, W and Gautier, F. (1975) *Eur. J. Biochem.* **54** p385-394  
 Interactions of heterochromatic compounds with nucleic acids. A-T specific non-intercalating ligands.
- Nadeau, J.G. and Crothers, D.M. (1989) *Proc. Natl. Acad. Sci. USA* **86**  
 p2622-2626  
 Structural basis for DNA bending.
- Neidle, S., Brown, D.G., Jenkins, T.C., Laughton C.A., Sanderson, M.R. and Skelly, J.V. (1990) *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions* Kluwer Academic Publishers p43-57  
 Drug-DNA Recognition: Sequence-specificity of the DNA minor groove binder berenil.
- Nelson, E.M., Tewey, K.M. and Liu, L.F. (1984) *Proc. Natl. Acad. Sci. USA* **81**  
 p1361-  
 Mechanism of antitumour drug action: poisoning of mammalian DNA topoisomerase II on DNA by 4'-(9-acridinylamino)methanesulfon-m-anisidide.
- Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987) *Nature* **330**  
 p221-226  
 The structure of an oligo(dA)•oligo(dT) tract and its biological implications.
- Nerdal, W., Hare, D.R. and Reid, B.R. (1988) *J. Mol. Biol.* **201** p717-739  
 Three-dimensional structure of wild-type lac Primbow promoter DNA in solution.
- Newton, B.A., (1975) *In Antibiotics III, mechanism of action of antimicrobial and antitumour agents.* Corcoran, I.W., and Hahn, F.G. (eds) Publ. Sringer-

*Verlag, Berlin.* **3** p34-47

Berenil: A trypanocide with selective activity against extranuclear DNA.

Ott, J. and Eckstein, F. (1985) *Biochemistry* **24** p2530-2535

<sup>31</sup>P NMR spectral analysis of the dodecamer d(CGCGAATTCGCG)

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J.* **9** p3085-3092

Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution.

Antennapedia homeodomain DNA complex

Pabo, C.O. and Lewis, M. (1982) *Nature* **298** p443-447

The operator-binding domain of  $\lambda$  repressor: structure and DNA recognition

Parkinson, J.A., Barber, J., Douglas, K.T., Rosamond, J. and Sharples, D. (1990) *Biochemistry* **29** p10181-10190

Minor groove recognition of the self-complementary duplex d(CGCGAATTCGCGC)<sub>2</sub> by Hoechst 33258: a high field NMR study.

Patel, D.J. (1978) *Eur. J. Biochem.* **99** p369-379

Netropsin · dG-dG-dA-dA-dT-dT-dC-dC complex. Antibiotic binding at Adenine-Thymine base pairs in the minor groove of the self complementary octanucleotide complex.

Patel, D.J. and Shapiro, L. (1985) *Biochimie* **67** p887-915

Molecular recognition in noncovalent antitumor agent-DNA complexes: NMR studies of the base and sequence dependent recognition of the DNA minor groove by netropsin.

Pearl, L.H., Skelly, J.V., Hudson, B.D. and Neidle, S. (1987)

*Nucleic Acids Research* **15** p3469-3477

The crystal structure of the DNA-binding drug berenil. Molecular modelling studies of berenil-DNA complexes.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.* **220** p457-479

Are time-averaged restraints necessary for nuclear magnetic resonance refinement?

Pelton, J.G. and Wemmer, D.E. (1988) *Biochemistry* **27** p8088-8096

Structural modeling of the Distamycin A-d(CGCGAATCGCG) complex using 2D NMR and molecular mechanics.

Pelton, J.G. and Wemmer, D.E. (1990) *J. Am. Chem. Soc.* **112** p1393-1399

Binding modes of distamycin A with d(CGCAAATTTGCG)<sub>2</sub> determined by two dimensional NMR.

Phillips, S.E.V. (1991) *Curr. Opinions Struct. Biol.* **1** p89-98

Specific  $\beta$ -sheet interactions.

Pjura, P.E., Grzeskowiak, G. and Dickerson, R.E. (1987) *J. Mol. Biol.* **197** p257-271

Binding of Hoechst 33258 to the minor groove of B-DNA.

Portugal, J. and Waring, M.J. (1986) *Nucleic Acids Research* **14** p8735-8754

Antibiotics which can alter the rotational orientation of nucleosome core DNA.

Portugal, J. and Waring, M.J. (1987) *Biochimie* **69** p825-840

Analysis of the effects of antibiotics on the structure of nucleosome core particles determined by DNAase I cleavage.

Portugal, J. and Waring, M.J. (1987) *Eur. J. Biochem.* **167** p281-289

Comparison of binding sites in DNA for benzenil, netropsin and distamycin. A footprinting study.

Postel, E.H., Flint, S.J., Kessler, D.J. and Hogan, M.E. (1991) *Proc. Natl. Acad. Sci. USA* **88** p8227-8231

Evidence that a triple helix-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels.

Powell, M.J.D. (1977) *Mathematical Programming* **12** p241-254

Quigley, G.J., Ughetto, G., van der Marel, G., van Boom, J.H., Wang, A.H.-J. and Rich, A. (1986) *Science* **232** p1255-1258

Non-Watson-Crick G-C and A-T base pairs in a DNA-antibiotic complex

Quigley, G.J., Wang, A.H.-J., Ughetto, G., van der Marel, G., van Boom, J.H., and Rich, A. (1980) *Proc. Natl. Acad. Sci. USA* **77** p7204-7208

Molecular structure of an anticancer drug-DNA complex: Daunomycin plus d(CGTACG)

Rafferty, J.B., Somers, W.S., Saint-Girons, I. and Phillips, S.E.V. (1989) *Nature* **341** p705-710

Three-dimensional crystal structures of *Escherichia coli* metJ repressor with and without corepressor.

Rajagopal, P., Gilbert, D.E., van der Marel, G.A., van Boom, J.H. and Feigon, J. (1988) *J. Magn. Reson.* **78** p526-537

Observation of exchangeable proton resonances of DNA in two-dimensional NOE spectra using a presaturation pulse; Application to d(CGCGAATTCGCG) and d(CGCGA<sup>m6</sup>ATTCGCG)

Ray, D. (1989) *Mol. Cell. Biol.* **9** p1365-1367

Conserved sequence blocks in kinetoplast DNA minicircles from diverse species of trypanosomes.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) *Nature* **311** p532-537

Structure of the nucleosome core particle at 7.0Å resolution.

Rinkel, L.J. and Altona, C. (1987) *J. Biomol. Str. Dynam.* **4** p621-649

Conformational analysis of the deoxyribofuranose ring in DNA by means of the proton-proton coupling constants: A graphical method.

Riou, G. and Bernard, J. (1980) *Biochem. and Biophys. Res. Comm.* **96** p350-354

Berenil induces the complete loss of kinetoplast DNA sequences.

Saenger, W. (1988) *Principles of Nucleic Acid Structure.* Springer-Verlag, Berlin.

- Sands, M., Kron, M.A., and Brown, R.B. (1985) *Rev. Infect. Diseases* **7**  
p625-634  
Pentamidine: A review.
- Sandstrom, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press,  
New York
- Sansom, C.E., Laughton, C.A., Neidle, S., Schwalbe, C.H. and Stevens,  
M.F.G. (1990) *Anti-Cancer Drug Design* **5**p243-248  
Structural studies on bio-active compounds. Part XIV. Molecular modelling of  
the interactions between pentamidine and DNA.
- Schultz, S.C., Shields, G.C. and Steitz, T.A. (1990) *J. Mol. Biol.* **213**p159-166  
Craystallisation of *Escherichia coli* catabolite gene activator protein with its  
DNA binding site. The use of modular DNA.
- Scott, M.P., Tamkun, J.W. and Hartzell, G.W. (1989) *BBA Rev. Cancer.* **989**  
p25-48
- Searle, M.S, and Embrey, K.J. (1990) *Nucleic Acids Research* **18**p3753-3762  
Sequence-specific interaction of Hoechst 33258 with the minor groove of an  
adenine-tract DNA duplex studied in solution by <sup>1</sup>H NMR spectroscopy.
- Shapiro, T.A. and Englund, P.T. (1990) *Proc. Natl. Acad. Sci. USA* **87** p950-954  
Selective cleavage of kinetoplast DNA minicircles promoted by  
antitrypanosomal drugs.
- Sklenár, V. and Feigon, J. (1990) *Nature* **345** p836-838  
Formation of a stable triplex from a single DNA strand.
- Steigemann, W. (1974) *Dissertation, Technische Universität, München.*
- Steitz, T.A. (1990) *Q. Rev. Biophys.* **23** p205-280  
Structural studies of protein-nucleic acid interaction: the sources of sequence-  
specific binding.
- Strobel, S.A. and Dervan, P.B. (1990) *Science* **249** p73-75  
Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-  
helix formation.
- Strum, N.R, and Simpson, L. (1991) *Nucleic Acids Research* **19** p6277-6281

*Leishmania tarentolae* minicircles of different sequence classes encode single guide RNAs located in the variable region approximately 150 bp from the conserved region.

Suck, D., Lahm, A. and Oefner, C. (1988a) *Nature* **322** p464-468

Structure refined to 2.0Å of a nicked DNA octanucleotide complex with DNase I

Sussman, J.L., Holbrook, S.R., Church, G.M. and Kim, S.-H. (1977)

*Acta Crstallogr.* **33** p800-804

A structure-factor least-squares refinement procedure for macromolecule structures using constrained and restrained parameters.

Swaminathan, S., Ravishanker, G. and Beveridge, D.L. (1991)

*J. Am. Chem. Soc.* **113**p5027-5040

Molecular dynamics of B-DNA including water and counterions: A 140-ps trajectory for d(CGCGAATTCGCG) based on the GROMOS force field.

Taylor, J.S., Schulz, P.G. and Dervan, P.B. (1984) *Tetrahedron* **40** p457-465

DNA affinity cleaving. Sequence specific cleavage of DNA by distamycin-EDTA-Fe(II) and EDTA-distamycin-Fe(II)

Teng, M., Usmann, N., Frederick, C.A. and Wang, A.H.-J. (1988) *Nucleic Acids Research* **16**p2671-2690

The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG)

Tidor, B., Irikura, K.K., Brooks, B.R. and Karplus, M. (1983)

*J. Biomol. Str. Dynam.* **1**p231-252

Dynamics of DNA oligomers.

Turnell, W.G., Satchwell, S.C., and Travers, A.A. (1988) *FEBS Lett.* **232**

p263-268

A decapeptide motif for binding to the minor groove of DNA.

Ughetto, G., Wang, A.H.-J., Quigley, G.J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1985) *Nucleic Acids Research* **13**p2305-2323

A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment.

van de Ven, F.J.M. and Hilbers, C.W. (1988) *Eur. J. Biochem.* **178** p1-38

Nucleic acids and nuclear magnetic resonance.

Van Dyke, M.W., Hertzberg, R.P. and Dervan, P.B. (1982)

*Proc. Natl. Acad. Sci. USA* **79** p5470-5474

Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA: DNA cleavage inhibition patterns with methidiumpropyl-EDTA·Fe(II).

Wagner, G. and Wütrich, K. (1979) *J. Magn. Reson.* **33** p675-680

Truncated driven nuclear overhauser effect (TOE). A new technique for studies of selected 1H-1H overhauser effects in the presence of spin diffusion.

Walzer, P.D., Kim, C.K., Foy, J., Linke, M.J. and Cushion, M.T.

(1986) *Antimicrobial agents and Chemotherapy* **32** p896-905

Cationic antitrypanosomal and other antimicrobial agents in the therapy of experimental *Pneumocystis carinii* pneumonia.

Wang, A.H.-J. and Teng, M.-K. (1990) in *Crystallographic and Modeling Methods in Molecular Design Springer-Verlag Publishers* p123-150

Molecular recognition of DNA minor groove binding drugs.

Wang, A.H.-J., Ughetto, G., Quigley, G.J. and Rich, A. (1986) *J. Biomol. Str. Dynam.* **4** p319-342

Interactions of a quinoxaline antibiotic and DNA: The molecular structure of triostin A-d(GCGTACGC) complex.

Wang, A.H.-J., Ughetto, G., Quigley, G.J. and Rich, A. (1987) *Biochemistry* **26** p1152-1163

The interaction between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to d(CGTACG) at 1.2Å resolution.

Wang, A.H.-J., Ughetto, G., Quigley, G.J., Hakoshima, T., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) *Science* **225** p1125-1121

The molecular structure of a DNA- triostin A complex.

- Ward, B., Rehfuss, R., Goodisman, J. and Dabrowiak, J.C. (1988)  
*Biochemistry* **27** p1198-1204  
 Determination of netropsin-DNA binding constants from footprinting data.
- Weber, I.T. and Steitz, T.A. (1987) *J. Mol. Biol.* **198** p311-326  
 Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5Å resolution.
- Weiner, P.K. and Kollman, P.A. (1981) *J. Comput. Chem.* **2** p287-295  
 "AMBER" assisted model building with energy refinement. A general program for modeling molecules and their interactions.
- Weisblum, B. and Haensler, E. (1974) *Chromosoma* **46** p255-260  
 Fluorimetric properties of the bisbenzimidazole derivative Hoechst 33258, a fluorescent probe specific for AT concentration in chromosomal DNA.
- Westhof, E., Dumas, P. and Moras, D. (1985) *J. Mol. Biol.* **184** p119-145  
 Crystallographic refinement of yeast aspartic transfer RNA.
- Wilbur, D.J., DeFries, T. and Jonas, J. (1976) *J. Chem. Phys.* **65** p1783-1788  
 Self-diffusion in compressed liquid heavy water.
- Williams, L.D, Elgi, M., Gao, Q., van der Marel, G van Boom, J.H., Rich, A. and Frederick, C.A. (1990) *Proc. Natl. Acad. Sci. USA* **87** p2225-2229  
 Structure of nogalamycin bound to a DNA hexamer d(<sup>m5</sup>CGTsA<sup>m5</sup>CG)
- Wilson W.D., Tanious, F.A., Barton, H.J., Jones, R.L., Fox, K., Wydra, R.L. and Streckowski, L. (1990) *Biochemistry* **29** p8452-8461  
 Binding of 4',6'-diamidino-2-phenylindole (DAPI) to GC and mixed sequences in DNA: Intercalation of a classical groove binding molecule.
- Wing, R.M., Pjura, P., Drew, H.R., and Dickerson, R.E. (1984) *EMBO J.* **3** p1201-1206  
 The primary mode of binding of cisplatin to a B-DNA dodecamer: CGCGAATTCGCG

Wispelwey, B and Pearson, R.D. (1991) *Infecti.Control Hosp. Epidemiol.* **12**  
p375-382

Pentamidine: A review.

Wolberger, C., Dong, Y., Ptashne, M. and Harrison S.C. (1988) *Nature* **335**  
p789-795

Structure of a phage 434 *cro*/DNA complex.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991)  
*Cell* **67** p517-528

Crystal structure of a MAT $\alpha$ 2 homeodomain-operator complex suggests a  
general model for homeodomain DNA interactions.

Wolffe, A.P. (1989) *EMBO J.* **8** p527-537

Dominant and specific repression of *Xenopus* oocyte 5S RNA genes and stellite  
I DNA by histone H1

Woynarowski, J.M., McHugh, M., Sigmund, R.D. and Beerman, T.A. (1989)  
*Molecular Pharmacology* **35** p177-182

Modulation of topoisomerase II catalytic activity by minor groove binding  
agents distamycin, Hoechst 33258 and 4',6-diamidino-2-phenylindole (DAPI).

Wu, H.-M. and Crothers, D.M. (1984) *Nature* **308** p509-513

The locus of sequence directed-directed and protein induced DNA bending

Yanagi, K., Prive', G.G. and Dickerson, R.E. (1991) *J. Mol. Biol.* **217** p201-214

Analysis of local helix geometry in three B-DNA decamers and eight  
dodecamers.

Yoshida, M., Banville, D.L. and Shafer, R.H. (1990) *Biochemistry* **29** p6585-6592  
Structural analysis of d(GCAATTGC)<sub>2</sub> and its complex with berenil by Nuclear  
Magnetic Resonance Spectroscopy.

Zarkezevska, K., Lavery, R. and Pullman, B. (1983) *Nucleic Acids Research* **11**  
p8825-8839

Theoretical studies of the selective binding to DNA of two non-intercalating  
ligands: netropsin and SN 18071

Zasedatelev, A.S., Mikhailov, M.V., Krylov, A.S., and Gursky, G.V. (1980)  
*Dokl. Akad. Nauk. SSSR.* **255** p756-760

Mechanism of recognition of A·T pairs in DNA by Hoechst 33258.

Zasedatelev, A.S., Mikhailov, M.V., Krylov, A.S., and Gursky, G.V. (1982)  
*Studia Biophys.* **87** p197-198

Ligand-DNA recognition: Molecular mechanism of AT-specific binding of dye  
"Hoechst 33258" to DNA.

Zimmer, CH. and Wähnert, U. (1986) *Prog. Biophys. Mol. Biol.* **47** p31-112

Nonintercalating DNA-binding ligands: Specificity of the interaction and their  
use as tools in biophysical, biochemical and biological investigations of the  
genetic material.

DNA - Chemistry

DNA - Drug effects

1' MOLECULAR CONFORMATION