

**Structural Studies in DNA  
by  
Ewan D. Booth**

**A thesis presented for the degree of Doctor of Philosophy**

**University of Edinburgh**

**1991**



To  
Mum and Dad

**Declaration**

I declare that this thesis is my own composition and that the work of which it is a record was carried out by myself unless otherwise acknowledged. No part of this thesis has been submitted in any previous application for a higher degree.

Ewan D. Booth 1991

### **Acknowledgements**

I would like to thank Dr. Tom Brown for his guidance and enthusiasm throughout the course of this project. I am greatly indebted to Dr Gordon A. Leonard for the refinement of crystal structures in addition to being an excellent friend and colleague. Thanks should also go to Mrs. Sandra Thomson, formerly of the Oswel DNA Service, for instruction in the use of DNA synthesisers and to Dr. Michaela Vorlíčková of the Czechoslovak Academy of Sciences for supplying the data from which Figs. 41-46 are constructed. I should also thank Dr. Alister G. Craig for hybridisation of genetic probes. Finally I must thank the SERC and The University of Edinburgh for funding and facilities to carry out this work.

## Abstract

Base pair mismatches may be formed during replication or genetic recombination. Not all mismatched base pairs are repaired with the same efficiency, in particular the A.G mismatch is consistently under repaired. Why the A.G mismatch is poorly repaired is not clear, although previous work has suggested the mismatch is conformationally flexible. In the present study the A.G mismatch in the sequence dCGCAAATTGGCG was examined by X-ray crystallography and pH dependant UV melting studies. Crystals were grown at pH 6.6. The base pair has an A (*anti*).G (*syn*) conformation. From the pH dependant UV melting studies it is likely this is a protonated AH<sup>+</sup> (*anti*).G (*syn*) base pair formed by protonation of the N(1)-atom of adenine. The G (*syn*) conformation is stabilised by hydrogen bonding to a network of solvent molecules in the major and minor grooves. A parallel investigation of the A.I mismatch was performed. The oligonucleotide dCGCAAATTIGCG was crystallised. An A (*anti*).I (*syn*) base pair may be indicated, pH dependant UV melting profiles suggest the existence of a protonated base pair.

5-Bromouracil is strongly mutagenic forming mismatches with G. The G.BrU mismatch has been reported to form a wobble base pair although other forms are possible. The oligonucleotide dCGCGAATTBrUGCG was synthesised and used in pH dependant UV melting studies. These suggest that there is little or no contribution to the base pair from non-wobble, ionised or enolised, forms.

Frame shift mutations may arise from nonmatched bases in DNA. A series of tridecamers was synthesised for analysis by X-ray crystallography; dCGCGTAATTCGCG, dCGCGBrUAATTCGCG, dCGCGTAATBrUCGCG and dCGCGTAATT~~B~~CGCG. It was hoped to determine the looped out, stacked or nonmatched nature of the additional bases.

2-Amino-2'-deoxyadenosine (dA') is an analogue of dA. Base pairs formed between A' and T contain three hydrogen bonds and are A.T like in the major groove and G.C like in the minor groove. The additional hydrogen bonding affords extra stability to A'.T base pairs relative to A.T base pairs. Z-DNA is formed by d(GC)<sub>n</sub> sequences but not by d(AT)<sub>n</sub> sequences. The mixed sequence oligonucleotide dCA'TA'TG was made for X-ray

crystallographic investigation and UV melting studies. It was hoped this would crystallise as Z-DNA.

Recently X-DNA has been investigated. This can be formed by poly(dA-dT) under stringent conditions due to the effects of salt or alcohols. X-DNA was first identified by its CD spectrum. Similar CD spectra have been observed for poly(dA'-dT) closer to physiological conditions. The oligonucleotides d(TA')<sub>n</sub> n = 2, 3, 4 were synthesised for X-ray crystallography, d(TA')<sub>4</sub> was also investigated by CD spectroscopy. This showed the ability of the oligonucleotide to form X-DNA.

The additional stability of dA' oligonucleotides gives them potential used as stable short genetic probes. A series of oligonucleotides was made, these hybridised successfully under conditions where the native sequences failed to.

A new phosphoramidite for the synthesis of dA' oligonucleotides was developed. This overcomes the problem of acid catalysed depurination of dA' residues during solid phase phosphoramidite DNA synthesis, di-n-butylformamide dimethylacetal was used to protect the N<sup>6</sup> function. This was shown to be stable to the detritylation conditions encountered in DNA synthesis.

## Abbreviations

A	adenine
A'	2-aminoadenine
BrU	5-bromouracil
BrC	5-bromocytosine
C	cytosine
EtU	5-ethyluracil
FU	5-fluorouracil
G	guanine
I	hypoxanthine
MeC	5-methylcytosine
Me <sub>2</sub> <sup>2</sup> G	N <sup>2</sup> -dimethylguanine
O <sup>6</sup> MeG	O <sup>6</sup> -methylguanine
T	thymine
U	uracil
dA	2'-deoxyadenosine
dA'	2-amino-2'-deoxyadenosine
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
dT	thymidine
dl	2'-deoxyinosine
CPG	controlled pore glass
DCA	dichloroacetic acid
DMT	4, 4'-dimethoxytrityl
DMA	dimethyl acetamide
DMF	dimethyl formamide
DMAP	4-dimethylaminopyridine
DNA	deoxyribonucleic acid
HPLC	high performance liquid chromatography
RP-HPLC	reverse phase HPLC
SAX-HPLC	strong anion exchange HPLC
SSC	15mM NaCitrate 0.5M NaCl
TCA	trichloroacetic acid
TEA	triethylamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
T <sub>m</sub>	oligonucleotide melting temperature
TPS-Cl	2, 4, 6-triisopropylbenzenesulphonyl chloride
NOE	nuclear overhauser effect
rA	adenosine
rU	uridine
rA'	2-amino-adenosine
RNA	ribonucleic acid
mRNA	messenger RNA
tRNA	transfer RNA

	<b>Contents</b>	<b>Page</b>
1.0	Introduction to DNA	1
2.0	Mutations	3
3.0	Structural Studies on Mismatched Base Pairs	10
3.1	The A.G Mismatch	11
3.1.1	The A.G Mismatch: Crystallisation, Data Collection and Refinement	16
3.1.2	A.G Mismatch Conformation	17
3.1.3	Structure of the A.G mismatch	20
3.1.4	Conformational Flexibility of the A.G Mismatch	24
3.1.5	Mismatch Stabilisation by Bifurcated Hydrogen Bonds and Solvent Interactions	25
3.1.6	Thermodynamic Parameters of the A.G Mismatch	27
3.2	The A.I Mismatch	29
3.2.1	The A.I Mismatch: Crystallisation, Data Collection and Refinement	31
3.2.2	A.I Mismatch Conformation	32
3.2.3	A.I Mismatch Conformation 2	34
3.2.4	Structure of the A.I Mismatch	35
3.2.5	Conformational Flexibility of the A.I Mismatch	37
3.3	Comparison of the A.G and A.I Mismatches	38
3.4	The G.BrU Mismatch	40
3.4.1	Structure of the G.BrU Mismatch	43
4.0	2-Amino-2'-deoxyadenosine and the DNA Duplex	45
4.1	The Effect of 2-Amino-2'-deoxyadenosine on DNA Duplex Stability	45
4.2	Factors Influencing Oligonucleotide Structure	47
4.3	Effect of 2-Amino-2'-deoxyadenosine on Oligonucleotide Structure	49
4.4	Conformational Flexibility of Poly(dA-dT)	53
4.4.1	Effect of Solvent on Poly(dA-dT)	55

4.4.2	Effect of CsF Alone on Poly(dA-dT)	56
4.4.3	Structural Transitions of Poly(dA'-dT)	58
4.4.4	Formation of Z-DNA in Mixed Sequence Oligonucleotides Containing 2-Amino-2'-deoxyadenosine	61
5.0	DNA Synthesis	63
5.1	Phosphoramidite Method	64
5.2	Phosphorites Protection	65
5.3	Base and Sugar Protection	67
6.0	Acid Catalysed Depurination	68
6.1	Influence of N <sup>6</sup> Protection on Depurination	72
7.0	Thermal Denaturation of Oligonucleotides (UV Melting)	75
8.0	2-Amino-2'-deoxyadenosine	77
8.1	2-Amino-2'-deoxyadenosine Nucleosides	79
8.2	Stability Studies on Model Nucleosides	82
9.0	Synthesis of 2-Amino-2'-deoxyadenosine Phosphoramidites	86
9.1	Oligonucleotide Synthesis Using dA' Phosphoramidites	87
9.2	Mixed Sequence Z-DNA dA' Oligonucleotides	93
9.3	Use of dA' in Stable Genetic Probes	96
9.3.1	Hybridisation of Trial Probes	97
9.3.2	UV Thermal Denaturation of Trial Probes	99
9.4	X-DNA and Oligonucleotides Containing dA'	103
9.4.1	Circular Dichroism of d(TA') <sub>4</sub>	108
9.5	Oligonucleotide Digestions	116
9.6	Improvements to dA' Phosphoramidite Design	123
10.0	DNA Tridecamers	124
10.1	Tridecamer Synthesis	126
10.2	Tridecamer Crystallisation	127
11.0	Materials and Methods	128

11.1	Experimental	131
12.0	References	148
	Appendix: Publications	162

## 1.0 Introduction to DNA

DNA (deoxyribonucleic acid) is a linear polymeric molecule formed from monomeric units, nucleotides. The nucleotides are in turn composed of three component parts; the phosphate, the sugar and a heterocyclic base. The sugar  $\beta$ -D-2'-deoxyribose as a cyclic furanose is joined via a  $\beta$ -glycosyl linkage to one of the four heterocyclic bases; adenine (A), guanine (G), cytosine (C) or thymine (T), to give the nucleosides 2'-deoxyadenosine (dA), 2'-deoxyguanosine (dG), 2'-deoxycytidine (dC) and thymidine (dT). The repeated phosphodiester junctions between the sugar units 3' and 5' hydroxyl groups gives the oligodeoxyribonucleotide (oligonucleotide) a directional nature with specified 5'-OH and 3'-OH ends. These features along with the base and sugar numbering systems are shown in Fig. 1a, b.

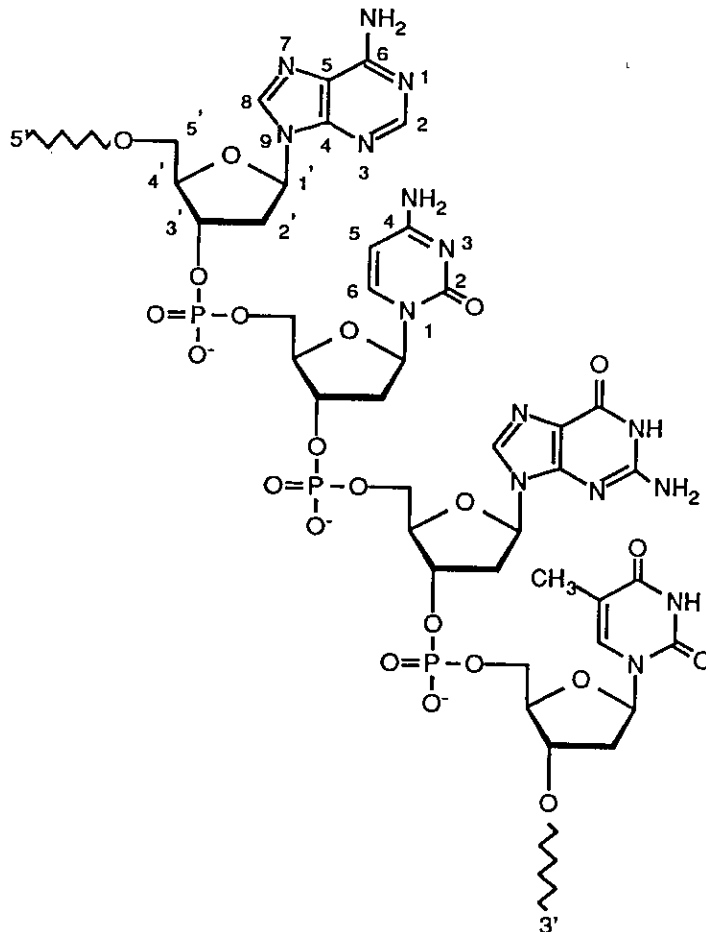


Fig. 1a. DNA and base numbering showing 3' and 5' ends of oligonucleotide.

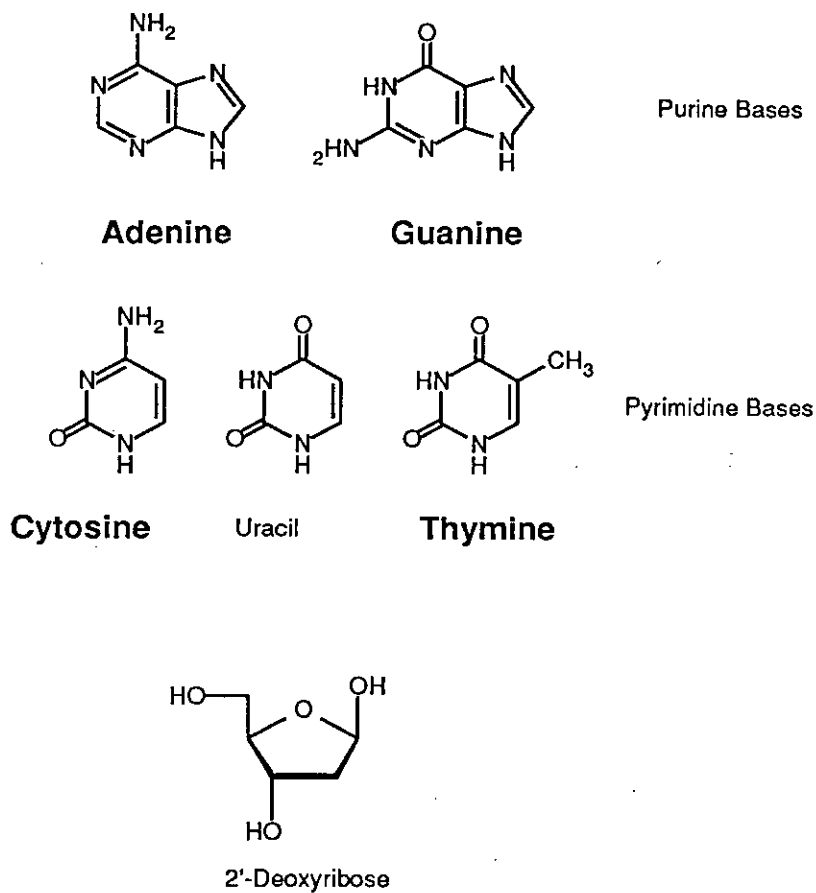


Fig. 1b. Major bases and sugar.

The double helix model for DNA that was introduced by Watson and Crick <sup>1</sup> in 1953 became a turning point in our understanding of the role of DNA and the way it carried out its biological functions. The idea of two complementary strands arose from the recognition that only certain allowed base pairings could take place if all the bases were in their major tautomeric forms giving base pairings between G and C and between A and T Fig. 2. This idea was supported by earlier work <sup>2, 3</sup> that had identified the ratios of adenine to thymine and guanine to cytosine to always be very close to 1.1.

A further requirement for the model was that the two strands should run in opposite directions with the bases approximately perpendicular to the helix direction. Hydrogen bonding between the bases hold the two strands together. Watson and Crick were not slow to realise that the complementary bases at once gave a mechanism for the propagation of

genetic information from one generation to the next. Prior to their model a stumbling block in this idea had been the requirement for some form of genetic template<sup>4</sup> to facilitate the formation of the next generations DNA. By restricting the allowed base pairing to G.C and A.T and having two complementary strands then each strand is by necessity the template for the other.

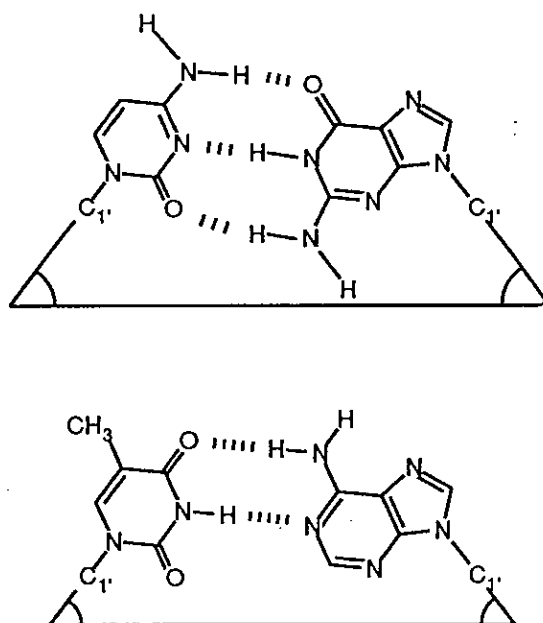


Fig. 2. C.G and T.A base pairing schemes, minor groove region indicated by solid line.

The initial concept of a double helix has now been extended to cover a wide range of structural variation. These include the A and B right handed forms which differ in their degree of hydration, the left handed Z form and the as yet not fully characterised X form.

## 2.0 Mutations

In order that genetic information is accurately passed from one generation or one cell to the next during cell division the cellular DNA must undergo replication. The replication needs to proceed with a high degree of fidelity. Any changes in the base sequence will alter the

genetic code in a way that may have no effect, a detrimental effect or beneficial effect depending on the precise nature of the change. Such changes are known as mutations.

Mutations occur when one or more bases in the daughter DNA strand is changed relative to the parent strand. Changes may be caused by the action of DNA damaging chemicals on the base structure, by the incorporation of the wrong base during nucleic acid biosynthesis giving rise to a mismatched base pair in the newly formed duplex, or by genetic recombination. The incorporation of the wrong base may cause a spontaneous substitution mutation if it should go undetected. However a range of repair enzymes has evolved to reduce the chances of a base mispairing being propagated<sup>5</sup>. There are three stages at which repair can take place:

- i. During base selection at replication by DNA polymerase, which travels in a 5'-3' direction incorporating the nucleoside triphosphates complementary to the template strand into the new DNA.
- ii. Proofreading by 3'-5' exonucleases which should recognise and excise incorrect bases.
- iii. After replication by postreplicative repair enzymes.

These repair systems rely on the differentiation between Watson-Crick and non Watson-Crick base pairs.

Non Watson-Crick base pairs can be formed to give eight different mismatched base pairs A.C, A.A, A.G, C.T, C.C, G.T, G.G and T.T. In order for hydrogen bonding to occur in all of these cases one may need to invoke the use of minor tautomers<sup>6</sup> Fig. 3 or wobble pairing<sup>7</sup>. Wobble pairing was first proposed to explain the degeneracy of the genetic code during translation in which  $4^3$  base triplets (64 triplets) code for only 20 amino acids. The first two bases in the mRNA codon form normal Watson-Crick base pairs with tRNA, while the third base is allowed to misalign and form a wobble pair as shown in Fig. 4. Wobble pairing has since been found in the structures of the G.T<sup>8-12</sup>, <sup>13</sup>, C.A<sup>14-16</sup>, BrU.G<sup>17</sup>, FU.G<sup>18</sup> and I.T<sup>19</sup> mismatches.

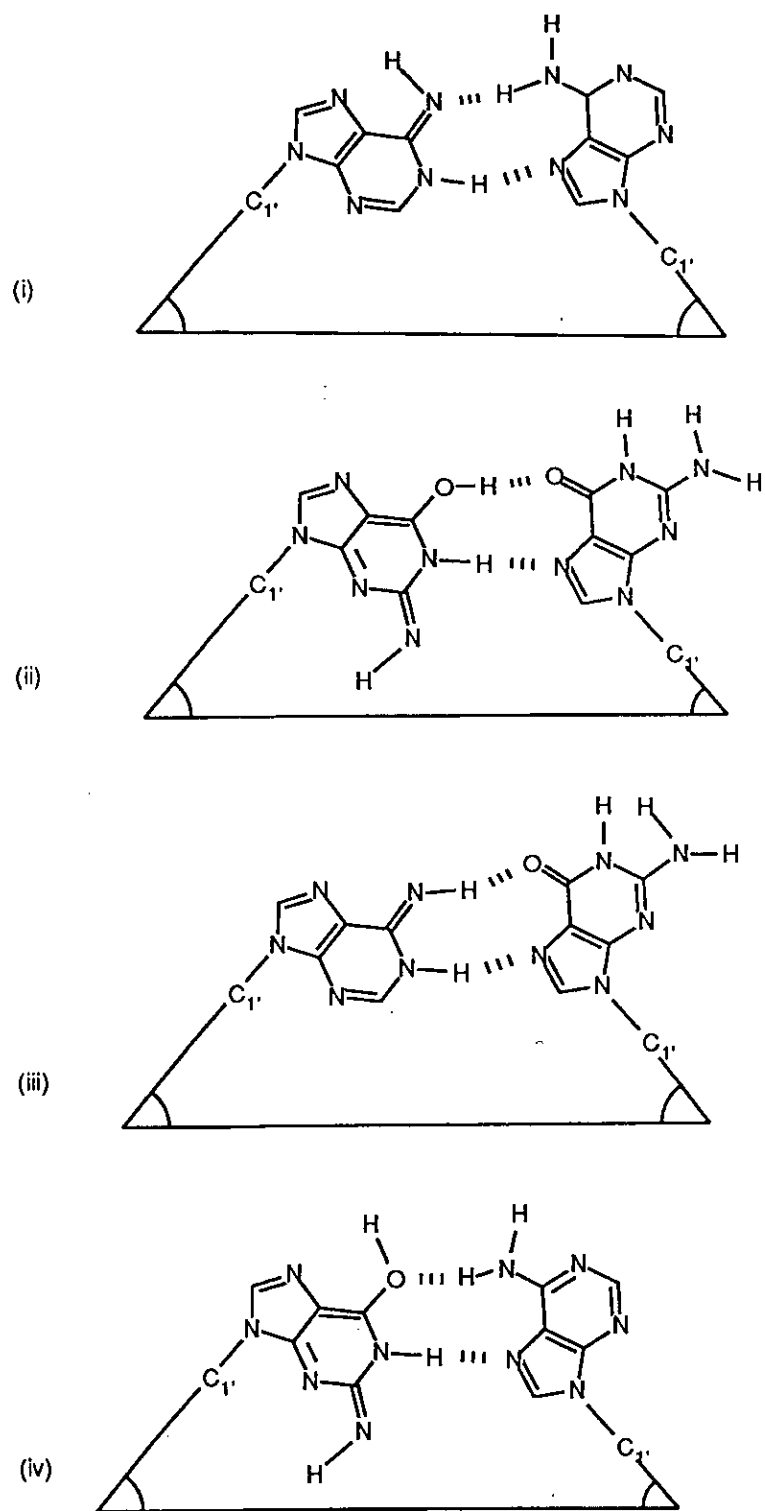


Fig. 3a. (i) A.A, (ii) G.G, (iii) A.G and (iv) G.A mismatches in minor tautomer forms.

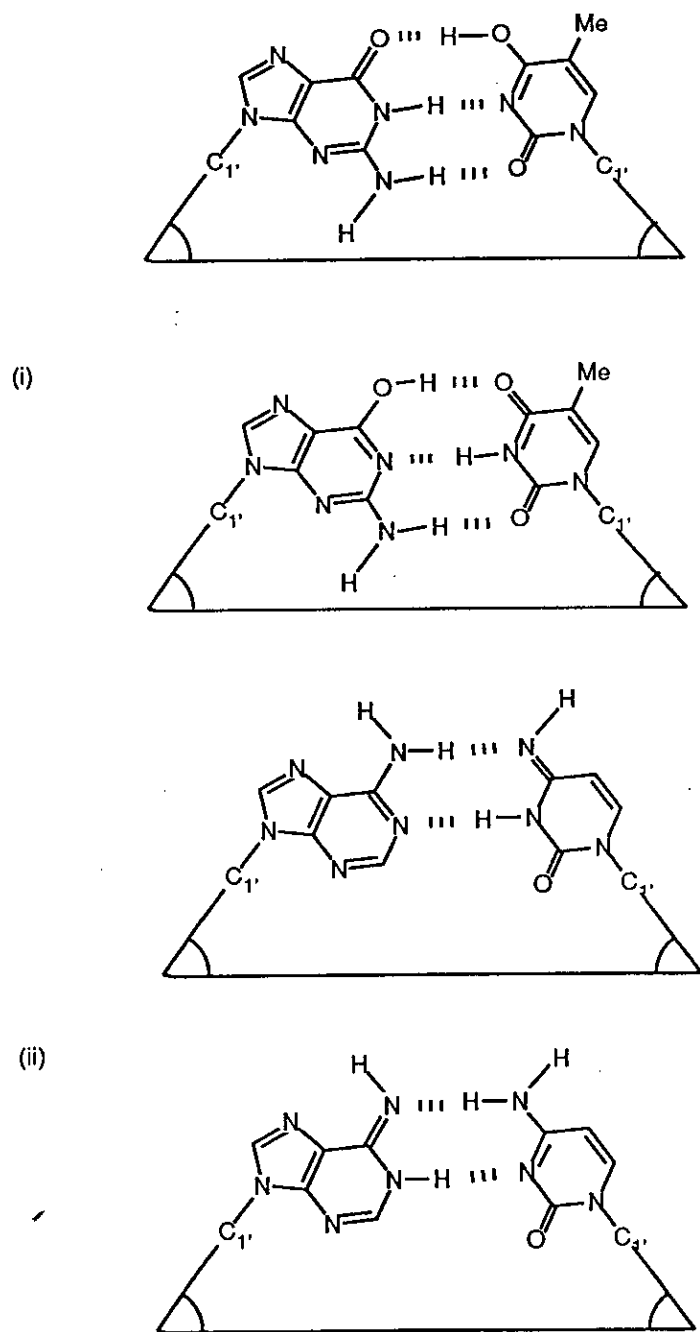


Fig. 3b. (i) G.T and (ii) A.C mismatches in minor tautomer forms.

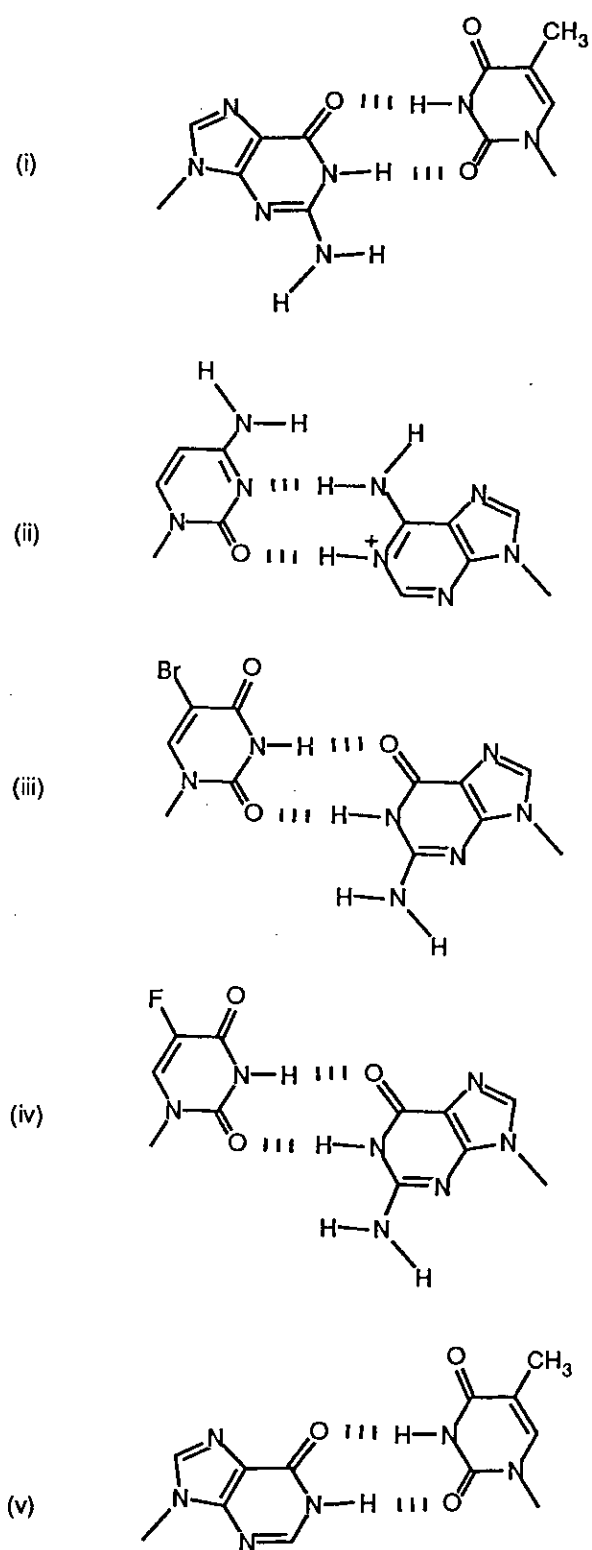


Fig. 4. Wobble base pairing, (i) G.T (ii) C.A (iii) BrU.G (iv) FU.G (v) I.T.

In the case of purine.purine pairing, which has the effect in normal Watson-Crick geometry of increasing the C1'-C1' distance, one needs also to consider the possibility of deviations from normal *anti.anti* base orientation with respect to the sugars. The C1'-C1' separation can be reduced if one of the purines adopts a *syn* conformation<sup>6</sup> and Hoogsteen type base pairing takes place.

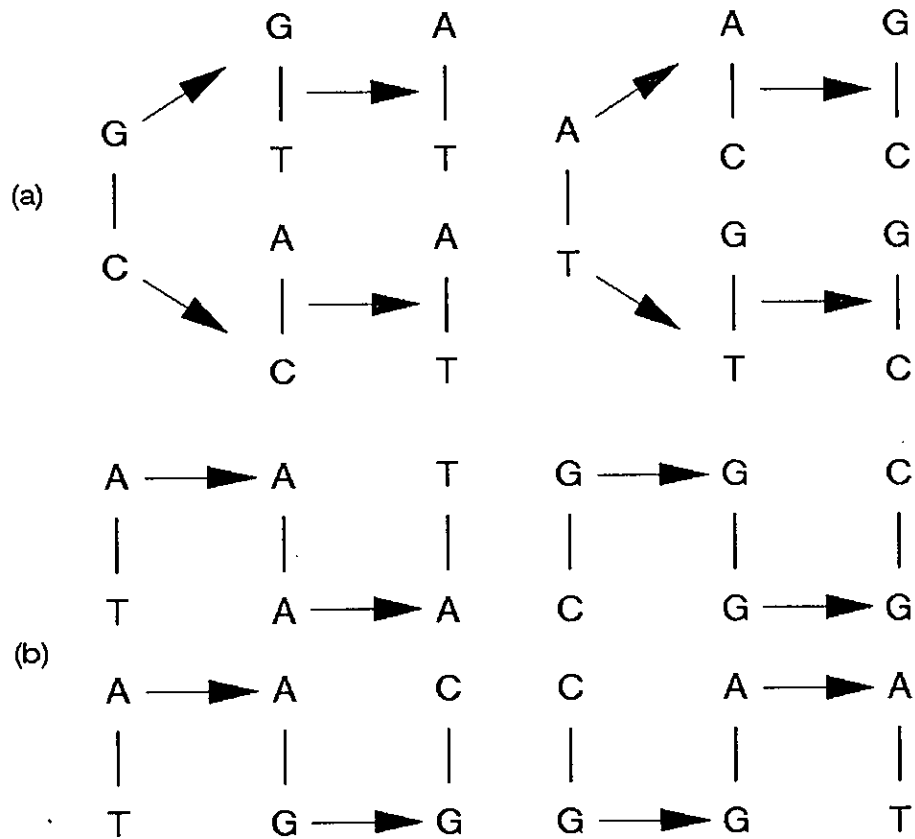


Fig. 5. Mutational pathways (a) Transitions (b) Transversions.

Mutations can be divided into two subsets, transitions and transversions Fig. 5. Transition mutations involve mispairing between purine.pyrimidine pairs giving the mispairs A.C and G.T. Transversion mutations involve a wider range of mismatches of purine.purine or pyrimidine.pyrimidine type giving the possible mispairs A.G, A.A, G.G, C.T, C.C and T.T.

In order for a mismatch to be repaired, it has first of all to be recognised as such. At the very least repair enzymes must be able to differentiate between Watson-Crick and non Watson-Crick pairings. The mechanisms by which base pair mismatches are recognised and repaired are not yet fully understood but may depend on a range of structural or thermodynamic factors or any combination of the two<sup>20</sup>. It is therefore important to examine mismatch repair. One of the more convenient examples to study is the post replicative repair system of *Escherichia Coli* (*E. Coli*). Consequently it has been the focus of much attention<sup>21-25</sup> and has provided insights into aspects of mismatch recognition.

It is possible for a mismatch to be repaired in either of two directions. Either the daughter strand can be repaired to give the complement of the template strand, or the template strand can be repaired to be the complement of the daughter strand. It is thought that the initial lack of methylation on the daughter strand just after replication is used as a signal to indicate that it is the one to be repaired<sup>26, 27</sup>. A series of studies has been carried out and this has led to the finding that different mismatches are repaired with differing efficiencies. This is hardly surprising since an enzymic repair system would not be expected to recognise purine.purine, purine.pyrimidine and pyrimidine.pyrimidine pairings equally well. All of these could reasonably be expected to have different conformational properties. It has been noted<sup>21</sup> that in *Streptococcus pneumoniae* the A.G and C.C mismatches are not repaired. Similarly Kramer *et al*<sup>22</sup> in experiments on the *E. Coli* repair system identified three efficiency classes for mismatch repair, high (T.G, C.A and G.G); intermediate (A.A) and low (G.A, A.G, T.T, C.T and T.C). In an examination of the repair efficiency of G.T and G.A mismatches Fazakerly *et al*<sup>24</sup> examined their rates of repair in a series of undecamers and used NMR techniques in an attempt to understand why the G.A mismatch should be less well repaired than the G.T mismatch. Their finding was that the G.A repair rate depended on whether the mismatch formed an intrahelical wobble pair or was looped out. Only the former was repaired. In an examination of all eight mismatches Dohet *et al*<sup>23</sup> supported their general findings and identified the A.G, C.T and C.C mismatches as being poorly repaired. However these studies are looking at repair classes in different environments. In order to assess

differential repair efficiencies more accurately a more systematic approach is required. Recently rates of repair of the eight different mismatches have been investigated where they are all located in the same position and the same environment in heteroduplexes<sup>25</sup>. Again it was found that repair efficiency was stratified into; high G.T, intermediate A.C, C.T, A.A, T.T and G.G, and low C.C and A.G. Further the A.G mismatch under went different repair routes to give either A.T or C.G pairs.

All these studies agree to the extent that they confirm different mismatches are repaired with a range of efficiencies. Noteable within this range is the A.G mismatch which is consistently poorly repaired. Why should this be so ? While one study<sup>24</sup> speculates on the structure of the mispairing a full structural analysis of the mispair is obviously required.

### **3.0 Structural Studies on Mismatched Base Pairs**

The examination of mismatched bases in synthetic oligonucleotides has become possible since the introduction of methods for the chemical synthesis of milligram quantities of DNA fragments of a precise base sequence and high purity using either the phosphotriester<sup>28</sup> or the more recent phosphoramidite<sup>29</sup> methods. The physical methods available for the study of DNA duplexes include X-ray crystallography of DNA single crystals, nuclear magnetic resonance spectroscopy, ultraviolet thermal denaturation studies (UV-melting) and circular dichroism. Use may also be made of biological structural probes. Of the physical methods, X-ray crystallography is the most powerful, when conditions are favourable for the preparation of DNA crystals, while the others being solution methods, do not have to rely upon the vagaries of crystal growth. X-ray crystallography is inherently difficult as it relies on the preparation of ordered single crystals of a size and quality good enough to allow the collection of high resolution data. It is often very difficult to grow good DNA crystals but the chances of obtaining them can be improved by using what has come to be known as crystal engineering.

In this one selects as a starting point a suitable parent sequence that is known to crystallise and diffract well. Analogues of the sequence are then made containing the

mismatch of interest. The intention is that the insertion of the mismatch does not upset the overall structure of the helix to such an extent that crystallisation will no longer occur. Indeed this is more optimistic than it might at first appear since energy minimisation calculations have shown that the sugar phosphate backbones of sequences containing mismatches are not greatly perturbed by their incorporation<sup>30</sup>. With this sort of strategy the chances of obtaining good quality crystals are greatly improved. This method was used to grow crystals of the first mismatch to be analysed by X-ray crystallography of a DNA single crystal<sup>10</sup>. The sequence dGGGGCTCC was crystallised as a modified octamer containing two symmetry related G.T mismatches<sup>31</sup>. The method has since been used for other mismatches with other parent sequences and has the dual advantages of increasing crystallisation probability and allowing one to look at a variety of mismatches in the same base stacking environment.

### 3.1 The A.G Mismatch

As previously mentioned, work on the postreplicative repair system of *E. Coli* has identified a series of efficiency classes for mismatch repair<sup>21-25</sup>. The A.G mismatch is noted to be consistently poorly repaired. In addition to this, experiments with DNA polymerase III have revealed another situation in which the A.G mismatch is not well repaired. In proofreading of the newly formed DNA, 10% of A.G mismatches escape repair while only 0.5% of G.T mismatches go undetected<sup>32</sup>. In an attempt to understand why the A.G mismatch is so poorly repaired, a large amount of effort has been directed towards identifying the structural features of the mismatch.

Prior to work on synthetic oligonucleotides, the A.G mismatch had been identified as being present in some tRNAs. In an X-ray crystallographic analysis of yeast tRNA<sup>Phe</sup> Rich *et al* identified a hydrogen bonding interaction between adenine 44 and N<sup>2</sup>-dimethylguanine 26<sup>33</sup>. It was not clear if this involved one or two hydrogen bonds. After considering a *syn* orientation for A44<sup>34</sup> further work defined it to be in an *anti* orientation<sup>35, 36</sup>, so in this instance both bases were incorporated as their major tautomers and were in an anti.anti orientation Fig. 6b. The base pair was held together by two hydrogen bonds between

NH<sub>2</sub>(6) A and O(6) G and N(1) A and N(1)H G. A later study which reappraised several earlier structures of *E. Coli* and related ribosomal RNAs identified many possible G.A or A.G base pairing sites<sup>37</sup>. Hence for RNA the concept of A.G base pairing or that of A with modified G is well established. Also known is the base pair between A and O<sup>6</sup>MeG which has been examined in DNA by NMR<sup>38</sup>. On the basis of weak NOEs between purine H8 and sugar H1' protons these were determined to be in an *anti.anti* orientation Fig. 6b.

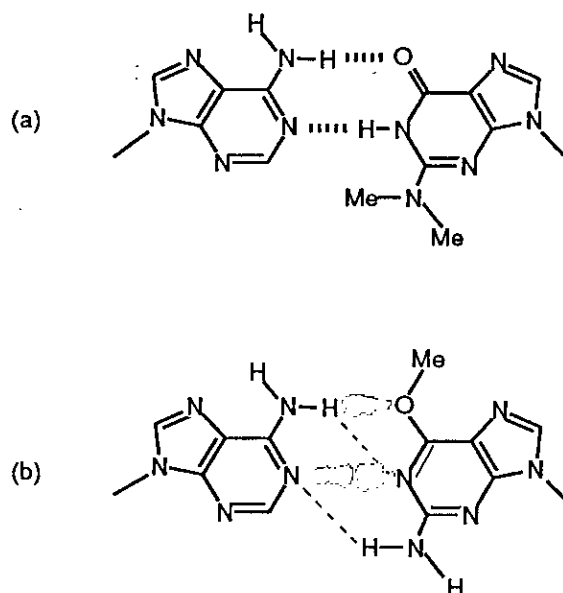


Fig. 6. (a) A.Me<sub>2</sub>G (b) A.O<sup>6</sup>MeG mismatches.

In their proposed mechanism for substitution mutations Topal and Fresco<sup>6</sup> set two conditions, one of which was that there exists more complementary base pairing than just the Watson-Crick pairings A.T and G.C and secondly that these extra base pairs fit the steric constraints of a double helix. For a purine.pyrimidine mismatch the actual size of the bases is not sufficient to cause distortion of the backbone. However, when contemplating purine.purine mismatches, consideration needs to be given to the relative orientations of the bases. If both purines lie in an *anti* orientation with respect to the sugar then the total C1'-C1' separation will be greater than if either of them lies in a *syn* orientation. The authors

suggested the use of minor tautomers and proposed that for steric reasons one base may lie *syn*.

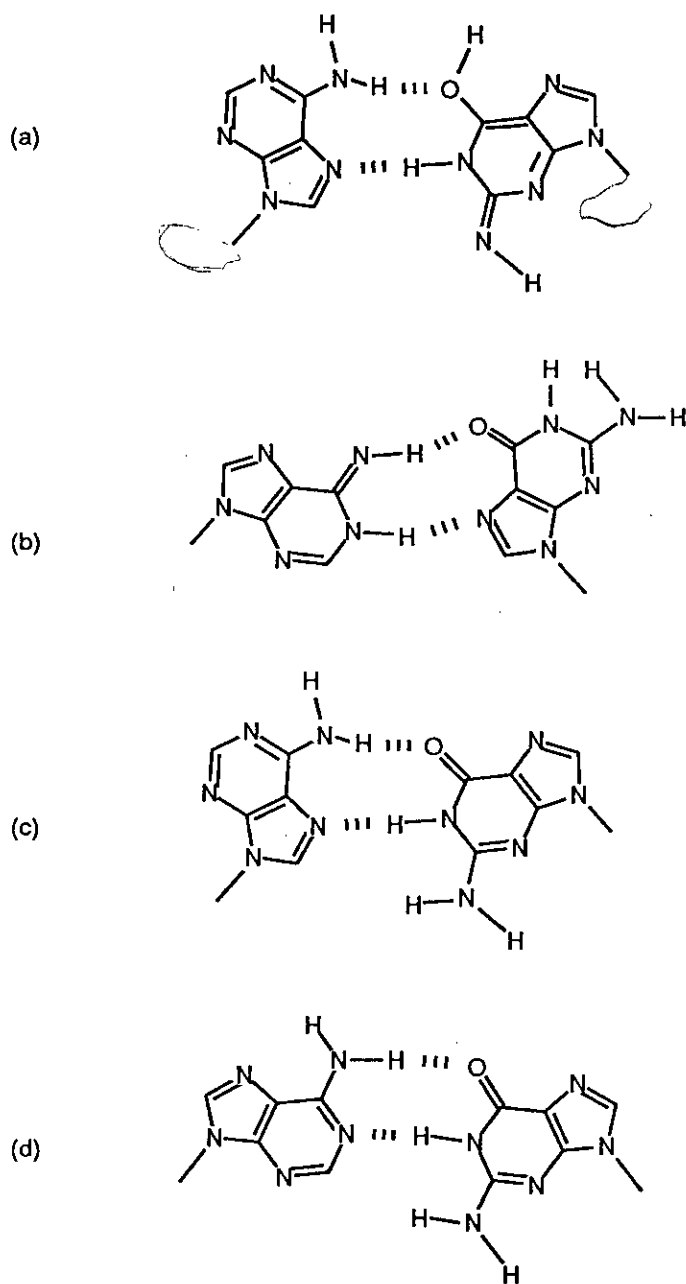


Fig. 7. Proposed minor tautomer forms of the A.G base pair. See text for details.

There are thus four possible conformations for a A.G base pair ;

Fig 7 (a) A (*syn* ).G (enol, imino) (b) A (*anti* , imino).G (*syn* ) (c) A (*syn* ).G (*anti* ) and (d) A (*anti* ).G (*anti* ).

In an attempt to resolve which of the possible base pairing mechanisms is utilised there have been several structural investigations of the A.G mismatch <sup>9, 39-45</sup>. The sequences studied are listed below

KAN <i>et al</i>	CCAAGATTGG	1983
PATEL <i>et al</i>	CGAGAATTCGCG	1984
HUNTER <i>et al</i>	CGCGAATTAGCG	1985
PRIVÉ <i>et al</i>	CCAAGATTGG	1987

The first of these studies by Kan *et al* <sup>39</sup> used NMR to examine the duplex formed by the partially self complementary decamer dCCAAGATTGG which contains a central pair of A.G mismatches. By the use of NOEs they observed that irradiation of the H(1) of G hydrogen bonded in the mismatch produced a negative NOE at the H(2) of A6 indicating that the NH of G is close to H(2) A. This can only be so if the two bases are *anti*. Hence the conclusion that the decamer contained a G (*anti* ).A (*anti* ) base pair.

An examination of the solution structure of the duplex formed by the dodecamer dCGAGAATTCGCG <sup>9, 40</sup>, an analogue of the known B-DNA dodecamer dCGCGAATTCGCG <sup>46-49</sup>, was performed by Patel and coworkers. The duplex contains two symmetrically displaced A.G mismatches three bases in from either end. They had previously determined the structure of the parent sequence by NMR and the use of NOEs <sup>50</sup> and this was in agreement with the known crystal structure. The orientation of the base pairs was here again identified as A (*anti* ).G (*anti* ) on the basis of observed NOEs. The adenine at position 3 was established as *anti* by observation of inter base NOEs at the adenine H(2) on saturation of the imino protons in the neighbouring G.C2 and G.C4 base pairs indicating that the adenosine H(2) is directed into the duplex and

stacked over the adjacent G.C pairs. Also intra base pair NOEs show both the G and the A in the base pair to be *anti*. Irradiation of the base pair imino proton produced an NOE at the adenine H(2) and nowhere else, indicating an *anti.anti* conformation for if G were *syn* then H(8) G would also show an NOE.

Thus the first two studies on the structure of the A.G mismatch in DNA as opposed to RNA determined the bases to be *anti.anti*. The first solid state structure determination of the mismatch came with the solution of the crystal structure of dCGCGAATTAGCG<sup>41-43</sup>. In this structure the bases were assumed to be in an *anti.anti* conformation but further refinement of the structure gave a much better fit to the electron density maps if the bases were allowed to be G (*anti*).A (*syn*). This was the first time that a *syn* orientation for a base had been observed in an oligonucleotide crystal structure other than in Z-DNA<sup>51</sup> or in a DNA triostin-A complex<sup>52</sup>. Allowing the adenine to lie *syn* reduces the C1'-C1' distance to a value comparable to the parent structure. The study also showed that it was not necessary to use the G (enol-imino) minor tautomer to allow G (*anti*).A (*syn*) base pairing to take place. The base pair could form using the major tautomers. In solving the X-ray crystal structure<sup>44</sup> of the decamer dCCAAGATTGG which had previously been examined by NMR, Privé and coworkers determined the mismatched bases to be in an *anti.anti* conformation in agreement with the earlier NMR results<sup>39</sup> but at odds with the dodecamer results<sup>41-43</sup>.

There was therefore some debate as to the conformational flexibility of the A.G base pair. We undertook to examine the mismatch in the dodecamer sequence dCGCAAATTGGCG which is the analogue of the previously studied dodecamer dCGCGAATTAGCG created by reversing the mismatch positions. While this work was in progress another structure of the A.G mismatch was published<sup>45</sup>. An NMR study of the dodecamers dCGGGAATTCACG and dCGAGAATTCGCG showed that the conformation of the mismatched bases could vary depending on the pH at which the structure was determined. The authors identified two quite different base pairs. At neutral pH the bases were thought to be in the G (*anti*).A (*anti*) orientation, while at acidic pH

they were found to be G (*syn* ).A<sup>+</sup> (*anti* ). This demonstrated that the earlier determinations of the base pair structure had not shown the full picture with regard to the flexibility of the system.

### 3.1.1 The A.G Mismatch: Crystallisation Data Collection and Refinement

It had been undertaken to examine the A.G mismatch in the sequence dCGCAAATTGGCG derived from the native dodecamer dCGCGAATTCGCG by the insertion of the mismatch bases at positions 4 and 9. This should give a symmetrical self complementary duplex with two related A.G mismatches between bases A4.G21 and G9.A16. The sequence numbering is shown in Fig. 8. This sequence is related to the previously crystallised A.G mismatch dCGCGAATTAGCG<sup>41-43</sup> by reversing the positions of the mismatched bases.

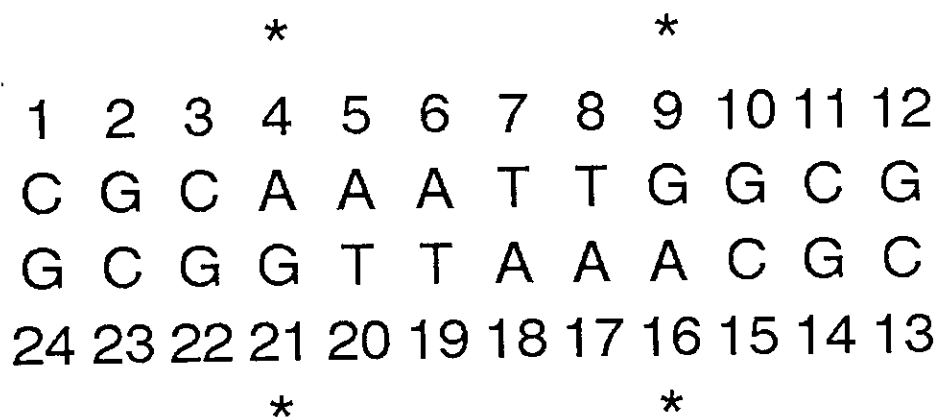


Fig. 8. A.G mismatch sequence numbering, mismatch positions indicated by \*.

The oligonucleotide dCGCAAATTGGCG was synthesised using standard phosphoramidite methods on an Applied Biosystems 380B DNA Synthesiser. In order to obtain sufficient material for crystallisations six 1µmole scale syntheses were performed and the products combined. Rigorous HPLC purification was carried out as described in the experimental section using Method 2. This produced about 10mg of pure lyophilised material. After a degree of iteration good crystallisation conditions were obtained. Crystals were grown using the sitting drop method at 4°C from a solution of the oligonucleotide

(0.5mM) in sodium cacodylate buffer (10mM pH 6.6), magnesium chloride (20mM), spermine tetrahydrochloride (1.0mM) and 2-methyl-2, 4-pentanediol (18% vol/vol). Crystals appeared over the course of several weeks. A suitable crystal approximately 2.0mm x 0.4mm x 0.3mm was mounted inside a sealed Lindemann capillary. The orthorhombic crystal had a space group of  $P2_12_12_1$  and unit cell dimensions of  $a = 25.23\text{\AA}$ ,  $b = 41.16\text{\AA}$  and  $c = 65.01\text{\AA}$ . Data was collected at  $4^\circ\text{C}$  on a Stoe-Siemens AED2 four circle diffractometer, fitted with a long arm and a helium path, using  $\text{CuK}\text{-}\alpha$  radiation. 2840 unique reflections with  $F \geq \sigma(F)$  to a resolution of  $2.25\text{\AA}$  were obtained. From the unit cell dimensions it was apparent the structure was quasi-isomorphous with the native B-DNA dodecamer dCGCGAATTCGCG<sup>46</sup>. Thus the native coordinates were used as the starting model in the structure refinement. Two refinements were performed<sup>53, 54</sup>. During the course of the structure refinements to avoid artificially biasing the results, the atoms of nucleotides G4, C9, G16 and C21 corresponding to the mismatch positions were given very low occupancy effectively omitting them from the structure factor calculations. This enabled the mismatched bases to move freely with respect to each other. As the refinement progressed an Evans and Sutherland PS300 graphics system with FRODO<sup>55</sup> was used to examine the  $2.5\text{\AA}$  electron density ( $2F_o - F_c$ ) and difference ( $F_o - F_c$ ) maps. The mismatch bases were manipulated to give the best fit to the maps and the refinement continued with the mismatched nucleotides given full occupancy. At this point it was already noticeable the A.G mispairs gave a significantly better fit to the density calculated in their absence in an A (*anti*).G (*syn*) orientation than A (*anti*).G (*anti*) or A (*syn*).G (*anti*) orientations. The refinement finally converged after the location of 62 solvent molecules with a crystallographic residual of  $R = 0.167$  for 2212 reflections with  $F \geq 2.0\sigma(F)$  in the region  $7.0\text{\AA}$  to  $2.25\text{\AA}$ .

### 3.1.2 A.G Mismatch Conformation

The overall form of the A.G dodecamer is the same as the native dodecamer Fig. 9a, b. From the  $2F_o - F_c$  and  $F_o - F_c$  fragment maps it is obvious the mismatch has an A (*anti*).G

(*syn*) conformation. Both the A4.G21 and G9.A16 mismatches have the same form. The base pairs have two good interbase hydrogen bonding distances of N(6) A to O(6) G  $2.6\text{\AA}$  and N(1) A to N(7) G of  $2.8\text{\AA}$  Fig. 10. Average C1'-C1' separation was found as  $10.8\text{\AA}$  about  $2\text{\AA}$  less than in an A (*anti*).G (*anti*) base pair<sup>38</sup> with an average propeller twist of  $15.5^\circ$ . These compare favourably with the native values of C1'-C1' separation  $10.5\text{\AA}$  and average propeller twist at the mismatch positions of  $15.9^\circ$ <sup>47</sup>.

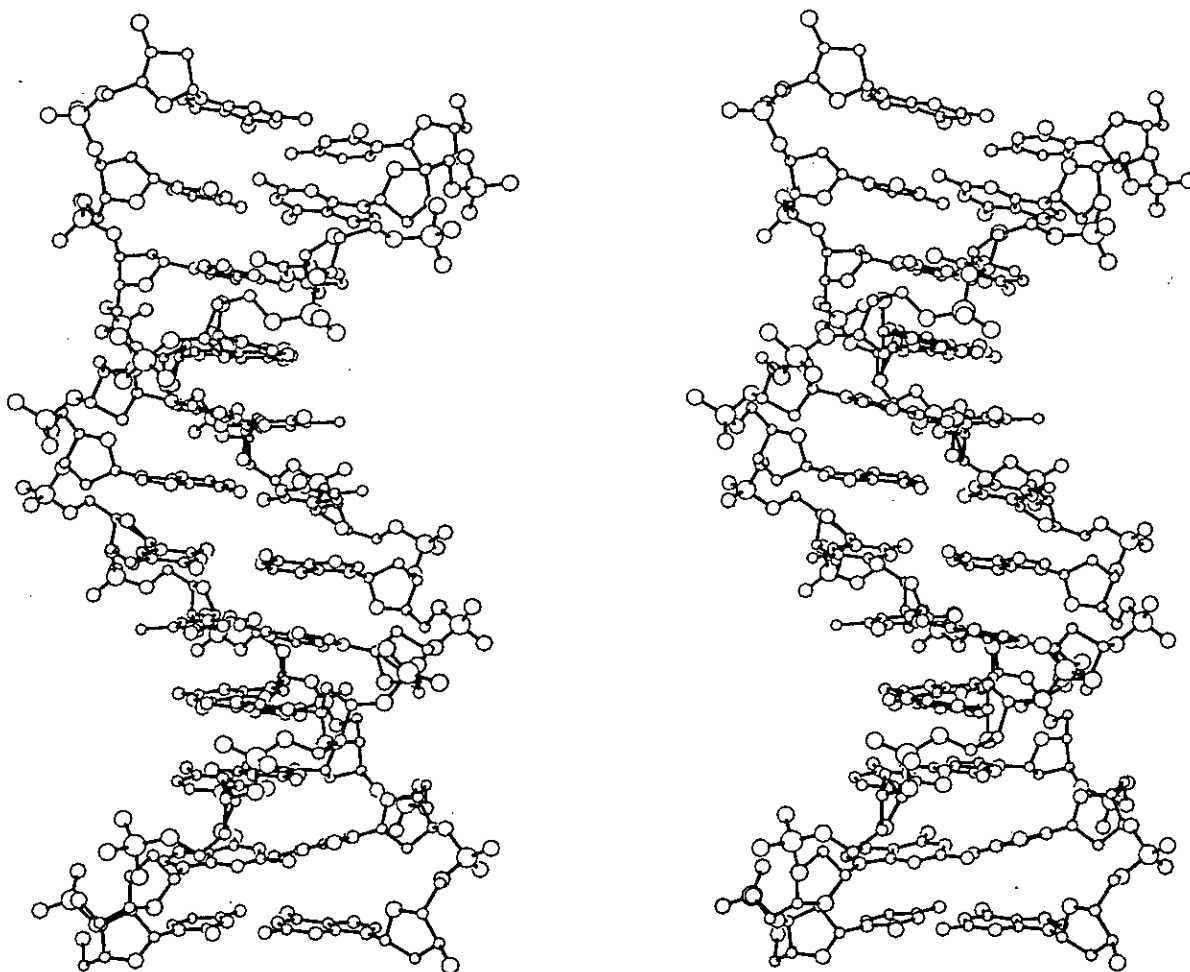


Fig. 9a. A.G 12mer duplex.

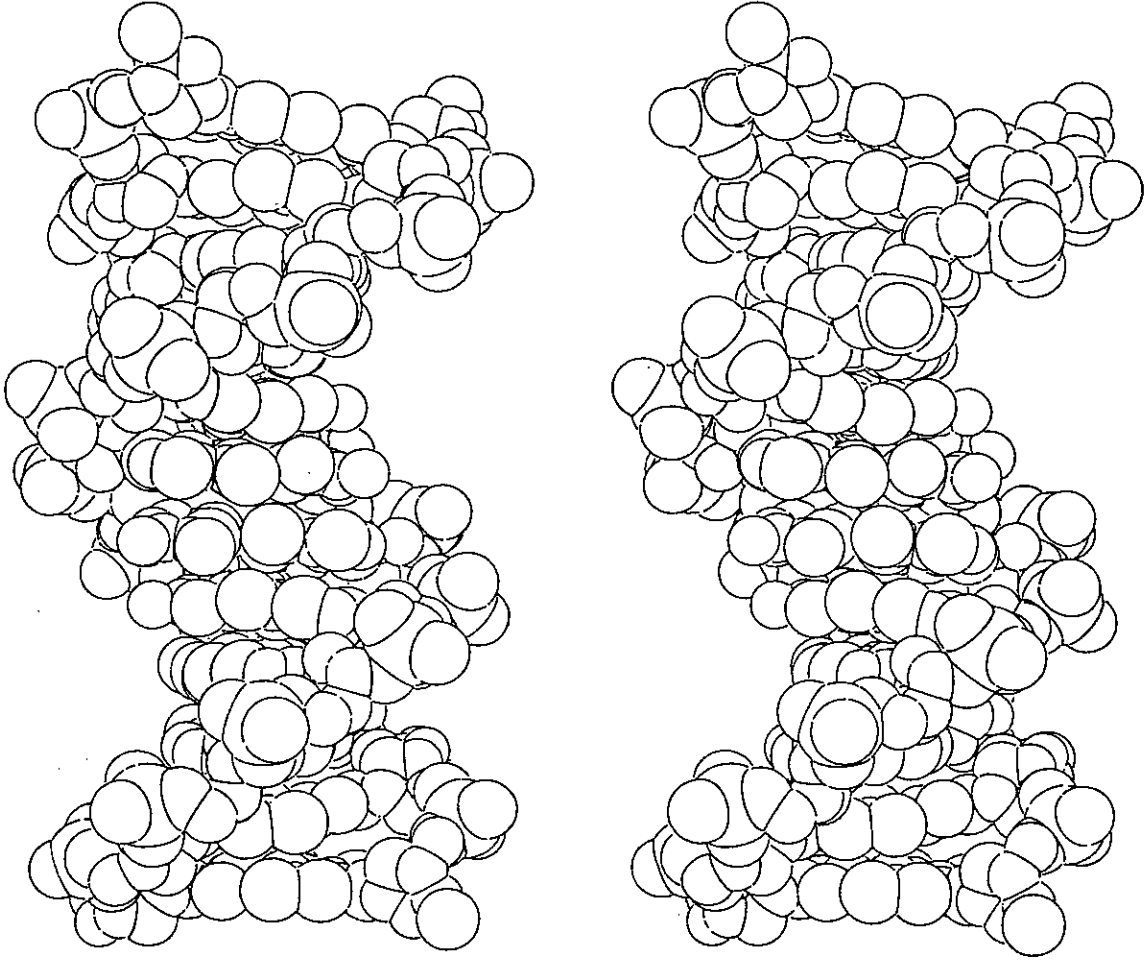


Fig. 9b. A.G 12mer duplex space filled.

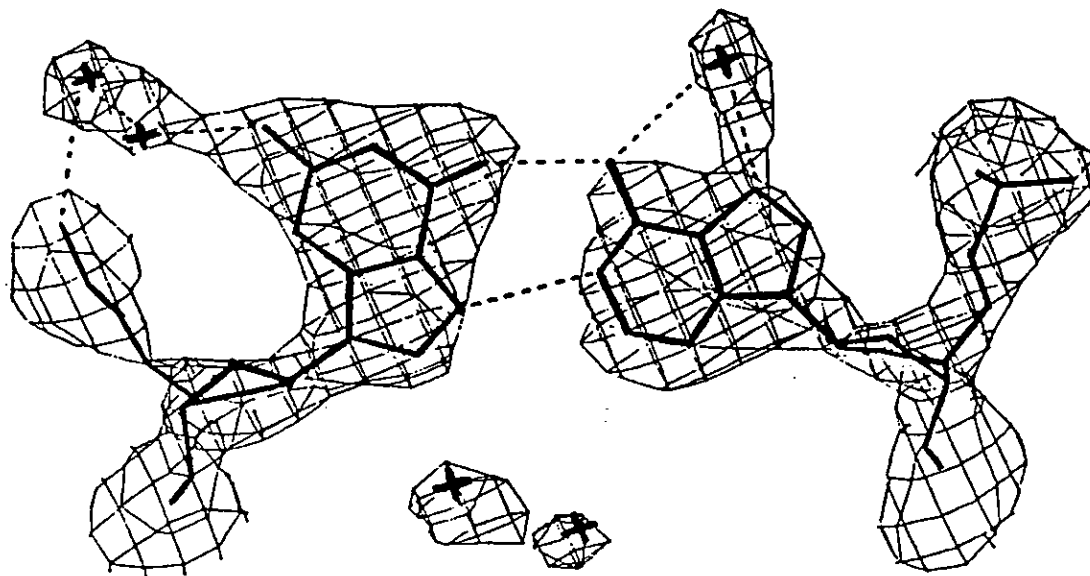


Fig. 10. The G9.A16 base pair 2Fo-Fc map. Dotted lines indicate hydrogen bonding between bases and with major groove solvent molecules.

### 3.1.3 Structure of the A.G Mismatch

The two interbase distances noted above are indicative of strong hydrogen bonding interactions. However if both bases are present in their major tautomeric forms two hydrogen bonds cannot form without either proton capture, in which a proton is shared, or base protonation taking place. Protonation of adenine may be expected to take place quite readily Fig. 11. The pKa of the N(1) atom of free adenine is about 4.5<sup>56</sup> although it should be noted that this may change in an oligonucleotide. Alternatively a base pair could be formed by a minor tautomer of adenine. Base pairings involving the minor imino tautomer of adenine have been proposed<sup>6</sup> but minor tautomers are thought to occur only rarely. The resolution to which the structure has been refined does not allow the location of hydrogen atoms. It is therefore not possible to tell from the crystal structure alone which of the possible base pairing mechanisms is being utilised.

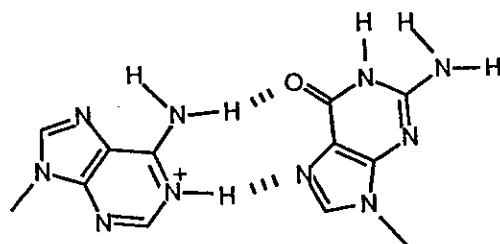


Fig. 11. Protonated  $AH^+$  (*anti*).G (*syn*) base pair.

The question of what type of base pair exists can be resolved by the use of UV thermal denaturation experiments performed over a wide pH range. A series of pH dependant UV thermal denaturation experiments has been performed<sup>53</sup> in the pH range 4.5 to 8.0. These indicate the existence of a protonated base pair of the form  $AH^+$  (*anti*).G (*syn*). From the graph it can be seen the stability of the A.G sequence increases sharply below pH 6.5 reaching a maximum  $T_m$  value of 311.5K at pH 5.2 Fig. 12. For comparison the behaviour of the native sequence is also shown. In this case the stability of the duplex decreases rapidly below pH 6.5. It should be noted that another protonated base pair is possible. Protonation of the guanine would give an A (*anti*). $GH^+$  (*syn*) base pair Fig. 13. In contrast to the  $AH^+$  (*anti*).G (*syn*) base pair in which both hydrogen donors/acceptors are on the same base, the A (*anti*). $GH^+$  (*syn*) conformation has the preferable cyclic hydrogen bonding arrangement. However this structure does not agree with the pH dependant experiments. The  $pK_a$  of guanine ( $pK_a = 3.2$ ) is lower than adenine. In addition pH dependant NMR studies appear to support adenine protonation<sup>45</sup>. Gao and Patel found in the acidic form of the mismatch the non exchangeable adenine H(8) resonates at 8.53ppm as opposed to 8.28ppm under neutral conditions. The downfield shift is assumed to be due to ring protonation<sup>57</sup>. A protonated guanine was thought less likely due to the neutral and acidic condition H(8) G resonances of 7.90ppm and 7.56ppm respectively. Further support for proposing a protonated  $AH^+$  species comes from a comparison to the A.C mismatch. X-ray

crystallographic analysis of the sequence dCGCAAATTCGCG identified an A.C wobble base pair <sup>14</sup>. Again pH dependant UV experiments have been used to decide on the structure of

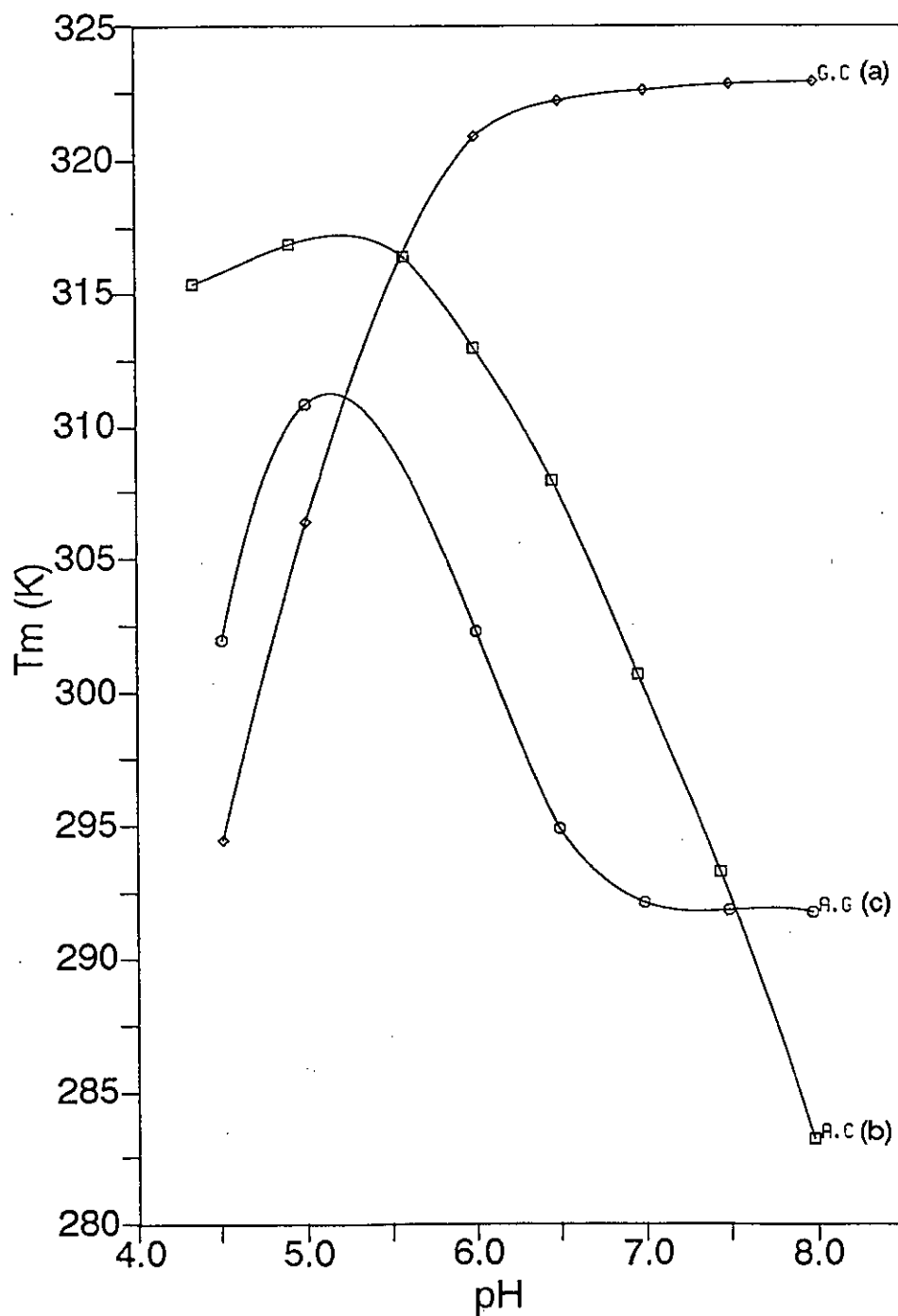


Fig. 12. Plots of  $T_m$  versus pH measured in 0.1M sodium phosphate buffer for duplexes (a) dCGCGAATTCGCG (b) dCGCAAATTCGCG (c) dCGCAAATTGGCG.

the base pair <sup>58</sup> Fig. 12. Three possible hydrogen bonding patterns could be envisaged Fig. 14. There is a clear correlation between pH and T<sub>m</sub> value over the pH range examined. The duplex has a very low stability at high pH T<sub>m</sub> = 284K. Upon going towards an acidic pH the duplex stability steadily increases reaching a maximum at pH 5.3 T<sub>m</sub> = 317K. This is strongly suggestive of an AH<sup>+</sup>.C base pair. In both cases the proposed AH<sup>+</sup> species is in the same base stacking environment.

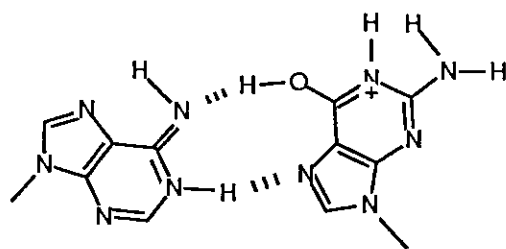


Fig. 13. Protonated A (*anti*).GH<sup>+</sup> (*syn*) base pair.

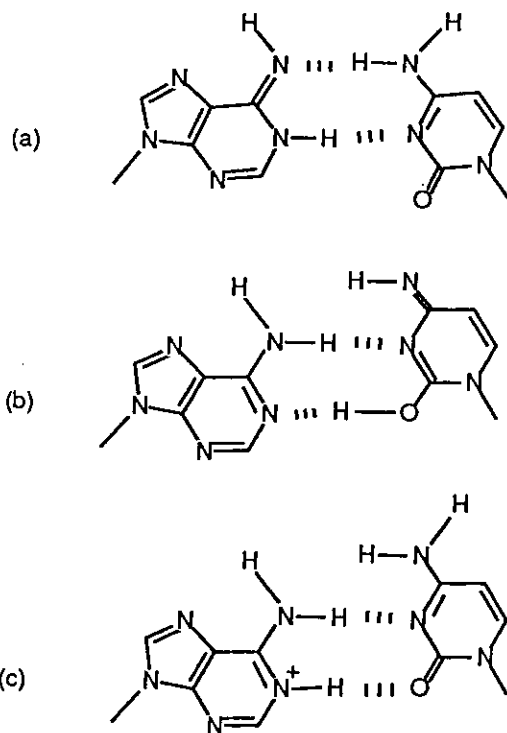


Fig. 14. Proposed forms of the A.C mismatch (a) A (imino).C (b) A.C (enol, imino) (c) AH<sup>+</sup>.C

The 2-amino group of the guanine in the AH<sup>+</sup> (*anti*).G (*syn*) base pair lies in the major groove unlike the other forms of the mismatch. In this position it is able to hydrogen bond to two water molecules. This cannot happen in a G (*anti*).A (*syn*) or (*anti*) base pair where it is prevented by steric hinderance in the crowded minor groove from achieving its full hydrogen bonding potential with the surrounding water molecules. In these base pairs there is a net loss of hydrogen bonding on going from single strand to duplex forms.

### 3.1.4 Conformational Flexibility of the A.G Mismatch

As stated there have been a number of investigations of the A.G mismatch by NMR and crystallographic methods<sup>9, 39-44</sup>. A recent NMR study of the sequence dCGGGAATTCACG showed the mismatch conformation to vary with pH being A (*anti*).G (*anti*) at neutral pH and AH<sup>+</sup> (*anti*).G (*syn*) at acidic pH. The present sequence dCGCAAATTGGCG has also been found to show a pH dependence in its NMR. At low pH an A (*anti*).G (*syn*) base pair was found while at high pH the bases are A (*anti*).G (*anti*)<sup>59</sup>. Comparing the A (*anti*).G (*syn*) base pair found in the present study sequence dCGCAAATTGGCG to the mismatch structures found in other sequences gives an insight into the sequence dependant nature of the A.G mismatch conformation. As an example of this sequence dependance the mismatch in the earlier X-ray analysis of dCGCGAATTAGCG in which the mismatched bases are transposed was found to be A (*syn*).G (*anti*)<sup>42, 43</sup>.

In the two sequences which have supported A (*anti*).G (*syn*) base pairs namely dCGCAAATTGGCG and dCGGGAATTCACG the guanine at the mismatch site is flanked by the guanine or guanines of normal G.C base pairs. The dipole moment of guanine is large, 7.5D<sup>60</sup>, and a GpG step in B-DNA with G (*anti*) bases might be expected to have an unfavourable base stacking interaction. In a B-DNA structure the dipole moments of the two bases would be similarly oriented with a relative rotation of about 33°. If the guanine of the mismatch is not constrained to lie *anti* but rather can lie *syn* then this unfavourable interaction should be reduced. In those sequences where the guanine of the mismatch is *anti*, dCGAGAATTCGCG and dCGCGAATTAGCG<sup>42, 43</sup> the guanine environment is different.

Here the guanine is flanked by cytosine and takes part in CpG base stacking. The dipole moment of cytosine is similar in size to that of guanine, 7.6D. Its direction however is  $102^\circ$  as opposed to  $-31^\circ$ . Thus in B-DNA this directional difference combined with the relative base rotation means the two dipoles act in virtually opposite directions. Hence the CpG base stacking interactions should be favourable for a G (*anti*) conformation. It can be proposed the much smaller dipole moment of adenine, 2.9D, will have less effect on the base pair conformation. In conclusion it would appear that for a purine.purine mismatch guanine stacked on cytosine is more stable in an *anti* form.

An indication of the stability to be gained from favourable base stacking interactions CpG (*anti*) and G (*anti*)pC as opposed to G (*anti*)pG steps can be obtained from the  $T_m$  values for the duplexes mentioned above<sup>58</sup>. Where the guanine is involved in two CpG steps, dCGAGAATTCGCG the duplex has a much higher melting temperature  $T_m = 308\text{K}$  than dCGCAAATTGGCG  $T_m = 292\text{K}$  at neutral pH. At acid pH there is no indication of a transition to an  $\text{AH}^+$  (*anti*).G (*syn*) mispair to further stabilise the duplex. It can be deduced the main stabilisation is probably derived from the CpG and GpC base stacking.

### 3.1.5 Mismatch Stabilisation by Bifurcated Hydrogen Bonds and Solvent Interactions

Bifurcated hydrogen bonds can form in A.T tracts when the bases have a high propellor twist<sup>61</sup>. This directs the N(6) of adenine towards the O(4) of a thymine base on its 3' side bringing the atoms within hydrogen bonding distance. Such bifurcated hydrogen bond networks have been identified in the A.T regions of dCGCAAAAAGCG<sup>62</sup> and dCGCAAATTTGCG<sup>63</sup>. The latter oligonucleotide is a native form of the A.G mismatch sequence. In the mismatch sequence a similar bifurcated hydrogen bond network appears to run through the AAATTG region of the major groove due to the relatively high propellor twist of the central A.T base pairs. There are also close contacts between the N(4) amino groups of C3 and C15 and the O(6) atoms of the mismatched guanines on their 3' sides Fig. 15. These additional interactions may assist in stabilising the duplex in the region of the A.G ..... mismatch .

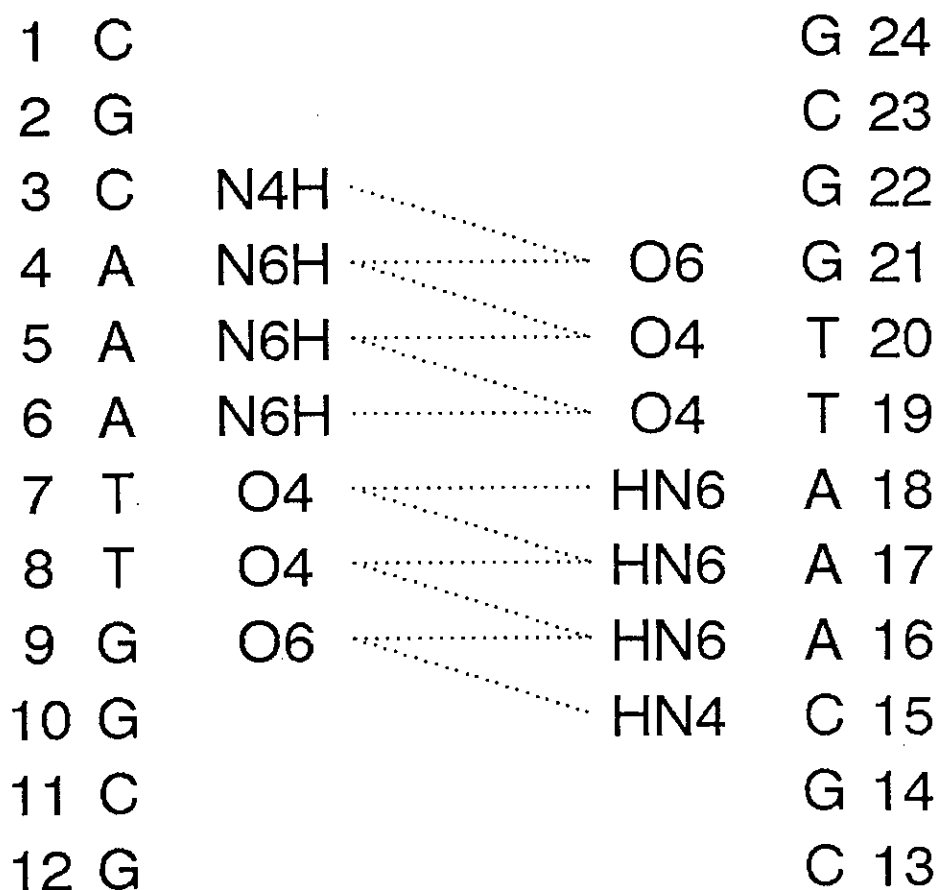


Fig. 15. Bifurcated hydrogen bond network found in the major groove of dCGCAAATTGGCG.

As stated, in the AH<sup>+</sup> (*anti*), G (*syn*) base pair the guanine 2-amino group is directed into the major groove. It can be seen there is a well ordered water structure around the mismatch Fig. 16. The *syn* conformation is stabilised by the presence of water molecules bridging the 2-amino group and its 5'-phosphate. In the minor groove a water molecule bridges the guanine H(8) and the O(4') atom of the deoxyribose sugar of the adjacent 3' cytidine nucleotide. This is not unlike the water structure seen around *syn*-guanine in Z-DNA <sup>64</sup>.

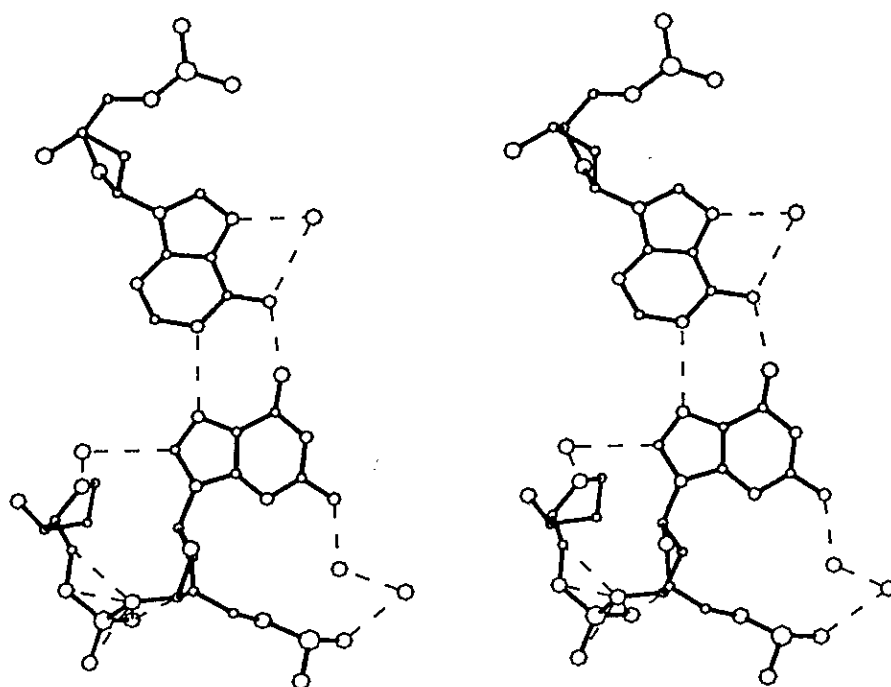


Fig. 16. The stabilising network of water molecules located around the  $AH^+$  (*anti*).G (*syn*) base pair.

### 3.1.6 Thermodynamic Parameters of the A.G Mismatch

The A.G mismatch has been found to be less stable than the equivalent G.C or A.T dodecamers<sup>54</sup>. In 1M NaCl at pH 7.0 they have respective  $\Delta G^\phi$  values of A.G = -10.3, G.C = -19.9 and A.T = -18.1 Kcalmole<sup>-1</sup>. It has also been found that in spite of its better repair efficiency the G.T mismatch was more stable ( $\Delta G^\phi = -12.5$  Kcalmole<sup>-1</sup>) than the A.G mismatch. This suggests mismatch stability is not the only criterion determining repair efficiency. A comparison of the thermodynamic parameters of the A.G and A.C mismatches, both of which contain an  $AH^+$  species, at pH 7.0 and 5.2 shows there is a significant increase in duplex stability with lowering of pH but this is dependant also on the salt concentration.  $\Delta\Delta G^\phi$  (A.G) = -2.2 Kcalmole<sup>-1</sup> in low salt (0.1M) where as at high salt (1.0M) there is no increase in stability with a reduction in pH. The A.C mismatch behaves, differently showing increased stability in acidic pH at both high and low salt

concentrations. This difference in behaviour may be attributed to the conformational flexibility of the A.G base pair, compared to the A.C base pair which has only one (protonated) stable form.

### 3.2 The A.I Mismatch

Inosine is an analogue of guanosine in which the 2-amino group is replaced by hydrogen Fig. 17. It is found naturally in ribonucleic acids where it is present in some tRNA anticodons<sup>65</sup> in which it can form stable base pairs with A, C or U in the mRNA codon. Additionally it has been found to be able to occupy the central anticodon position and pair with A<sup>66</sup>. Its ability to form base pairs with A, C and U has led to its use as a universal base in hybridisation probes for cloning genes for proteins containing degenerate codons<sup>67-69</sup>.

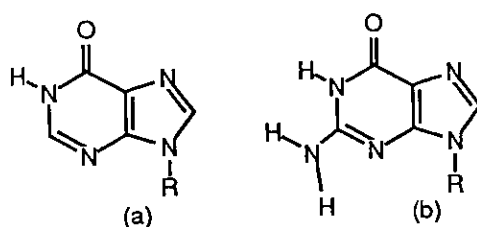


Fig. 17. (a) 2'-Deoxyinosine (b) 2'-deoxyguanosine. R = 2'-deoxyribose.

It has been found previously<sup>66, 68</sup> that inclusion of inosine does not significantly affect the stability of a DNA duplex as it forms stable base pairs although it has been noted that I.C pairs are more stable than I.A, I.G or I.T base pairs<sup>69</sup>. Inosine can be formed by the deamination of adenosine, this may result in a transition mutation<sup>70</sup> via the formation of an I.C base pair to give a G.C base pair. Since it is potentially mutagenic hypoxanthene is efficiently removed by the enzyme hypoxanthene DNA glycolase<sup>71a, b, 72</sup>.

A previous X-ray crystallographic study<sup>73</sup> has examined the structure of the I.A base pair in the B DNA dodecamer sequence dCGCIAATTAGCG. This sequence is the inosine analogue of the previously examined A.G mismatch dodecamer<sup>42</sup>. Solution of the I.A structure showed the bases to be in an I (*anti*).A (*syn*) orientation similar to the G (*anti*).A (*syn*) base pair found before Fig. 18a. This result was different to an earlier NMR study<sup>74</sup> on the sequence dGGIACC which was the first study to show direct

evidence for the formation of I.A base pairs. The authors of this report compared the I.A mismatched sequence to the structure of the oligonucleotide dGGICCC. By use of NOE measurements they concluded that the bases were in an I (*anti*).A (*anti*) conformation Fig. 18b. Additional support for this was offered by analysis of the  $^{31}\text{P}$  NMR spectra which showed the sugar-phosphate backbone of the I.A sequence was distorted to accommodate the larger ( $12.7 \text{ \AA}$ ) C1'-C1' distance of the *anti.anti* base pair.

It was decided to look at the inosine analogue of the A.G mismatch sequence dCGCAAATTGGCG namely dCGCAAATTIGCG. It was hoped this would show what the effects of the 2-amino group and base environment are when compared to the A.G mismatch in the same sequence and base stacking environment.

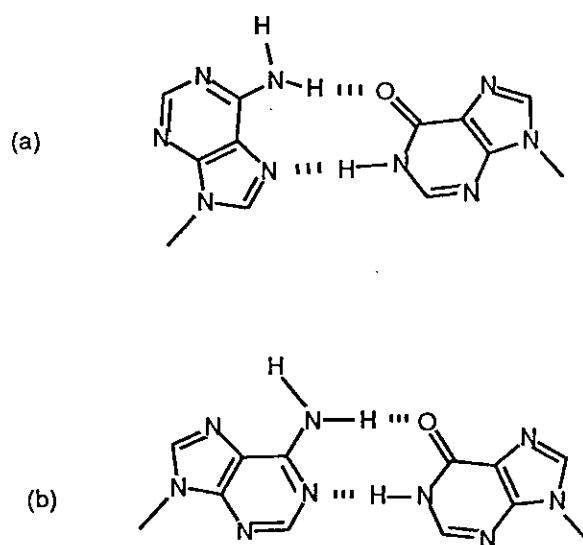


Fig. 18. The A.I base pair (a) A (*syn*).I (*anti*) (b) A (*anti*).I (*anti*)

### 3.2.1 The A.I Mismatch: Crystallisation Data Collection and Refinement

An examination of the sequence dCGCAAATTIGCG containing two symmetrical A.I mismatches in positions 4.21 and 9.16, the same as the A.G mismatch was undertaken Fig. 19. This would provide information on the factors governing the conformational flexibility of the A.G mismatch. In particular the role of major groove water structure on the stabilisation of *syn* guanine.

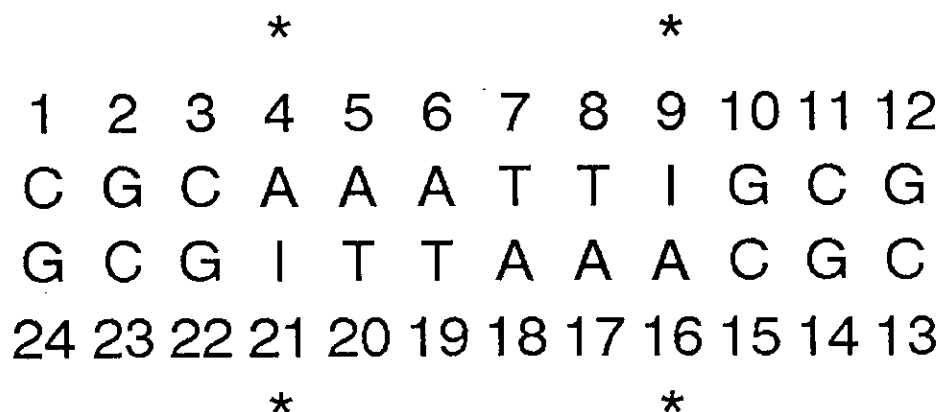


Fig. 19. A.I mismatch sequence numbering, mismatch positions indicated by \*.

The oligonucleotide dCGCAAATTIGCG was synthesised and purified in the same way as the A. G mismatch sequence to give about 10mg of pure lyophilised material. An iterative process was used to determine suitable crystallising conditions. Initial crystallisations used the sitting drop method with central well vapour diffusion. Crystals were grown from a solution containing the oligonucleotide (0.5mM), sodium cacodylate buffer (13mM, pH 6.5), magnesium chloride (40mM), spermine tetrahydrochloride (0.5mM) and 2-methyl-2, 4-pentanediol (15% vol/vol) by vapour diffusion against 2-methyl-2, 4-pentanediol (50%) at 4°C. After six weeks growth a crystal of a suitable size was mounted in a sealed Lindemann capillary. Data was collected in the same way as for the A.G mismatch, again at 4°C. The crystal was orthorhombic with space group  $P2_12_12_1$  and unit cell dimensions of  $a = 25.12\text{Å}$ ,  $b = 41.18\text{Å}$ , and  $c = 65.42\text{Å}$ . Thus as expected the A.I mismatch sequence dCGCAAATTIGCG was quasi-isomorphous with the native dodecamer

dCGCGAATTCGCG<sup>47</sup>

The structure refinement was performed using Konnert-Hendrickson techniques<sup>75</sup> with refined coordinates for the native dodecamer used as the starting model. As before the atoms of nucleotides G4, C9, G16 and C21 were given very low occupancies effectively omitting them from the structure factor calculations allowing the bases to move freely with respect to each other. The electron density (2Fo-Fc) and difference density (Fo-Fc) maps were examined once this part of the refinement had converged with an R factor of R = 0.42. FRODO<sup>55</sup> was used to manipulate the mismatched bases to give the best fit to the maps where upon the atoms of the mismatched A4, I9, A16, and I21 nucleotides were given full occupancy and the refinement continued. Resolution was increased to 2.5Å<sup>o</sup> at which point the refinement converged with a crystallographic residual of R = 0.20 for 1861 reflections with  $F_o \geq 2.0 \sigma(F)$  and the inclusion of 99 solvent molecules. A second data set was collected giving 1745 unique reflections. Merging both data sets gave 2462 reflections with  $F \geq \sigma F$ . The refinement was repeated. This converged at an R factor of 24% with no solvent, and 19.8% with the inclusion of 26 solvent molecules.

### 3.2.2 A.I Mismatch Conformation

From the 2Fo-Fc fragment maps calculated initially in their absence during the first refinement, it appears the mismatch has an A (*anti*).I (*syn*) conformation as the bases gave a much better fit to the available density in these positions. Both the A4.I21 and I9.A16 mispairs have the same overall shape. In the second refinement the 16.9 mismatch was better defined than the 4.21 base pair. However both mismatches gave a good fit to the available density and even in the case of the less well defined 4.21 site an I (*anti*).A (*syn*) mismatch could not be fitted to the density. The base pairs have two interbase hydrogen bonding distances of N(6) A to O(6) I 2.57Å<sup>o</sup> and N(1) A to N(7) I 3.24Å<sup>o</sup>, obtained after constraining hydrogen bond lengths Fig. 20. Use of constrained hydrogen bonding reduced the R factor to 22.9% with no solvent molecules included.

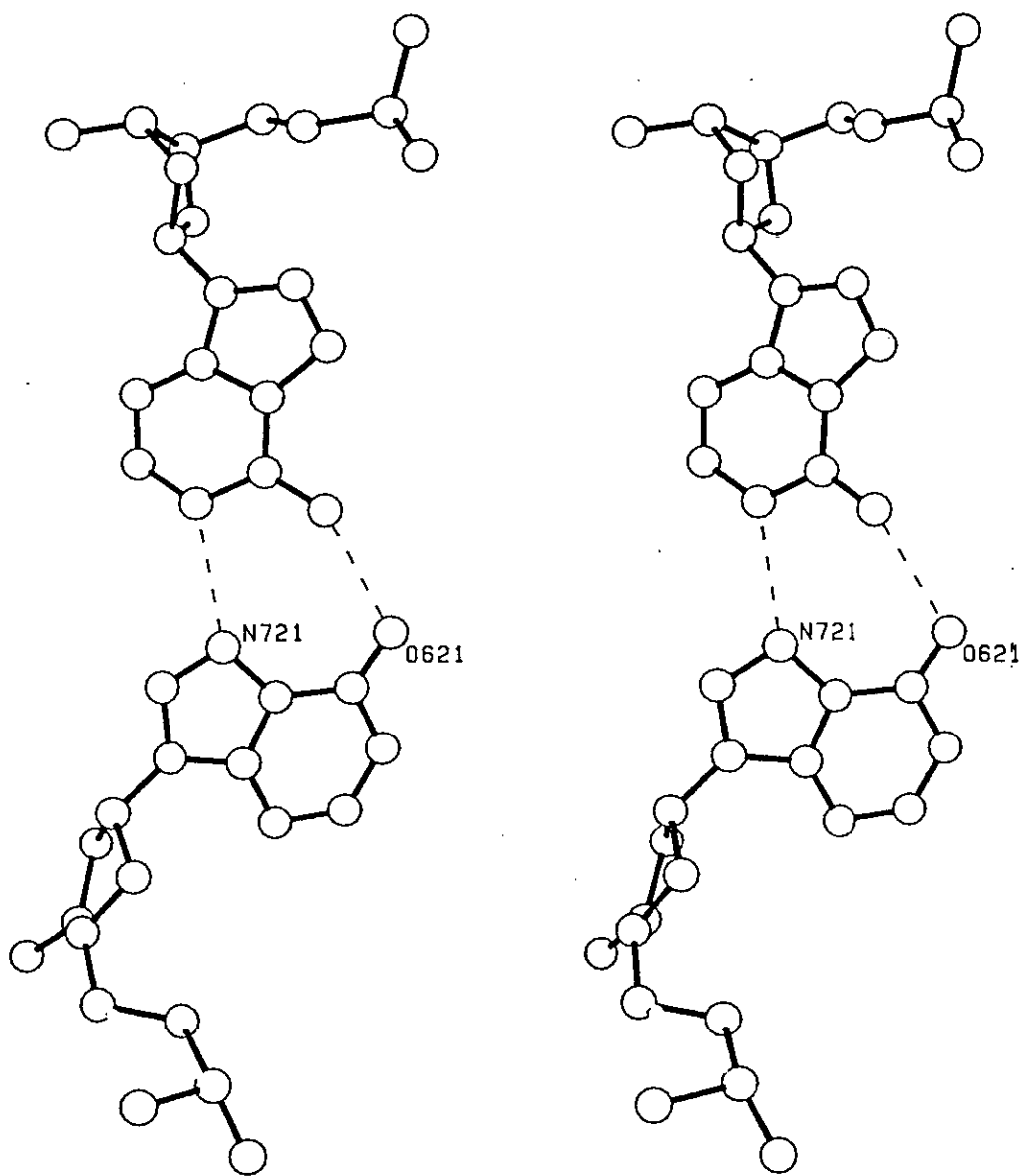


Fig. 20. The A4.I21 base pair, hydrogen bonding interactions indicated by dotted lines.

### 3.2.3 A.I Mismatch Conformation 2

Examination of the 2Fo-Fc maps showed there was a large amount of electron density associated with the inosine. It was possible to manipulate the base to fit in an I (*anti*).A (*anti*) orientation. Refining the structure with this in place gave an R factor of 23% but with a significant difference in the base pair geometries in terms of their hydrogen bonding distances. The N(6) A.O(6) I distances were 2.64Å and 3.07Å for the 4.21 and 16.9 base pairs respectively. There was little difference in the N(1) A.N(1) I distances of 2.48Å and 2.46Å Fig. 21a,b.

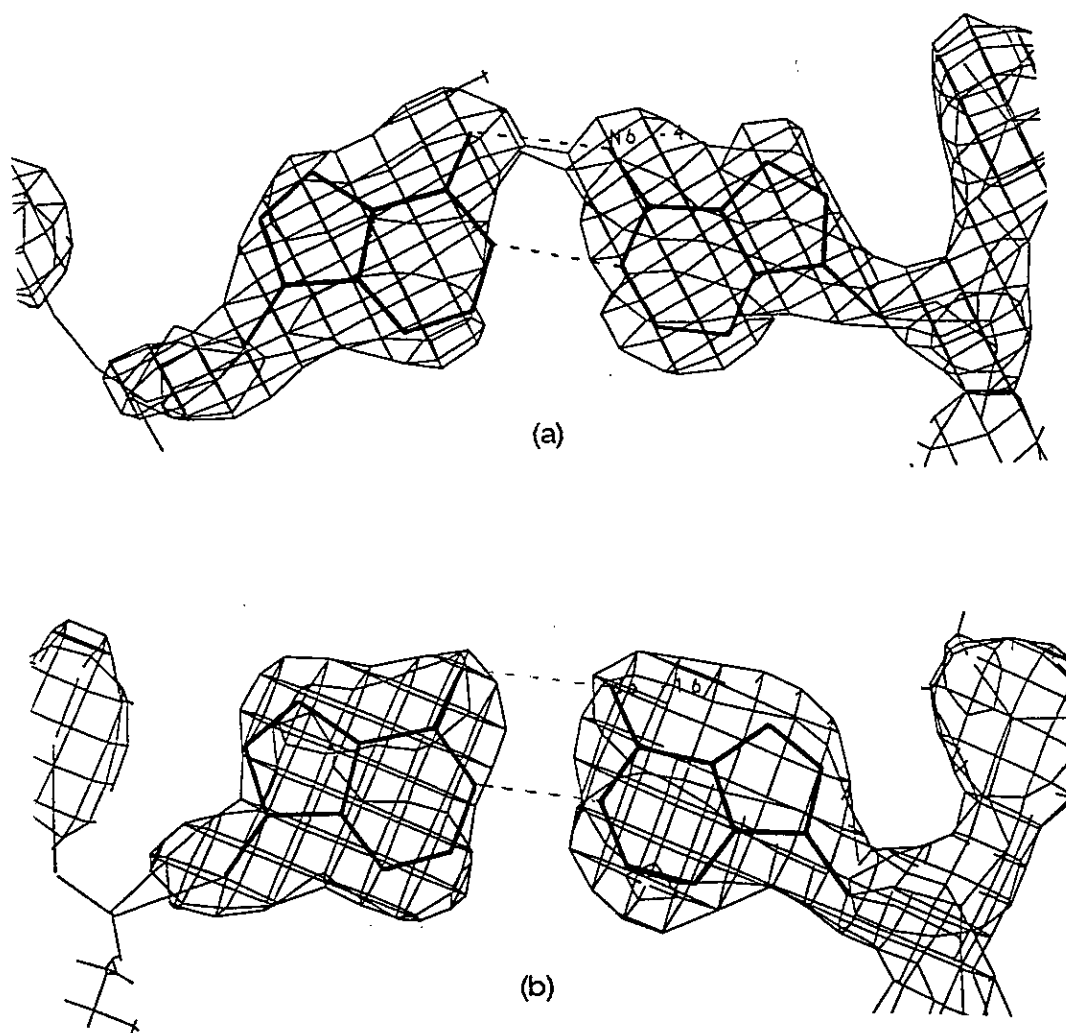


Fig. 21. (a) The 121.A4 2Fo-Fc map showing possible I (*anti*).A (*anti*) base pair.(b) The 19.A16 2Fo-Fc map showing possible I (*anti*).A (*anti*) base pair. In both cases R = 0.23.

### 3.2.4 Structure of the A.I Mismatch

The interbase distances found in the A (*anti*).I (*syn*) conformation are indicative of hydrogen bonding interactions suggesting the base pair is held together by two hydrogen bonds. As is the case with the A.G mismatch these cannot be formed if both bases are present as their major tautomers. While a hydrogen bond can be readily formed between N(6) A and O(6) I, N(1) A and N(7) I cannot hydrogen bond in their major tautomeric forms. The N(1) A.N(7) I hydrogen bond can only be formed by base protonation or if one of the bases is a minor tautomer. The resolution of the X-ray crystal determination is unable to differentiate which of these possibilities is taking place. The use of pH dependant UV thermal denaturation experiments resolves which hydrogen bonding arrangement is present <sup>76</sup>.

A series of pH dependant experiments in the pH range 4.5 to 8.0 was performed Fig. 22. From the graph the presence of a protonated species is indicated. The stability of the dCGCAAATTIGCG duplex increased below pH 6.5 with a maximum  $T_m = 308K$  at a pH of 5.25. This suggests protonation of adenine at N1 is taking place to give an  $AH^+$  (*anti*).I (*syn*) base pair Fig. 23. The base pair could also be formed by protonation of inosine to form a base pair with two cyclic hydrogen bonds. This is less likely due to its unfavourable  $pK_a$  (hypoxanthene  $pK_a = 2.0$ ).

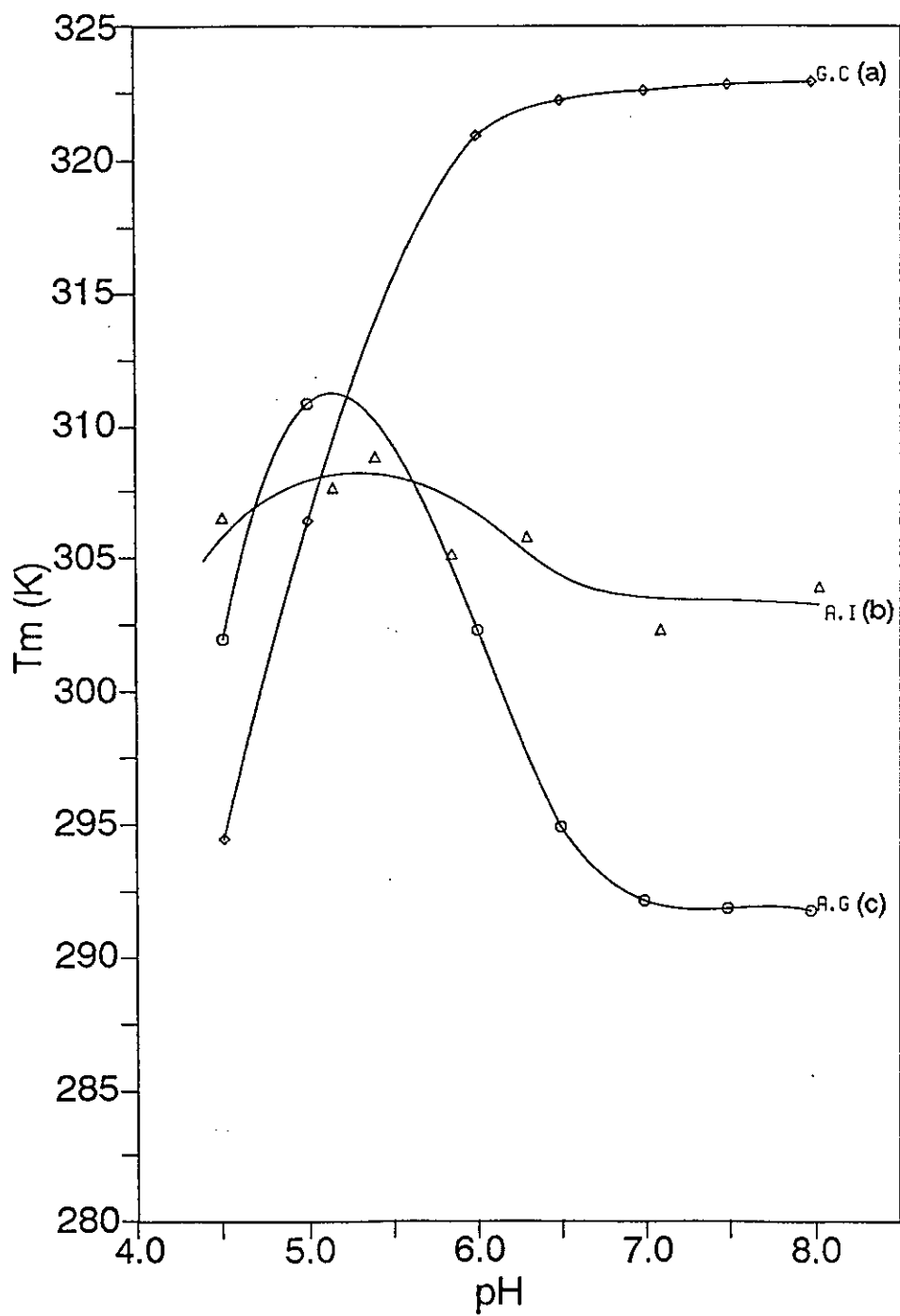


Fig. 22. Plots of  $T_m$  versus pH measured in 0.1M sodium phosphate buffer for duplexes (a) dCGCGAATTCGCG (b) dCGCAAATTIGCG (c) dCGCAAATTGGCG.

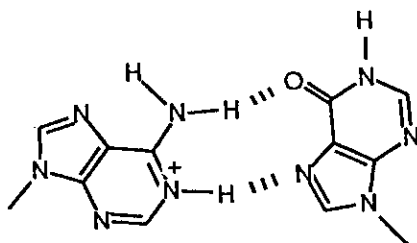


Fig. 23. Protonated AH<sup>+</sup> (*anti*).I (*syn*) base pair.

### 3.2.5 Conformational Flexibility of the A.I Mismatch

Previous examinations of the A.I mismatch have determined different conformations for the base pair. While it was found to be A (*anti*).I (*anti*) by NMR in dGGIACC<sup>74</sup> an X-ray crystal structure of dCGCIAATTAGCG gave a different result<sup>73</sup>. The A.I mismatch was in this case A (*syn*).I (*anti*). The A.I mismatch therefore displays conformational flexibility depending on its sequence environment. Examination of the base stacking interactions in the steps around the mismatch sites in the two related sequences dCGCIAATTAGCG and dCGCAAATTIGCG can be rationalised by looking at the stabilising and destabilising effects at each mismatch position. In the earlier sequence the mismatch site has a Cpl step while in the latter one there is an lpG step. Inosine has a dipole moment (5.6D, -16°) intermediate in size between adenine and guanine. This will participate in destabilising base stacking interactions with guanine (dipole 7.5D, -31°) in the lpG step of dCGCAAATTIGCG as both dipoles have a similar orientation. The strain of this interaction can be relieved if the inosine is able to adopt a *syn* orientation. The dipoles of inosine and guanine will now have roughly opposite directions. The magnitude of this effect can be expected to be less than in the GpG step of the A.G mismatch sequence. Similar considerations for the Cpl step show with an *anti* conformation the dipoles of inosine and cytosine have anti-parallel directions. The interactions in the Tpl step (thymine dipole 4.13D, 40°)<sup>77</sup> will not be destabilising in a B-DNA base structure. As was discussed for the A.G mismatch, adenine with its smaller dipole moment (2.9D, 64°) would not exhibit this unfavourable base stacking in the ApG step.

### 3.3 Comparison of the A.G and A.I Mismatches

The two analogous sequences dCGCAAATTGGCG and dCGCAAATTIGCG have been examined by X-ray crystallography and UV thermal denaturation experiments. In both cases the mismatched bases gave a better fit to the calculated fragment maps in A (*anti*).G or I (*syn*). The pH dependence of the duplex melting temperatures indicated the presence of protonated base pairs at acidic pH. These are thought to be of the form AH<sup>+</sup> (*anti*).G or I (*syn*).

That the two mismatches gave effectively the same structure was initially surprising. It had been expected that different conformations might have been found highlighting the role of the 2-amino group of guanine in mismatch conformation. In the A.G mismatch the *syn* orientation of the guanine directs the 2-amino group into the major groove where it is able to fully satisfy its hydrogen bonding potential with the surrounding water molecules. The well ordered water structure around the A.G mismatch is expected to contribute to its stability. As has been mentioned an A (*anti*).G (*anti*) base pair prevents the 2-amino group from hydrogen bonding to solvent due to its crowded position in the minor groove and this has been proposed to have a destabilising effect on the A.G mismatch<sup>7</sup>. The 2-amino group is not present in inosine so it cannot benefit from increased hydrogen bonded solvent interactions in a *syn* position compared to an *anti* one and conversely does not suffer a steric clash in the minor groove of an *anti.anti* base pair. This may be a contributing factor to the greater stability of A.I mismatches compared to A.G mismatches. At neutral pH the A.I mismatch has a much higher melting temperature than the A.G mismatch Fig. 22. It could have been predicted that an A (*anti*).I (*anti*) base pair would be found on the basis of net hydrogen bonding considerations since there is no loss of overall hydrogen bonding in going from single strand to duplex, due to minor groove crowding. Also the absence of the 2-amino group prevents the base from stabilising a *syn* conformation by forming solvent bridged interactions in the major groove. Other factors may govern the conformation of purine.purine mismatches. These will include the effect of base stacking and dipole-dipole interactions at the mismatch site as determined by the sequence environment of the mismatched bases. In

the oligonucleotide dGGIACC both mismatched bases were found to be *anti*. However the NMR spectra used to determine the structures were collected at pH 7.1 which may have prohibited the formation of a protonated  $AH^+$  (*anti*).I (*syn*) base pair. Such pH dependent conformational flexibility has been found before in NMR studies of the A.G mismatch <sup>45</sup>. Reference to the pH dependent UV melting data shows the percentage increase in stability at acidic pH for the A.I is much less than for the A.G mismatch Fig. 22. Less potential stabilisation can be gained in an I (*syn*).A<sup>+</sup> (*anti*) form due to the lack of stabilising influences in the duplex with I (*syn*) as opposed to I (*anti*). Crystals were grown at pH 6.5. This, from the  $T_m$  values, appears to be intermediate between the two stable forms of the mismatch. Hence at this pH there may be a coexistence of the two forms of the mismatch formed with A (*anti*).

The final crystallographic residual of the A.I mismatch structure refinement was higher than desired. This may in part be due to a more disordered structure at the mismatch sites. As discussed, inosine cannot benefit from stabilising solvent interactions in the same manner as guanine. Due to their smaller magnitude, dipolar interactions will not be as strong in the A.I mismatch and so the structure is less constrained to a specific conformation. It may therefore be that there exists an equilibrium of *syn* and *anti* inosine either within a duplex or within a crystal. This led to the need to collect further crystallographic data. Repeated efforts made to obtain more good quality crystals by the sitting drop method were unsuccessful due to their tendency to aggregate making them impossible to mount. Crystals were grown by the hanging drop method. These were mounted on the diffractometer but due to defects in the crystals, did not diffract well. This may again reflect disorder within the duplex and its effect on crystal packing forces.

### 3.4 The G.BrU Mismatch

5-Bromouracil is an analogue of thymine having a bromine substituent replacing the methyl group at the 5 position Fig. 24. Like thymine it is able to form stable base pairs with adenine and is incorporated as a substrate by DNA polymerase opposite template adenine at about the same rate. 5-Bromouracil is strongly mutagenic. The mutagenicity has been proposed to be due to a greater tendency to form mismatches with guanine, than will thymine 78.

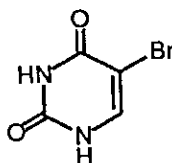


Fig. 24. 5-Bromouracil

When BrU is present on the template strand, mispairing with guanine induces A.T to G.C transition mutations. These have been detected in the BrU mutagenesis of lambda phage 79. The reverse of this, a G.C to A.T transition may also occur. In this instance the BrU is present as a substrate for DNA polymerase. In addition to its ability to mispair with guanine, its mutagenicity has been attributed to its inhibition of diphosphate reduction preventing the reduction of cytidine diphosphate to deoxycytidine diphosphate thus upsetting the dCTP pool concentration 80. Competitive incorporation experiments *in vitro* with T4 DNA polymerase have shown that BrU is incorporated in direct competition to dC with a frequency of about 1% 81. After proof reading by exonucleases this is reduced to 0.16% leading to G.C to A.T mutations. Under similar conditions the incorporation of thymine to give G.T mismatches was not measurable.

Regardless of the overall mechanism for the mutagenicity of 5-bromouracil the formation of stable G.BrU base pairs is a fundamental and key step. Three possible base pairing structures for G.BrU mismatches can be proposed Fig. 25. A stable Watson-Crick base pair could be formed by the rare enol tautomer of BrU. This would give a three hydrogen bond

G.C like base pair. Enolisation may be enhanced by halogen substitution at the 5 position<sup>82</sup> but will still be unlikely to occur to any significant extent. This base pairing arrangement will probably not represent a major form of the mismatch. Alternatively an ionised base pair can be envisaged. Ionisation of the N(3)H of BrU is induced by the presence of the electronegative bromine. The N(3)H atom of BrU has a pKa of 8.5<sup>76</sup> compared to a pKa of 10 for thymine. This base pair might therefore be expected to form fairly readily at physiological pH. The third possibility is the formation of a wobble base pair<sup>7</sup>.

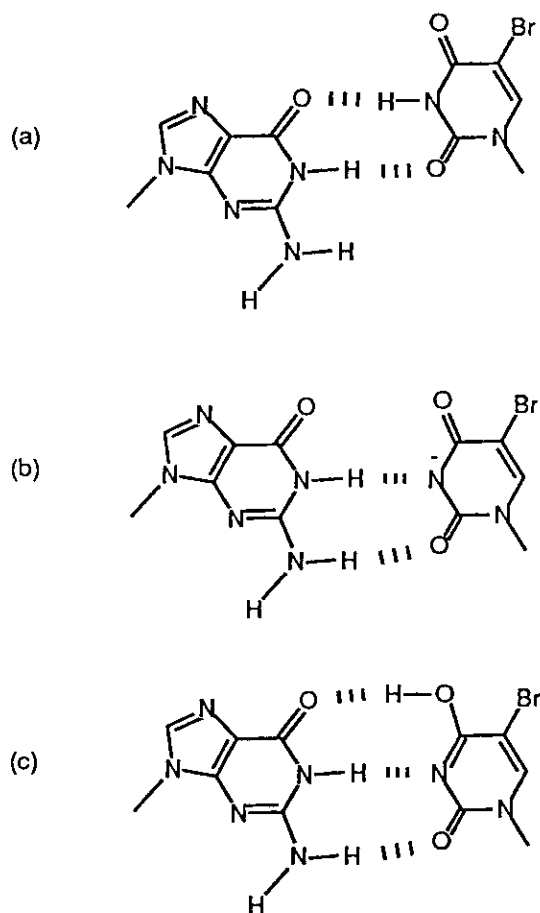


Fig. 25. Possible forms of the G.BrU mismatch (a) G.BrU (wobble) (b) G.BrU<sup>-</sup> (c) G.BrU (enol).

The G.BrU mismatch has been identified as a wobble base pair in the Z-DNA structure of dBrUGCGCG<sup>17</sup>. The crystals used in the structure solution were grown at pH 8.5 at which

BrU should be 50% ionised. It might be expected that under these conditions the ionised base pair would form or at least have some contribution to the mismatch structure. No direct evidence for the formation of either an enolised or ionised G.BrU base pair was found in the electron density maps at the mismatch sites.

In order to provide further insight into the detailed structure, and to determine if there was any contribution from the ionised base pair to the form of the G.BrU mismatch, it was decided to look at the effect of pH on the duplex melting temperature of an oligonucleotide containing G.BrU mispairs. If there was any contribution to the mismatch from the ionised base pair this would be expected to be reflected in the pH dependence of the melting temperature. To remain consistent with investigations into other mismatches a variation on the dodecamer dCGCGAATTCGCG would again be used. The mismatch sequence chosen was dCGCGAATTBrUGCG giving mismatches at positions 4.21 and 9.16 in the duplex Fig. 26. The oligonucleotide was to be synthesised on solid support using phosphoramidite methods. The phosphoramidite of BrU was not commercially available and so would be prepared as the 5'-OH trityl protected N, N-diisopropylamino 2-cyanoethyldeoxynucleoside phosphoramidite.

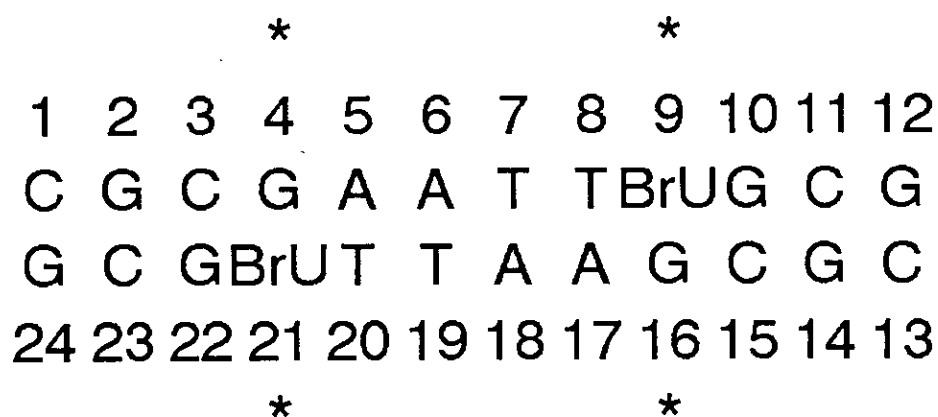


Fig. 26. G.BrU mismatch sequence numbering, mismatch positions indicated by \*.

### 3.4.1 Structure of the G.BrU Mismatch

The phosphoramidite of 5-bromo-2'-deoxyuridine was prepared as outlined. The sugar was protected as the 5'-O 4, 4'-dimethoxytrityl ether using 4, 4'-dimethoxytrityl chloride and converted to the reactive phosphoramidite with 2-cyanoethyl N, N-diisopropylchlorophosphoramidite. The product was difficult to obtain as a solid and three precipitations into hexane were required to prevent the formation of an oily product. The oligonucleotide was synthesised using standard coupling cycles (CE103A) and ending procedures (deprce03 Trityl OFF, Auto) on derivatised CPG with the phosphoramidite dissolved in acetonitrile and placed on port 5 of the DNA synthesiser. After purification by HPLC, Method 2, the oligonucleotide was used in pH dependant UV melting experiments over the pH range 5.5 to 8.5<sup>76</sup>. It can be seen that there is no marked increase in the melting temperature over the pH range tested Fig. 27. The formation of the ionised base pair would be favoured by a high pH ( $pK_a \text{ BrU N}(3)\text{H} = 8.5$ ) so if the base pair did have an ionised contribution changes in  $T_m$  might have been expected. However any formation of an ionised base pair would also have a destabilising component. In the ionised mismatch there are close contacts between two hydrogen bond acceptors. The 5-bromouracil O(4) and guanine O(6) would be adjacent to each other. This deleterious acceptor.acceptor interaction could be stabilised by protonation of either carbonyl group. This would only be possible at very low pH. There is an incompatibility between the the conditions required for the formation of the 5-bromouracil N(3).guanine N(1) hydrogen bond and the 5-bromouracil O(4). guanine O(6) hydrogen bond. This incompatibility disfavors the ionisation base pair.

From pH dependant UV melting studies there is no clear evidence for any hydrogen bonding arrangement other than the wobble base pair found previously by X-ray crystallography. The wobble base pair is further favoured by the  $e^-$  withdrawing properties of the bromine which reduces the hydrogen bonding potential of the O(4) atom and increases that of the O(2) atom<sup>83</sup>. While both the G.BrU wobble and A.BrU Watson-Crick base pairs have two hydrogen bonding interactions, those in the A.BrU base pair will be weakened due to the effect of the 5-bromine atom on O(4) hydrogen bonding potential in the 5-bromouracil

O(4).adenine N(6)H hydrogen bond. The net effect will be to offer relative stabilisation to the G.BrU mismatch and concomitant enhancement of its mutagenicity.

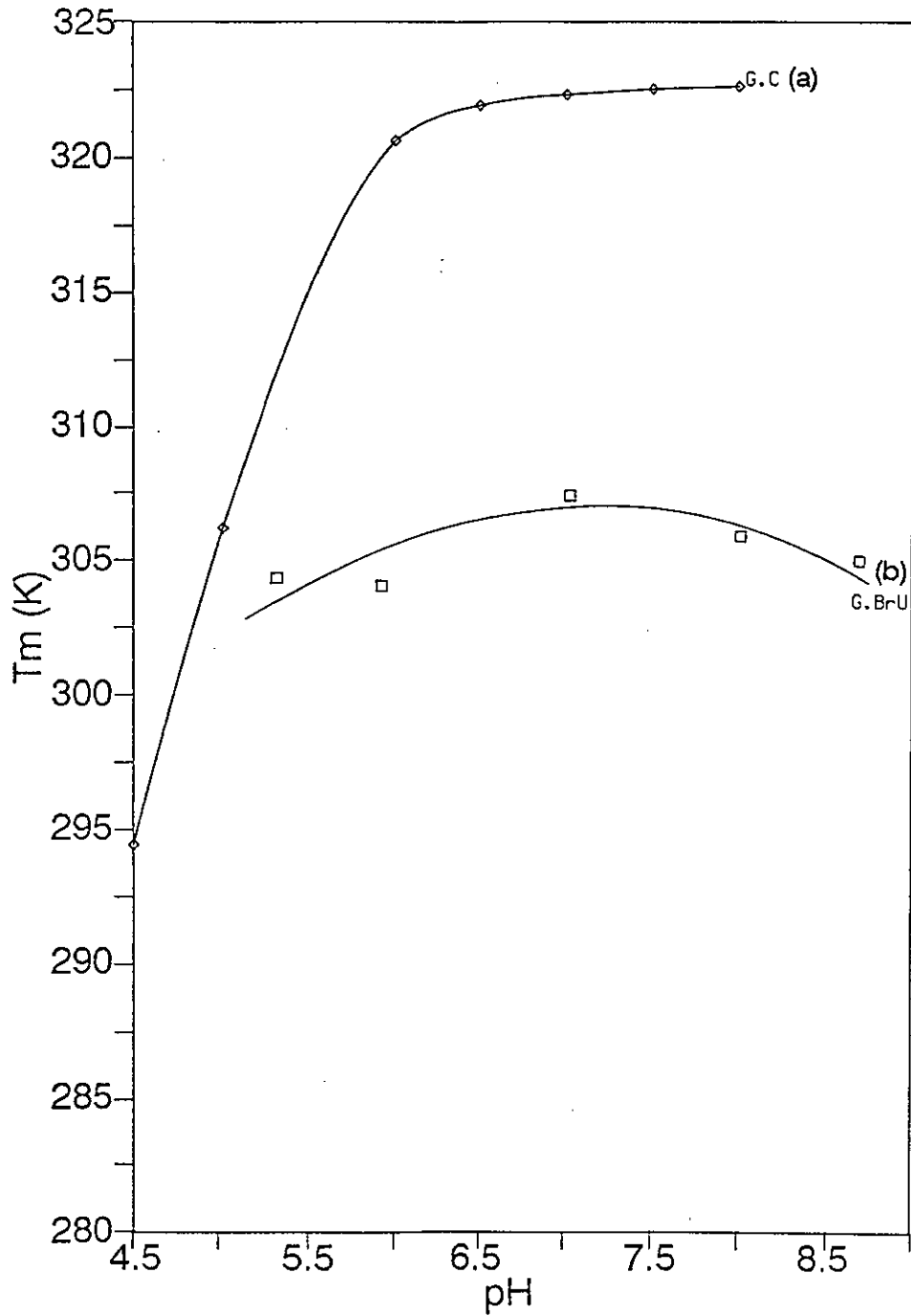


Fig. 27. Plots of  $T_m$  versus pH measured in 0.1M sodium phosphate buffer for duplexes (a) dCGCGAATTCGCG (b) CGCGAATTBrUGCG.

#### 4.0 2-Amino-2'-deoxyadenosine and the DNA Duplex

The study of DNAs and RNAs has shown that they may contain bases other than the major ones A, G, C and T (U). Among these is the purine analogue 2, 6-diaminopurine or 2-aminoadenine. As the name indicates this is an analogue of adenine with a second amino group in the 2 position. It has been identified as occurring naturally in the DNA of S-21 cyanophage<sup>84, 85</sup> where 2'-deoxyadenosine is fully replaced by 2-amino-2'-deoxyadenosine forming stable base pairs with T. The reason for this substitution was not investigated by the authors and is still the subject of speculation but it may be for structural, thermodynamic or biochemical reasons. The introduction of the amino function at the 2 position inserts an NH<sub>2</sub> group into the minor groove of the base pairs it forms. This allows the formation of three hydrogen bonds with T in DNA or U in RNA<sup>86</sup> Fig. 28. The resultant three hydrogen bond base pairs are A.T like in the major groove and G.C like in the minor groove. Indeed it has been found that in at least one case A' is read as G rather than A due to recognition of the minor groove amino group<sup>87</sup>. Hence the incorporation of diaminopurine into DNA is of interest due to its effect on DNA stability, structure and biological uses as genetic probes.

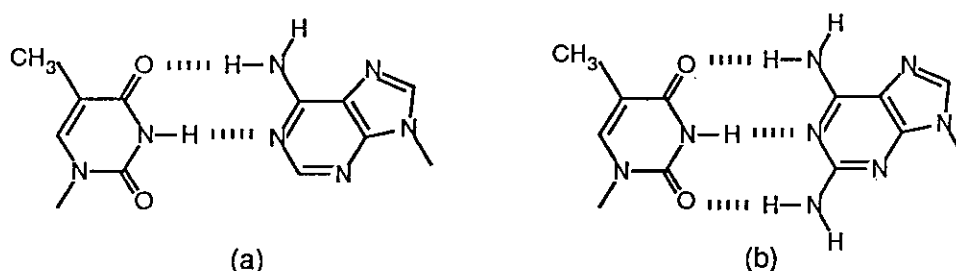


Fig. 28. (a) T.A and (b) T.A' base pairs.

#### 4.1 The Effect of Diaminopurine on DNA Duplex Stability

The formation of three hydrogen bond base pairs between A' and T (U) should be seen to increase the thermal stability of RNA and DNA when compared to the native sequences with

the normal two bond A.T (U) base pairs. This has been investigated in both RNA structures and DNA duplexes. Howard and coworkers <sup>86</sup>, who had previously identified increased thermal stability of monomer polymer helices by incorporation of A' into RNA, examined the duplex formed by poly(rA.rU) and poly(rA'.rU) and found the diaminopurine duplex to be about 25<sup>o</sup> more stable than the native polymers. This increase was rationalised to be due to a greater enthalpy of base-base interaction since the entropy changes on duplex melting should be similar for both and the authors proposed that a three bond A'.U base pair was formed.

Shortly after this, Cerami and coworkers used DNA polymerase to prepare an alternating DNA copolymer poly(dA'.dT) containing 90% replacement of A by A' <sup>88</sup>. By a similar method they also made the native DNA copolymer (dA.dT). Comparing the melting temperatures for the two copolymers showed the native sequence to be much lower melting ( $\Delta T_m = 20^{\circ}\text{C}$ ). This was thought to be consistent with the previous proposal for three hydrogen bond base pairs in RNA. These two studies had examined the thermal stability of synthetic polynucleotides. The opportunity of examining natural sequences containing diaminopurine came with the isolation of the DNA from S-2L cyanophage <sup>84, 85</sup>. Prior to this it had not been identified in natural DNA and its biological function had not been considered. Determination of the  $T_m$  values for the cyanophage DNA and ordinary DNA of similar G.C content showed the cyanophage DNA to have a higher  $T_m$  value by about 3<sup>o</sup>C. The identification of diaminopurine occurring naturally in DNA promoted its further study. Different enzymatic and chemical methods were developed to allow the preparation of oligonucleotides or polynucleotides with A' inserted into them. For example Scheit and coworkers enzymatically made poly dA' and in common with others noted the elevating effect on  $T_m$  of dA' as opposed to dA in duplex formation of poly(dA').poly(dT) and poly(dA).poly(dU)<sup>89</sup>.

The first examination of short synthetic oligonucleotide fragments containing dA' was performed by Gaffney and coworkers who reported the stabilising effect of dA' on  $T_m$  such that the hexamer d(TA')<sub>3</sub> formed a stable duplex at temperatures where the native oligonucleotide did not <sup>90</sup>. They later prepared other hexamers and examined their stabilities.

This led them to classify the A'.T base pair as being more stable than A.T base pairs, but less stable than G.C base pairs<sup>91</sup>. It was early on noted that the increase in T<sub>m</sub> values upon inclusion of A' were much less in DNA than had been found in RNA<sup>92</sup>. Howard and coworkers suggested that the lower increases might be due to a change in the hydration pattern of DNA which was disrupted by the minor groove amino group<sup>92, 93</sup>. The overall gain in stability was equal to the beneficial gain from forming three hydrogen bonds less the loss in stability due to changes in hydration patterns. This was thought to be particularly detrimental in the case of A.T tracts<sup>94</sup> where Dickerson had previously reported finding a spine of hydration. This was in part supported by the work of Chollet and coworkers who examined the use of diaminopurine in synthetic probes paying particular attention to the stability of the duplexes they formed with native DNA<sup>95-97</sup>. While in most cases the diaminopurine probes were more stable than the native ones it was noted that substitution of A' in A.T tracts such as in the duplex dCAAAA'AAAG destabilised the duplex. This is likely to be due to the disruption of a network of bifurcated hydrogen bonds formed between dA N<sup>6</sup> and dT O<sup>4</sup> on opposite strands in the adjacent base pair<sup>62, 63, 98</sup>. In this case T<sub>m</sub> native = 45°C and T<sub>m</sub> (A') = 41°C.

#### **4.2 Factors Influencing Oligonucleotide Structure.**

The structures of a substantial number of DNA fragments have been determined by solution or X-ray methods of both natural and synthetic DNAs. It has become apparent that the double helix can show a diverse range of structural variations. These can be grouped into family types A, B, C, D, Z, and X. The actual structure an oligonucleotide or polynucleotide adopts depends on any or all of the following.

- i. degree of hydration.
- ii. counter ion type and concentration.
- iii. base sequence.
- iv. base composition.

## Degree of hydration

Natural random polynucleotides are heavily hydrated *in vivo* due to the DNA containing many hydration sites on the base sugar and phosphate components. Of these the strongest interaction occurs at the phosphorus<sup>99</sup>. In conditions of high humidity the duplex is fully hydrated and adopts the B-form, the *in vivo* default structure. However when the humidity is reduced to below some critical value the conformation may change to an A or a C form. Crystal structures have suggested that at least in the A form this reflects a change in the hydration pattern such that the ordered spine of hydration is lost from the B form and replaced by a ladder like arrangement of water molecules bridging phosphates on opposite strands across the minor groove of the A form<sup>100</sup>. These hydration dependant changes are usually reversible, raising the humidity will cause an A to B re-ordering.

## Counterion Type and Concentration

The phosphate backbone of DNA is highly charged and able to interact with other ions eg. Na<sup>+</sup>, Mg<sup>2+</sup> or Cs<sup>+</sup>. The effect of counterions and their concentration is in some respects another way of considering hydration, since at high salt concentration water activity is reduced and B to A type transitions occur. Salt concentration also affects B-Z or X type changes and this will be considered later.

## Base Sequence

DNAs of similar overall composition but different sequence may produce different structures. For example poly(dA).poly(dT) exists as a B form at >70% humidity and below this changes to a metastable A form which disproportionates into a triple stranded A type poly(dA).2poly(dT) and a single nucleotide strand<sup>101</sup>. Its compositional isomers poly(dA-dT).poly(dA-dT) and poly(dA-dA-dT).poly(dA-dT-dT) behave differently. The latter exists only in B, and at low humidity D forms<sup>102</sup> while the former can adopt also a metastable A form<sup>103</sup>.

## Base Composition

DNAs of similar purine pyrimidine alternating sequence but with different composition may show different structures. For example poly(dG-dC) readily forms a Z helix while poly(dA-dT) does not <sup>104</sup>.

### 4.3 Effect of Diaminopurine on Oligonucleotide Structure

One of the reasons it is desirable to make oligonucleotides containing A' is to see what effect its inclusion will have on oligonucleotide structure. The base has both A like and G like features and one could expect it to behave like either base or a mixture of the two. Like G it has a 2-amino group, the role of which in structural transitions is not fully understood. The most dramatic transition it may influence is the change from B to Z (R to L) forms. The first evidence of the Z form of DNA came from the observation of salt induced changes in the CD spectrum of poly(dG-dC) <sup>104</sup>. On going from low to high salt concentrations (NaCl or MgCl<sub>2</sub>), an inversion in the CD spectrum was obtained. This was loosely termed an L-form to indicate that it was probably left handed. After being initially assigned as an alternating B-form <sup>105</sup>, the precise structure of the high salt form was determined by the of Z-DNA structures obtained from X-ray analyses of single crystals <sup>51, 106</sup> of dCGCGCG and dCGCG in which it was found that the base residues adopted alternating C *anti* and G *syn* orientations giving rise to a dinucleotide repeat zig-zag phosphate backbone the authors termed as Z.

In order to assess whether A' could influence B-Z transitions it is first of all necessary to look at what influences the formation of Z-DNA. The initial discovery that poly(dG-dC) could adopt different conformations in different salt concentrations promoted the interest in looking at other sequences which may show the potential for transitions to Z-DNA. It has become apparent that not only high salt induces isomerisation into Z-DNA. Much milder conditions can be used depending on the precise oligonucleotide. It has been found that bromination of poly(dG-dC) allows it to form a Z helix at low salt concentrations <sup>107</sup>. The CD spectrum of poly(dG-dBrC) was characteristic of Z-DNA in 150 mM NaCl. Low concentrations of

Mg(ClO<sub>4</sub>)<sub>2</sub> have also been found to stabilise Z-DNA and promote the B-Z transition <sup>108-114</sup>. Methylation of C to give poly(dG-dMeC) <sup>108-113</sup> also promotes the transition and allows milder conditions for Z-DNA formation. In the presence of even trace amounts of Mg<sup>2+</sup> ions, only physiological salt concentrations are required. Methylation of G has the effect of making the transition more rapid as well as reducing the salt required to physiological levels <sup>114</sup>.

Apart from poly(dG-dC) can other sequences undergo a transition to Z-DNA? The dodecamer dCGCGAATTCGCG, which was crystallised by Dickerson and coworkers, was an attempt to explore the Z-forming potential of other oligonucleotides <sup>46-49</sup>. By combining Z-favouring dCGCG ends with a dAATT Z-incompatible middle the authors hoped to obtain Z-DNA tetramers at either end in conjunction with either a right handed or melted AATT middle region. However, they obtained a right handed B-DNA helix. This would suggest that Z-DNA has a low stability when associated with non Z regions. Shortly after the publication of this structure Quadrifoglio and coworkers looked at the the sequences d(AT)<sub>3</sub>(CG)<sub>3</sub> <sup>115</sup> and d(TA)<sub>3</sub>(CG)<sub>3</sub> <sup>116</sup>. These two sequences, both being non self-complementary, form concatamers in solution with alternating (AT)<sub>3</sub> and (CG)<sub>3</sub> or (TA)<sub>3</sub> and (CG)<sub>3</sub> hexamer duplex regions. In the case of d(AT)<sub>3</sub>(CG)<sub>3</sub>, both at low salt and high salt (5M NaCl) concentrations, the duplex failed to form Z-DNA and remained as a right handed B-DNA. The d(TA)<sub>3</sub>(CG)<sub>3</sub> sequence on the other hand formed concatamers in which the (CG)<sub>3</sub> parts showed Z-like CD properties, namely a negative long wavelength band.

Another potential candidate for Z-DNA is poly(dG-dT).poly(dA-dC) which is found throughout eukaryotic genomes <sup>117</sup>. This has also been examined as its C-methylated and brominated analogues poly(dG-dT).poly(dA-dMeC) <sup>118</sup> and poly(dG-dT).poly(dA-dBrC) <sup>119, 120</sup>. Again these all have an alternating purine pyrimidine hierarchy suggesting that they should behave in a like way to poly(dG-dC).

In similar CD studies Vorlíčková <sup>121</sup> and coworkers and Zimmer <sup>122</sup> and coworkers looked at the effect of differing concentrations of Cs<sup>+</sup>, Ca<sup>2+</sup> and ethanol on the polynucleotide structure. It was found that both CsF and CsCl induced a change in the sign of the longwave region of the CD spectrum with the former having the greater effect, perhaps

due to the greater binding affinity of  $F^-$  over  $Cl^-$  for amine groups. The CD inversion was also highly variable with temperature, its intensity being greatest at low temperatures. It was further found that the longwave CD inversion could be induced by trace amounts of  $Ca^{2+}$  in high ethanol concentrations. In their examination of the brominated and methylated sequences, Jovin and coworkers found that the B to Z (R to L) isomerisation was highly temperature dependent, unlike the true B to Z isomerisation of poly(dG-dC).

These results would suggest that a strictly alternating pyrimidine purine sequence is an essential requirement for Z-DNA, perhaps because purines can adopt a *syn* conformation more readily than pyrimidines<sup>94</sup>, and that any alternating pyrimidine purine sequence can undergo the transition. In agreement with this it has been found that the CG regions in d(CG)<sub>3</sub>TATA(CG)<sub>3</sub> are in the Z-form in 4M NaClO<sub>4</sub><sup>123</sup>.

However the conditions that determine Z formation cannot be as simple as this, for poly(dA-dT) fails to show any signs of forming Z-DNA though it is conformationally flexible. In addition the sequences dBrCGATBrCG<sup>124</sup> dBrCGTABrCG<sup>125</sup> and dMeCGTAMeCG<sup>125</sup> have been crystallised as Z-DNA while the sequences dBrCGBrCGATBrCGBrCG<sup>126</sup> and dMeCGCAMECGTGCG<sup>127</sup> have been analysed as Z-DNA in solution. In all these sequences Z-DNA was obtained despite the alternating pyrimidine purine pattern not being present. The T was found to lie *syn* under strongly Z-forcing conditions of high NaCl and methanol concentrations (3M NaCl, 33% MeOH). Also Spiro and coworkers<sup>128</sup> recently found that there was no evidence to support a B-Z transition of poly(dG-dT).poly(dA-dC). Furthermore, Vorlíčková and coworkers now believe that what they had previously interpreted as a B to Z transition may in fact have been a B to X transition<sup>129-131</sup>. Hence the factors promoting Z-DNA are more involved than simple pyrimidine purine alternation.

All of the sequences so far mentioned that are able to form Z- or L-DNAs have had a common structural feature. They all contain purines with a 2-amino group or a five-functionalised dC. Where no purine 2-amino group is present functionalisation of dC helps favour Z-DNA. Is this essential for Z-DNA? Two purines which do not possess a 2-amino group are A and I. An examination of the conformational flexibility of poly(dA-dT) and



poly(dI-dC) would be appropriate. Both of these have been investigated but neither has been found to undergo a salt induced B to Z or R to L transition <sup>104, 105</sup>, although structural transitions have been induced under specific conditions <sup>129-131</sup> and the related polymer poly(dI-dBrC) can adopt the Z conformation <sup>105, 132</sup>. It would therefore appear that the 2-amino group has a key role in at least one structural transition, that of right handed B-DNA to left handed Z-DNA.

The 2-amino group may influence DNA structure by two mechanisms. It may either be involved directly in stabilising interactions within the duplex or it can disrupt other stabilising properties such as hydration. Effects of the direct interaction type have been observed in the crystal structure of dCGCGCG <sup>51</sup>. Here, a hydrogen bonding interaction was observed between the 2-amino group of G and a water molecule which bridged the amino group and one of the phosphate oxygens. Furthermore, the high salt crystal structure of dCGCG showed the bridging water to be absent, it being replaced by a chloride ion <sup>106</sup>. This had the effect of repelling the phosphate group and altering the the G sugar geometry from C3'-endo to C1'-exo. The 2-amino group can also influence structure by its absence. The crystal structure of the dodecamer dCGCGAATTCGCG was solved with the inclusion of 72 water molecules <sup>49</sup>. The majority of the ordered waters were found to be located in the major or minor grooves, bound to either N or O on the heterocyclic base edges. The ordering of the water structure was particularly strong in the AATT region of the minor groove. Here a first shell spine of hydration was located which zig-zagged down the groove bridging non-hydrogen bonded bases in adjacent base pair steps, on opposite strands from N(3) A to O(2) T. These two atoms are located in almost exactly the same position with respect to the sugar linkage. These waters were in turn bridged by a second layer of water to give them a virtual tetrahedral coordination. The water structure is different in the G.C rich region further along the minor groove. The hydrogen bonding interaction with the water now takes place with the N(2) of the purine. This has the effect of disrupting the ordered spine and lifting it out of the minor groove.

It is thought that much of the stability of B-DNA can be attributed to the hydration spine, theoretical calculations have shown that the hydration spine can control the base conformation and selectively favour certain conformations of the double helix <sup>132</sup>. This highly ordered water structure cannot form if the purine 2-amino group is present to perturb it. Also DNA conformation may be determined by phosphate hydration, this being different in A, B and Z-DNAs <sup>133a, b</sup>. Hence by preventing an extra stabilising contribution, which would favour the B form in sequences such as poly(dG-dC), the 2-amino group may allow other factors to favour Z-DNA. Molecular mechanical calculations show that A' substitution for A should increase the Z philicity of the sequence dTA'TA'TA' <sup>134</sup>. It would be reasonable to expect this to be the case for any alternating pyrimidine purine sequence where the purines bear 2-amino groups such as would be the case for alternating oligonucleotides in which A was replaced by A'.

#### 4.4 Conformational Flexibility of Poly(dA-dT)

It is necessary to look at what is known of poly(dA-dT) structure. Initial fibre diffraction patterns suggested a D-DNA structure <sup>135</sup> which has been interpreted in different ways including a left handed form containing a seven fold arrangement of Hoogsteen base pairs <sup>136</sup>. Another interesting anomaly with poly(dA-dT) is the observation that it is cut into dinucleotides of pTpA when exposed to DNase 1 rather than mononucleotides <sup>137</sup>. The field was further opened by the publication of the first tetranucleotide crystal structure <sup>138a, b</sup>. Rather than forming, as one might expect a self complementary double helix with four base pairs, the oligonucleotide dATAT crystallised as two dinucleotide minihelices which were right handed. A significant part of this finding was that the sugar pucker altered depending on what base was present on the sugar. The deoxyribose was C3' *endo* when carrying a purine and C2' *endo* when carrying a pyrimidine. When the dinucleotide helical fragments were used to generate a polynucleotide model this by necessity also contained the alternating C3' *endo* , C2' *endo* features and different glycosidic torsion angles at A (0°) and T (70°). This structure was further developed in a model for alternating B-DNA <sup>139</sup>. In this the base stacking was

allowed to vary such that the dApdT overlap was improved and maximised at the expense of the slight dTp dA base stacking. Support for this comes also from computer modelling which shows dApdT and dTp dA to favour different stacking arrangements <sup>140</sup>. Recently the crystal structure of dCGCATATATGCG has been solved to 2.2Å<sup>0</sup>. The central d(AT)<sub>3</sub> region crystallised as alternating B-DNA <sup>141</sup>.

Several NMR studies of poly(dA-dT) have been performed <sup>111, 130, 142-151</sup>. A 2D NOE study of poly(dA-dT) <sup>149</sup> showed both bases to be in an *anti* conformation by observation of strong interactions of H(8) A and H(6) T with the sugar H(2') and that the helix is likely to be right handed due to the weaker interactions with neighbouring nucleotide H(2''). The authors concluded the structure was a right handed B-DNA. They further found that the requirement for the H(8) A, H(6) T-H(2') internucleotide interaction to be stronger than the H(8) A-H(2'') T or H(6) T-H(2'') A interaction for A-DNA with C3' *endo* was not fulfilled. The lack of a C3' *endo* form also partly ruled out an alternating B-DNA which was further excluded by the longer H(8) A-H(2'') T and shorter H(8) A-H(2') distances than the alternating B-DNA model requires. Additionally the model predicts H(6) T to be close to H(2') T (1.7Å<sup>0</sup>) but distant from H(2'') A (2.9Å<sup>0</sup>). These distances were found to be similar as were the H(8) A-H(1') T and H(6) T-H(1') A distances *i.e.* no alternation. However an alternating structure was seen in its <sup>31</sup>P NMR <sup>150</sup>. This showed up as different NOEs from neighbouring sugar protons H(3'), H(5') and H(2'), with the dTpA step having a greater NOE indicating different phosphorous environments. This is in agreement with earlier findings of two <sup>31</sup>P NMR signals <sup>142, 144</sup> and splitting of chemically equivalent sugar <sup>13</sup>C atoms <sup>145</sup> proposed to be due to different phosphorous conformations and different sugar puckers in the alternating hierarchy.

#### 4.4.1 Effect of Solvent on Poly(dA-dT)

Poly(dA-dT) will undergo structural transitions when exposed to different solvent environments 130, 152-161. These changes have been looked at in ethanolic solutions with  $\text{Na}^+$ ,  $\text{Cs}^+$  152, 154, 155, 159 and  $\text{Ca}^{2+}$  130, methanol solutions 156, 157 peptide binding 160 and with  $\text{Hg}^{2+}$  161. Poly(dA-dT) exhibits a B to A isomerisation induced by ethanol 130 in sodium phosphate buffers. When followed by circular dichroism a major change in the spectrum was seen on the addition of ethanol to a test solution of poly(dA-dT) in EDTA/Na phosphate. After a gradual lessening of the positive longwave signal a strong 260nm band was seen, this being typical of an A-type DNA. Also it was found that the  $^{31}\text{P}$  NMR taken in ethanolic solutions did not give a double  $^{31}\text{P}$  NMR signal typical of an alternating B-DNA but rather a single peak as would be expected for regular A-DNA.

In certain cases the effect of salt on the structure of poly(dA-dT) can be quite marked. The observation of a B to Z transition in poly(dG-dC) was first made with  $\text{MgCl}_2$  and  $\text{NaCl}$  51, 104, 106, but these salts have little or no effect on the structure of poly(dA-dT) and do not induce a conventional B to Z transition. It has been found that poly(dA-dT) can undergo a transition with certain other salts and DNA binding agents 130, 152-161. This difference is not surprising since it is well established that A.T and G.C base pairs are differently hydrated and that the first layer of hydration is essentially impervious to salt cations 106, 162, 163. With this in mind interest has turned to those salts in which the ions can insert themselves into the ordered water structure.

$^1\text{H}$  and  $^{31}\text{P}$  NMR showed that poly(dA-dT) will alter its structure with  $\text{Cs}^+$ ,  $\text{Me}_4\text{N}^+$  and 3, 5, 17-dipyranidinium cations but not with  $\text{Na}^+$  146. Observation of the H(8) A, H(6) T and sugar H(1') resonances showed the H(8) A signal was unaffected by  $\text{NaCl}$  but shifted upfield in  $\text{CsF}$  reflecting changes in base pair overlap, glycosidic torsion angles and phosphate geometry. Similarly the  $^{31}\text{P}$  NMR signal was split in two by  $\text{CsF}$  0.6ppm upfield changing progressively with addition of 2, 4 and 6M  $\text{CsF}$ . It was not determined with which step the changes were associated. One of the sugar H1' signals was slightly altered in both salts while the other was unchanged at 6.1ppm. The binding of the steroid or the addition of 4M  $\text{Me}_4\text{NCl}$

also resolved the  $^{31}\text{P}$  dTpdA and dApdT phosphodiester signals as does CsF <sup>147</sup>. This demonstrated that a change to a dinucleotide repeat was being observed and further that due to the nature of the progressive change in the  $^{31}\text{P}$  NMR signals with increasing CsF indicative of a low activation energy process, the change was likely to be within one helix family.

Kypr *et al* and Vorlickova *et al* have examined the effect of CsF on alternating and non-alternating A.T sequences <sup>143, 152</sup>. Their choice of salt arose from the previously mentioned need to find a salt which could mimic closely the water structure around the DNA. CsF is unique in its ability to adopt water structure <sup>164</sup>. The CD spectra of the alternating polynucleotide was strongly affected by NaCl and CsF whereas the non-alternating homopolymer CD spectra were not. This prompted further investigation of the effect of CsF on poly(dA-dT) both in the presence <sup>130</sup> and absence <sup>153</sup> of ethanol. The effect of the alcohol is to much reduce the amount of  $\text{Cs}^+$  required to bring about an observable isomerisation in the CD spectrum.

#### 4.4.2 The Effect of CsF Alone on Poly(dA-dT)

The effect of CsF on the CD spectrum was recorded over a range of concentrations. In general terms the effect was to invert the longwave region of the spectrum in a way that could not be accounted for by polynucleotide aggregation and light scattering <sup>165</sup>. A lowering of the  $\text{Cs}^+$  concentration reversed the process. The changes observed could be divided into two subsets; an initial reduction in the intensity of the whole CD spectrum was noted up to 3M CsF followed by a rise in the shortwave spectrum and the formation of a more deeply negative longwave region. The changes seen in the concentration range 0 to 3M CsF have previously been associated with DNA winding <sup>166</sup>, increasing  $\tau$  decreases the longwave intensity, and the absence of any isodichroic point in the spectrum implied the change was within one conformation. The authors proposed this to be a gradual winding within an alternating B-DNA. In the region beyond 3M CsF the CD spectra have an isodichroic point at 257nm suggesting a transition between two different thermodynamically stable DNA

conformations. The net effect of CsF addition to poly(dA-dT) in solution as observed through the CD spectrum was to bring about a transition from the proposed alternating B-DNA to some new conformation. The authors elected to call this new conformation X-DNA.

The transition can also be observed with trace amounts of Cs<sup>+</sup> in ethanolic solutions of poly (dA-dT) <sup>130</sup>. The addition of millimolar quantities of CsF to 50% ethanol solutions containing 0.05mM EDTA 0.2mM Na phosphate of poly(dA-dT) again produced changes in the longwave CD region with an isodichroic point present at 257nm indicative of the alternating B to X transition with an  $\Sigma_{275}$  of -8, qualitatively the same as that seen with 6.3M CsF alone. In the absence of EDTA in the buffer solution the midpoint of the transition occurred at lower ethanol concentrations indicating that trace amounts of divalent cations present may stabilise the X-DNA. Addition of CaCl<sub>2</sub> to the solutions beyond that which could be complexed with the EDTA present again produced the transition at a lower ethanol concentration. However Ca<sup>2+</sup> alone could not produce the alternating B to X change. A-form solutions of poly(dA-dT) in ethanol and Na<sup>+</sup> could be transformed to X-DNA by addition of CsCl but CsCl alone was unable to produce an X-form. In the absence of ethanol CsCl could only induce an ellipticity at 275nm of  $\Sigma = -2$  <sup>153</sup>.

These findings suggested that the CsF has a two part effect. The Cs<sup>+</sup> begins the process by gradual reordering of the helix while the F<sup>-</sup> brings about the transition to the X-form proper. This was supported by measuring the UV absorbance as a function of CsF concentration <sup>153</sup>. The UV absorbance increased sharply upto 3M CsF and then remained essentially constant. This would imply an initial decrease in the overall base stacking followed by no further loss during the actual transition.

These changes can also be followed in the <sup>31</sup>P NMR spectrum. As noted before by others <sup>142-147</sup> the <sup>31</sup>P NMR signal in the absence of CsF consists of two signals at -4.2ppm and -4.45ppm indicative of two P environments. The changes seen with increasing CsF parallel those seen in the CD spectrum again falling into two processes. Up to 3M CsF the upfield signal shifted more upfield while the other was unchanged. This has been related <sup>167</sup> to increased stacking within one step towards a more *gauche gauche* environment with an

associated rotation of the dA residues and destacking in the other <sup>143, 146, 153</sup>. The situation can be further clarified by reference to phosphorothioate substitution experiments <sup>148</sup>. The NMR spectra of poly(dAs-dT) and poly(dTs-dA) each show two signals in low salt at 51.80 and -4.25ppm and 51.51 and -4.43ppm respectively. Hence the downfield resonance at -4.25ppm relates to the dTpdA step and the upfield resonance at -4.4ppm arises from the dApdT step. Beyond 3M CsF this signal was invariant while the downfield signal moved further downfield. This change is associated with the true isomerisation indicating it involves an increase in a non *gauche* <sup>167</sup> conformation of the step associated with the downfield resonance the dTpdA step. Hence, in 0-3M CsF dApdT steps become more stacked and by inference dTpdA less stacked emphasising the alternating B structure from which the B to X transition itself takes place.

#### 4.4.3 Structural Transitions of Poly (dA'-dT)

Although poly(dA'-dT) has been studied less there have been a number of NMR and CD studies at the polymer and oligomer level <sup>91, 120, 129, 168-173</sup> as well as a comparative study of poly(dA-d5xU) where x is I or Br <sup>174</sup>. The initial conclusion drawn from collecting the CD spectrum of poly and oligo(dA'-dT) in low and high NaCl concentrations was that the high salt form was probably Z-DNA <sup>90, 91, 120</sup>. This conclusion was reached due to the general inversion of the CD spectrum, when going from low to high salt, in a manner analagous to that which had been seen for poly(dG-dC). In these studies it was noted that the CD at high salt showed some differences from other left handed CD spectra so this alone could not be taken as evidence for Z-DNA being the high salt form of poly(dA'-dT). Indeed one analysis went so far as to note that the low and high salt CD spectra of d(TA')<sub>3</sub> and d(CG)<sub>3</sub> while being generally similar were sufficiently different as to suggest they were related to different left handed structures <sup>91</sup>. Furthermore the observation that dCGTA'TA'CG showed no salt dependent CD changes suggested that the two structures might be mutually exclusive. Two later salt dependent CD studies also proposed a Z-DNA structure for poly(dA'-dT) at high salt <sup>168, 169</sup> though again it was noted the CD spectra were not

identical to that seen before for Z-DNA.  $^{31}\text{P}$  NMR appears to support a high salt Z-DNA structure <sup>120, 168, 170</sup>. In low concentrations of NaCl the polynucleotide has two clearly resolved signals. These, increase their separation from an initial 32 to 71Hz in 4M NaCl. The doublet in the  $^{31}\text{P}$  NMR was clearly indicative of two different phosphorus environments and was likened to that seen for poly(dA-dT), although it was felt the downfield signal was closer to that for poly(dG-dC).

This was taken to indicate a Z-DNA structure for poly(dA'-dT). A different conclusion was reached in a later 2D NMR and  $^{31}\text{P}$  NMR experiment <sup>170</sup>. Spectra were recorded in 0.1 and 4.0 M NaCl. At high salt concentration the most important change noted was the observation of two new cross peaks not seen at low salt, corresponding to a base proton (H(6) T or H(8) A at high salt these signals overlapped) H(3') interaction and a H(3')-(Me) T interaction. Here again the  $^{31}\text{P}$  NMR doublet was seen in 4.0M NaCl at -3.73 and -4.25ppm. No cross peaks were observed for H(8) A'-H(1') interaction as would be required for Z-DNA with a *syn* oriented purine in 4.0M NaCl. However a cross peak corresponding to an aromatic (H(8) or H(6)) - H(3') was seen, this being indicative of a C3' *endo* sugar pucker. An aromatic-H(2'), H(2'') cross peak possibly due to a C2' *endo* sugar pucker was also seen and was proposed to be due to H(6) T-H(2'') A' interactions, which are close (1.73 Å<sup>0</sup>) in A-DNA.

It was therefore concluded that the high salt structure belonged to the A-DNA family probably with alternating C2' *endo* pyrimidine residues. Similar results were found for an alternating dA'.d5haloU polynucleotide <sup>174</sup>. In summary it was concluded the high salt form of poly(dA'-dT) had Z like features but was not Z-DNA and may have been some new structure belonging to the A-DNA family. This conformational change did not appear to be restricted to the effect of high salt concentrations but could also be induced by alcohols <sup>172</sup>. If the transition was a B to A one then this would not be surprising as alcohols and low water activity are known to favour the less hydrated A-DNA.

Vorlíčková and coworkers <sup>172</sup> recorded the CD spectra of poly(dA'-dT) under the same conditions as they had previously found to induce a B to A transition in poly(dA-dT) <sup>130</sup>. The CD spectra of the high salt and ethanol conformers of poly(dA'-dT) were the same. The

alcohol CD spectrum was recorded into the vacuum region and showed very sharp changes in sign indicating a change of handedness or relative base pair displacement <sup>175</sup>. The final spectrum had similarities to those seen before for A to X transitions <sup>130, 172</sup> and it was concluded the ethanolic or high salt form of poly(dA'-dT) was not A-DNA but X-DNA. The effect of TFE on poly(dA'-dT) was the same as ethanol and methanol including a deeply negative band at around 278nm with an ellipticity of about  $\Sigma = -10$ . At the same concentrations TFE and methanol did not affect poly(dA-dT), 80% and 70% respectively, unless trace amounts of cesium cations were added. This produced the X-DNA spectrum. The cesium cations and the 2-amino group appear to have similar conformational roles in affecting the hydration of the DNA <sup>49</sup>. Indeed, cesium ions have been found in the minor groove of an X-ray fibre diffracton study of T2 DNA <sup>176</sup>.

Recently it has been shown that poly(dA'-dT) will form X-DNA under very mild conditions <sup>173</sup>. At low concentrations of the poly nucleotide, submillimolar amounts of magnesium ions will cause a transition. This suggests that X-DNA may have a biological significance since under the same conditions poly(dG-dMeC) isomerised into Z-DNA. This also establishes further that X and A-DNA are not the same, A-DNA is destabilised by divalent cations in aqueous solution <sup>177</sup>. It is perhaps worth noting that the 2-amino modification of dA is not the only change that enhances the ability of poly(dA-dT) to isomerise into X-DNA. Pyrimidine substitution at the 5 position has also been found to enhance it in the series poly(dA-dU) <sup>131</sup>, poly(dA-dT), poly(dA-dEtU) and poly(dA-dBrU) <sup>178</sup>. This was also interesting because at low salt poly(dA-dU) is a regular helix and not an alternating B-DNA showing only one signal in its <sup>31</sup>P NMR <sup>144</sup>. Increasing the size of the 5 substituent reduced the amount of CsF required to isomerise the polynucleotide.

It appears that Z-DNA cannot be formed simply by the presence of a 2-amino group on all alternating purines. Rather poly(dA-dT) and poly(dA'-dT) isomerise into the novel X-DNA. This conformation is known not to be equivalent to A or Z-DNA though it shares the feature with Z-DNA of a dinucleotide repeat, albeit with a reversed character. The dTpdA or dTpdA' in X-DNA mimics the conformational role of the dGpdC step in Z-DNA. It is most likely right

handed and a derivative of the A-DNA family. In all cases it is formed *via* an alternating B-DNA. Both Cs<sup>+</sup> and F<sup>-</sup> ions are essential requirements of X-DNA for poly(dA-dT) though poly(dA'-dT) will form X-DNA under physiological conditions with millimolar amounts of Mg<sup>2+</sup> ions. This extends its interest into the biological region .

No full structural analysis of X-DNA has yet been achieved by X-ray crystallography. It would be of great interest to obtain single crystals of X-DNA. By analogy with Z-DNA the most suitable candidate sequences for this are d(TA')<sub>2</sub>, d(TA')<sub>3</sub> and d(TA')<sub>4</sub>. d(TA')<sub>3</sub> has already been examined by CD <sup>91, 179</sup> and shown to undergo a conformational change with high salt but not one that could be attributed to Z-DNA. The authors at this time did not consider X-DNA.

#### **4.4.4 Formation of Z-DNA in Mixed Sequence Oligonucleotides Containing 2-Amino-2'-deoxyadenosine**

There has been some success in obtaining Z-DNA with mixed nonalternating sequences containing all four bases <sup>124-127</sup>. These have all relied on having a 5 functionalised dC which favours the formation of Z-DNA. The only case in which a mixed sequence DNA has formed Z-DNA with no dC functionalisation has been where wobble pairing has taken place <sup>13, 17</sup> although in one case this still involved a 5 functionalised uridine in the wobble pair. What then is the situation for forming Z-DNA in mixed sequence oligonucleotides containing 2-amino 2'-deoxyadenosine? There have been several preliminary examinations of this <sup>91, 127, 179</sup>, only one of which was successful in obtaining accurate data by X-ray crystallography <sup>179</sup>. The sequence dCA'CGTG has been studied by IR, CD and NMR <sup>127, 178</sup> as well as by X-ray crystallography <sup>179</sup>. The IR spectrum was very similar to that seen before for d(GC)<sub>n</sub> as Z-DNA and in 4.0M NaCl the <sup>1</sup>H NMR showed two resonances for each proton suggesting the coexistence of two DNA forms with slow exchange between them. From integration the proportion of Z-DNA was determined as 22±3%. This was thought to account for the lack of evidence of Z-DNA in the CD spectrum also seen by others <sup>179</sup> although a slight reduction in the intensity of the negative band at 260nm was observed. The

crystal structure has been solved to 1.3Å<sup>o</sup> 179. The overall structure was very similar to d(CG)<sub>3</sub> and like this retained the alternating purine *syn* C3' *endo*, pyrimidine *anti* C2' *endo* steps. The base stacking was similar also with *anti-p-syn* steps showing slight stacking of pyrimidines on opposite strands. In the *syn-p-anti* steps the 2-amino group of dA' stacks over the pyrimidine below it on the same strand in the same way as the 2-amino group of guanine, allowing heteronitrogen atom-aromatic π electron interactions *ie.* the dA' may stabilise Z-DNA in the same way as dG, which dA cannot. Both of these amino groups were hydrogen bonded to water molecules which bridged to nearby phosphate groups. This was complemented by a hydration spine linking the pyrimidine O(2) atoms. The oligonucleotide dCGTA'CG also crystallised as Z-DNA. Attempts to observe Z-DNA in the high salt CD spectra of dCGTA'CG, dCGUA'CG and dCGTA'TA'CG<sup>91</sup> failed to show any sign of Z-DNA though comparing the CD spectra to those obtained for dCA'CGTG and dCGTA'CG<sup>179</sup> there is a similar reduction in intensity of the negative band at 260nm so the authors may have overlooked a slight amount of Z-DNA formation in 5.0M NaCl. It is desirable to extend the range of mixed sequences that have been crystallised and determined as Z-DNA.

## 5.0 DNA Synthesis

In order to examine the effects of dA' on DNA duplex structure and stability a method was needed for the synthesis of the required oligonucleotides. There are two different ways in which oligonucleotides can be made, either by solution or solid phase synthesis. Solution phase synthesis is usually reserved for situations where greater than 50mg of DNA is required. This amount is far in excess of the quantity needed for the majority of biological and physical experiments, typically less than 1mg and 10-15mg respectively. In addition solid phase synthesis by its very nature lends itself to automation of the repeated steps required to build up an oligonucleotide, whereas solution synthesis requires repeated isolation and purification steps between each addition reaction.

Solid phase methods can further be divided into phosphotriester<sup>28</sup> and phosphoramidite<sup>29</sup> strategies. The phosphoramidite method is an improved version of the phosphite triester method pioneered by Letsinger and coworkers in 1975<sup>180</sup>. For the current work solid phase synthesis is the method of choice having the advantages of speed and ease of product separation. All the required reagents are delivered to the growing oligonucleotide anchored to an inert support. Thus all the chemistry takes place within the reaction vessel, work up between steps in the synthesis requires little more than filtering off excess reagents and byproducts, washing and introduction of the reagents for the next step. At the end of the synthesis the product is obtained by cleaving it from the support. Such reactions require to be very high yielding<sup>181</sup> and so it is normal to add an excess of reagents to ensure all reactions are driven to virtual completion. This should produce a product of high quality. The four main stages of the solid phase method are: derivatisation of a suitable support; oligonucleotide assembly; deprotection of the final oligonucleotide and purification of the oligonucleotide after cleavage from the support.

In most cases the phosphoramidite method is the preferred solid phase method with the benefits of good yields, high product purity, the ready commercial availability of standard protected nucleosides and derivatised supports of the four major bases.

## 5.1 Phosphoramidite Method

The first steps in the development of this method were taken with the solution synthesis of  $d(T)_2$ <sup>180</sup>. This utilised the reactivity of an *o*-chlorophenyl phosphorodichloridite towards a 5' protected dT nucleoside to yield a reactive monochloridite which on condensation with a 3' protected dT nucleoside gave the desired dinucleotide as a phosphate triester. Oxidation to the phosphodiester with  $H_2O/I_2$  and deprotection with NaOH gave the natural phosphodiester linkage. This worked well for the synthesis of dinucleotides and was also applied to the synthesis of  $d(T)_5$ <sup>182</sup>. For this the phosphorodichloridite protecting group was changed from *o*-chlorophenyl to 2, 2, 2-trichloroethyl due to instability of the *o*-chlorophenyl phosphotriester of the former to the phosphorodichloridite. Although the synthesis was done in solution the authors noted the rapidity of the coupling reactions (<5mins) and their high yields (95%) gave the method great potential for exploitation on a solid support. This application was realised where the support was amino-derivatised HPLC grade silica gel to which a nucleoside had been attached by a 3'-O succinate linkage<sup>183, 184</sup>. The reaction strategy, which initially involved the condensation of a trichloroethyl phosphorodichloridite derivatised nucleoside with the support bound nucleoside, followed by oxidation, as in the solution method was modified to use a methoxy-protected phosphorodichloridite  $CH_3OPCl_2$ . This had previously been used in the solution synthesis of ribonucleotides<sup>185</sup>. The coupling efficiencies were not satisfactory for repeated solid phase synthesis. Trials with triazole, tetrazole and 4-nitroimidazole as replacement leaving groups for  $Cl^-$  revealed the tetrazole derivative gave >95% coupling with a shorter condensation time. Unlike the present situation the tetrazolide was prepared in advance and added as the active condensing agent. The tetrazolides are extremely reactive species, very sensitive to hydrolysis and oxidation by air and have relatively short lifetimes. This requires them to be kept stored under an inert gas at  $-20^\circ C$  until used<sup>184</sup>. Additionally, the synthetic route to the nucleotide tetrazolides uses a bifunctional phosphitylating agent which as a byproduct yields significant amounts of nucleotide 3'-3' dimers. These do not interfere with the coupling reaction but reduce the efficacy of the coupling agent.

There was therefore a need to develop a more stable phosphorous reagent which was achieved by the introduction of a new form of phosphite, a phosphoramidite<sup>29</sup>. These were N, N-dimethylaminophosphoramidites of the nucleosides protected as before and activated *in situ* with a weak acid such as N, N-dimethylaniline hydrochloride to give the reactive monochloridite species. More importantly it was observed the phosphoramidite could be activated with tetrazole. Their stability is much greater than the tetrazolides previously used such that they may be stored under nitrogen at room temperature for at least a month. As an additional benefit they contain little 3'-3' dinucleotide phosphite as they are prepared from a monofunctional phosphitylating agent chloro-N, N-dimethylaminomethoxyphosphine (MeOPCINMe<sub>2</sub>). The phosphoramidites were further refined by increasing the steric hindrance at the labile P-N bond. There is a direct correlation between increasing steric hindrance (methyl, ethyl, isopropyl) and stability in solution<sup>186</sup>. Also an examination of deoxynucleoside diisopropylamino-, morpholino-, pyrrolidino- and 2, 2, 6, 6-tetramethylpiperidino-phosphoramidites, highlighted the utility of the N, N-diisopropylamino and N, morpholino phosphoramidites<sup>187</sup>. These have low volatility and are therefore less reactive towards atmospheric moisture. These were again prepared from monofunctional phosphitylating agents avoiding the formation of significant amounts of 3'-3' dimers<sup>186, 187</sup>.

A further development was the introduction of an *in situ* method for the phosphoramidite preparation which avoided altogether exposure of the nucleoside to chloridite reagents<sup>188-190</sup>. Instead, a bis-dialkylamine phosphine was reacted directly with a 5'-O protected nucleoside.

## 5.2 Phosphorous Protection

After oxidation the phosphorous triester was in most cases still protected with a methyl group. This is removed at the end of the synthesis with a thiophenol, triethylamine, dioxane treatment<sup>191</sup>, t-butylamine<sup>192, 193</sup> or concentrated ammonia<sup>194</sup>. These removal processes have associated problems. The thiophenol treatment needs to be done while the oligonucleotide is still bound to the solid support. The procedure for fully deprotecting the

oligonucleotide thus becomes a two step one, a lengthy thiophenol treatment to remove the phosphorous protection, followed by base deprotection and cleavage of the oligonucleotide from the support. Alternatively the thiophenol treatment may be replaced by refluxing the support bound oligonucleotide in t-butylamine for 15 hrs <sup>193</sup>. A simpler procedure involves only the use of concentrated ammonia solution. Although only one reagent is used the actual process is a two step one involving a prolonged overnight treatment to partly remove the methyl esters and solubilise the oligonucleotide followed by heating at 55°C for several hours to remove the base protection and finish the methyl ester cleavage. A side reaction in the ammonia treatment is the methylation of thymine at the N3 position <sup>195, 196</sup> or possible formation of an internucleotide phosphoramidate linkage <sup>197</sup>.

A range of alternative phosphorous protecting groups have been investigated <sup>188, 198-203</sup>. To date the most useful has been the 2-cyanoethyl group <sup>204, 205</sup> which had previously found use in the phosphodiester <sup>206</sup> and phosphotriester <sup>207</sup> approaches in addition to being a trial protecting group for the dichloridite synthesis of ribonucleosides <sup>198</sup>. Trials with the 2-cyanoethyl monochlorophosphoramidites of N, N-dimethylamine, N-morpholine and N, N-diisopropylamine indicated they were suitable monofunctional phosphitylating agents for the preparation of the corresponding nucleoside phosphoramidites. Of the three, the N, N-diisopropylaminomonochloridite had the most favourable physical characteristics and could readily be obtained pure after distillation under reduced pressure. The N, N-dimethylamino and N-morpholino monochlorophosphoramidites were less amenable, the former requiring much more stringently anhydrous handling conditions while the latter was more difficult to purify due to a tendency to decompose when distilled. These properties were reflected in the nucleoside phosphoramidites derived from the monochloridites. Thus the N, N-diisopropylamino 2-cyanoethyldeoxynucleoside phosphoramidite was favoured due to its stability and purity. It has since been observed that the 2-cyanoethyl dT phosphoramidite hydrolyses six times more slowly than the corresponding methoxy phosphoramidite <sup>208</sup>. The cyanoethyl group is easily removed with a short ammonia treatment. A simple one step treatment with ammonia will cleave the oligonucleotide from its support, remove the

phosphate triester protection and deprotect the exocyclic amine functions to yield the fully deprotected oligonucleotide. N, N-diisopropylamino-2-cyanoethyldeoxyribonucleoside phosphoramidites have become the monomers of choice for phosphoramidite oligonucleotide synthesis Fig. 29.

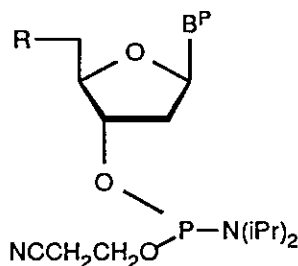


Fig. 29. General phosphoramidite, R = 5'-OH protection, B<sup>P</sup> = protected base.

### 5.3 Base and Sugar Protection

While much effort has been expended in the development of the phosphoramidite reagents the protecting groups used for base and sugar protection are generally those introduced by Khorana and coworkers<sup>209</sup> although many others have been investigated. Thus the 5'-hydroxyl functions are usually protected as an acid labile trityl ether, most often the 4, 4'-dimethoxytrityl ether, although the monomethoxytrityl ether may be used instead but is less acid labile. The purine and pyrimidine amino groups are normally protected as N<sup>6</sup>-benzoyl-2'-deoxyadenosine, N<sup>4</sup>-benzoyl-2'-deoxycytosine and N<sup>2</sup>-isobutyryl-2'-deoxyguanosine. The acyl groups are removed by alkaline hydrolysis during the ammonia treatment. There is therefore an orthogonality in the protecting strategy with the permanent protection being base labile while the sugar protection, which needs to be removed for each coupling cycle, is acid labile. The 4, 4'-dimethoxytrityl group is favoured over the monomethoxytrityl or triphenylmethyl protecting group due to its much greater acid lability which reduces the problem of acid catalysed depurination during deprotection of the 5'-OH at the start of each coupling cycle.

## 6.0 Acid Catalysed Depurination

The problem of depurination of adenosine residues during acidic deprotection of the 5'-OH was noted early on <sup>210</sup>. This is particularly bad if the adenosine is N<sup>2</sup>-acyl protected with a benzoyl amide, and has in some cases led to the adoption of a deprotection strategy where the base was deprotected before removal of the 4, 4'-dimethoxytrityl group. However this is inappropriate to the synthesis of oligonucleotides where the base must remain protected throughout the synthesis. The earlier in a sequence the adenosine is placed the worse the depurination will be by virtue of its exposure to more acid detritylation steps during the course of the synthesis.

The obvious protecting group strategy for a phosphoramidite synthesis of oligonucleotides incorporating 2-amino-2'-deoxyadenosine is to use the acyl protection relevant to dG and dA in conjunction with the standard 4, 4'-dimethoxytrityl protection of the sugar moiety. This will however leave the base susceptible to depurination in the same way as dA. Two previous attempts to synthesise 2-amino-2'-deoxyadenosine oligonucleotides were severely handicapped by the effects of depurination <sup>90, 91</sup> even with the use of ZnBr<sub>2</sub> deprotection. Once depurination has occurred the oligonucleotide, apart from having an undesired incomplete base sequence, will be cleaved at the apurinic site by ammonia treatment during work up at the end of the synthesis. The oligonucleotide is cleaved on the 3'-OH side of the apurinic site in a manner analogous to Maxam-Gilbert sequencing <sup>211a, b</sup>. Whilst degradation causes the apurinic material to behave chromatographically quite differently from the product allowing its easy separation, it will reduce the yield. The product obtained from the synthesis will thus be a combination of the desired oligonucleotide, apurinic and fragmented material. Any strategy for synthesis of 2-amino-2'-deoxyadenosine oligonucleotides should take cognisance of these problems. There are in essence three approaches that could be adopted to avoid problems of depurination.

- i. Use of reagents other than TCA to remove the 4, 4'-dimethoxytrityl ether.
- ii. Adoption of a different form of 5'-OH protection.
- iii. Re-design the base protection in order to make the C1'-N9 bond less acid labile.

i. At present the reagent used to remove the 4, 4'-dimethoxytrityl-ether is a solution of TCA in dichloromethane. Other reagents have previously been used for its removal such as benzenesulphonic acid <sup>212, 213</sup> which again causes significant amounts of depurination. Toluene sulphonic acid in acetonitrile has also been tried with some success <sup>183</sup> but when used in conjunction with N-benzoyldeoxyadenosine residues 50% depurination occurred in 20 minutes <sup>184</sup>. TCA causes less depurination than benzene sulphonic acid <sup>197, 214</sup>. This prompted the investigation of aprotic reagents for de-dimethoxytritylation. Heterogeneous mixtures of zinc bromide in dichloromethane gave quantitative detritylation <sup>215, 216</sup>. There are obvious disadvantages in applying a heterogeneous reagent to a solid phase system. Saturated solutions of zinc bromide in nitromethane (0.1M) were equally effective. Depurination did not occur as long as the nitromethane was scrupulously anhydrous <sup>216</sup>. However the deprotected 5'-OH is unreactive towards condensation reactions until it has been demetalated with mild alkaline hydrolysis. When incorporated into a solid phase cycle the deprotection becomes a two step process of (1) a 30 minute treatment with saturated ZnBr<sub>2</sub>/nitromethane followed by (2) a 5 minute basic wash with n-butanol in THF:2,6-lutidine to regenerate the 5'-OH <sup>184</sup>. This method could be used to prevent depurination but the complexity of the reagents used and the increase in cycle time incurred compared to a TCA cycle make it an unattractive option.

ii. Other groups that have been used for 5'-OH protection include: trityloxyacetyl, removed with dilute NH<sub>4</sub>OH <sup>217</sup>; levulinyl, removed with hydrazine <sup>218</sup>; o-dibromomethylbenzoyl, removed with silver perchlorate <sup>219</sup>; t-butyl dimethylsilyl, removed with F<sup>-</sup> or HOAc <sup>220</sup>; β-methoxyethoxymethyl, removed with ZnBr<sub>2</sub> or TiCl<sub>4</sub> <sup>221</sup>; 4, 4', 4''-tris(benzoyloxy)trityl, removed with NaOH <sup>222</sup>; 2-(methylthiomethoxy)ethoxycarbonyl, removed with Hg<sup>2+</sup> perchlorate/OH<sup>-</sup> <sup>223</sup>; α-naphthyl diphenylmethyl, removed by reductive radical cleavage <sup>224</sup> and the 9-phenylxanthen-9-yl group (pixyl) which is removed by mild acid hydrolysis <sup>225</sup>.

Whilst not exhaustive the preceding examples illustrate the problems that a change in 5'-OH protection would incur. Cleavages involving basic conditions or fluoride ions are not acceptable as they mean a loss of the orthogonality of base and sugar protection if base labile sugar protection is adopted, while the same problems of depurination will be encountered if an acid labile group such as the pixyl group is adopted. In view of this no real advantages can be offered by deviating from the established 4, 4'-dimethoxytrityl protection whilst many other problems would be introduced *eg.* phosphate cleavage by fluoride ion. Indeed 4, 4'-dimethoxytrityl protection has several advantages of its own. It is relatively cheap to use and 4, 4'-dimethoxytrityl protected phosphoramidites of the four major bases are readily available commercially. The 4, 4'-dimethoxytrityl group is highly coloured and when cleaved with acid its cation gives a bright orange colour. This can be used to accurately assay the coupling efficiency if collected and measured spectrophotometrically at each cycle in a synthesis <sup>226</sup>.

iii. This is the most appropriate course of action since it should not necessitate any change in the phosphate or sugar protecting groups. It is desirable to keep to a minimum the number of changes, if any, which need to be made to the well established and reliable cycles for phosphoramidite solid phase synthesis.

Many different groups have been tried for the protection of the base exocyclic amines with several aimed specifically at reducing the amount of acid catalysed depurination of deoxyadenosine on detritylation. Among those examined a modification of N<sup>6</sup>-benzoyl protection an N<sup>6</sup>-dibenzoyl protected nucleoside has been found to eliminate depurination during moderate treatment with TCA <sup>227</sup>. A modified form of the trityl group has also been proposed for the N<sup>6</sup> position <sup>228</sup>. The preparation of N<sup>6</sup>-4, 4', 4''-tris(benzoyloxy)trityl-dA was found to offer up to eight times the stability of dAbz. This was dependent on the detritylation conditions being the same in 1% TFA/dichloromethane, four times as stable in 2% DCA/dichloromethane and eight times as stable in 80% AcOH. The phthaloyl group offered better depurination resistance in 2% DCA/dichloromethane and 80% AcOH <sup>229a, b</sup>.

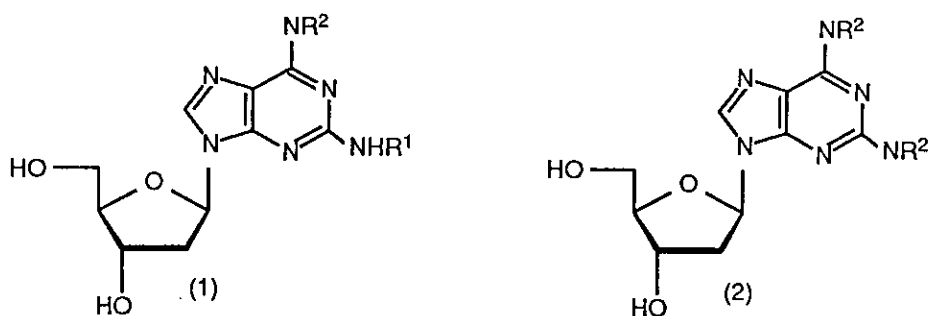
At the same time it was more base labile than the benzoyl group being easily removed with 0.5M hydrazine hydrate/pyridine/AcOH solution in one minute. N<sup>6</sup>-phthaloyl protected deoxyadenosine has been successfully used in triester syntheses. Later dichloro- and tetrachloro-substituted phthaloyl groups were proposed<sup>230</sup>. These were more effective than benzoyl protection but less effective than phthaloyl alone. N<sup>6</sup>-succinyl dA was also used showing improvements over phthaloyl protection<sup>230</sup>. In an effort to change the site of protonation of N<sup>6</sup>-benzoyl dA a different approach was proposed by Morin<sup>231</sup>. N1 oxidation reduced the depurination during detritylation with benzene sulphonic acid<sup>232</sup>. Other N<sup>6</sup> protection that has been tried include the levulinyl<sup>233</sup>, 9 fluorenylmethoxycarbonyl (Fmoc)<sup>234</sup>, 2-nitrophenylsulphenyl (Nps)<sup>235</sup>, and 2, 2, 2-trichloro-t-butyloxycarbonyl (TCBOC)<sup>236</sup> groups. The most promising improvements in N<sup>6</sup> protecting group choice came with the application of amidine reagents to solid phase synthesis monomers<sup>237</sup>.

Amidine protection had been used for ribonucleotide<sup>238</sup> protection before being reintroduced for deoxyribonucleotides<sup>239</sup>. This took the form of the N-dimethylamino group which was easily placed on the N<sup>6</sup> of dA with dimethylformamide dimethylacetal in DMF to yield only N<sup>6</sup>-protected material in 90% yield. However when exposed to the conditions of oligonucleotide synthesis in particular those used for detritylation (ZnBr<sub>2</sub> in nitromethane/methanol or 3% TCA in dichloromethane) the amidine was rapidly hydrolysed<sup>240</sup>. This led to the proposal by McBride and Caruthers of the N<sup>6</sup>-(N-methyl 2-pyrrolidine amidine)-dA as a suitably protected nucleoside for phosphitylation and use in oligonucleotide synthesis<sup>240</sup>. The half life of this compound in 1M TCA was three times that of the benzoyl protected dA (60 mins as opposed to 20 mins). In an effort to extend the range of usable amidines and to increase their hydrolysis resistance a series of N<sup>6</sup>-dialkylformamidines has been prepared with increasing steric bulk in the dialkyl portion<sup>237</sup>. During this work it was noted the N<sup>6</sup>-dimethylformamide group was sensitive to methanol. This may have contributed to its hydrolysis during the detritylation test with ZnBr<sub>2</sub><sup>240</sup>.

N<sup>6</sup>-diisopropylformamide-dA was resistant to depurination in 2% DCA / dichloromethane at 25°C with a half life of about 2 hrs but its deprotection with ammonia at 60°C took 12 hrs.

More suitable was N<sup>6</sup>-di-n-butylformamidino-dA which was stable to all oligonucleotide synthesis conditions readily deprotected with concentrated ammonia in 3 hrs at 60°C and offering an approximate 20 fold suppression of depurination.

Thus formamidino protection of the N<sup>6</sup> position of dA offers the following advantages. The protecting groups are easy to prepare from cheap starting materials. Unlike traditional protecting strategies the unprotected nucleoside can be selectively N<sup>6</sup> protected without the need for transient protection of the 3' and 5'-OH functions<sup>241</sup>. The formamidino is easily removed under the standard deacylation conditions used for the other base protecting groups *ie.* concentrated ammonia treatment. Most importantly they can greatly increase the resistance of dA to acid catalysed C1'-N9 cleavage during the cyclical detritylation procedure. These features make it an extremely attractive option for the protection of the N<sup>6</sup> position of 2-amino-2'-deoxyadenosine. The N<sup>2</sup> position may be protected with either by analogy to dG, the isobutyryl group or with a second amidino group (1) (2).



R<sup>1</sup> = isobutyryl, R<sup>2</sup> = CH-N(n-C<sub>4</sub>H<sub>9</sub>)<sub>2</sub>

### 6.1 Influence of N<sup>6</sup> Protection on Depurination

The mechanism by which depurination takes place does not make it obvious that the choice of N<sup>6</sup> protection will influence its rate. Depurination is known to proceed by an A1 type mechanism involving the the protonation of the purine in a pre-equilibrium step followed by the rate determining step of C1'-N9 glycosyl bond cleavage to form a cyclic glycosyl

oxocarbenium ion<sup>242-245</sup>. This mechanism is supported by linear pH rate profiles due to the rate constants for the reactions via the mono and dication species being almost equal, a secondary deuterium isotope effect and the influence of changes in the glycosyl moiety on the hydrolysis rates as observed for the hydrolysis of ribosides versus deoxyribosides<sup>244</sup>. Deoxyadenosine has two reactive nitrogens as sites for protonation, N1 and N7 so the pre-equilibrium step may be the formation of either a mono or dication both of which lead to C1'-N9 bond cleavage. 1 Methyl-adenosine depurinates at almost the same rate as adenosine implying the leaving group in the latter case is an N1, N7 dication species<sup>243</sup>. What then is the effect of N<sup>6</sup> protection on the depurination rate?

A series of 9-(2-deoxy-β-D- *erythro* -pentafuranosyl) purines was examined by Oivanen and Lonnberg<sup>245</sup>. The effect of the 6-substituent on the observed rate was only slight with both electron donating and electron withdrawing groups being rate reducing. This can be rationalised in the case of electron withdrawing groups for example as due to the reduction in pre-equilibrium base protonation while at the same time the purine is an improved leaving group. These two effects on the rate limiting glycosyl cleavage work against each other and in the main cancel each other out. This makes it difficult to account for the observation that N<sup>6</sup>-benzoyl dA undergoes hydrolysis with a half life one eighth that of dA in 0.2M HCl/dioxan at 20°C<sup>246</sup>. At  $[H^+] < 0.01M$  the hydrolysis rate is about one order of magnitude greater while at  $[H^+] > 0.2M$  it is only about twice as great<sup>245</sup>. Plots of rate profiles for N<sup>6</sup>-benzoyl dA deviate from linear with an inflection at  $pH = pK_1$  indicating  $k_1/K_1$  is greater than  $k_2/K_2$ <sup>247</sup>. Experimental evidence supports this with  $k_1/K_1$  for N<sup>6</sup>-benzoyl dA being fifteen times that for dA. It would therefore appear that N<sup>6</sup>-acyl derivatives increase the rate of depurination via the mono-cation species.

To explain this the relative basicities of the N1 and N7 sites are considered. By observing changes in the N1 and N7 <sup>15</sup>N NMR shifts in TFA the protonation sites can be determined. 2'-Deoxyadenosine showed a change in only the N1 shift indicating N1 protonation in agreement with the greater basicity of N1<sup>248</sup> while N<sup>6</sup>-benzoyl dA was protonated at N1 and N7. In other words the relative basicity of the nitrogens is changed in the latter case probably

due to a reduction in electron density of the pyrimidine ring caused by the electron withdrawing N<sup>6</sup>-substituent. It has been proposed that N7 protonation of dA leads to more rapid C1-N9 cleavage. The N<sup>6</sup>-benzoyl dA appears to depurinate more readily due to its preferential N7 protonation. This may be due to N7 protonated adenine being a more stable leaving group than the N1 tautomer <sup>249</sup>. This change in protonation site had been observed previously in C8 hydrogen exchange experiments of N<sup>6</sup>-substituted dA where protonation occurred at N7 preferentially <sup>250</sup> with the degree of change depending on the acyl group present, benzoyl having a greater effect than an acetyl group <sup>251</sup>. Changes in nitrogen basicity influence depurination rates via the monocation route. The use of dialkylformamide for dA (2-amino-dA) N<sup>6</sup> protection such as the sterically hindered di-n-butylformamide can be expected to reduce the changes in protonation and N1, N7 relative basicity by eliminating suppression of the adenosine amidine resonance.

Oligonucleotides were to be synthesised using commercially available solid supports and cyanoethyl phosphoramidites of the four major bases. In order to exploit the efficiencies of both time and reagents these offer, 2-cyanoethyl phosphoramidites of 2-amino-dA were to be prepared. These would take two forms. To avoid problems during the detritylation of the in house and commercial phosphoramidites di-n-butylformamide protection of the N<sup>6</sup> position of 2-amino-dA was to be used. In one phosphoramidite this would be in conjunction with N<sup>2</sup>-isobutryl protection (10) while a second would be prepared with both the N<sup>2</sup> and N<sup>6</sup> protected as the formamide group (13). When this work was initiated 2-cyanoethyl reagents were not commercially available. 2-Cyanoethylphosphorodichloridite was to be prepared from PCl<sub>3</sub> and 2 cyanoethanol by the method of Claesen *et al* <sup>200</sup>. This would then be used in the synthesis of 2-cyanoethyl N, N-diisopropylchlorophosphoramidite from which the nucleotide monomers can be obtained. Synthesis of d(TA)<sub>n</sub> sequences requires the solid support to be functionalised with 2-amino-dA. It was proposed to prepare CPG derivatised with 5'-O-4, 4'-dimethoxytrityl protected 2-amino-dA bound by a succinate linkage <sup>194</sup>.

## 7.0 Thermal Denaturation of Oligonucleotides (UV Melting)

In solution DNA exists as an equilibrium between double helix and single stranded (or coil) forms. Under favourable conditions which allow duplex formation the equilibrium lies towards double helix DNA. In this form all hydrogen bonding potential between bases is satisfied and the bases are highly stacked. When a solution of double helix DNA is heated it converts to two single strands if enough energy is supplied to overcome the free energy of duplex formation. When this occurs the DNA duplex is said to have melted. The melting process is usually co-operative and rapid, and is accompanied by a reduction in hypochromicity. Prior to melting a small amount of base stacking may be lost, while the melting temperature itself is characterised by a sudden destacking and unzipping of the helix to give two single strands. After melting only small increases in the UV absorbance of the sample are seen. Thus DNA double helix melting is an all or nothing process. The mid point of the process, the sharp increase in UV absorption, can be taken as the melting temperature  $T_m$ . Alternatively a differential melting curve may be plotted in which  $\Delta A/\Delta T$  is plotted against  $T$ . The maximum of this curve will occur at the value of  $T_{1/2}$ , the steepest point of the melting profile. For direct comparison between two or more similar oligonucleotides under the same conditions  $T_m$  can be taken as the maximum of the first differential of the melting profile. A more rigorous analysis of melting profiles can also be followed <sup>252, 253</sup>.

For an all or nothing model the equilibrium constant for the formation of a double stranded helix from two single strands can be expressed as

$$K = \theta/2(1-\theta)^2C_t$$

where  $\theta$  is the fraction of strands in double helices and  $C_t$  is the total strand concentration.

The enthalpy of helix formation is given by

$$d\ln K/d(1/T) = -\Delta H/R.$$

The relationship between the shape of the differential melting curve and the enthalpy of the transition from helix to coil is given by the equation

$$d\theta/d(1/T) = -\theta(1-\theta)/(1+\theta) \cdot \Delta H/R.$$

When  $\Delta A/\Delta T$  versus  $T$  is at a maximum  $d^2\theta/d(1/T)^2 = 0$  and this gives  $\theta = \sqrt{2}-1$  and

$d\theta/d(1/T) = -0.17\Delta H/R$ . At the half height value of  $d\theta/d(1/T)$  the relationship gives  $\theta = 0.11$  or 0.81.  $K$  can be found for these two values of  $\theta$  to give

$$\Delta H = -4.37/(1/T_{1/2} - 1/T_{3/4}) \text{ calories}$$

$$\Delta H = -18.24/(1/T_{1/2} - 1/T_{3/4}) \text{ joules.}$$

At  $T = T_{1/2}$ ,  $\theta = 0.414$  and  $K_{1/2} = 2.4/C_1$

$$\Delta G = -RT_{1/2} \ln K_{1/2}.$$

The entropy of the transition is given by

$$\Delta S = (\Delta H - \Delta G)/T_{1/2}.$$

Corrected values of  $T_m$  when  $\theta = 0.5$  can be calculated from the approximate relationship

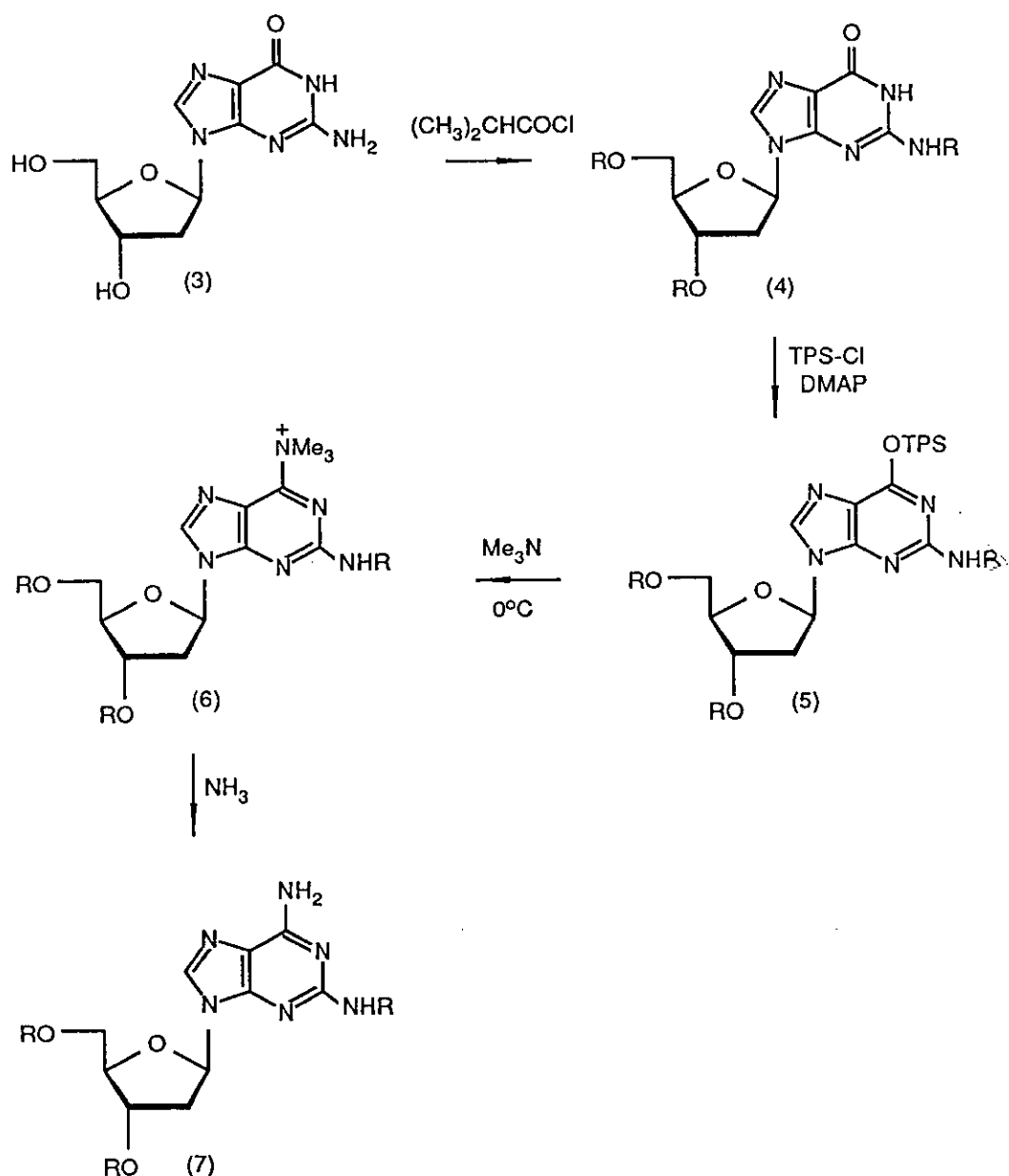
$$T_m = T_{1/2}(1 - T_{1/2}/|\Delta H|).$$

## 8.0 2-Amino-2'-deoxyadenosine

This has previously been prepared from dA by utilisation of of an N1-oxide <sup>251</sup>. This when treated with cyanogen bromide yielded a cyclic N1-N<sup>6</sup> product which with methanolic ammonia gave N<sup>6</sup>-cyano-dA-N1-oxide. Reaction of this with methyl iodide gave the N1-methoxy-oxide which upon reaction with NaOH and heating in ethanol rearranged to 2-amino-6-methylamino-2'-deoxypurine. This may be converted to 2-amino-dA by catalytic hydrogenation.

The observation by Sung <sup>254a, b</sup> that thymidine could be converted to 5-methyl-dC using relatively mild conditions provided another approach to the synthesis of 2-amino-dA <sup>255</sup>. The conversion of dT to 5-methyl-dC was achieved via an intermediate 4-triazolopyrimidone. This was prepared from the reaction of suitably protected dT with p-chlorophenyl phosphorodichloridate and 1, 2, 4-triazole in pyridine. It is of interest to note this is similar to a known side reaction in phosphoramidite chemistry which converts dG to 2-amino-dA <sup>256</sup>. After work-up, treatment with ammonia gave 5-methyl-dC. The ammonia treatment can be performed after oligonucleotide synthesis to convert the base *in situ*. This method was used in the synthesis of oligonucleotides with either mixed dG and dA' or exclusively dA' sites <sup>255</sup>. However in another case this method did not work <sup>95</sup>. When the reaction was done as before only highly polar N<sup>6</sup>-pyridinium chloride salts were obtained. This prompted the authors to use the previously described N<sup>6</sup>-N-methyl-2-pyrrolidine-amidine protection <sup>95, 248</sup> on 2-amino-dA prepared by the method of Gaffney *et al* <sup>91</sup>. This route utilises the reactivity of the O<sup>6</sup> position of dG towards sulphonating and alkylating agents which which may be exploited as a general synthesis for 6 substituted dG <sup>257</sup> or O<sup>6</sup> protection <sup>258</sup>. Sulphonation of the O<sup>6</sup> position of dG has been reported to proceed easily <sup>259a, b, 260</sup> to give a sulphonated nucleoside which readily undergoes displacement reactions with nitrogen nucleophiles to introduce an N<sup>6</sup> function. The actual method involves the sulphonation of triisobutyl-dG with triisopropylbenzenesulphonyl chloride in dichloromethane TEA solutions using a DMAP catalyst. The sulphonated derivative obtained may then be converted directly to 2-amino-dA by a liquid ammonia displacement of the sulphonyl group, although as this reaction is

reported to be slow, it is more convenient to exploit the facile displacement of an intermediate trimethylamino group which is known to easily undergo nucleophilic displacement<sup>257, 262, 263</sup>. Thus the O<sup>6</sup>-sulphenyl-dG is converted to N<sup>6</sup>-trimethylamino-dG with trimethylamine. This upon displacement with liquid ammonia yields triisobutyryl-2-amino-dA Scheme 1.



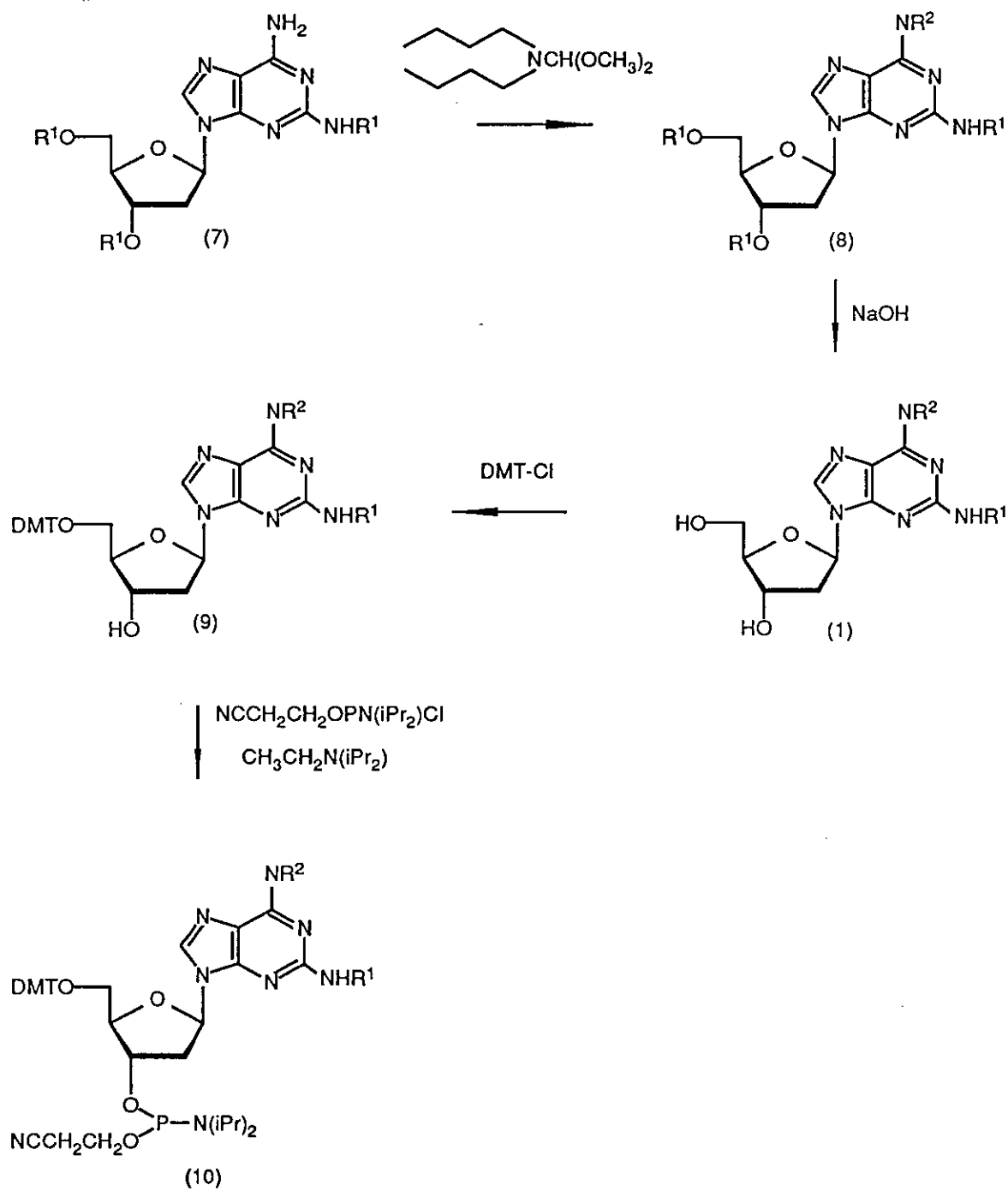
Scheme 1. R = isobutyryl.

At this point N<sup>6</sup> protection may be introduced if desired to give a fully acylated and N<sup>6</sup> protected nucleoside. Alternatively the 3' and 5'-OH protection may be selectively removed with NaOH or full deprotection can be achieved by incubating with hot concentrated ammonia solution.

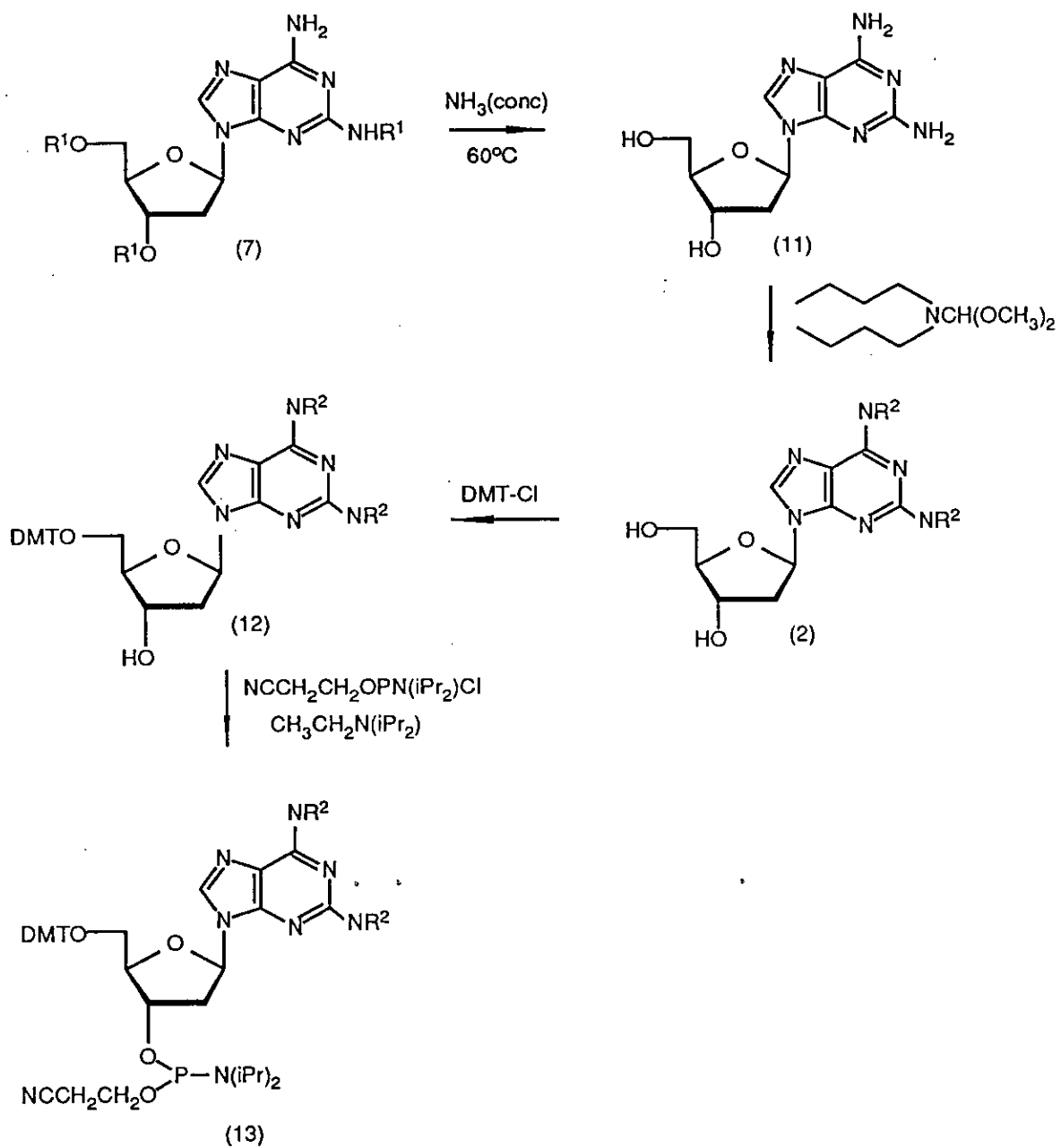
For the synthesis of the two formamidine protected nucleosides both of the above strategies were used. The triisobutyl protection of dG not only protects the N<sup>2</sup> and 3' and 5'-OH functions but has the additional advantage of greatly increasing the solubility of dG in organic solvents. The removal of the 3' and 5'-OH isobutyl groups with NaOH precludes the use of dibenzoyl protection for the N<sup>6</sup> position since this readily reverts to the monobenzoyl form when the nucleoside is treated with dilute NaOH<sup>241</sup>. The N<sup>2</sup>-isobutyl and fully deprotected 2-amino-dA can then be protected at the N<sup>6</sup> and N<sup>2</sup>, N<sup>6</sup> positions respectively by reaction with di-n-butylformamide dimethylacetal to give the protected nucleosides suitable for phosphorylation with 2-cyanoethyl N, N-diisopropylchlorophosphoramidite. Thus through this route the two different 2-amino-2'-dA oligonucleotide synthesis monomers should be obtained Scheme 2, 3.

### 8.1 2-Amino-2'-deoxyadenosine Nucleosides

2-Amino-2'-deoxyadenosine was prepared *via* its 2-N, 3'-O, 5'-O triisobutyl form by the method of Gaffney *et al*<sup>91</sup>. This method utilised the reactivity of the O<sup>6</sup> of guanine, after suitable protection of the other reactive positions on the nucleoside, towards sulphonation by TPS-Cl. 2'-Deoxyguanosine was 2-N, 3'-O, 5'-O protected by reaction with isobutyl chloride in dry pyridine. Sulphonation gave the O<sup>6</sup>-TPS derivative. This underwent nucleophilic displacement with triethylamine followed by liquid ammonia to yield the protected nucleoside 2-N, 3'-O, 5'-O-triisobutyl 2-amino-2'-deoxyadenosine. This was used as starting material for the synthesis of the two phosphoramidites.



Scheme 2.  $R^1$  = isobutyryl,  $R^2$  =  $\text{CH-N(n-C}_4\text{H}_9)_2$ .



Scheme 3. R<sup>1</sup> = isobutyryl, R<sup>2</sup> = CH-N(n-C<sub>4</sub>H<sub>9</sub>)<sub>2</sub>.

It had previously been found that the use of di-n-butylformamide dimethyl acetal as a protecting agent for the N<sup>6</sup> of adenine significantly reduced its susceptibility to depurination in acid media<sup>237</sup>. This protecting group was prepared from di-n-butylamine and N, N-dimethylacetal dimethylformamide. In this way it fulfils a major requirement in the design of a protecting group strategy for oligonucleotide synthesis being very cheap to prepare. Protection of the N<sup>6</sup> of the tri-isobutyrylated nucleoside was readily achieved due to the solubilising properties of the isobutyryl groups. Stirring the nucleoside in CH<sub>2</sub>Cl<sub>2</sub> with di-n-butylformamide dimethylacetal gave the fully protected nucleoside. Once in place the di-n-butylformamide group has an additional advantage of being fluorescent under UV light on silica TLC. This may be used to monitor subsequent reactions. Cleavage of the isobutyryl esters with dilute NaOH freed the 3' and 5'-OH groups allowing subsequent reaction with DMT-Cl to protect the 5' position giving the precursor to the first 2-amino-2'-deoxyadenosine phosphoramidite.

The second target phosphoramidite was both N<sup>2</sup> and N<sup>6</sup> protected using di-n-butylformamide. This necessitated removing all the isobutyryl protection from the nucleoside which was achieved using more forcing conditions. The nucleoside was incubated in 0.88 ammonia solution for five days. The free nucleoside has very poor solubility and this hindered initial attempts to convert it to the N<sup>2</sup>, N<sup>6</sup>-bis(di-n-butylformamido)nucleoside. After trials with a range of solvents, stirring a solution of the nucleoside in DMA with di-n-butylformamide dimethyl acetal for 18 hours under nitrogen with the exclusion of light gave the product. Reaction with DMT-Cl gave the precursor to the second 2-amino-2'-deoxyadenosine phosphoramidite.

## 8.2 Stability Studies on Model Nucleosides

Previous efforts to synthesise oligonucleotides containing dA' residues have been hampered by the ease with which depurination of the protected nucleoside takes place during synthesis. This is particularly bad with an N<sup>6</sup>-acyl protected nucleoside. With reference to previous work on the depurination of dA the di-n-butylformamide group had been adopted for

N<sup>6</sup> protection and N<sup>2</sup>, N<sup>6</sup> protection of dA'. In order to confirm that this protection strategy would also be suitable for dA', a series of depurination studies of model nucleosides was performed. To this end the five model nucleosides (1), (2), (11), (14), (15) were exposed to a commercial detritylation solution of 3% TCA in CH<sub>2</sub>Cl<sub>2</sub>. Decomposition of the nucleosides was followed by TLC or HPLC methods. The two unprotected nucleosides dA (14) and dA' (11) were included as references. The results are shown in Table 1.

It can be seen that, as expected, the N<sup>6</sup>-benzoyl protection has a severe effect on the susceptibility of the nucleoside to acidic depurination. This is in agreement with previous findings<sup>91</sup> where attempts at oligonucleotide synthesis were made using N<sup>6</sup>-benzoyl protected dA'. By analogy with studies on depurination of dA it may be proposed that benzoyl protection has the effect of altering the N1 and N7 relative basicities *via* its electron withdrawing properties to favour N7 protonation. This species then acts as a good leaving group in the monocation depurination route. The half life is short enough to render the nucleoside unuseable in standard DMT-phosphoramidite oligonucleotide synthesis. The N<sup>6</sup>-di-n-butylformamide nucleoside has a much increased half life. Although not examined before as a protecting group for dA' this is in agreement with the findings of studies on dA<sup>237</sup>. The formamidine protection is an electron donating group which will not promote N7 protonation in the same way as acyl protection does. In this nucleoside the basicity of the N1 position over that of N7, will be maintained. Furthermore the monocation formed by N1 protonation can be resonance stabilised by delocalisation into the formamidine moiety giving a reduction in C1'-N9 bond cleavage and depurination.

Protection of both the exocyclic amines with the formamidine gives a similar half life to that with N<sup>2</sup>-isobutyryl protection. The N<sup>2</sup>-protecting group does not exert a modifying influence on depurination. In this position there is no resonance stabilisation of the N1 monocation by delocalisation. No improvement in nucleoside stabilisation is offered by the use of the second formamidine group. Thus choices between these two nucleosides would have to rest on other features such as their relative ease of synthesis or their physical characteristics.

Both nucleosides have a similar half life, of a magnitude to suggest they would be stable during oligonucleotide synthesis with no marked base loss.

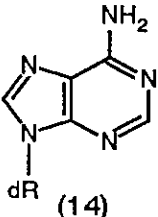
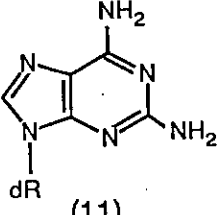
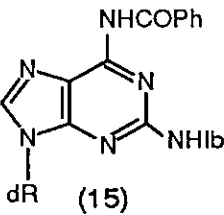
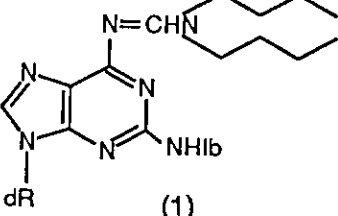
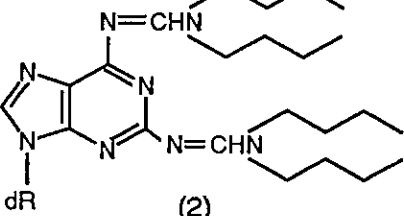
Nucleoside	Half Life
 <p>(14)</p>	45mins
 <p>(11)</p>	35mins
 <p>(15)</p>	2mins
 <p>(1)</p>	1.5-2hrs
 <p>(2)</p>	1.5-2hrs

Table 1. Depurination of nucleosides in 3%TCA/CH<sub>2</sub>Cl<sub>2</sub> solution, dR = 2'-deoxyribose, lb = isobutryl.

From this it was decided phosphoramidites of the two formamidine protected nucleosides should be suitable for DNA synthesis.

Deprotection of nucleoside (1) was examined by incubating the nucleoside with 0.88 ammonia solution at 60°C for 3 hours, the normal deprotection method. Monitoring the reaction by TLC showed the nucleoside was fully and cleanly base deprotected. This was in agreement with the results for dA<sup>237</sup>. Hence it was concluded the di-n-butylformamide group met the required properties of offering depurination resistance and ease of removal after synthesis using standard phosphoramidite synthesis and deprotection strategies.

## 9.0 Synthesis of 2-Amino-2'-deoxyadenosine Phosphoramidites

Oligonucleotides were synthesised using 2-cyanoethyl N, N-diisopropyl nucleoside phosphoramidites with tetrazole activation. The general synthesis cycle for oligonucleotide assembly is shown in Fig. 30.

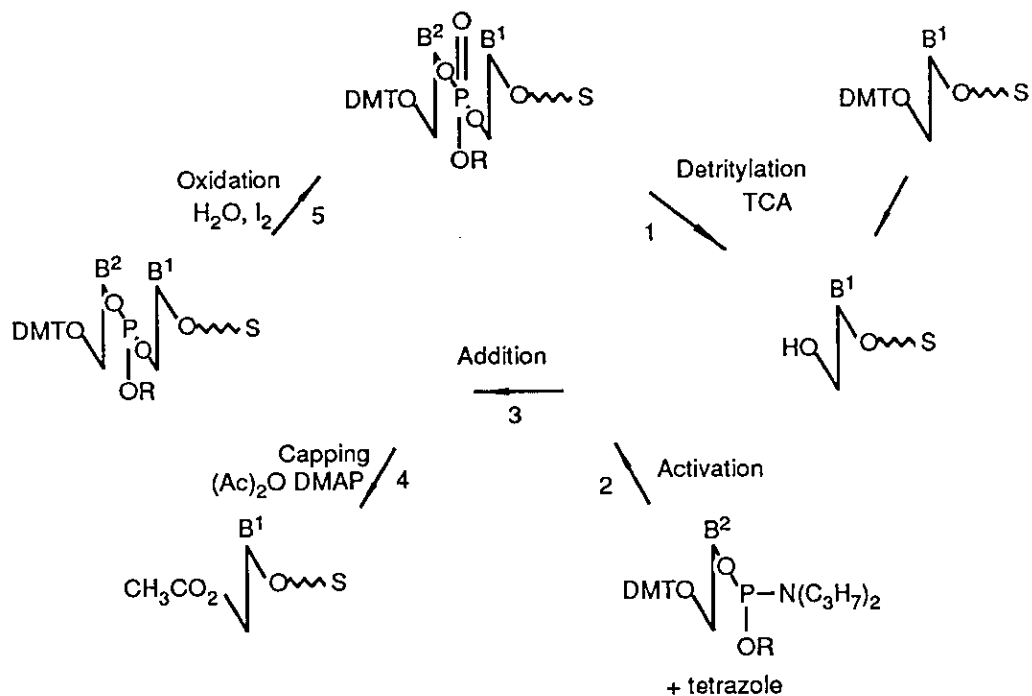


Fig. 30. DNA synthesis cycle, S = solid support, R = NCCH<sub>2</sub>CH<sub>2</sub>O  
B<sup>1</sup>, B<sup>2</sup> = protected base

2-Cyanoethyl N, N-diisopropylchlorophosphoramidite was required to prepare the reactive nucleoside phosphoramidites where these were not commercially available. At the start of this project 2-cyanoethyl N, N-diisopropylchlorophosphoramidite was not available. This was prepared by the methods of Claisen *et al* <sup>200</sup> and Sinha *et al* <sup>205</sup> with slight modification.

The reaction of phosphorus trichloride with 2-cyanoethanol readily gave 2-cyanoethyl phosphorodichloridite which was obtained in good yield after distillation. Conversion to the monochloridite was more problematic. It was found that even with very short reaction times the reaction between N, N-diisopropylamine and the phosphorodichloridite gave predominantly one product, 2-cyanoethyl N, N, N', N'-tetraisopropylphosphorodiamidite. This has been prepared before by others <sup>264a, b</sup> and used in the *in situ* method <sup>265</sup>. The desired product was obtained by using much more dilute solutions of dichloridite and amine and working up the reaction immediately addition of the amine solution was complete. In this way it was possible to obtain reasonable yields of 2-cyanoethyl N, N-diisopropylchlorophosphoramidite after distillation. Shortly after optimisation of this synthesis 2-cyanoethyl N, N-diisopropylchlorophosphoramidite became available from commercial sources. This was used for the synthesis of nucleoside phosphoramidites.

The phosphoramidite (**10**) was prepared from the 5'-OH DMT protected base protected nucleoside (**9**) using the method of Sinha *et al* <sup>205</sup> by reaction with 2-cyanoethyl N, N-diisopropylchlorophosphoramidite. Repeated precipitation of the crude product into hexane at -78°C failed to yield solid material. Purification of the product by chromatography on silica with ethyl acetate:acetonitrile 50:50 gave the nucleoside phosphoramidite as an immobile oil. TLC showed partial resolution of the asymmetric phosphorous diastereomers and two <sup>31</sup>P NMR signals were observed at 149.1452 and 149.036. Phosphoramidite (**13**) was prepared in the same manner from the protected nucleoside (**12**). Precipitation from ethyl acetate into hexane at -50°C gave the phosphoramidite as an oil. Both phosphoramidites were stored frozen under nitrogen.

### 9.1 Oligonucleotide Synthesis using dA' Phosphoramidites

Initial syntheses used the phosphoramidite (**10**). This was used as a 0.1M solution in oligonucleotide synthesis grade acetonitrile with syntheses performed using the standard small scale (0.2 μmole) cyanoethyl phosphoramidite coupling cycle SSCE103A. The first oligonucleotide chosen for synthesis was the sequence dCGCA'AATTTGCG an analogue of

the Dickerson dodecamer<sup>46</sup>. Monitoring the coupling efficiency by the released DMT cation gave a coupling efficiency for the phosphoramidite of 95%. This was slightly lower than is desirable for optimum performance where coupling efficiencies in excess of 98% are required. The oligonucleotide was deprotected according to standard methods and examined by HPLC. Reverse phase HPLC of the crude product showed two peaks eluting with a retention time of the right order for oligonucleotidic material. The presence of the two peaks suggested incomplete deprotection of the oligonucleotide with the later eluting, and much smaller, of the two peaks being the more hydrophobic protected material. In order to test this assumption the oligonucleotide was incubated for a further 24 hours in the deprotection solution with the addition of a trace amount of  $\text{NH}_4\text{OAc}$ <sup>237</sup>. Reverse phase HPLC analysis after this time showed only one major peak, the magnitude of the secondary peak was much reduced. Depurination of nucleotides leaves oligonucleotides very susceptible to chain cleavage at the apurinic sites when exposed to concentrated ammonia solution in a manner that is similar to Maxam-Gilbert sequencing<sup>211a, b</sup>. Thus if any depurination had taken place during oligonucleotide assembly this would show up, post ammonia deprotection, as strand cleavage. Strong anion exchange HPLC will separate oligonucleotides of different chain length with good resolution. Analytical SAX-HPLC of the crude oligonucleotide did not show any signs of degradation of the oligonucleotide which could be attributed to depurination and ammonia treatment. The result obtained was that of a good synthesis with a small amount of n-1 material as expected Fig. 31. The results of this initial synthesis indicated di-n-butylformamide gave good protection against depurination of dA' in oligonucleotide synthesis.

The oligonucleotide dCA'TCTC was synthesised to allow investigation of the deprotection conditions. Coupling was again assayed by the released DMT cation and found to be >99%. RP-HPLC analysis of the crude oligonucleotide deprotection solution following incubation for 20 hours showed two DNA peaks in a 1:2 ratio. Heating at 70°C for a further 4 hours with the addition of a trace of  $\text{NH}_4\text{OAc}$  reduced the peak ratios to 1:1 which after a further 4 hours became 2:1. The deprotection appeared complete after 48 hours Fig. 32, 33. Repeating the synthesis to confirm the deprotection again gave a coupling efficiency >99%. The monomer

appears to work well under normal oligonucleotide synthesis conditions but longer deprotection times are required.

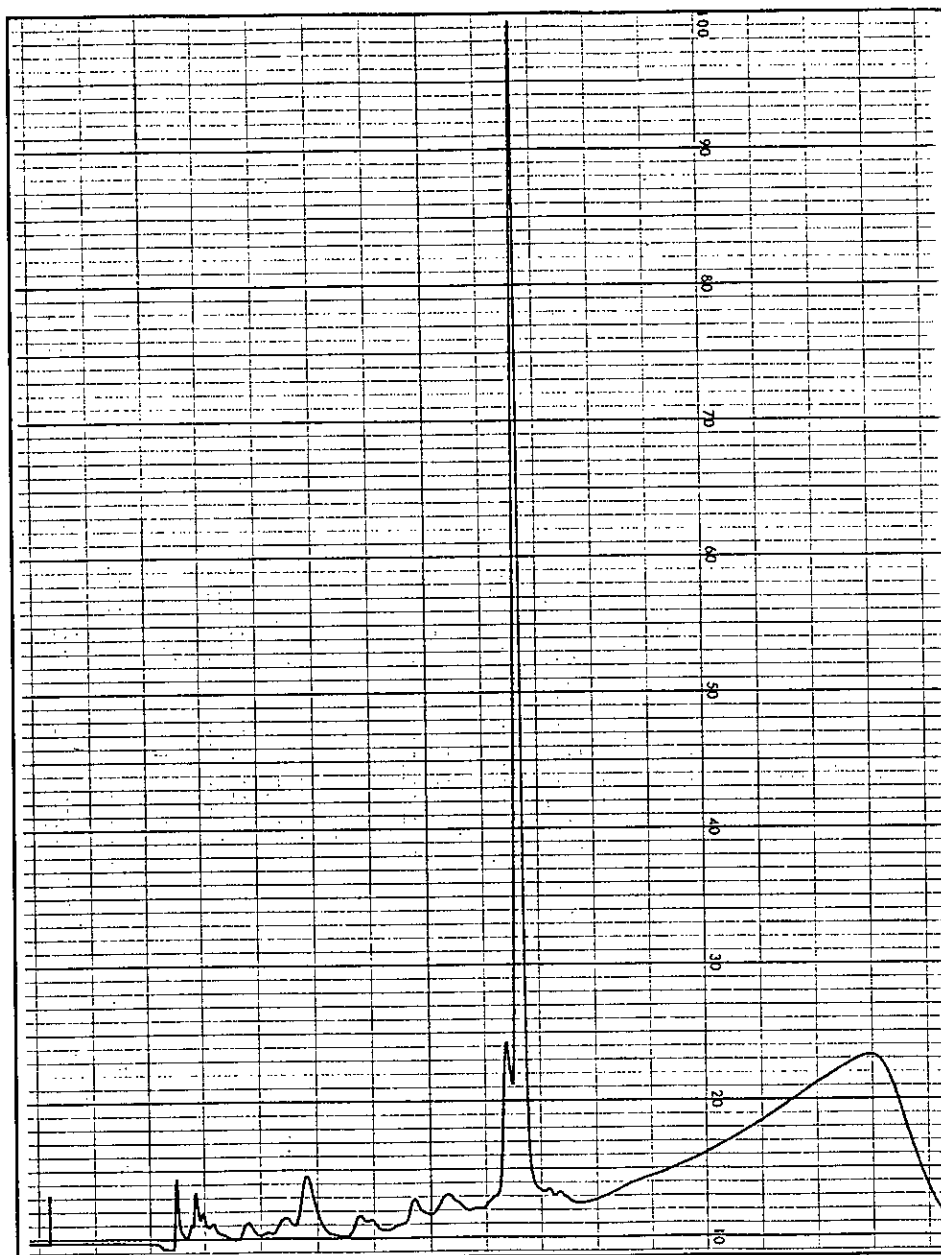


Fig. 31. SAX-HPLC of unpurified deprotected dCGCA'AATTTGCG 0.64Aufs. Buffer A; 0.03M  $\text{KH}_2\text{PO}_4$  in 20%  $\text{CH}_3\text{CN}$ , Buffer B; 0.66M  $\text{KH}_2\text{PO}_4$  in 20%  $\text{CH}_3\text{CN}$  pH 6.4, gradient 10-100% B in 22mins.

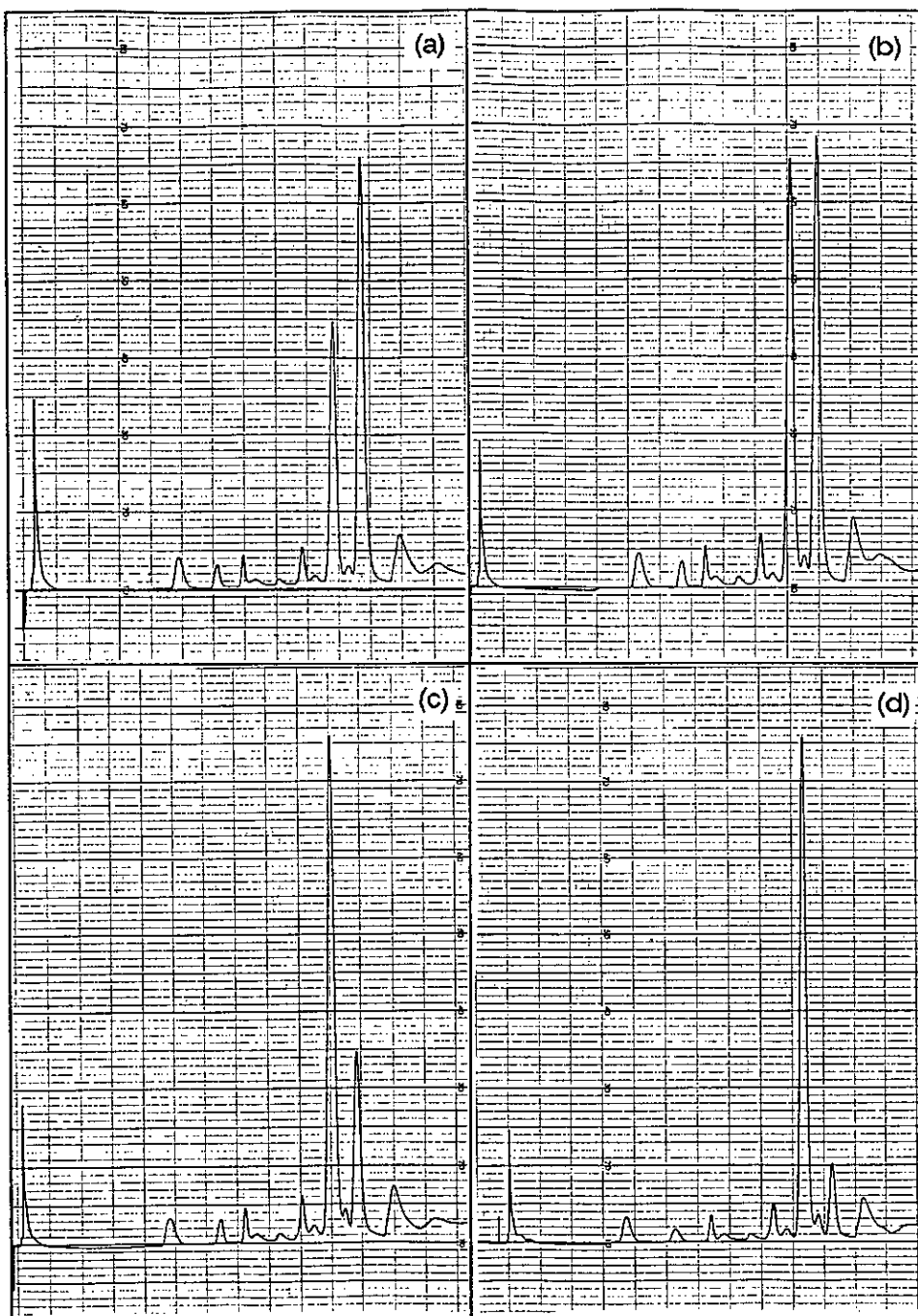


Fig. 32. Deprotection of dCA'TCTG in ammonia monitored by RP-HPLC. (a) 20hrs 60°C (b) as (a) plus 4hrs 70°C 100 $\mu$ l 0.1M NaOAc (c) as (b) plus 4hrs (d) as (c) plus 20hrs, 0.64Aufs Buffer A; 0.1M NH<sub>4</sub>OAc, Buffer B 0.1M NH<sub>4</sub>OAc in 20% CH<sub>3</sub>CN, gradient 0-45% B in 27mins.

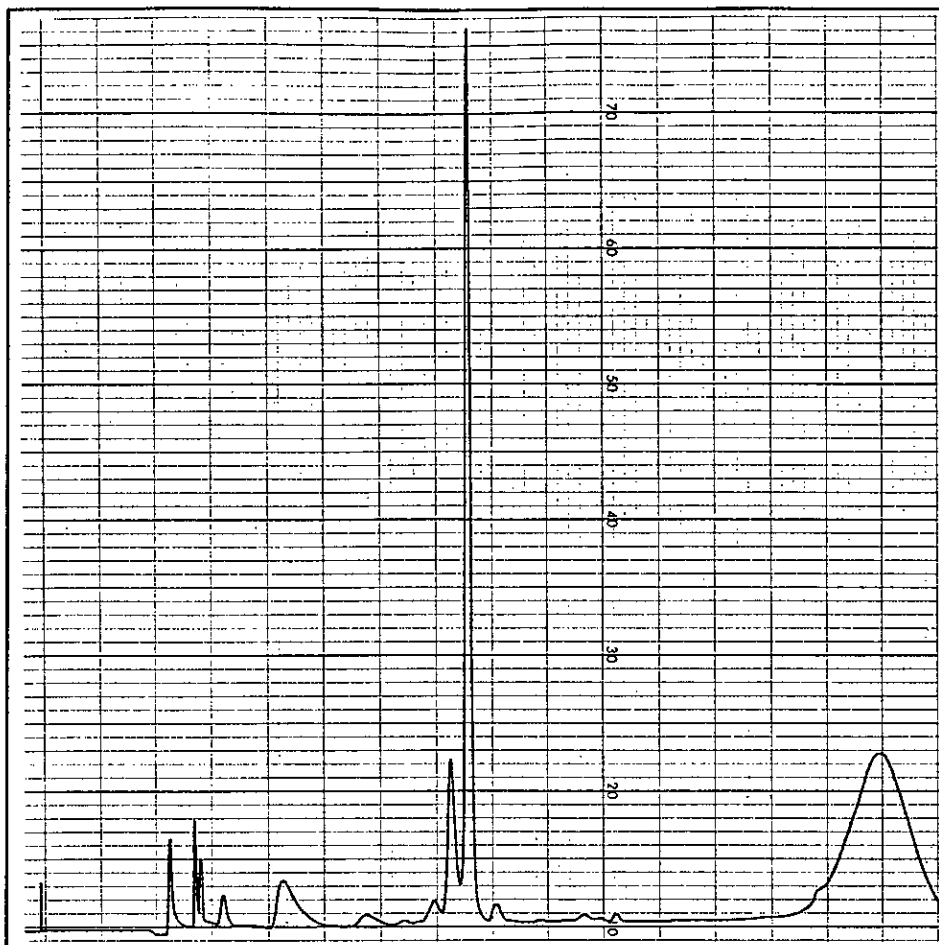


Fig. 33. SAX-HPLC of unpurified deprotected dCA'TCTG 0.64Aufs. Buffer A; 0.03M  $\text{KH}_2\text{PO}_4$  in 20%  $\text{CH}_3\text{CN}$ , Buffer B; 0.66M  $\text{KH}_2\text{PO}_4$  in 20%  $\text{CH}_3\text{CN}$  pH 6.4, gradient 10-35% B in 4-20mins.

The second dA' phosphoramidite (**13**) prepared in the same way was used in an initial synthesis to test its efficacy. A  $1\mu\text{mol}$  synthesis of the oligonucleotide dCGCA'AATTTGCG was attempted. As with previous syntheses the coupling efficiency at each step was assayed by the released DMT cation. Coupling of the dA' phosphoramidite to the previous residue dA was only 80%. Coupling of the next residue dC to the dA' was effectively zero. The synthesis failed. The initial coupling of the dA' phosphoramidite was low but still significant. However addition to the 5'-OH of this residue after detritylation did not take place. This implies the phosphoramidite is viable in its own coupling reactions but does not allow reaction to take

place at the 5'-OH. Comparing the two dA' phosphoramidites suggests a possible reason for this. The two di-n-butylformamidine groups are bulky and flexible. This is not a problem in the N<sup>6</sup> position but may be so in the N<sup>2</sup> position. The N<sup>2</sup> protecting group will be able to dictate the accessibility of the 5'-OH to the incoming activated nucleotide. In particular the replacement of the N<sup>2</sup>-isobutyryl protection with the formamidine group may cause steric hindrance at the 5'-OH.

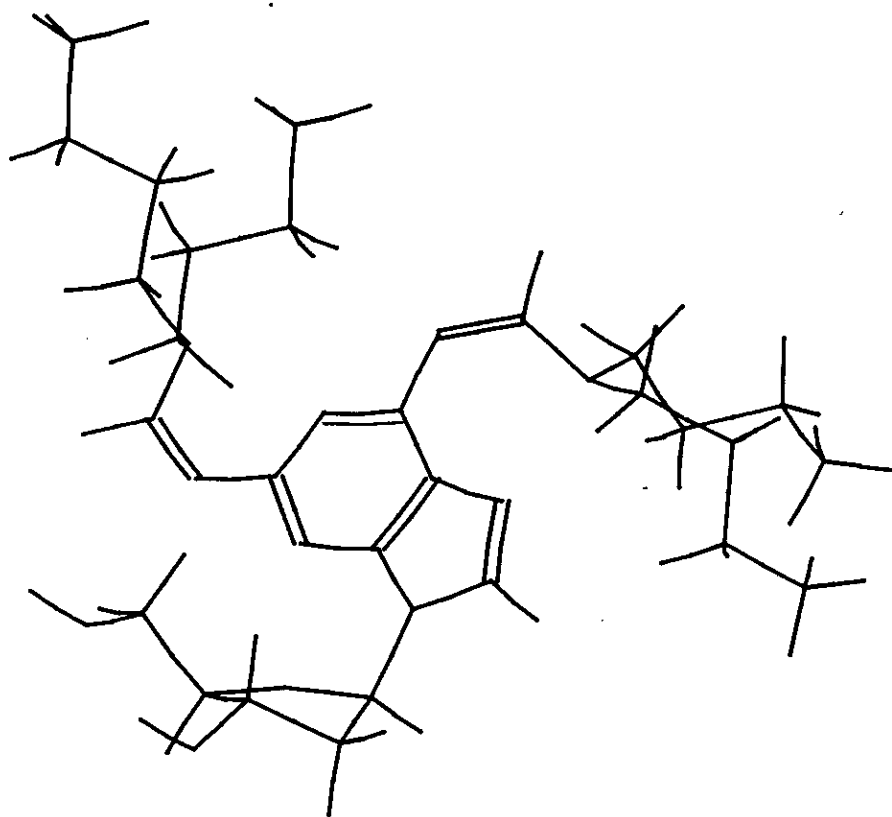


Fig. 34. Computer model of phosphoramidite (13) with  $\chi = 60^\circ$  showing hinderance of 5'-OH. Model constructed with Alchemy II.

2-Amino-2'-deoxyadenosine may be considered to be like dG with regard to the 2-amino group. When not hydrogen bonded in a base pair the base is not constrained in its movements to adopt an *anti* conformation. It has been found that the *syn/anti* equilibrium for the 2'-monophosphate of guanine favours the *syn* form<sup>266</sup>. If dA' in the growing oligonucleotide single strand can be considered with reference to its 2-amino group to be like guanine-2'-monophosphate this implies the base orientation is towards the *syn* form where the 2-amino protecting group is brought towards the 5'-OH. Computer model building of the two phosphoramidites highlights the difference in accessibility of the 5'-OH in the two cases. With the torsion angle  $\chi$  in the *syn* region the 5'-OH group is hindered. Where di-*n*-butylformamide protection is used for both amine functions this effect is greater. The mobile arms of the group may prevent the incoming phosphoramidite-tetrazole complex from approaching the 5'-OH. Fig. 34 shows the arrangement around the 5'-OH site with the  $\chi$  torsion angle fixed in a *syn* conformation. The model was constructed with  $\chi = 60^\circ$ .

## 9.2 Mixed Sequence Z-DNA dA' Oligonucleotides

One of the reasons it was desirable to synthesise oligonucleotides with dA' was to look at the effect of this base residue on DNA structures, in particular those of Z and X-DNA. Some previous attempts have been made to crystallise mixed sequence oligonucleotides as Z-DNA. These have met with differing degrees of success. The crystal structures of the sequences dCA'TGTG and dCGTA'CG have been refined as Z-DNA<sup>179</sup>. To extend this series the sequence dCA'TA'TG was synthesised. This sequence is an analogue of the known Z-DNA hexamer dCGCGCG<sup>51</sup> but has four pseudo A.T base pairs in its middle. The A'.T base pairs are A.T like but also have the 2NH<sub>2</sub> group which is thought to favour Z-DNA. Thus the sequence is an extreme example of what might under forcing conditions be Z-DNA. (There have been no reports in the literature of AT analogues of the sequence crystallising as Z-DNA or indeed AT runs crystallising as Z-DNA.) The oligonucleotide was synthesised as six 1mmole syntheses, rigorously deprotected and purified by HPLC Method 1. Attempts were made to crystallise the oligonucleotide using the sitting drop method with central well

vapour diffusion. The conditions used initially were those found previously for the crystallisation of other hexamers. To a solution of DNA in sodium cacodylate (50mMolar, pH 6.5) was added  $MgCl_2$  (100mMolar) and spermine tetrahydrochloride (30mMolar). The wells containing DNA were diffused against 20% iPrOH. Attempts were also made using high salt concentrations to force a Z-DNA conformation and aid crystallisation using 2Molar  $MgCl_2$ . Crystallisation was unsuccessful. The added effect of two extra A'.T base pairs in the hexamer may be enough to inhibit crystal growth.

UV melting experiments were performed on the hexamer and its native equivalent dCATATG. These highlighted the increase in duplex melting temperature afforded by the formation of three hydrogen bond A'.T base pairs. At the DNA concentrations used a difference in melting temperature of 17K was observed. The results of these experiments are shown in Table 2 and Fig. 35a, b.

<u>Oligonucleotide</u>	$KJmole^{-1}$ <u><math>\Delta H</math></u>	$KJmole^{-1}$ <u><math>\Delta G</math></u>	$Jmole^{-1}$ <u><math>\Delta S</math></u>	K <u><math>T_m</math></u>	K <u><math>T_m</math> (corr.)</u>
CATATG	-153.2	-30.0	-437.2	281.75	281.23
CA'TA'TG	-192.2	-32.2	-534.4	299.38	298.91

Table 2. UV melting results for oligonucleotides shown, measured at 254nm in SSC buffer.  $T_m$  = maximum of first derivative of the melting profile,  $T_m$  (corr.) = adjusted value.

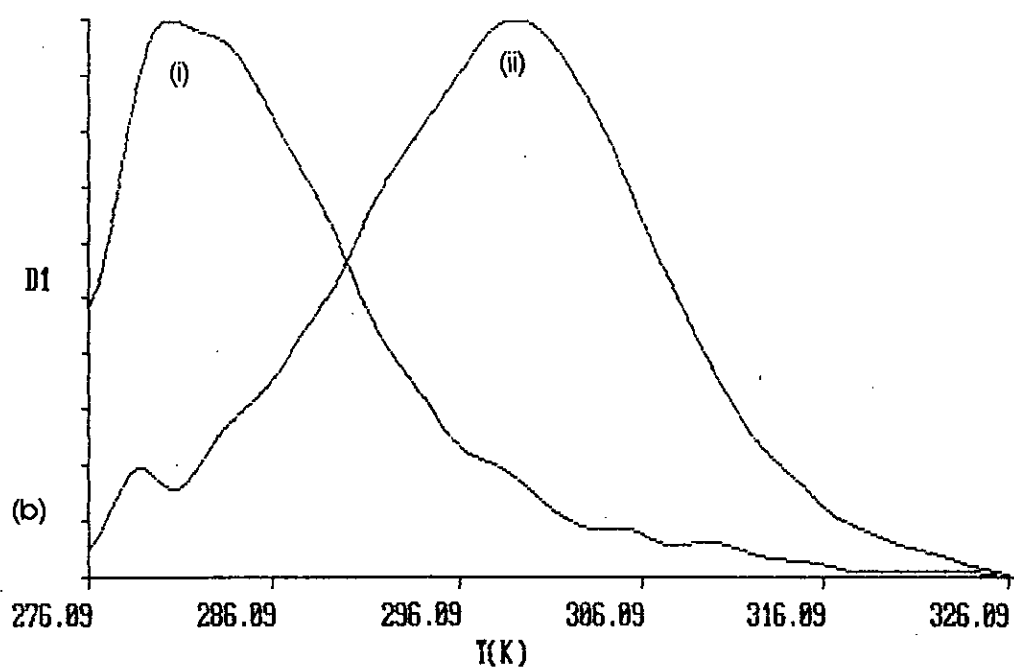
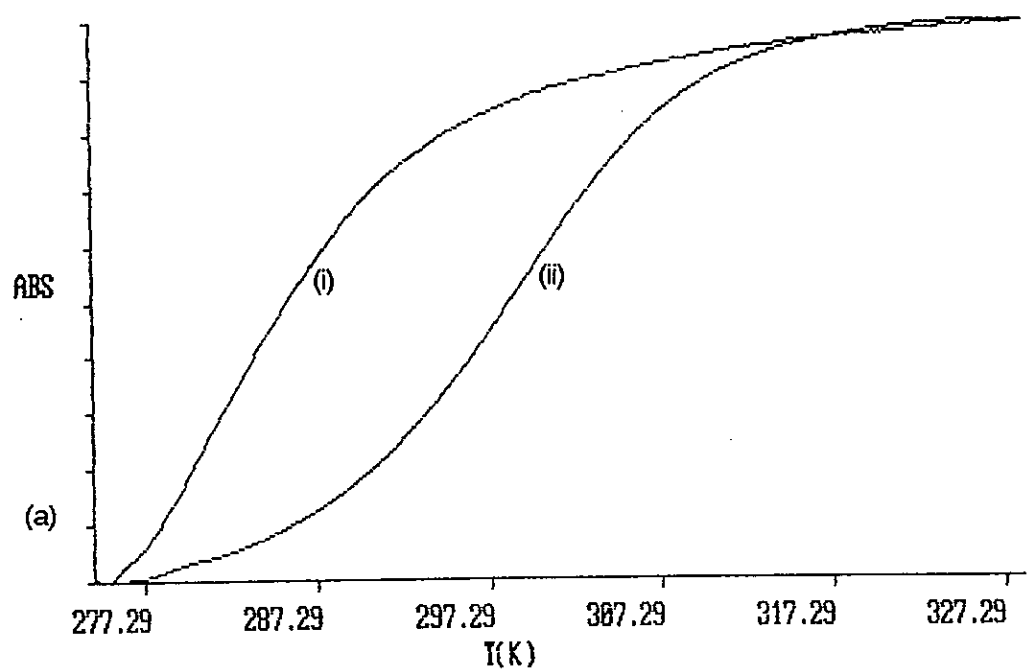


Fig. 35. (a) UV melting profiles of (i) dCATATG and (ii) dCA'TA'TG measured in SSC buffer at 254nm, arbitrary vertical scaling. (b) First derivatives of (i) dCATATG and (ii) dCA'TA'TG melting profiles, arbitrary vertical scaling.

### 9.3 Use of dA' in Stable Genetic Probes

There is currently much interest in and effort being expended on the mapping out of whole genomes, with interest focusing on the human genome. As an essential step in this process it is necessary to bridge the gap between genetic and molecular distances in the molecular analysis of mammalian genomes. There are problems in analysing and manipulating lengths of thousands of kilo base pairs which may separate what are genetically closely linked regions in mammalian genomes. At present methods have been developed that allow either the high resolution analysis of short regions surrounding a mutation or lower resolution physical mapping of large regions, or entire chromosomes. It has been proposed that ordered clone libraries designed to cover whole chromosomes will be an essential prerequisite to the high resolution molecular analysis of large parts of the human genome 267.

Theoretical considerations indicate that genetic probes of 8 or 9 bases in length would be suitable tools for probing genetic structures<sup>268</sup>. Sequences of this length occur about once in every 100 000 base pairs. However an inherent problem with such short sequence probes is their low melting temperature and their frequent occurrence in the *E. Coli* genome. This gives a high background hybridisation to DNA prepared by *in situ* lysis. These problems can be avoided by constructing 10mer probes with sequences of biased dinucleotide frequencies which occur only rarely in the *E. Coli* genome and in only 25% of mammalian cosmids. During work with such probes the wash protocol after hybridisation has to be designed so that perfectly matched hybrids give much stronger signals than partially mismatched hybrids. Even when these precautions are taken some probes misbehave and melt off too early or cross hybridise at high stringency. Attempts to hybridise 10mer probes to *in situ* cosmid colony lifts however do not work. Oligonucleotides containing dA' are known to have higher melting temperatures than native oligonucleotides. It was therefore proposed to synthesise a series of short genetic probes in which dA was replaced by dA'. This should reduce the problems encountered during washing of hybrids and allow better 10mer hybridisation. The following sequences were synthesised on a 0.2  $\mu$ mole scale.

<u>Probe</u>	<u>Sequence</u>
HLP-9tet	dA'CTA'TCGA'C
HLP-10tet	dCA'CTA'TCGA'C
51P-10	dGA'A'GA'GA'GTT

### 9.3.1 Hybridisation of Trial Probes

Hybridisations were performed in 0.5M NaPi pH 7.2, 7% SDS at 27°C overnight to 10ng of DNA immobilised onto nylon membranes. The hybrids were then washed with 3.0M tetramethylammonium chloride/50mM Tris-HCl, pH 7.6/2mM EDTA/0.1% SDS at a range of temperatures (2 x 2 minutes). Prior to washing unhybridised material was removed by rinsing the membranes in 6 x SSC at room temperature. Binding was assayed using an Ambis plate scanner after which membranes were exposed to X-ray film. Results of a typical film developed from the DNA dot blot experiments are shown for HLP-10tet Fig. 36. The HLP-10tet sequence contains dA' residues while the TET-10 sequence is a conventional oligonucleotide. It can be clearly seen that there is a greater find for hybridisation of the HLP-10tet sequence. The DNA dots are laid out in the form of a 7 and a 3. The 3 contains the TET sequence while the 7 does not. This shows that there is discrimination in the hybridisation of the dA' sequence and that it will only hybridise to it's compliment. It can be concluded that the diaminopurine sequence has hybridised and is stable to the washing conditions. The non-diaminopurine oligonucleotide however, has failed to hybridise or has been removed from the membrane during the washing procedure.

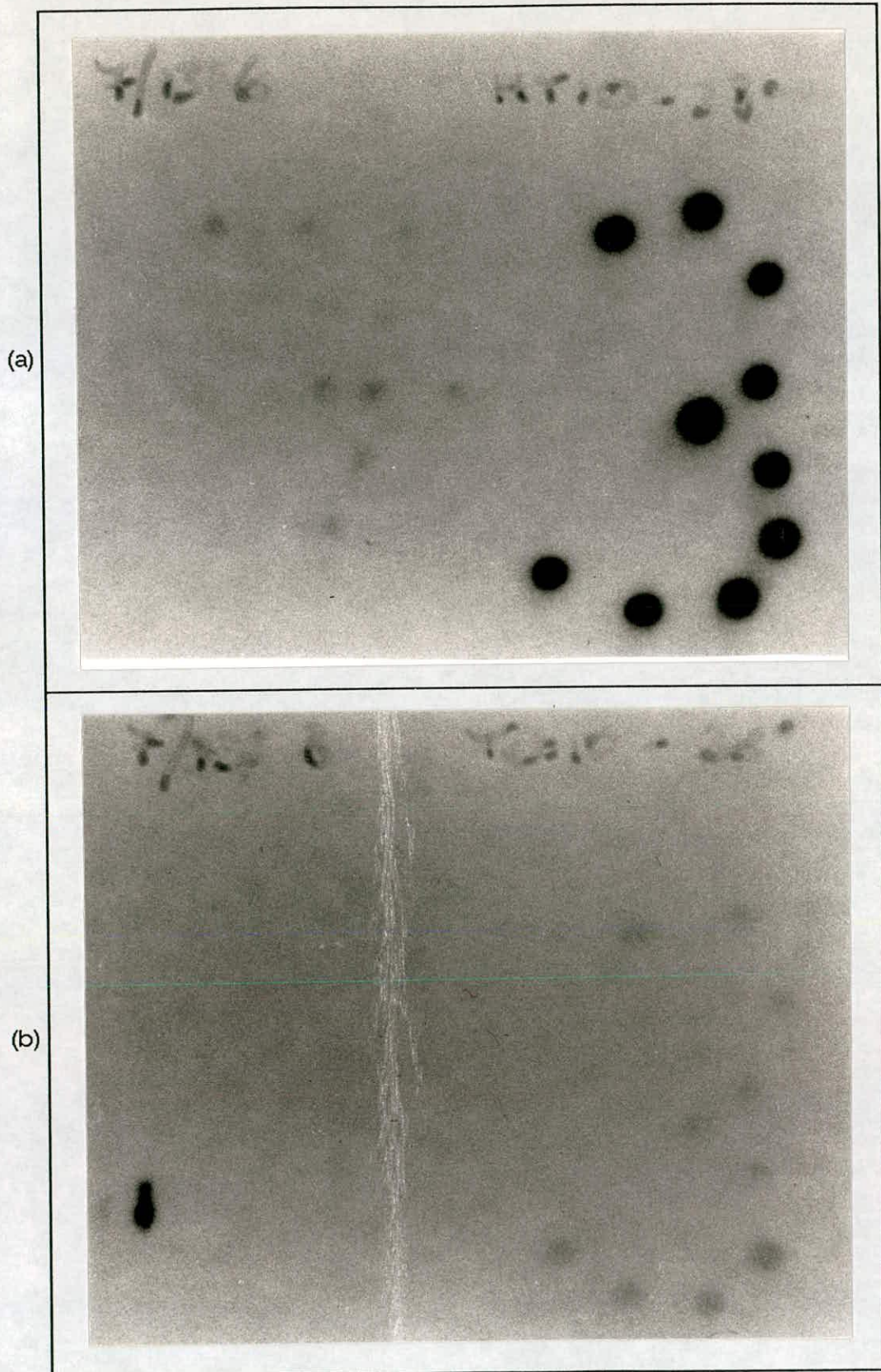


Fig. 36. DNA dot blots of (a) HLP-10tet (b) Tet10. See text for details.

### 9.3.2 UV Thermal Denaturation of Trial Probes

To complement the hybridisation work UV melting experiments were performed on the trial dA' genetic probes and their native sequence equivalents. These were performed in SSC buffer at an oligonucleotide concentration of 0.36 $\mu$ molar in duplex. The results of these experiments are shown in Fig. 37a-f and Table 3. As had been predicted even at low concentrations of duplex there is a marked increase in the thermal stability of the dA' containing probes over the native ones. The stability increase is greatest in the oligonucleotide with four dA' substitutions as would be expected from implying that increased  $T_m$  is due to diaminopurine incorporation.

<u>Oligonucleotide</u>	KJmole <sup>-1</sup> <u><math>\Delta H</math></u>	KJmole <sup>-1</sup> <u><math>\Delta G</math></u>	Jmole <sup>-1</sup> <u><math>\Delta S</math></u>	K <u><math>T_m</math></u>	K <u><math>T_m</math> (corr.)</u>
HLP9-tet	-226.9	-38.8	-604.5	311.10	310.67
HLP9-Ntet	-188.2	-37.9	-494.7	303.65	303.10
HLP-10tet	-308.2	-40.0	-837.8	320.13	319.80
HLP-10Ntet	-269.5	-39.2	-734.4	313.62	313.26
51P-10	-317.9	-40.2	-863.0	321.78	321.45
51P-10N	-231.1	-38.4	-626.5	307.57	307.16

Table 3. UV melting of trial probes and their native equivalents.  $T_m$  = maximum of first derivative,  $T_m$  (corr.) = adjusted value.

Fig. 37. (Over)(a), (c), (e) UV melting profiles of trial probes and native equivalents measured in SSC buffer at 260nm. Solutions 0.36 $\mu$ molar in duplex. (b), (d), (f) First derivatives of UV melting profiles. Arbitrary vertical scaling.

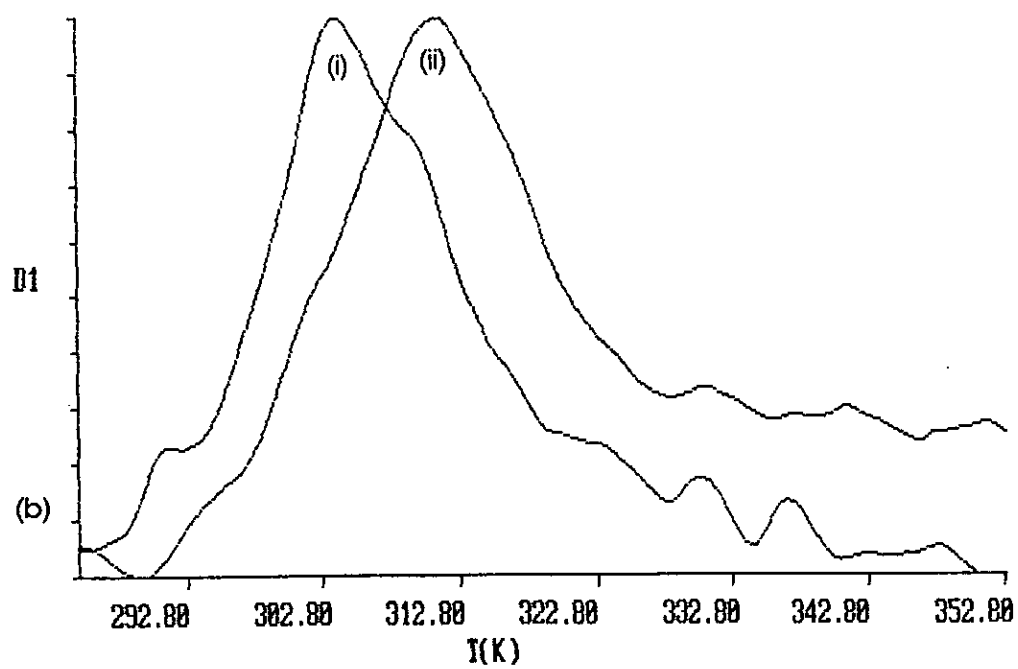
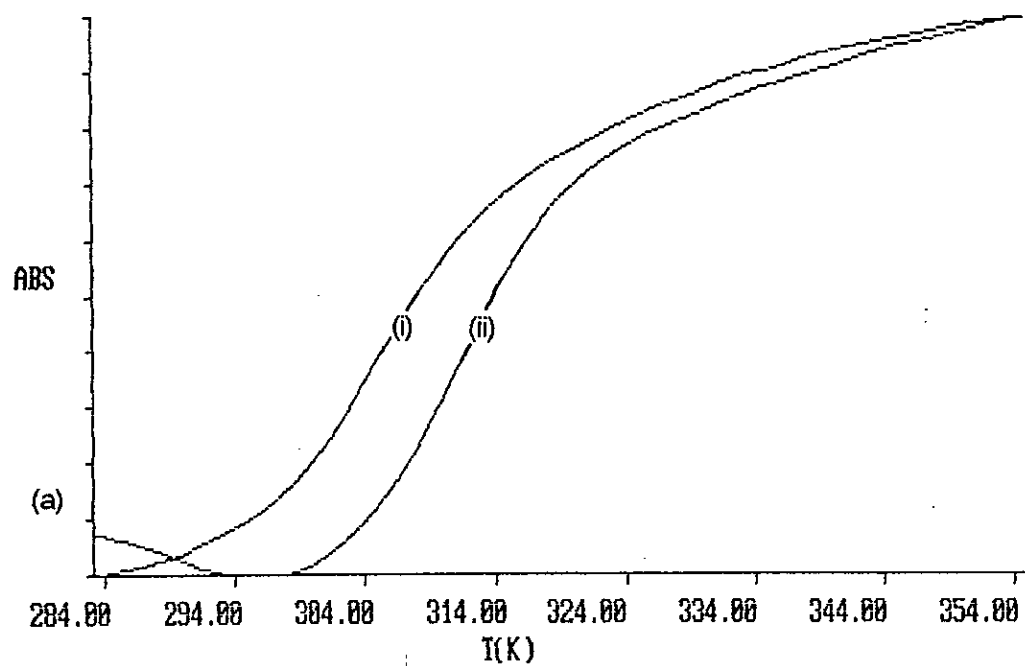


Fig. 37(a) (i) dACTATCGAC (ii) dA'CTA'TCGA'C. (b) (i) dACTATCGAC (ii) dA'CTA'TCGA'C.

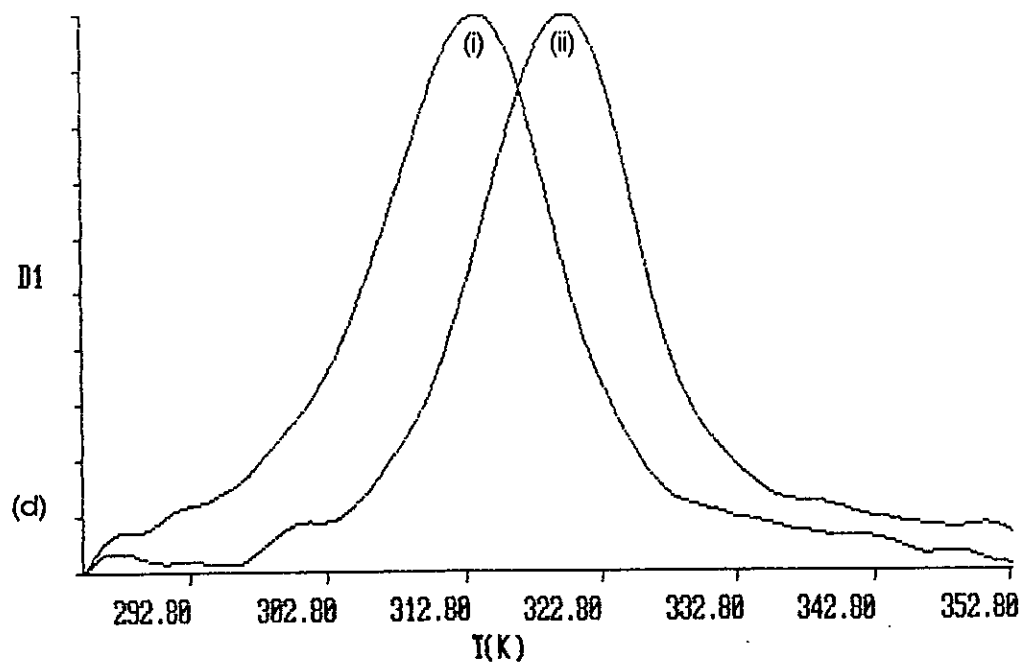
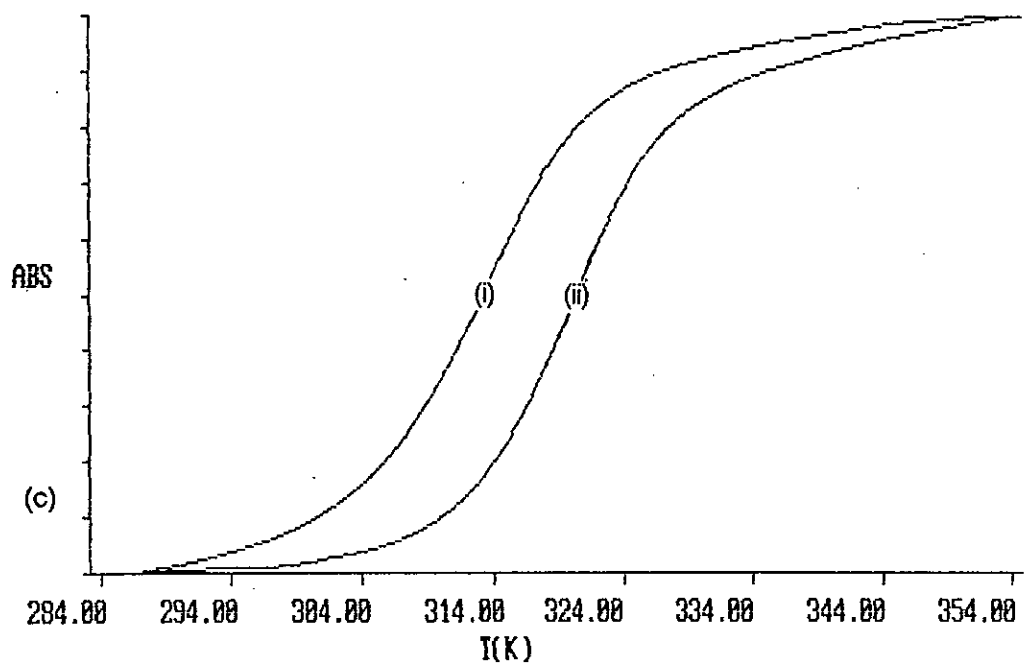


Fig. 37. (c) (i) dCACTATCGAC (ii) dCA'CTA'TCGA'C  
 (d) (i) dCACTATCGAC (ii) dCA'CTA'TCGA'C

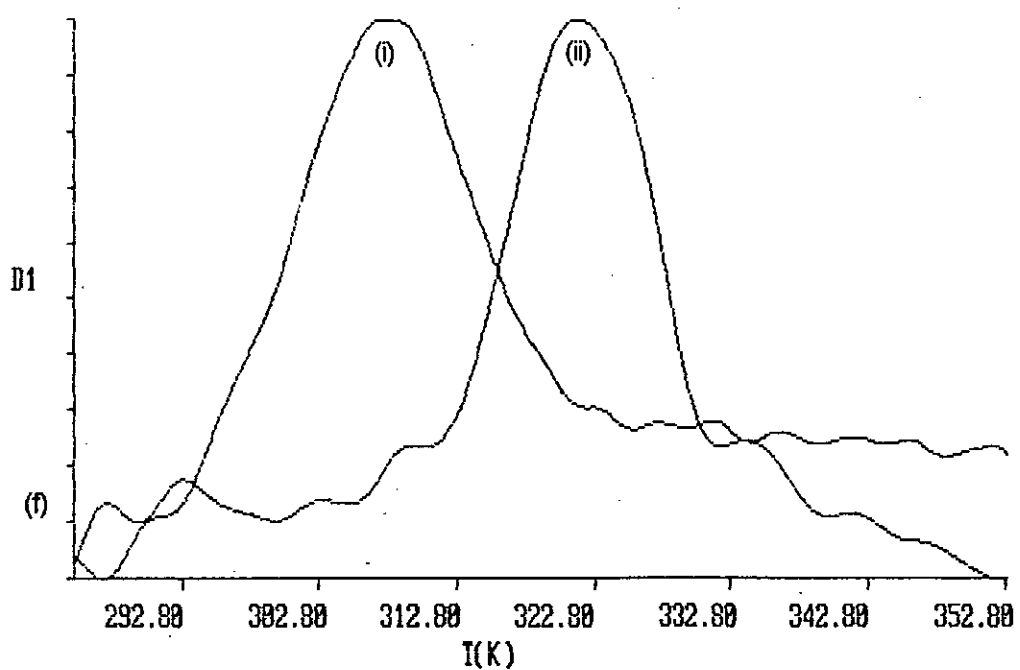
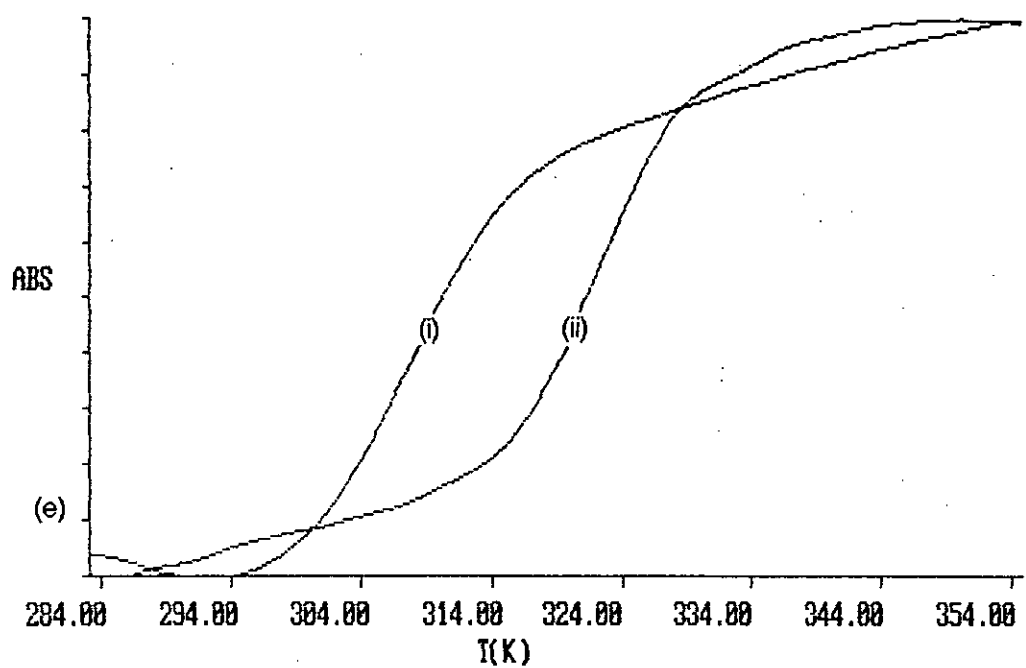


Fig. 37 (e) (i) dGAAGAGAGTT (ii) dGA'A'GA'GA'GTT  
 (f) (i) dGAAGAGAGTT (ii) dGA'A'GA'GA'GTT.

#### 9.4 X-DNA and Oligonucleotides Containing dA'

In order to gain information on the structure of short sequence oligonucleotides that might be able to adopt the novel and uncharacterised X-form the three oligonucleotides dTA'TA', dTA'TA'TA' and dTA'TA'TA'TA' were synthesised. Proprietary CPG is not available functionalised with dA' so this had to be prepared. The chosen support was CPG derivatised with a long chain alkyl group terminating with a free amine. Protected 2-amino-2'-deoxyadenosine nucleoside was linked to this *via* a succinate linkage by a series of sequential coupling reactions until no further increase in resin loading could be detected. Assaying the CPG loading by release of DMT cation from the support bound nucleoside gave a loading value of  $17\mu\text{mole g}^{-1}$ . Free sites on the resin were capped by reaction with acetic anhydride. The free flowing resin was loaded into standard  $1\mu\text{mole}$  synthesis columns and used in the normal way. The oligonucleotides were synthesised as six  $1\mu\text{mole}$  syntheses. Very lengthy exposure to concentrated ammonia was required for deprotection. Purification was achieved by HPLC, Method 1.

Crystallisation may be expected to take place most readily in those sequences where stable duplexes can form although other factors such as local DNA concentration and crystal packing forces may have a strong influence. UV melting experiments were performed using each oligonucleotide to establish whether stable duplexes or aggregates formed. For the tetramer dTA'TA' there was no indication of duplex melting above 273K suggesting the oligonucleotide was substantially premelted. It has been estimated from NMR measurements that the oligonucleotide, even at millimolar concentrations, has a duplex population of only 40% at 273K<sup>171</sup>. This is effectively a lower population than would be predicted at  $T_m$  according to the all or nothing DNA melting theory<sup>252</sup>. Similarly no UV melting data could be obtained on the native oligonucleotide dTATA which has a duplex population under the same conditions of 26%<sup>171</sup>. UV melting experiments with dTA'TA'TA' and the native sequence dTATATA gave a  $T_m$  value for the diaminopurine oligonucleotide of 292K at a duplex concentration of 0.011 millimolar in SSC Fig. 38, while the native hexamer failed to show any temperature dependence in its UV absorbance indicating it was not present in a duplex form.

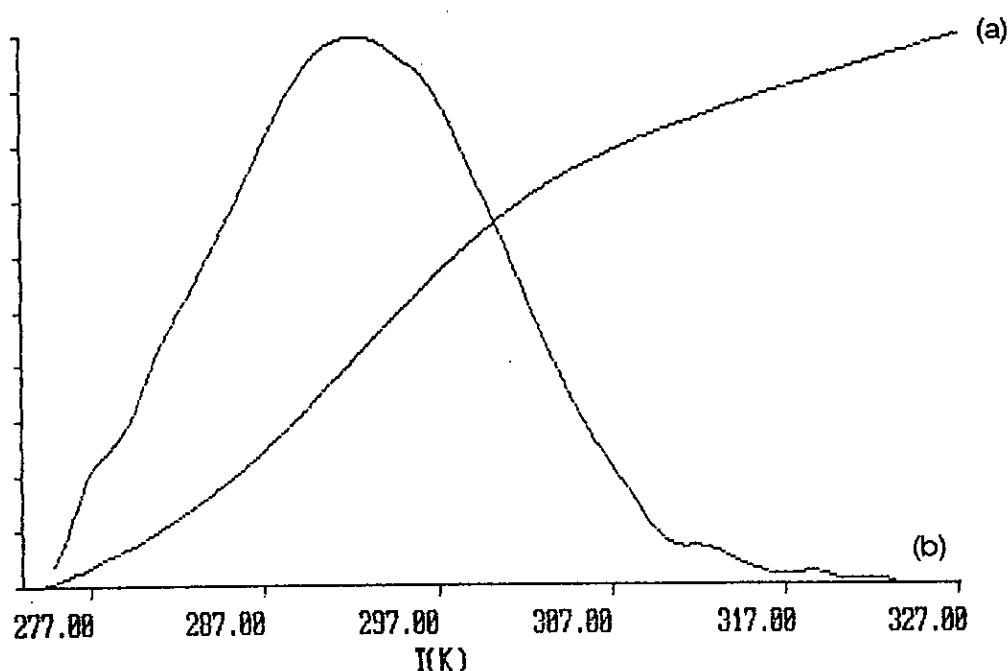


Fig. 38. (a) UV melting profile of  $d(TA')_3$  in SSC buffer at 260nm. (b) First derivative of melting profile. Arbitrary vertical scaling.

Comparative melting studies of  $dTA'TA'TA'TA'$  and  $dTATATATA$  indicated both sequences were capable of forming duplexes as shown by their temperature dependent UV absorbance Fig. 39a, 40a. The increase in duplex stability afforded by  $dA'$  was 28K Table 4. Thus both the hexamer and octamer should be good candidates for crystallisation while the tetramer might crystallise due to other stabilising influences.

	KJmole <sup>-1</sup>	KJmole <sup>-1</sup>	Jmole <sup>-1</sup>	K	K
<u>Oligonucleotide</u>	<u>ΔH</u>	<u>ΔG</u>	<u>ΔS</u>	<u>T<sub>m</sub></u>	<u>T<sub>m</sub> (corr.)</u>
TATATATA	-126.1	-28.6	-345.7	282.00	281.37
TA'TA'TA'TA'	-182.6	-30.5	-491.2	309.68	309.34

Table 4. UV melting results for  $d(TA)_4$  and  $d(TA')_4$  in SSC at 260nm.  $T_m$  = first derivative of the melting profile,  $T_m$  (corr.) = adjusted value.

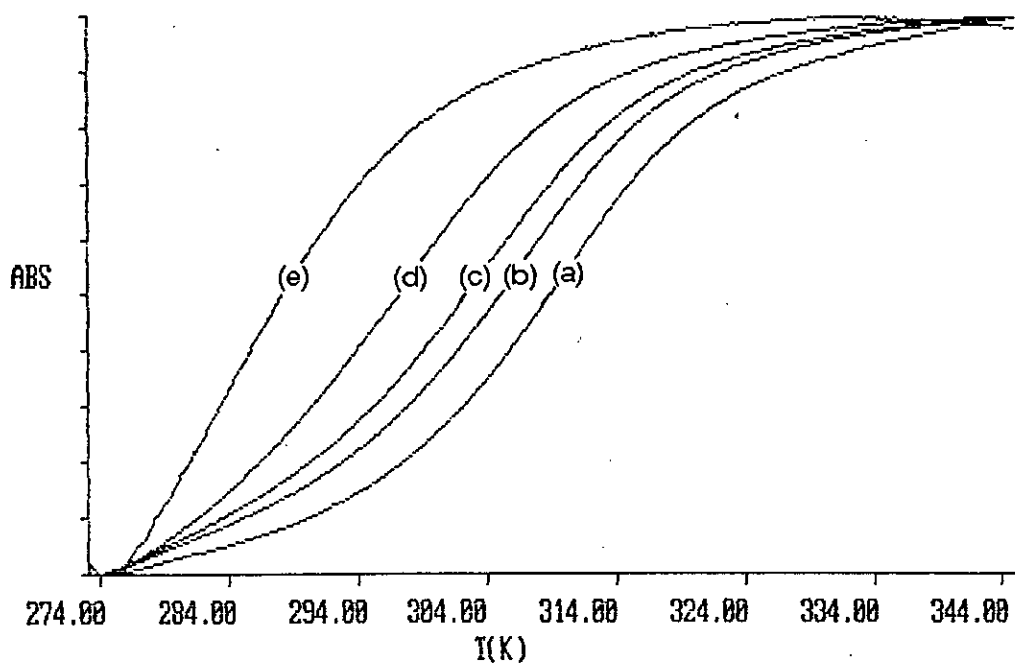


Fig. 39. UV melting profiles of d(TA)<sub>4</sub> at 260nm in (a) SSC buffer (b) SSC + 0.25M MgCl<sub>2</sub> (c) SSC + 0.50M MgCl<sub>2</sub> (d) SSC + 1.0M MgCl<sub>2</sub> (e) SSC + 2.0M MgCl<sub>2</sub>

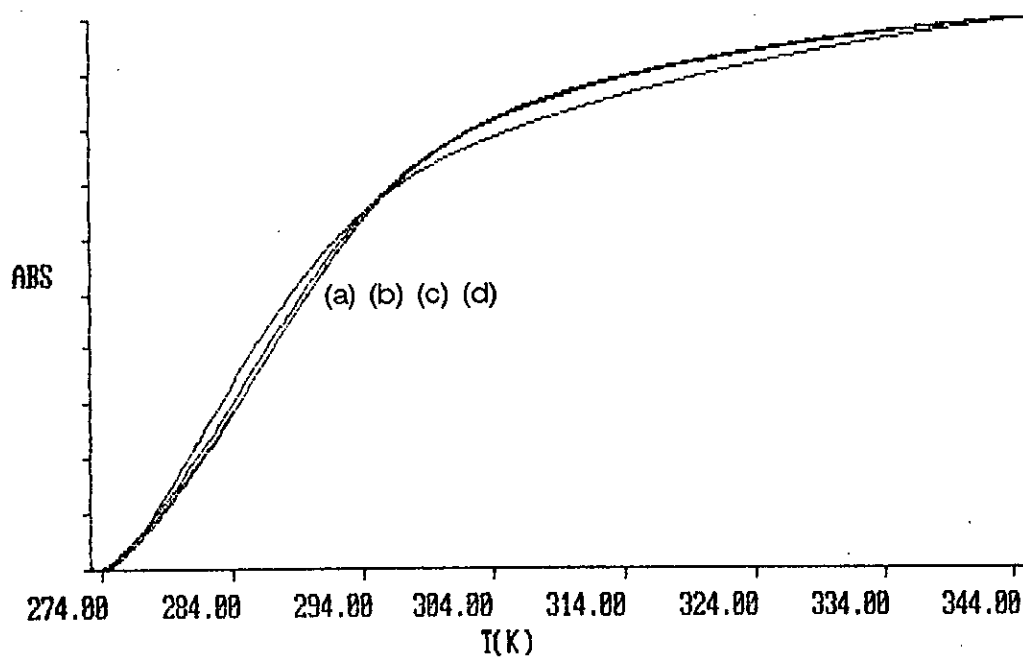


Fig. 40. UV melting profiles of d(TA)<sub>4</sub> at 260nm in (a) SSC buffer (b) SSC + 0.25M MgCl<sub>2</sub> (c) SSC + 0.50M MgCl<sub>2</sub> (d) SSC + 1.0M MgCl<sub>2</sub>.

It was hoped to crystallise the oligonucleotides in their X-form. No X-ray crystallographic analysis of this has been performed. X-DNA can be induced in poly(dA'-dT) by the addition of  $MgCl_2$ <sup>173</sup>. The effect of  $MgCl_2$  on the melting temperatures of dTA'TA'TA'TA' and dTATATATA was investigated. The two oligonucleotides showed quite different behaviour Fig. 39, 40. It is known from considerations of the electrolyte theory of DNA melting that addition of salt to DNA solutions should have the effect of elevating  $T_m$  due to mass action favouring the double helix form which can bind more salt<sup>269</sup>. The salt concentration of solutions of both oligonucleotides in SSC was increased stepwise by addition of  $MgCl_2$ . The effect of salt concentration on  $T_m$  is shown in Table 5. For dTA'TA'TA'TA' increasing salt concentration caused a reduction in the oligonucleotide  $T_m$ . The native oligonucleotide showed more normal behaviour with addition of  $MgCl_2$  raising the melting temperature.

<u>Oligonucleotide</u>	<u>Buffer</u>	<u><math>T_m</math> K</u>
d(TA') <sub>4</sub>	SSC	309.7
	SSC + 0.25MMgCl <sub>2</sub>	306.4
	SSC + 0.5MMgCl <sub>2</sub>	303.9
	SSC + 1.0MMgCl <sub>2</sub>	297.3
	SSC + 2.0MMgCl <sub>2</sub>	282.8
d(TA) <sub>4</sub>	SSC	281.8
	SSC + 0.25MMgCl <sub>2</sub>	286.5
	SSC + 0.5MMgCl <sub>2</sub>	286.4
	SSC + 1.0MMgCl <sub>2</sub>	285.5

Table 5. Effect of salt on  $T_m$  for d(TA')<sub>4</sub> and d(TA)<sub>4</sub>.

These results were difficult to rationalise without further information on the behaviour of the duplexes in solution. An initial analysis of the results would suggest that the dA' oligonucleotide is showing some strange behaviour which has not previously been considered

in UV melting theory. The amount of material available meant that NMR investigation of this phenomenon was not possible. X-DNA was first identified by virtue of its CD spectrum and this remains one of its identifying characteristics. Samples of the two oligonucleotides were submitted for investigation by CD spectroscopy.

#### 9.4.1 Circular Dichroism of d(TA')<sub>4</sub>

Circular dichroism experiments were performed on samples of d(TA')<sub>4</sub>. These were designed to parallel the UV melting experiments in MgCl<sub>2</sub>. CD spectra were recorded in a range of MgCl<sub>2</sub> concentrations in SSC over a range of temperatures. This gives information about the salt required for the formation of X-DNA and its thermostability. Similar procedures were followed to obtain results for NaCl. As previously mentioned it was found in the UV melting experiments that addition of MgCl<sub>2</sub> to samples of d(TA')<sub>4</sub> had the unexpected effect of reducing the melting temperature. The effect of MgCl<sub>2</sub> concentration on the CD spectra is shown in Fig. 41. The transition to X-DNA is accompanied by a shift in the minimum of ellipticity towards 280nm. It can be clearly seen that this shift is present with increasing MgCl<sub>2</sub> concentration indicating the onset of X-form DNA. It should be noted that these CD spectra were all recorded at low temperatures.

A second feature of the CD spectrum of X-DNA is a dramatic increase in the minimum ellipticity value at 280nm. While this has been observed with X-form poly(dT-dA')<sup>172, 173</sup> no increase in minimum ellipticity is seen for d(TA')<sub>4</sub>. This may be rationalised in light of the UV melting results obtained under similar conditions. These showed that T<sub>m</sub> decreased with increasing MgCl<sub>2</sub> concentration. This suggests that while poly(dT-dA') is stabilised by its transition to the X-form the behaviour at the oligomer level is different. By plotting graphs of ellipticity at a set wavelength against temperature, for a set concentration of salt the processes taking place can be highlighted. Ideally these plots should appear similar in shape to those obtained in UV melting experiments for plots of absorbance versus temperature. These are shown in Fig. 42. Although it is not as clear as in a UV experiment, it can be seen that in 1.0M MgCl<sub>2</sub> denaturation is starting at the lowest temperature and in 2.0M MgCl<sub>2</sub> the oligonucleotide appears fully denatured even at low temperature, as indicated by the absence of any low temperature plateau. Values of the expected minimum ellipticities calculated from the UV melting T<sub>m</sub> results are also shown for 1.0M and 2.0M MgCl<sub>2</sub> for fully duplex states. These are much lower than observed and indicate that d(TA')<sub>4</sub> has a complex conformational behaviour in MgCl<sub>2</sub>.

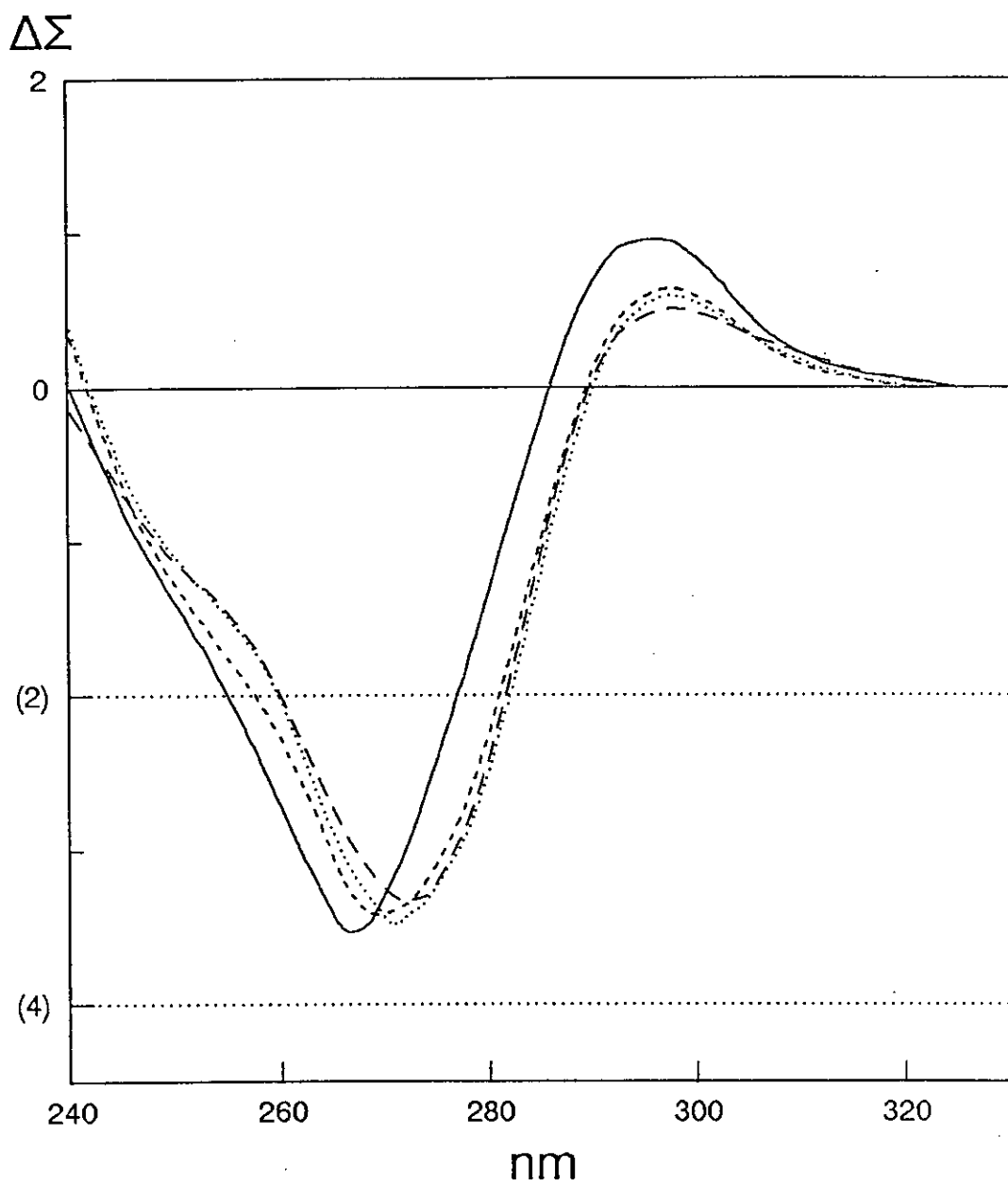


Fig. 41. CD spectra of d(TA')<sub>4</sub> in SSC under the following conditions; — 275.9K 0.0M MgCl<sub>2</sub>, ---- 275.9K 0.5M MgCl<sub>2</sub>, ..... 277.4K 1.0M MgCl<sub>2</sub>, and - - - 275.8K 2.0M MgCl<sub>2</sub>.

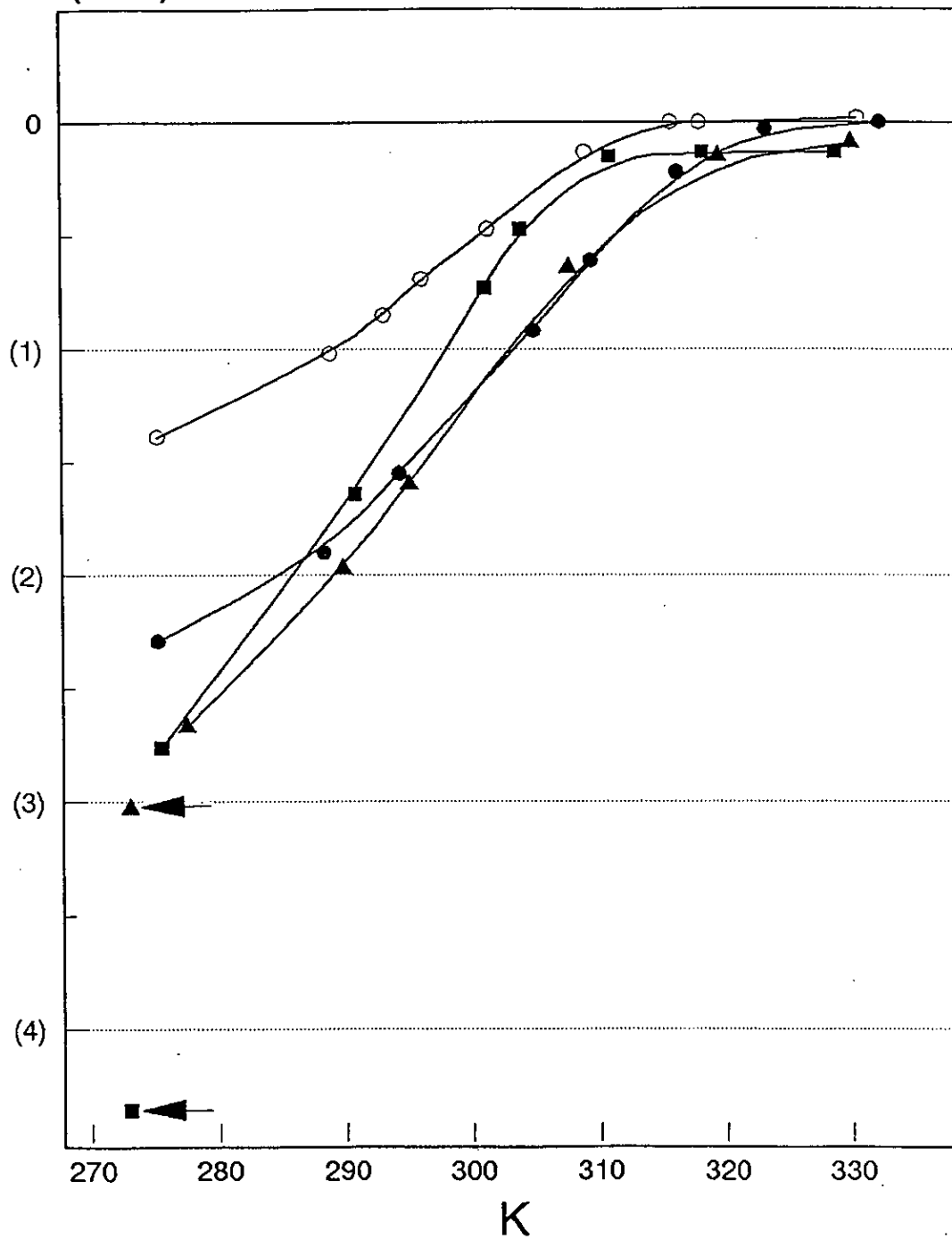
$\Delta\Sigma(280)$ 

Fig. 42. Plots of  $\Delta\Sigma_{280}$  against  $T(K)$  for  $d(TA)_4$  in SSC under the following conditions;  $\circ$  0.0M  $MgCl_2$ ,  $\bullet$  0.5M  $MgCl_2$ ,  $\blacktriangle$  1.0M  $MgCl_2$ , and  $\blacksquare$  2.0M  $MgCl_2$ . Arrows indicate calculated values.

The results of the CD experiments performed in high concentrations of NaCl are much clearer Fig. 43 although the CD spectra in SSC+2Molar salt solutions were similar for both salts Fig.44d, 45b. Increasing NaCl concentration brought about a shift in the minimum ellipticity to 280nm as well as increasing its intensity. Again the spectra were recorded at low temperature. Plotting the ellipticity values at 280nm for each NaCl concentration against temperature was also clearer Fig. 46. In this case in SSC plus 4.0M and 4.5M NaCl the curves plateau at about  $\Sigma_{280} = -4$ . This is close to the theoretical minimum for 2.0M  $MgCl_2$  and would suggest that in 4.5M NaCl at 273K  $d(TA)_4$  is fully X-DNA. It would appear that X-form  $d(TA)_4$  is more stable in high NaCl concentrations, than in SSC with the addition of 2.0M  $MgCl_2$ , as judged by the relative intensities of the negative band at 280nm.

The UV melting behaviour can be explained by proposing that oligo  $d(TA)_4$  is less thermally stable in the X-form than in the B-form. There is a significant temperature dependence in the CD spectrum of  $d(TA)_4$  as the  $MgCl_2$  concentration is increased and B to X-DNA transitions take place Fig. 44b, c, d. Where X-DNA may be indicated 0.5, 1.0, and 2.0M  $MgCl_2$  the CD spectrum changes back to that of B-DNA when the temperature is raised much above 273K and the minimum ellipticity at *c.a.* 273nm relaxes back towards 260nm Fig. 44d. Similar but much more marked changes were observed in the higher NaCl concentrations. Here the X-form is more pronounced and two quite different regions can be seen in the temperature dependent CD spectra. Raising the temperature above 283K causes a change in the spectra back towards that associated with B-DNA, while below 283K X-DNA appears to be the predominant species Fig. 45c, d. The lack of a major increase in the minimum intensity at 280nm of  $d(TA)_4$  in  $MgCl_2$  even though this would be expected from previous work<sup>173</sup>, may be because the onset of X-DNA is accompanied by an increase in the proportion of melted oligomer. As seen  $d(TA)_4$  as X-DNA appears to be less stable than B-DNA. There is thus in solution an equilibrium of B, X and coil DNA.

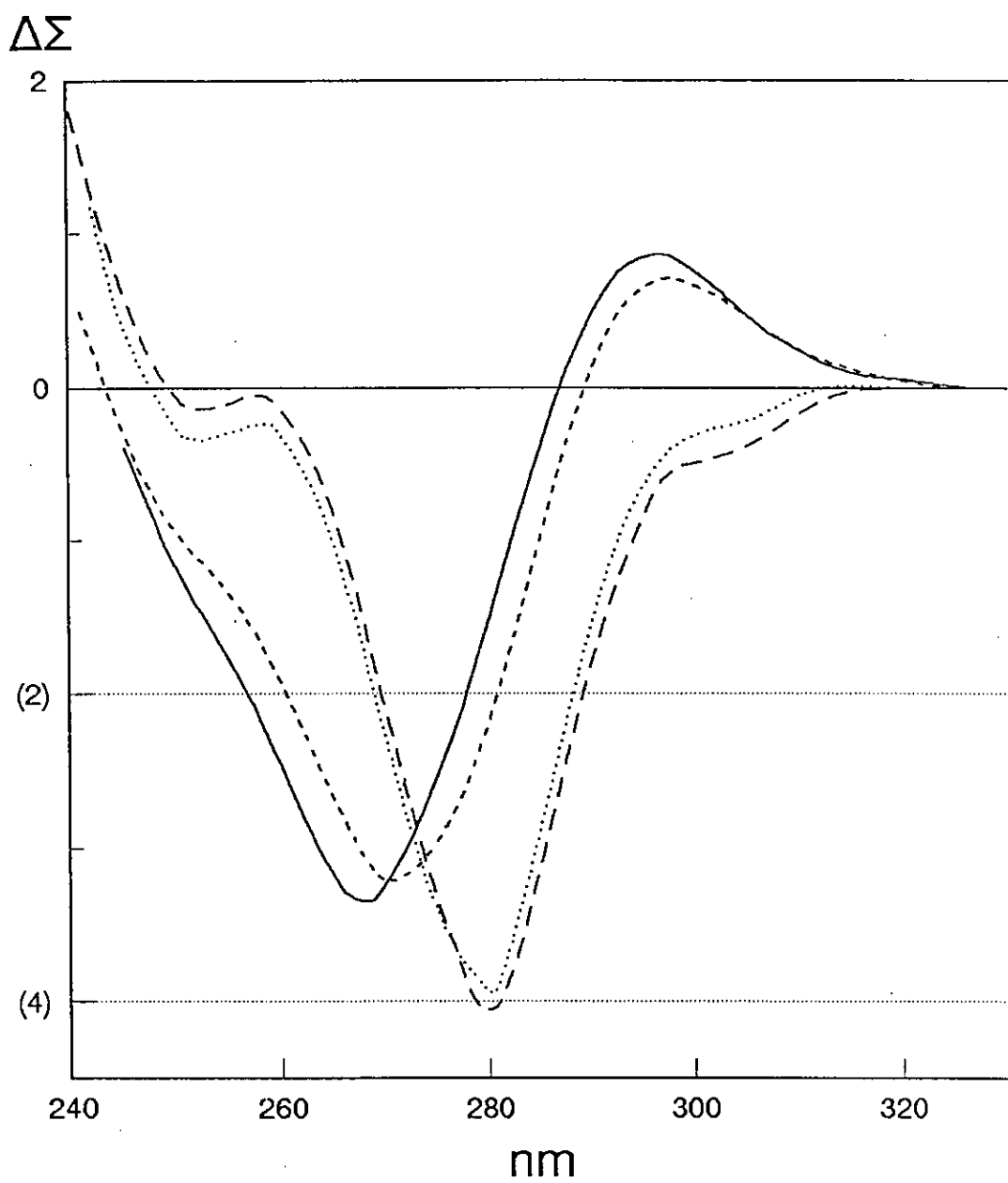


Fig. 43. CD spectra of d(TA')<sub>4</sub> in SSC under the following conditions; — 273.2K 0.55M NaCl ---- 273.7K 2.0M NaCl, ..... 272.6K 4.0M NaCl, and -.-.- 273.2K 4.56M NaCl.

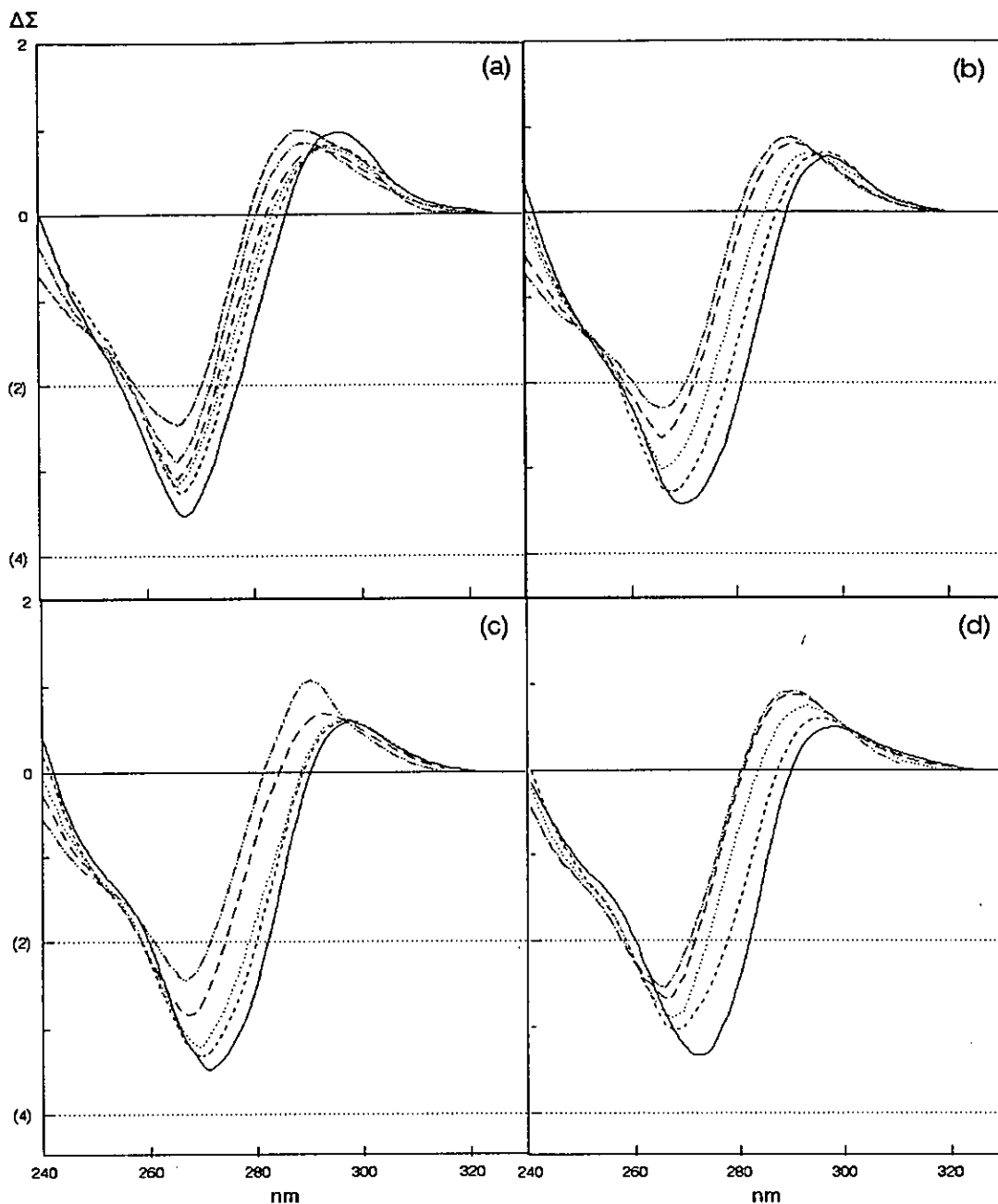


Fig. 44. CD spectra of  $d(TA')_4$  in SSC under the following conditions; (a) SSC + 0.0M  $MgCl_2$ ; — 275.9K, ---- 293.1K, ..... 296.0K, — — — 301.0K, \_\_\_ .. \_\_\_ .. \_\_\_ 308.5K, and — — — — — 330.6K. (b) SSC + 0.5M  $MgCl_2$ ; — 275.3K, ---- 293.5K, ..... 304.5K, — — — 315.6K, and \_\_\_ .. \_\_\_ .. \_\_\_ 331.3K. (c) SSC + 1.0M  $MgCl_2$ ; — 277.4K, ---- 290.2K, ..... 295.0K, — — — 318.9K, and \_\_\_ .. \_\_\_ .. \_\_\_ 329.1K. (d) SSC + 2.0M  $MgCl_2$ ; — 275.8K, ---- 290.6K, ..... 300.6K, — — — 310.4K, and \_\_\_ .. \_\_\_ .. \_\_\_ 318.0K.

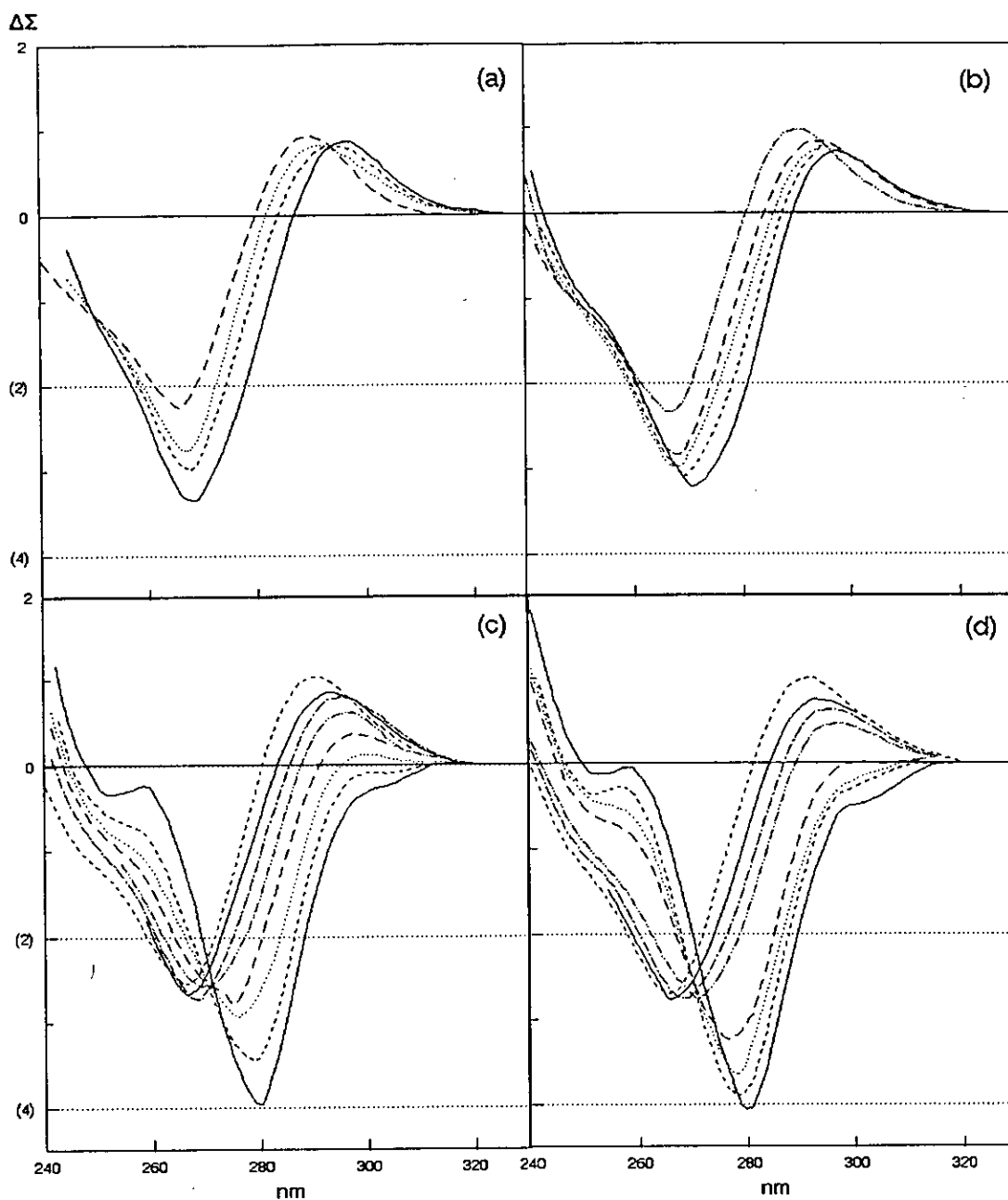


Fig. 45. CD spectra of  $d(TA')_4$  in SSC under the following conditions; (a) SSC + 0.0M NaCl; — 273.2K, ---- 302.2K, ..... 309.9K, and — — — 333.0K. (b) SSC + 2.0M NaCl; — 273.7K, ---- 283.0K, ..... 293.9K, — — — 302.7K, and \_\_\_.. \_\_\_.. \_\_\_ 327.9K. (c) SSC + 4.0M NaCl; — 272.6K, ---- 276.5K, ..... 282.8K, — — — 284.5K, \_\_\_.. \_\_\_.. \_\_\_ 288.8K, — — — — 294.5K, — 296.5K, and ---- 310-324K. (d) SSC + 4.56M NaCl; — 273.2K, ---- 277.5K, ..... 279.6K, — — — 282.3K, \_\_\_.. \_\_\_.. \_\_\_ 285.9K, — — — — 288.4K, — 293.9K, and ---- 304.4K.

$\Delta\Sigma(280)$

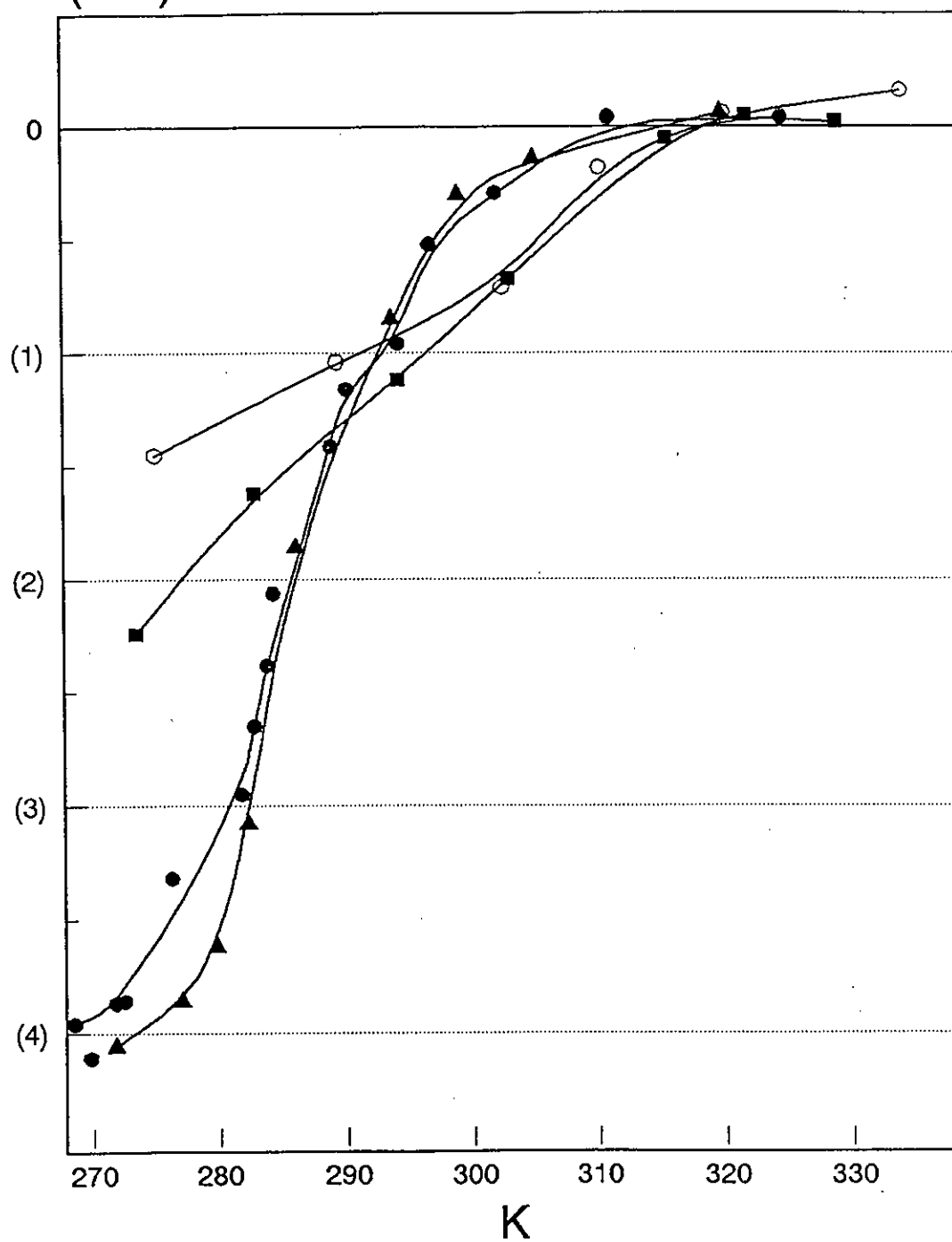


Fig. 46. Plots of  $\Delta\Sigma_{280}$  against  $T(K)$  for  $d(TA')_4$  in SSC under the following conditions;  $\circ$  0.55M NaCl,  $\blacksquare$  2.0M NaCl,  $\bullet$  4.0M NaCl, and  $\blacktriangle$  4.5M NaCl.

## 9.5 Oligonucleotide Digestions

The overall base composition of the dA' oligonucleotides was determined by digesting a sample of each oligonucleotide to its constituent nucleosides. Initial digestion to nucleotides with phosphodiesterase I and further to nucleosides with alkaline phosphatase was followed by RP-HPLC analysis. Peaks corresponding to the relevant nucleosides were noted as determined by comparison to authenticated standards. The results for digestions of d(TA')<sub>4</sub>, d(TA')<sub>3</sub>, dCA'TA'TG, dA'CTA'TCGA'C, dCA'CTA'TCGA'C and dGA'A'GA'GA'GTT are shown Figs. 47a-f.

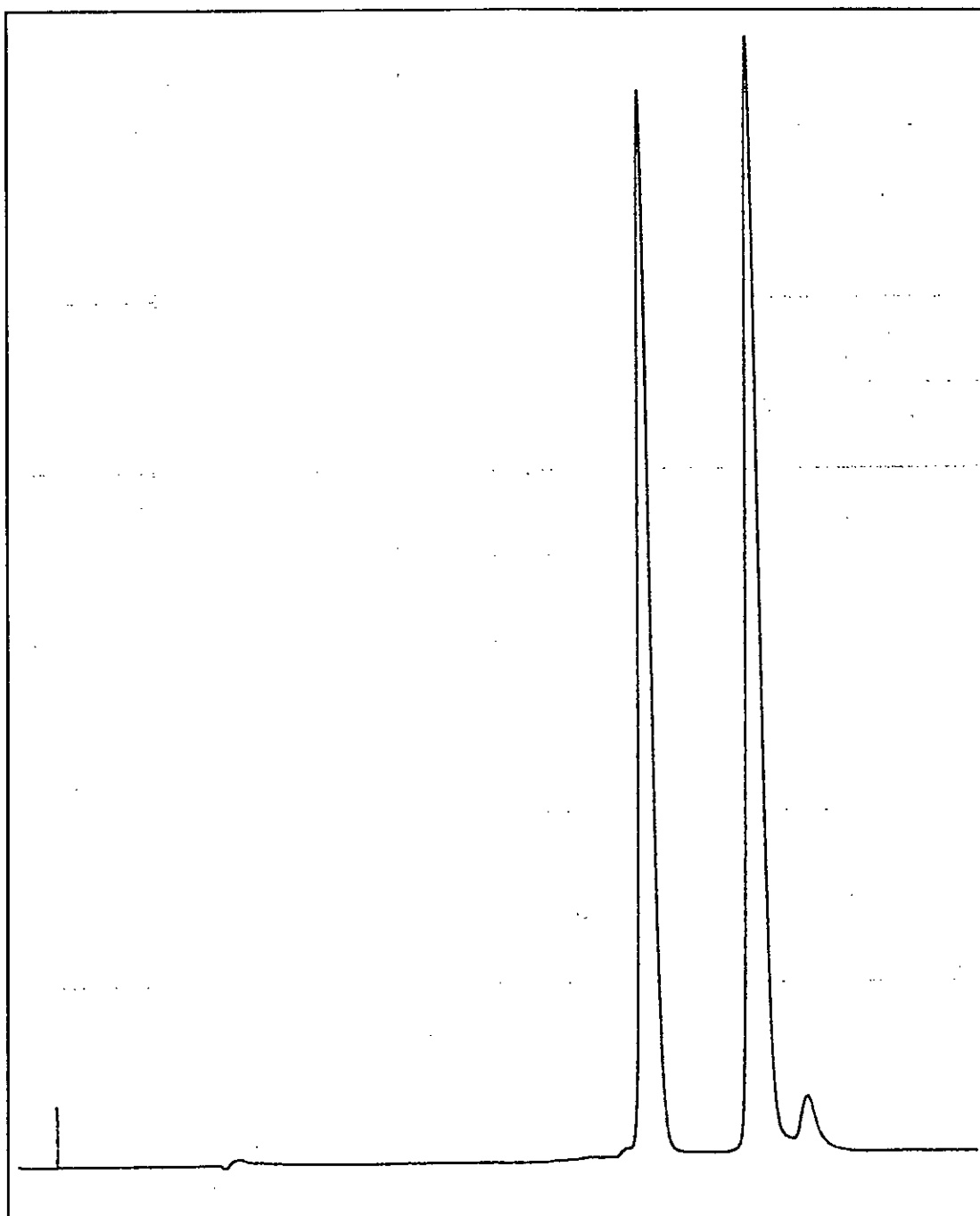


Fig. 47. RP-HPLC of oligonucleotides digested with phosphodiesterase I and alkaline phosphatase. Buffer A; 0.1M  $\text{NH}_4\text{OAc}$ , Buffer B 0.1M  $\text{NH}_4\text{OAc}$  in 20%  $\text{CH}_3\text{CN}$ , gradient 0-100% B over 5-25mins. (a)  $\text{d}(\text{TA})_4$  0.16Aufs,

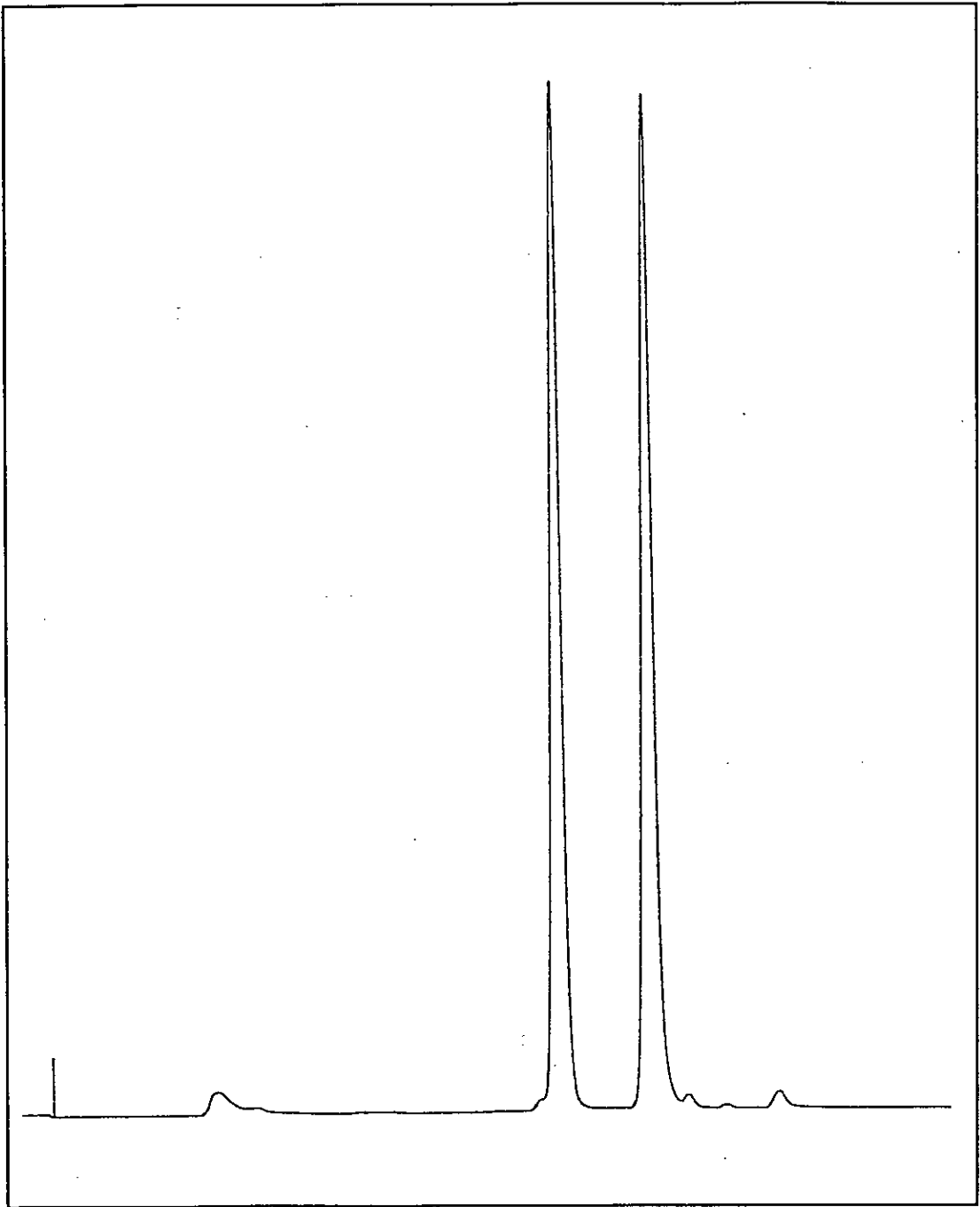


Fig. 47b.  $d(TA')_3$  0.64Aufs.

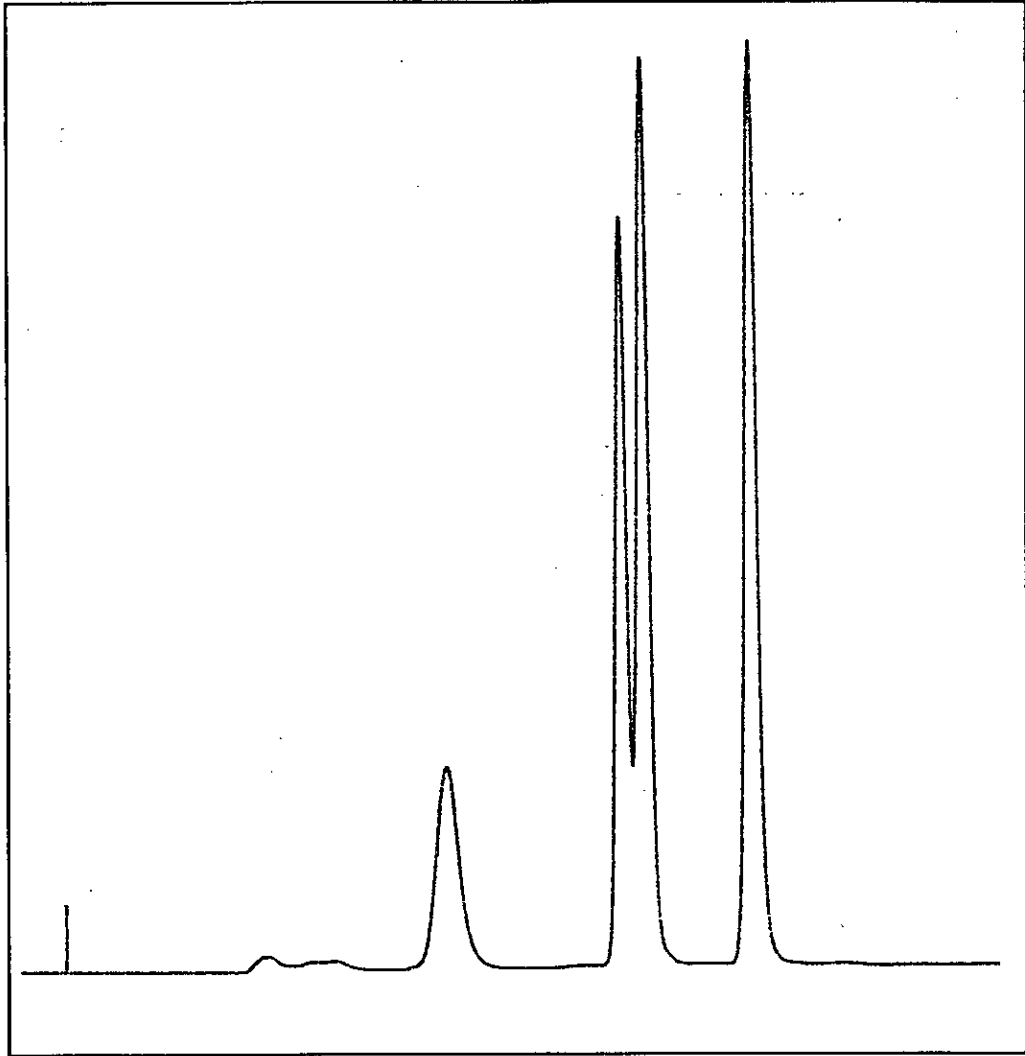


Fig. 47c. dCA'TA'TG 0.32Aufs.

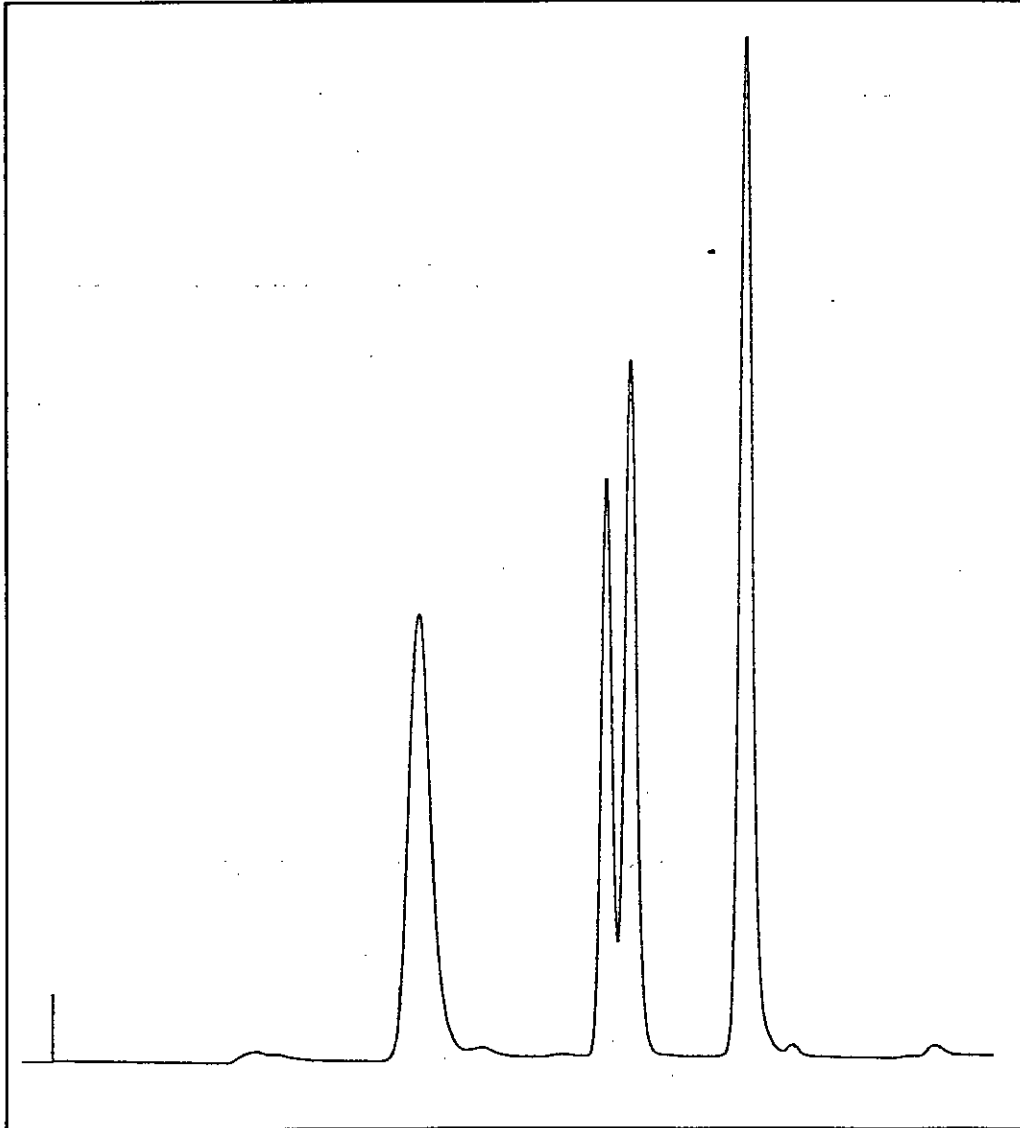


Fig. 47d. dA'CTA'TCGA'C 0.32Aufs.

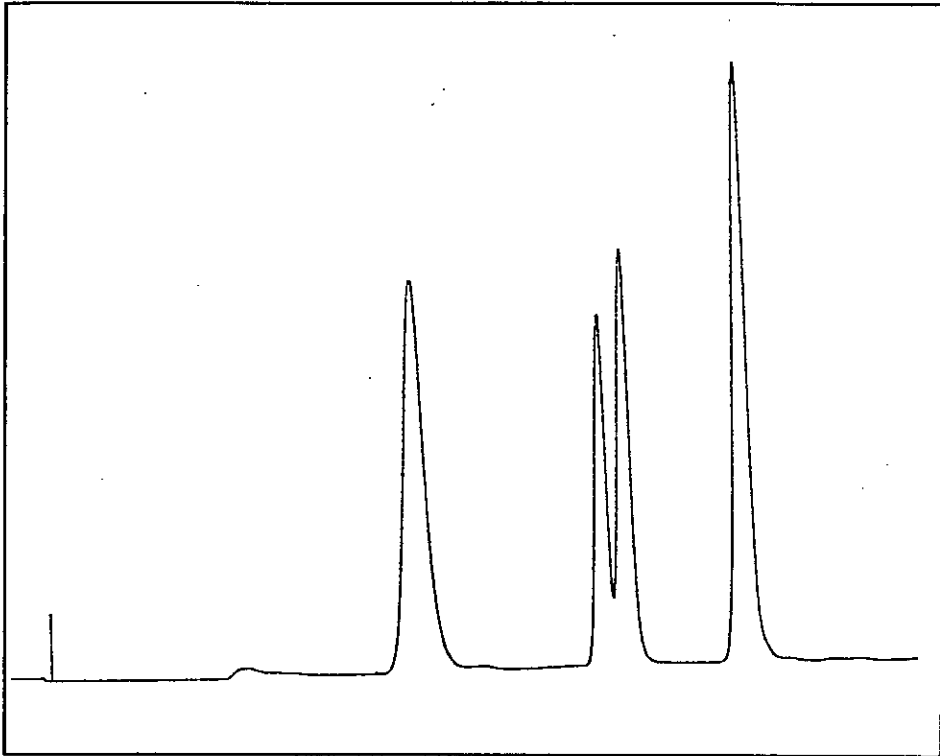


Fig. 47e. dCA'CTA'TCGA'C 0.32Aufs.

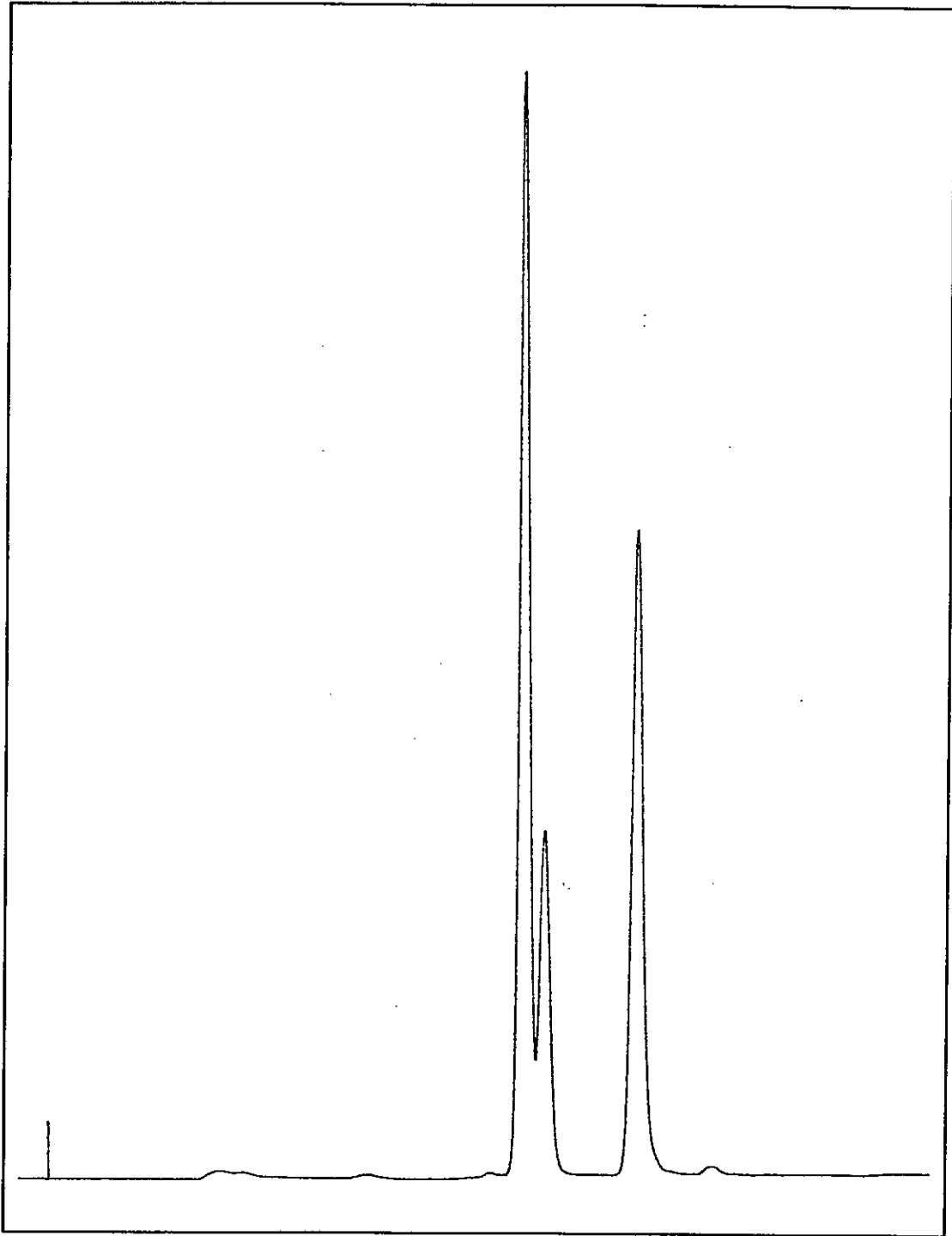


Fig. 47f. dGA'A'GA'GA'GTT 0.32Aufs.

## 9.6 Improvements to dA' Phosphoramidite Design

It was noted that the lengthy deprotection procedure required for oligonucleotides with dA' residues was a significant hold up in the synthesis turn around time. This hold up, which is not seen in the synthesis of normal sequence oligonucleotides, must therefore be attributed to the dA' residues. The question arises of which protecting group, if indeed only one, is responsible for the stringent deprotection procedure. The isobutyryl group used for the protection of the 2NH<sub>2</sub> position and originally introduced by Khorana<sup>209</sup> for dG is probably stable to a degree far in excess of that required for modern coupling cycles. Recent findings suggest that incomplete cleavage of isobutyryl groups may be responsible<sup>270</sup>. Although it appeared from TLC experiments the N<sup>6</sup>-formamidine-N<sup>2</sup>-isobutyryl nucleoside could be readily deprotected in 3 hours at 60°C with concentrated ammonia it should be noted that deprotection of the free nucleoside may be quite different from that when it is incorporated into DNA<sup>91</sup>.

Recently a protecting group has been selected for the 2NH<sub>2</sub> of dG that is compatible with all the standard methodology of phosphoramidite DNA synthesis and appears to be much more labile than the traditional isobutyryl protection. Schulhof and coworkers<sup>270</sup> have used the phenoxyacetyl (pac) group in DNA synthesis. It is readily introduced *via* reaction with phenoxyacetyl chloride in the presence of 1-hydroxybenzotriazole. The half life for deprotection of the N<sup>2</sup>-phenoxyacetyl-2'-deoxyguanosine at 20°C in ammonia is reported as 15 minutes with complete deprotection in 4 hours. This is significantly less than was required to deprotect the dA' oligonucleotides. Thus a suitable second generation 2-amino-2'-deoxyadenosine phosphoramidite could be envisaged with di-n-butylformamidine protection of the N<sup>6</sup> position and phenoxyacetyl protection of the 2NH<sub>2</sub>. Very recently the dimethylformamidine group has been introduced for amino protection of dG<sup>271</sup>. This had previously been reported to be unstable in solution<sup>205</sup> and had been rejected for DNA synthesis.

## 10.0 DNA Tridecamers

The majority of structural analyses performed on short sequences of DNA have examined the structure of base pair mismatches or short complementary sequences. While mismatches may account for many mutational events other sources of mutation exist. These include frame shift mutations. Frame shift mutations may occur in monotonous runs of base pairs where one DNA strand is able to slip to form base pairs one or two positions removed from their original position. Models have been proposed to explain how this mechanism might work. After strand slippage non-matched bases may be looped out into solution to form a bulge structure or alternatively may be stacked into the helix. This, after replication, will give rise to either deletion (-1) or addition (+1) frame shift mutations<sup>272</sup>. DNA secondary structure may also be implicated. The formation of hairpin loops or cruciform structures, within quasipalindromic sequences, containing mismatched or nonmatched bases which again may be looped out into solvent or stacked into the helix has been proposed<sup>273</sup>. The secondary double stranded structure acts as an aberrant template for replication. Frame shift mutations of the T4 *rII* gene have been mapped to the deletion of looped out bases formed in secondary hairpin structures, thus lending support to the model<sup>274</sup>. These findings indicate the need to understand the precise structure of the DNA duplex in the region of nonmatched bases. Extrahelical bases disrupt the regular flow of the DNA duplex and, as described, may be present in two different forms. A nonmatched base may be stacked into the surrounding duplex or it may be directed away from the helix and bulged out.

Nonmatched base structures have been examined by physicochemical methods. These have shown the surrounding sequence may influence the behaviour of the nonmatched base. Further more, nonmatched purines and pyrimidines display different behaviour. NMR studies of tridecamers containing nonmatched adenosine residues dCGCAGAATTCGCG<sup>275</sup>, dCGCAGAGCTCGCG<sup>276</sup>, dCCGGAATTCACGG and dCCGAGAATTCCGG<sup>277</sup> have shown that regardless of the surrounding sequence, and independent of temperature, the extra adenosine always stacks into the duplex. For pyrimidines a different situation has been observed. The noncomplementary polymers polyd(TCTC).polyd(GGA)<sup>278</sup> and polyd(TTC).polyd(GA<sub>n</sub>)<sup>279</sup> exist with the extra nonmatched thymidines looped out and the flanking bases stacked on one

another. Oligonucleotides have also been investigated. The additional dC residues in dCA<sub>3</sub>CA<sub>3</sub>.dCT<sub>6</sub>G<sup>280</sup> and dGATGGGCAG.dCTGCCCCATC<sup>281</sup> were found to be looped out, while in the tridecamer dCCGCGAATTCCGG<sup>282</sup> the extra dC was present as an equilibrium between looped out bulge structure (low temperature) and stacked (elevated temperature) forms. This contrasts with an earlier proposal that the stacking interactions of a base will determine whether it is stacked or looped out<sup>280</sup> on the basis of which it was proposed pyrimidines, with lower stacking interactions than adenosine, would always form bulge structures. The situation becomes further complicated from the observation that dCTGGTGCGG.dCCGCCCAG<sup>283</sup> stacks the extra thymidine internally, but in rCUGGUGCGG.rCCGCCCAG<sup>284</sup> the unpaired uridine is looped out. Very recently NMR structures of the two tridecamers dCCGTGAATTCCGG and dCCGGAATTCTCGG have been reported to contain additional thymidine bulges<sup>285</sup>. In these sequences the conformation adopted by the non paired thymidine depended on the flanking sequences. When flanked by purines the thymidine was looped out at low temperature but stacked into the duplex at elevated temperature (prior to onset of melting). Where flanking pyrimidines were present no temperature dependence was observed with only the looped out form being observed.

Thus for nonmatched bases, which may be involved in the mechanism of frame shift mutations, the following generalisations can be made. Unpaired adenosine stacks into the duplex regardless of flanking sequence or temperature, while cytidine and thymidine display conformational flexibility with conformation depending on neighbouring bases, temperature and whether DNA or RNA (U). Because of their obvious importance with regard to understanding the genesis of addition (+1) or deletion (-1) frame shift mutations a fuller analysis of nonmatched base structures is desirable. To date only one X-ray crystal structure of nonmatched bases in DNA has been reported<sup>286</sup>. The tridecamer dCGCAGAATTCGCG containing an additional adenosine has been crystallised and unit cell dimensions obtained (C<sub>2</sub> space group) although no full structure has been reported. It was decided to try and obtain accurate detailed structural information on nonmatched pyrimidine sequences by X-ray crystallography.

As with previous investigations<sup>275, 286</sup> sequences based on the Dickerson dodecamer dCGCGAATTCGCG were used with additional pyrimidines inserted at position 5. The sequences chosen for investigation are shown below. The first sequence is in effect the parent of the series.

dCGCGTAATTCGCG

dCGCGBrUAATTCGCG

dCGCGTAATBrUCGCG

dCGCGTAATTBrCGCG

With these short self complementary sequences there is the possibility of forming pyrimidine.pyrimidine mismatches to give a conventional 13mer duplex. However in order to form such a mismatch a distortive narrowing of the C1'-C1' distance in the duplex is required to allow hydrogen bonding. Given that other structures, not requiring this distortion, can form more readily this is unlikely to occur. Brominated nucleosides are included both as spatial analogues, as in the case of 5-bromouracil and thymine, and to assist with crystal structure solution. It was hoped to glean information on the stacked/looped out nature of the structures and the effect of the nonmatched structure on the neighbouring sequence and solvent environment.

### 10.1 Tridecamer Synthesis

Phosphoramidites of the brominated nucleosides were not available commercially. BrU was treated as described previously to give the phosphoramidite. 5-Bromo-2'-deoxycytosine was treated similarly with the additional initial step of protection of the 4NH<sub>2</sub> by benzylation with 2, 3, 4, 5, 6-pentafluorophenyl benzoate. Oligonucleotides were synthesised using standard coupling cycles as six 1µmole syntheses. Pure lyophilised material was obtained after HPLC purification, Method 3. Coupling efficiency for the in house phosphoramidites was assayed by the released DMT cation and was 98% or greater.

## 10.2 Tridecamer Crystallisation

Attempts were made to crystallise the oligonucleotides using the sitting drop method from solutions of DNA ( $20\text{mgml}^{-1}$ ) in sodium cacodylate ( $50\text{mM}$  pH 6.95). Solutions were added to droplets of the DNA solutions ( $5\mu\text{l}$ ) in individual wells to give crystallisation mixtures of DNA plus; strontium chloride ( $5\text{-}25\text{mM}$ ), spermine tetrahydrochloride ( $0\text{-}3.5\text{mM}$ ) and 2-methyl-2, 4-pentenediol ( $12.5\text{-}25\%$  vol/vol). Crystals of a size and quality for data collection were not obtained. Microcrystals were readily grown in all cases while larger disordered non-prismatic crystals were also obtained for the first two tridecamers. These failed to extinguish in a satisfactory manner under polarising filters. That crystals of a suitable size and order could not be obtained perhaps reflects the inherent disorder of the structures being examined. It is possible that a complex equilibrium of looped out dodecamer, mismatched tridecamer, and stacked tetradecamer exists which inhibits the crystallisation process.

## 11.0 Materials and Methods

### Melting Points

These were obtained using a Kofler Hotstage and are uncorrected.

### Elemental Analysis

Performed by Mrs E. McDougal on a Perkin Elmer 240 Elemental Analyser.

### Nuclear Magnetic Resonance Spectra

$^1\text{H}$  NMR were collected on a Bruker WP200SY spectrometer by Mr. J. R. A. Miller and Mrs. H. Grant. Chemical shifts are quoted in ppm against <sup>TMS</sup> methylsilane.

$^{31}\text{P}$  NMR were collected on a Jeol JMN FX90Q spectrometer by various members of the research group.

### Mass Spectra

Mass Spectra were obtained on a Kratos MS50 mass spectrometer by positive or negative fast atom bombardment techniques (FAB) from a thioglycerol matrix by Mr. A. Taylor.

### UV Spectra

UV spectra were recorded on a Perkin Elmer Lambda 15 spectrometer.

### Flash Chromatography

Fluka silica 60mesh (230-400) was used under slight nitrogen pressure.

### Oligonucleotide Synthesis

- This was performed on an Applied Biosystems 380B DNA Synthesiser using 0.1 molar solutions of cyanoethyl phosphoramidites in acetonitrile (DNA Synthesis Grade, ABI). Commercial phosphoramidites, where available, were purchased from ABI or BDH.

## Thin Layer Chromatography

TLCs were run on aluminium backed plates Merk Art 5554 Keiselgel 60 F<sub>254</sub> using the following systems

- A. CH<sub>2</sub>Cl<sub>2</sub>:EtOH:ammonia, 9:1:1.
- B. CHCl<sub>3</sub>:MeOH:acetic acid, 10: 3:1.
- C. EtOH:CHCl<sub>3</sub>, 1:9.
- D. EtOH:CHCl<sub>3</sub>, 4:6.
- E. EtOAc:CH<sub>2</sub>Cl<sub>2</sub>:TEA, 45:45:10.
- F. EtOAc:CH<sub>3</sub>CN, 50:50.
- G. MeOH:CHCl<sub>3</sub>, 2:98.
- H. EtOAc:MeOH:ammonia, 9:3:1.
- I. CH<sub>2</sub>Cl<sub>2</sub>:MeOH:TEA, 10:2:1.
- J. EtOH:CHCl<sub>3</sub>:Pyridine, 9.5:90:0.5.

Following examination under UV light, development was achieved in one or more of the following ways. Exposure to HCl vapour. This imaged all trityl components as bright orange. Exposure to iodine vapour imaged organic materials. Development with a spray containing acetic acid 2%, anisaldehyde 2%, sulphuric acid 5% in ethanol. Sugar components were imaged as a dark green or blue colour on heating the sprayed plates.

## HPLC

HPLC was performed as outlined later using the following equipment.

Perkin Elmer Series 410 LC Pump, LC 90 UV detector and R50 recorder.

Gilson Pump Model 303 (x2), Manometric Module 803C, Dynamic mixer 811, Shimadzu SPD-6A UV detector, Perkin Elmer R50 recorder and IBM AT with Gilson 714 software.

Gilson Pump Model 303 (x2), Manometric Module 802C, Dynamic mixer 811B, Model 116 UV detector, Perkin Elmer R50 recorder and IBM PS2 Model 30 with Gilson 714 software.

All systems were fitted with Rheodyne 7125 injectors.

## Solvents

Solvents used were HPLC grade if available or were purified by standard methods. Pyridine was dried by distillation twice from KOH pellets. The bright clear distillate was stored under  $N_2$  in amber glass bottles. Dichloromethane was dried by distillation after reflux from  $CaH_2$  and stored over 4A<sup>o</sup> molecular sieves. Ethylacetate used for phosphoramidite precipitation was washed with 5%  $Na_2CO_3$ , saturated  $CaCl_2$  and distilled from  $P_2O_5$ . Toluene and diethyl ether were dried by the addition of sodium wire. THF was distilled from benzophenone/sodium metal.

## 11.1 Experimental

*Oligonucleotide Purification.* Oligonucleotides were obtained trityl-off after synthesis and were fully base and phosphate deprotected by heating with concentrated ammonia. The crude oligonucleotides obtained were then purified using the following protocols.

*Method 1.* Reverse phase HPLC purification only. The crude oligonucleotide was dissolved in distilled H<sub>2</sub>O (2ml) and applied in four portions to a preparative scale C<sub>18</sub> reverse phase column. A gradient system was used to elute the oligonucleotide. Buffer A, 0.1molar NH<sub>4</sub>OAc; Buffer B, 0.1molar NH<sub>4</sub>OAc in 20% CH<sub>3</sub>CN. The column eluent was monitored at 260nm and the major product peak collected. Lyophilisation removed the majority of the volatile buffer to leave a solid. This was reconstituted with H<sub>2</sub>O (2ml) and loaded onto a water equilibrated Sephadex G10 column (16x500mm). The UV absorbing peak corresponding to the fully desalted oligonucleotide was collected. Lyophilisation gave the oligonucleotide as a white solid.

*Method 2.* As *Method 1* but with an additional ion exchange HPLC purification prior to reverse phase HPLC. The crude oligonucleotide was dissolved in distilled H<sub>2</sub>O (2ml) and applied in four portions to a preparative scale SAX ion exchange column. A gradient system was used to elute the oligonucleotide. Buffer A, 0.03molar KH<sub>2</sub>PO<sub>4</sub> in 20% CH<sub>3</sub>CN; Buffer B, 0.66molar KH<sub>2</sub>PO<sub>4</sub> in 20% CH<sub>3</sub>CN. The pH of both buffers was adjusted to pH6.4 with KOH. The oligonucleotide product peak which eluted later than failed sequences was collected, pooled, taken to dryness and purified by Method 1.

*Method 3.* As *Method 1* but with a pre-HPLC step. The crude oligonucleotide was loaded onto a DEAE cellulose column (16x700mm) equilibrated in 0.08molar TEAB. A gradient of 0.08molar to 0.8molar TEAB was run over 16Hrs. Monitoring the column effluent at 260nm showed the oligonucleotide to be in the late eluting fractions. These were pooled, taken to dryness and purified by Method 1.

### *Oligonucleotide Crystallisations*

*Sitting Drop.* Nine well dropping plates (Dow Corning USA) were ground to give a flat sealing surface and treated with a 10% solution of  $(\text{CH}_3)_2\text{SiCl}_2$  in  $\text{CH}_2\text{Cl}_2$  for 15 minutes. They were then rinsed with acetone and distilled water and dried. On selected plates channels were incised connecting sets of three wells designated 1, 2 and 3; 4, 5 and 6; and 7, 8 and 9. These were cut at a level above that of the liquid content to allow  $\text{H}_2\text{O}$  vapour diffusion to a central reservoir well containing a solution of aqueous 2-methyl-2, 4-pentanediol. In all cases the wells were sealed with microscope slides and vacuum grease.

*Hanging Drop.* Linbro tissue culture plates with 24  $2\text{cm}^2$  wells (Flow Laboratories, Inc.) were used as supplied. Microscope glass cover slips  $24\times 24\text{mm}$  (Chance Proper Ltd.) were treated with a 10% solution of  $(\text{CH}_3)_2\text{SiCl}_2$  in  $\text{CH}_2\text{Cl}_2$  by immersion for 15 minutes and rinsed with acetone and distilled water. After air drying they were stored free from dust and handled only by the edges to avoid contaminating the surface with grease. The oligonucleotide-buffer solution was placed on the cover slip which was then gently rotated and inverted over a solvent reservoir well in the tissue culture plate, sealing with vaseline. To allow room for manipulation only offset wells were used giving 12 crystallisations per plate.

*2-N, 3'-O, 5'-O-Triisobutryl-2'-deoxyguanosine (4).* 2'-deoxyguanosine (5.0g 18mmol) was dried by coevaporation from dry pyridine (x 3) and suspended in the same solvent (200ml). After cooling on ice to  $0^\circ\text{C}$  isobutryl chloride (19.2g 180mmol) was added and the mixture stirred for 2 hrs then poured into water (400ml) containing sodium hydrogen carbonate (26g 300mmol). The solution was reduced in volume until crystallisation started and the product was collected by filtration.

7.91g, 92% Mp.  $105-111^\circ\text{C}$  (lit <sup>91</sup>  $104-112^\circ\text{C}$ ),  $\lambda_{\text{max}}$  ( $\text{CH}_3\text{CN}$ ) 257 ( $18600\text{dm}^3\text{mol}^{-1}\text{cm}^{-1}$ )  
280nm (13700)  $\delta_{\text{H}}$  ( $\text{CDCl}_3$  200MHz) d 1-1.25 (m, 18,  $3[\text{CH}_3]_2\text{C}$ ), 2.3-2.7 (m, 3,  $\text{Me}_2\text{C-H}$ ),  
2.7-3.1 (dq, 2,  $\text{H}_{2'-2''}$ ), 4.2-4.4 (q, 2,  $\text{H}_{5'-5''}$ ), 4.6-4.85 (q, 1,  $\text{H}_4'$ ), 5.3-5.4 (d, 1,  $\text{H}_3'$ ), 6.1-6.25

(t, 1, H<sub>1'</sub>), 7.75 (s, 1, H<sub>8</sub>), 9.7 (brs, 1, N<sup>2</sup>H), 12.15 (brs, 1, N<sup>1</sup>H). (Found: C, 54.4; H, 6.46; N, 14.4. Calc for C<sub>22</sub>H<sub>31</sub>N<sub>5</sub>O<sub>7</sub>: C, 55.3; H, 6.54; N, 14.7%). m/z m<sup>+</sup> 478.23017 C<sub>22</sub>H<sub>32</sub>N<sub>5</sub>O<sub>7</sub> requires 478.23015 rf. 0.41 (A).

*2-N, 3'-O, 5'-O-Triisobutyryl-2-amino-2'-deoxyadenosine (7)*. To 2-N, 3'-O, 5'-O-triisobutyryl-2'-deoxyguanosine (4.8g 10mmol) dissolved in freshly distilled methylene chloride (100ml) was added dry triethylamine (5.6ml 40mmol) 2,4,6-tri-isopropylbenzenesulphonyl chloride (6.0g 20mmol) and 4-dimethylaminopyridine (60mg) as catalyst. The mixture was stirred at room temperature for 2 hrs cooled to 0°C and trimethylamine (20ml 13.22g 22mmol) was added. Anhydrous ammonia gas was then condensed into the mixture until 400ml had been added and the reaction refluxed under a dry ice cold finger for 5 hrs before allowing the ammonia to evaporate overnight. The residue was filtered and the filtrate washed with water (2 x 100ml) dried over MgSO<sub>4</sub> and taken to an oil which was dissolved in dry diethyl ether. After standing at 4°C for 24 hrs the product was filtered off as a white precipitate. Two further crops were obtained in a similar manner.

2.72g, 57% Mp 147-150 °C λ<sub>max</sub> (CH<sub>3</sub>CN) 224 (26896dm<sup>3</sup> mol<sup>-1</sup> cm<sup>-1</sup>), 271nm (14052) δ<sub>H</sub> (CDCl<sub>3</sub> 200MHz) d 1.05-1.3 (m, 18, 3[CH<sub>3</sub>]<sub>2</sub>C), 2.4-3.5 (m, 5, 3Me<sub>2</sub>C-H, H<sub>2'-2''</sub>), 4.25-4.7 (m, 3, H<sub>4'</sub>, H<sub>5'-5''</sub>), 5.3-5.45 (d, 1, H<sub>3'</sub>), 6.25-6.45 (d of t, 3, H<sub>1'</sub>, N<sup>6</sup>H<sub>2</sub>), 7.9 (s, 1, H<sub>8</sub>), 9.1 (brs, 1, N<sup>2</sup>H). (Found: C, 54.9; H, 6.66; N, 17.5. Calc for C<sub>22</sub>H<sub>32</sub>N<sub>6</sub>O<sub>6</sub>: C, 55.1; H, 6.7; N, 17.5%). m/z m<sup>+</sup> 477.24613 C<sub>22</sub>H<sub>33</sub>N<sub>6</sub>O<sub>6</sub> requires 477.24614 rf. 0.61 (A).

*2-Amino-2'-deoxyadenosine (11)*. 2-N, 3'-O, 5'-O-Triisobutyryl-2-amino-2'-deoxyadenosine (375mg 0.8mmol) was divided equally between three Schott tubes threequarters filled with concentrated ammonia (20ml) The tubes were sealed and incubated at 70°C for five days. Removal of the ammonia gave the product as a precipitate from a mixture of toluene and chloroform.

185mg, 87% Mp 147-151°C (Lit <sup>251</sup> 146-148°C) (Found: C, 42.7; H, 5.37; N, 31.57. Calc. for C<sub>10</sub>H<sub>14</sub>N<sub>6</sub>O<sub>3</sub>·0.75H<sub>2</sub>O: C, 45.1; H, 5.30; N, 31.60%). λ<sub>max</sub> (CH<sub>3</sub>CN) 215 (22864dm<sup>3</sup> mol<sup>-1</sup>

$^1\text{cm}^{-1}$ ) 256 (8739) 278nm (9520)  $\delta_{\text{H}}$  (DMSO- $d_6$  200MHz) 2.1-2.25 & 2.5-2.75 (m & m, 1 & 1,  $\text{H}_{2'-2''}$ ), 3.4-3.7 (m, 2,  $\text{H}_{5'-5''}$ ), 3.9 (m, 1,  $\text{H}_{4'}$ ), 4.35 (m, 1,  $\text{H}_{3'}$ ), 5.24-5.30 (m, 2, 5'-OH & 3'-OH), 5.7 (s, 2,  $\text{N}^2\text{H}_2$  exchangeable), 6.2 (t, 1,  $\text{H}_{1'}$ ), 6.8 (s, 2,  $\text{N}^6\text{H}_2$  exchangeable), 7.9 (s, 1,  $\text{H}_8$ )  $m/z$   $m^+$  267.12054  $\text{C}_{10}\text{H}_{15}\text{N}_6\text{O}_3$  requires 267.12055  $r.f.$  0.51 (B).

*di-n-Butylformamidedimethylacetal.* Di-n-butylamine (42ml 250mmol) and N, N-dimethylformamide dimethylacetal (38ml 286mmol) were protected from moisture and heated together at 100°C for five days to give a straw coloured solution. Reduced pressure distillation gave the desired product which was stored in the dark under nitrogen.

12.687g, 25% bp. 97°C 13 mm Hg  $^1\text{H}$  NMR ( $\text{CDCl}_3$  200MHz) 0.85 (t, 6,  $\text{CH}_3\text{CH}_2$ ), 1.1-1.5 (m, 8,  $\text{EtCH}_2$ ,  $\text{MeCH}_2$ ), 2.52 (t, 4,  $\text{PrCH}_2$ ), 3.25 (s, 6,  $\text{CH}_3\text{OCH}_3$ ), 4.45 (s, 1,  $\text{CHOCH}_3$ ).

*2-N, 3'-O, 5'-O-triisobutyryl-N<sup>6</sup>-di-n-butylformamide-2-amino-2'-deoxyadenosine (8).*

2-N, 3'-O, 5'-O-triisobutyryl-2-amino-2'-deoxyadenosine (5.0g 10.5mmol) was dissolved in freshly distilled dichloromethane (25ml). Di-n-butylformamidedimethylacetal (2.56g 2.94ml 12.6mmol 1.2 equiv) was added and the solution stirred at room temperature for 3 hrs filtered and taken to an oil. Dissolving the oil in dry diethyl ether (20ml) and standing at 4°C gave the product as a white solid. 77% Mp 75-78°C,  $\lambda_{\text{max}}$  ( $\text{CH}_3\text{CN}$ ) 201 ( $12672\text{dm}^3\text{mol}^{-1}\text{cm}^{-1}$ ) 258 (20369) 317nm (23875)  $\delta_{\text{H}}$  ( $\text{CDCl}_3$  200MHz) d 0.8-1.0 (2t, 6, formamide  $\text{CH}_3$ ), 1.1-1.25 (m, 18,  $3[\text{CH}_3]_2\text{C}$ ), 1.25-1.75 (m, 8, formamide  $\text{EtCH}_2$ ,  $\text{MeCH}_2$ ), 2.4-2.7 (m, 3,  $3\text{Me}_2\text{CH}$ ), 2.7-3.1 (q, 2,  $\text{H}_{2'-2''}$ ), 3.25-3.75 (2t, 4, formamide  $\text{PrCH}_2$ ), 4.2-4.5 (m, 3,  $\text{H}_{4'}$ ,  $\text{H}_{5'-5''}$ ), 5.3-5.4 (d, 1,  $\text{H}_{3'}$ ), 6.3-6.4 (t, 1,  $\text{H}_{1'}$ ), 7.92 (s, 1,  $\text{H}_8$ ), 7.99 (s, 1, formamide CH exchangeable), 9.07 (s, 1,  $\text{N}^2\text{H}$  exchangeable). (Found: C, 60.8; H, 8.26; N, 16.0. Calc for  $\text{C}_{31}\text{H}_{49}\text{N}_7\text{O}_6$ : C, 60.47; H, 8.02; N, 15.93%).  $m/z$   $m^+$  616.38226  $\text{C}_{31}\text{H}_{50}\text{N}_7\text{O}_6$  requires 616.38223  $r.f.$  0.66 (C).

*2-N-isobutyryl, 5'-O-(4, 4'-dimethoxytrytyl)-N<sup>6</sup>-di-n-butylformamide-2-amino-2'-deoxyadenosine (9).* To 2-N, 3'-O, 5'-O-triisobutyryl-N<sup>6</sup>-di-n-butylformamide-2-amino-2'-deoxyadenosine (4.8g 7.8mmol) dissolved in dry thf:pyridine 3:4 (40ml) stirred at 0°C was

added dropwise sodium hydroxide (2Molar 35ml) After stirring for 50 minutes the mixture was slurried with DOWEX 50X8-200 hydrogen form (20ml) and passed through a column containing more of the resin (30ml) eluting with pyridine:water 1:4 (250ml). The neutral eluent was extracted with dichloromethane (3 x 150ml) dried over Na<sub>2</sub>SO<sub>4</sub> and foamed by coevaporation with toluene and THF. To this dissolved in dry dichloromethane (50ml) was added 4, 4'-dimethoxytrityl chloride (3.17g 9.36mmol 1.2 equiv) triethylamine (1.1g 1.4 equiv) and 4-dimethylaminopyridine (47mg). The reaction was stirred for 24 hrs washed with water (2 x 50ml) dried over Na<sub>2</sub>SO<sub>4</sub> and foamed. Wet flash chromatography on silica gel eluting with ethylacetate:acetonitrile 50:50 gave the product.

Mp. 88-90°C  $\lambda_{\max}$  (CH<sub>3</sub>CN) 205 ( 58931dm<sup>3</sup> mol<sup>-1</sup> cm<sup>-1</sup>) 233 (27660) 259 (21646) 318nm (21727)  $\delta_{\text{H}}$  (CDCl<sub>3</sub> 200MHz) d 0.75-1.0 (d of t, 6, formamidine CH<sub>3</sub>), 1.1-1.25 (d of t, 6, [CH<sub>3</sub>]<sub>2</sub>C), 1.3-1.75 (m, 8, formamidine EtCH<sub>2</sub> MeCH<sub>2</sub>), 2.6-2.7 (m, 3, H<sub>2'</sub>-2'' Me<sub>2</sub>CH), 3.3-3.4 (d of t, 4, formamidine PrCH<sub>2</sub>), 3.7 (s, 6, MeO), 3.8 (s, 1, OH), 4.1-4.3 (m, 2, H<sub>5'</sub>-5''), 4.7 (m, 1, H<sub>3'</sub>), 6.6 (t, 1, H<sub>1'</sub>), 6.7-7.5 (m, 13, aromatic), 7.99 (s, 1, formamidine CH), 8.05 (s, 1, H<sub>8</sub>), 9.07 (s, 1, N<sup>2</sup>H). (Found: C, 66.5; H, 7.10; N, 12.2; Calc for C<sub>44</sub>H<sub>55</sub>N<sub>7</sub>O<sub>6</sub>: C, 67.9; H, 7.17; N, 12.61%). m/z m<sup>-</sup> 776.41352 C<sub>44</sub>H<sub>54</sub>N<sub>7</sub>O<sub>6</sub> requires 776.41353 rf 0.65 (F).

*2-N-Isobutyryl, 5'-O-(4, 4'-dimethoxytrityl)N<sup>6</sup>-di-n-butylformamido-2-amino-2'*

*deoxyadenosine-3'-O-(2-cyanoethyl)-di-iso-propylamido-phosphite (10).* To vacuum dried 2-N-isobutyryl, 5'-O-(4, 4'-dimethoxytrityl)-N<sup>6</sup>-di-n-butylformamide-2-amino-2'-deoxyadenosine (2.0g 2.6mmol) dissolved in freshly distilled THF (15ml) under nitrogen was added dry diisopropylethylamine (1.33g 1.79ml 10.4mmol 4 equiv) and, dropwise, 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (1.22g 1.15ml 5.2mmol 2 equiv). After stirring at room temperature for 45 minutes the reaction was filtered under nitrogen pressure and the solvent removed to leave an oil which was dissolved in nitrogen-saturated ethyl acetate (50ml), washed with ice cold 10% Na<sub>2</sub>CO<sub>3</sub> solution (2 x 50ml), dried over Na<sub>2</sub>SO<sub>4</sub> and again taken to an oil. The oil was purified by wet flash chromatography on silica eluting with nitrogen

saturated ethyl acetate: acetonitrile 50: 50. The product fractions were collected and foamed from anhydrous acetonitrile to give an immobile oil.

2.39g, 93%  $^{31}\text{P}$  NMR ( $\text{CDCl}_3$ ) 149.1452, 149.0362 rf 0.90 (E).

*2-N,N6-Bis(di-n-butylformamido)-2-amino-2'-deoxyadenosine (2)*. 2-Amino-2'-deoxyadenosine (238mg 0.9mmol) was dried by coevaporation from dimethylacetamide and redissolved in fresh solvent (7ml). Di-n-butylformamide dimethyl acetal (53ml 2.25mmol 2.5equiv) was added and the mixture stirred at room temperature in the dark under nitrogen for 18 hours after which it was evaporated to an oil, dissolved in dichloromethane (20ml), washed with 5%  $\text{NaHCO}_3$  (2 x 50ml) and dried over anhydrous  $\text{Na}_2\text{SO}_4$ . The dried organic extract was evaporated to an oil and purified by wet flash chromatography on silica (methanol: dichloromethane: triethylamine, 5-10% MeOH in dichloromethane with 0.1% triethylamine) The product fractions were pooled and foamed before precipitation from dichloromethane (2-3ml) into 40-60 light petrol (250ml). The product was collected by filtration as an oily solid. 198mg, 40% m/z  $m^+$  545.39271  $\text{C}_{28}\text{H}_{49}\text{N}_8\text{O}_3$  requires 545.39274 rf 0.64 (D).

*2-N,N6-Bis(di-n-butylformamido)-5'-O-(4, 4'-dimethoxytrityl)-2-amino-2'-deoxyadenosine-3'-O-(2-cyanoethyl)-di-iso-propylamido-phosphite (13)*. To 2-N, N<sup>6</sup>-bis(di-n-butylformamido)-2-amino-2'-deoxyadenosine (1g 1.88mmol) dissolved in dry pyridine (15ml) was added 4, 4'-dimethoxytrityl chloride (700mg 2.07mmol 1.1 equiv). After stirring at room temperature for 5 hours water (20ml) was added and the mixture was extracted with chloroform (2 x 75ml). The organic extract was dried over  $\text{Na}_2\text{SO}_4$ . Removal of the solvent gave a mobile oil which was purified by wet flash chromatography on silica (ethanol: dichloromethane: triethylamine; 7.9: 92: 0.1.) to yield a mobile orange oil (580mg 0.68mmol) rf 0.34 (J). The oil was dried by coevaporation from pyridine, toluene and THF and dissolved in fresh THF (10ml) under nitrogen. N, N-diisopropylethylamine (252mg 0.48ml 2.72mmol 4equiv) was added followed by 2-cyanoethyl N, N-diisopropylchlorophosphoramidite (320mg 0.301ml 1.36mmol 2equiv) dropwise. After stirring at room temperature under nitrogen for 1 hour the reaction was

filtered under nitrogen pressure and taken to an oil. The oil was dissolved in nitrogen saturated ethylacetate (20ml) washed with ice cold 10% Na<sub>2</sub>CO<sub>3</sub> (2 x 25ml) and again taken to an oil. The oil was precipitated from ethylacetate (5ml) into hexane (800ml) at -50°C. The precipitate was filtered off and dessicated at 4°C to give an oil which was stored frozen under nitrogen.

525mg, 27% <sup>31</sup>P NMR (CDCl<sub>3</sub>) δ 148.078 (t, 1, P=O), 0.72 (E).

*2-Cyanoethylphosphorodichloridite.* To phosphorous trichloride (30.6ml 350mmol) dissolved in dry acetonitrile (20ml) under nitrogen was added dropwise from a second flask via a double ended needle a solution of 3-hydroxypropionitrile (2-cyanoethanol) (3.554g 3.41ml 50mmol 0.14equiv) in dry acetonitrile (20ml) with further solvent (5ml) used to rinse the delivery flask and needle. After stirring for 15 minutes at room temperature the solvent was removed to leave an oil. Distillation at reduced pressure yielded the product as a colourless oil.

6.48g, 76% Bp. 81°C 0.01mmHg δ<sub>H</sub> (CDCl<sub>3</sub> 80MHz.) δ 2.7 (t, 2, NCCH<sub>2</sub>, J<sub>6,13</sub>), 4.4 (d of t, 2, CH<sub>2</sub>OP, J<sub>6,13</sub>) <sup>31</sup>P NMR δ 178.8.

*2-Cyanoethyl N,N-diisopropylchlorophosphoramidite.* To 2-cyanoethyl phosphorodichloridite (0.787g 0.5ml 4.58mmol) stirred in dry diethyl ether (5ml) at 0°C under nitrogen was added dropwise from a second flask a solution of N, N-diisopropylamine (0.927g 1.284ml 9.16mmol 2equiv) in dry diethyl-ether (5ml). As soon as the addition was complete (2-3 minutes) the reaction mixture was filtered under nitrogen pressure and the solvent removed to leave an oil. The oil was purified by distillation on a Kugelrohr apparatus to give 2-cyanoethyl N,N-diisopropylchlorophosphoramidite as a colourless oil.

0.648g, 60% Bp. 150°C 0.8mm Hg. δ<sub>H</sub> (CDCl<sub>3</sub> 80MHz) δ 1.22 (d, 12, N[C(CH<sub>3</sub>)<sub>2</sub>]<sub>2</sub>, J<sub>6,82</sub>), 2.69 (t, 2, NC-CH<sub>2</sub>, J<sub>6,3</sub>), 3.60-3.91 (m, 2, N[CH]<sub>2</sub>), 4.05, 3.95 (2t, 2, P-OCH<sub>2</sub>, J<sub>6,40</sub>).

*2-N-Isobutyryl, 5'-O-(4, 4'-dimethoxytrityl)-N6-di-n-butylformamido-2-amino-2'-deoxyadenosine-3'-O-succinate*. To 2-N-isobutyryl, 5'-O-(4,4'-dimethoxytrityl)-N<sup>6</sup>-di-n-butylformamido-2-amino-2'-deoxyadenosine (311 mg 0.4mmol) in dry pyridine (5ml) was added 4-dimethylaminopyridine (24mg 0.5 equiv 0.2mmol) and succinic anhydride (32mg 0.32mmol). The reaction was stirred for 48 hrs at room temperature in the dark in a nitrogen atmosphere after which time it was evaporated to a gum by coevaporation with toluene. The residue was dissolved in dichloromethane (20ml), washed with ice cold 10% aqueous citric acid (2 x 15ml), dried over sodium sulphate and evaporated to a foam. The entire product was used for the functionalisation of CPG.

0.345g, 97% rf. 0.41 (G)

*Functionalisation of controlled pore glass with 2-N-isobutyryl, 5'-O-(4, 4'-dimethoxytrityl)-N6-di-n-butylformamido-2-amino-2'-deoxyadenosine-3'-O-succinate*. Controlled pore glass (Pierce CPG/long chain alkylamine) (1g) was dried in vacuo over P<sub>2</sub>O<sub>5</sub> for three days placed in a resin functionaliser under nitrogen and washed with a 10% solution of diisopropylethylamine in DMF (4 x 5ml) and DMF (5 x 10ml). While the washing steps were being performed the symmetric anhydride of the protected 2'-deoxynucleoside-3'-O-succinate was prepared. To the protected 2'-deoxynucleoside-3'-O-succinate (165mg 0.2mmol) in dichloromethane (1ml) was added a 200 microlitre aliquot of a DCCI stock solution (100mg) in dichloromethane (1ml). The mixture was stirred for 15 minutes, taken to dryness, the residue dissolved in DMF (2ml) and poured onto the CPG. Further DMF (1ml) was used to rinse the flask onto the CPG. The activated solution was left over the CPG for 20 hrs, gently agitating with nitrogen before filtration and washing of the CPG with DMF (5 x 10ml) and diethyl ether (10 x 10ml) to give a dry, free flowing powder. At this point the resin loading was assayed and the functionalisation repeated until a satisfactory level had been achieved after which the CPG was treated with a capping solution of acetic anhydride:pyridine (5ml 1:9v/v) for 1hr, washed with pyridine (5 x 10ml) and diethyl ether (10 x 10ml) and dried *in vacuo*.

*General method to assay dimethoxytrityl (nucleoside) loading of derivatised CPG.* This was measured as the amount of dimethoxytrityl cation released upon treatment of the CPG with acid. Dried derivatised CPG (2-3mg) was accurately weighed into a 25ml volumetric flask and made up to the mark with 60% HCl. After shaking for 5-10 minutes the absorbance of the solution was measured at 495nm in 1cm cells. The reading obtained was multiplied by 25 to give the total dimethoxytrityl absorbance for the CPG sample and divided by 71.7 (the absorbance of 1 micromole of dimethoxytrityl cation at 495nm) to convert it to the number of micromoles<sup>287</sup>. From this the number of micromoles per gram of CPG was calculated which is equivalent to the nucleoside loading. This was assayed as 17  $\mu\text{molg}^{-1}$ .

*Depurination of model nucleosides.* A small amount of each of the nucleosides (~1mg) was dissolved in a 3% solution of trichloroacetic acid in dichloromethane (100 microlitres). Samples were taken at time zero and at appropriate intervals thereafter, diluted with the appropriate eluent and injected onto an analytical ion exchange HPLC system eluting isochratically with 5mM TBAP at pH 4.5 containing either 5% acetonitrile v/v (2-N-isobutyryl-N<sup>6</sup>-benzoyl-2-amino-2'-deoxyadenosine and 2-N,N<sup>6</sup>-bis(di-n-butylformamide)-2-amino-2'-deoxyadenosine) or 10% acetonitrile v/v (2'-deoxyadenosine and 2-amino-2'-deoxyadenosine). The loss of the initial nucleoside peak and the growth of secondary peaks were monitored at 260nm and an approximate half life for the stability of the nucleosides in acid solution was obtained.

<u>Nucleoside</u>	<u>Half Life</u>
2'-deoxyadenosine	45 minutes
2-amino-2'-deoxyadenosine	35 minutes
2-N-isobutyryl-N <sup>6</sup> -benzoyl-2-amino-2'-deoxyadenosine	2 minutes
2-N,N <sup>6</sup> -bis(di-n-butylformamido)-2-amino-2'-deoxyadenosine	1.5-2 hours

*Stability of 2-N-Isobutyryl-N6-di-n-butylformamide-2-amino-2'-deoxyadenosine.* A small amount of the protected nucleoside (~1mg) was dissolved in a 3% solution of trichloroacetic acid in dichloromethane (100microlitres). Samples were taken at time zero and at intervals thereafter and examined by thin layer chromatography (ethanol: chloroform, 4: 10) for signs of degradation against the starting material. Examination of the plates revealed loss of starting material and build up of other products. From this an approximate half life for the stability of the nucleoside of in acid solution, of 1.5-2 hours, was obtained.

*General method for base composition analysis.* A solution of Tris acetate (0.025molar pH8.8) prepared from Tris base and magnesium acetate was adjusted to the required pH with acetic acid and used to make up a solution of the relevant oligonucleotide (0.5-1.0 OD units in 2.5ml). From the sample absorbance, the approximate number of micromoles of oligonucleotide and phosphate present were calculated. An appropriate number of units of phosphodiesterase 1 (5'-Exonuclease; Oligonucleate 5'-nucleotidohydrolase; EC 3.1.4.1) and alkaline phosphatase (Orthophosphoric-monoester phosphohydrolase [alkaline optimum]; EC 3.1.3.1) were added to fully digest the oligonucleotide, and the solution incubated at 37°C for 2-3 hrs or overnight. The digest solutions were analysed by HPLC using a preparative scale C8 reverse phase column and a gradient of : Buffer A 0.1 molar ammonium acetate and Buffer B 20% acetonitrile in 0.1 molar ammonium acetate.

<u>Time</u>	<u>%B</u>
0	0
5	0
25	100
27	100
28	0
34	0

This produced one peak per nucleoside with peak areas in the appropriate ratio of oligonucleotide composition. The peak authenticities were checked by comparison with a

series of nucleoside standards run under the same HPLC conditions both individually and as mixtures giving the following retention times: C, 9.5; G, 13.7; T, 14.4; A', 18 and A, 18.6 minutes.

#### Unit Definitions

Alkaline phosphatase: One unit will hydrolyse 1.0 micromole of p-nitrophenyl phosphate per minute at pH 10.4 at 37°C.

Phosphodiesterase 1: one unit will hydrolyse 1.0 micromole of bis(p-nitrophenyl) phosphate per minute at pH 8.8 at 37°C.

*2,3,4,5,6-Pentafluorophenylbenzoate*. To 2,3,4,5,6-pentafluorophenol (1.46g 7.96mmol) in dry DMF was added benzoic acid (0.96g 7.85mmol 0.98equiv) and DCCl (1.76g 8.53mmol 1.08equiv). After stirring for 1 hour at room temperature the mixture was filtered and evaporated to dryness to give a solid which was dissolved in ethyl acetate and filtered.

Removal of the solvent left an oil which crystallised on standing. Recrystallisation (ethanol:water;99:1) gave the product as a cream coloured solid.

1.79g, 71% Mpt 74-76°C (lit <sup>288</sup> 75-77°C).

*N4-Benzoyl-5-bromo-5'-O-(4, 4'-dimethoxytrityl)-2'-deoxycytosine*. 5-Bromo-2'-deoxycytosine (2g 6.53mmol) was dried by co-evaporation from dry pyridine and then suspended in fresh dry pyridine (20ml). To this was added 2,3,4,5,6-pentafluorophenylbenzoate (2.882g 9.79mmol 1.5equiv). The reaction was stirred for 24 hours at room temperature after which 4, 4'-dimethoxytrityl chloride (3.98g 11.7mmol 1.8equiv) 4-dimethylaminopyridine (72mg 0.5mmol 0.08equiv) and triethylamine (925mg 9.14mmol 1.27ml 1.4equiv) were added. After stirring for a further 18 hours at room temperature an equal volume of water was added and the mixture extracted with diethyl-ether (3 x 250ml). The solvent was removed and the oil produced purified by wet flash chromatography on silica (methanol: dichloromethane: triethylamine, 1: 99: 0.1) to give an orange gum.

991mg, 22% rf. 0.57 (H).

*N*<sup>4</sup>-Benzoyl-5-bromo-5'-O-4, 4'-dimethoxytrityl-2'-deoxycytosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite. To *N*<sup>4</sup>-benzoyl-5-bromo-5'-O-4, 4'-dimethoxytrityl-2'-deoxycytosine (991mg 1.4mmol) dried by coevaporation from pyridine toluene and THF dissolved in THF (15ml) in the presence of N,N-diisopropylethylamine (0.723g 0.971ml 5.6mmol 4 equiv) under nitrogen was added dropwise 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (0.663g 0.625ml 2.8mmol 2 equiv). After stirring for 1 hour the reaction was filtered under nitrogen pressure and taken to an oil which was dissolved in nitrogen saturated ethyl acetate (50ml) washed with ice cold 10% Na<sub>2</sub>CO<sub>3</sub> (2 x 75ml) dried over Na<sub>2</sub>SO<sub>4</sub> and again taken to an oil. The oil was precipitated from toluene (5ml) into hexane (800ml) at -45°C. The precipitate was collected, desiccated at 4°C for 18hrs and reprecipitated from toluene (8ml) into hexane (800ml) at -45°C to give a cream coloured solid which was stored under nitrogen at 4°C.  
1.035g, 81% <sup>31</sup>P NMR (CDCl<sub>3</sub>) δ 149.8768, 149.3809. m/z m<sup>+</sup> 910.25813 C<sub>46</sub>H<sub>50</sub>N<sub>5</sub>O<sub>8</sub>PBr requires 910.25807 rf. 0.65 (E).

*5'*-O-4, 4'-Dimethoxytrityl-5-bromo-2'-deoxyuridine. 5-Bromo-2'-deoxyuridine (2.0g 6.5mmol) was dried by coevaporation from pyridine and dissolved in fresh solvent (20ml). To this was added 4-dimethylaminopyridine (40mg 0.325mmol 0.05equiv) triethylamine (0.921g 1.268ml 9.1mmol 1.4equiv) and 4, 4'-dimethoxytrityl chloride (2.65g 7.8mmol 1.2equiv). After stirring for 2 hours at room temperature, TLC (ethylacetate: methanol: ammonia, 90: 15: 10) showed the reaction to be incomplete so more 4, 4'-dimethoxytrityl chloride (1.14g 3.38mmol 0.5equiv) and triethylamine (0.73g 7.2mmol 1.0ml 1.1equiv) were added. After 20 hours an equal volume of water was added and the mixture extracted with diethyl ether. The extract was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and taken to a foam which was purified by wet flash chromatography on silica (methanol: dichloromethane: triethylamine, 0.5-1.0: 99: 0.1) to give the product as a dry foam on removal of the solvent.  
3.96g, 87% λ<sub>max</sub> (CH<sub>3</sub>CN) 205 (46322dm<sup>3</sup> mol<sup>-1</sup> cm<sup>-1</sup>) 233 (14121) 276nm (6573)

$\delta_H$  (CDCl<sub>3</sub> 200 MHz) d 2.1-2.3 (m, 2, H<sub>2'-2''</sub>), 3.3 (m, 2, H<sub>5'-5''</sub>), 3.8 (s, 6, 2[CH<sub>3</sub>O]), 4.1 (m, 1, H<sub>4'</sub>), 4.55 (s, 1, H<sub>3'</sub>), 5.7 (brs, 1, OH), 6.25-6.4 (t, 1, H<sub>1'</sub>), 6.75-7.6 (m, 13, aromatic), 8 (s, 1, NH) m/z 609 (M+H<sup>+</sup>) 305 (M+H -trityl) rf. 0.26 (I).

*5-Bromo-5'-O-4, 4'-dimethoxytrityl-2'-deoxyuridine-3'-O-(2-cyanoethyl)-diiso-propylamido-phosphate.* To 5-bromo-5'-O-4, 4'-dimethoxytrityl-2'-deoxyuridine(3.37g 5.53mmol) dried by coevaporation from pyridine toluene and THF dissolved in THF (30ml) in the presence of N, N-diisopropylethylamine (2.859g 22mmol 3.85ml 4 equiv) was added dropwise 2-cyanoethyl N, N-diisopropylchlorophosphoramidite (2.62g 2.47ml 11.06mmol 2 equiv). After stirring at room temperature for 50 minutes the mixture was filtered under nitrogen pressure and the solvent removed to leave an oil which was dissolved in nitrogen saturated ethyl acetate (250ml), washed with ice cold 10% Na<sub>2</sub>CO<sub>3</sub> (2 x 100ml) dried over Na<sub>2</sub>SO<sub>4</sub> and again taken to an oil. The oil was precipitated from toluene (15ml) into hexane (600ml) at -78°C and the precipitate collected as a gum which was reprecipitated from toluene (15ml) into hexane (600ml) at -78°C and reprecipitated again from toluene (15ml) into hexane (600ml) at -45°C. The precipitate was collected by filtration, washed with hexane (400ml) at -60°C and dessicated at 4°C to give a white powder which was stored under nitrogen at 4°C. 2.761g, 61.7% <sup>31</sup>P NMR (CDCl<sub>3</sub>) d 149.5928, 149.2291 m/z m<sup>+</sup> 809.21391 C<sub>39</sub>H<sub>46</sub>N<sub>4</sub>O<sub>8</sub>PBr requires 809.2139 rf. 0.7 (E)

*Ultraviolet Thermal Denaturation of Oligonucleotides.* This was carried out using a Perkin Elmer Lambda 15 UV/Vis Spectrophotometer fitted with Digital Controller C570-0701, Temperature Programmer C570-0710 and a nitrogen gas stream to prevent the condensation of atmospheric moisture on cell windows thus allowing low temperature absorbances to be measured. The system was controlled via , and data stored on, an IBM XT Model 286 personal computer. Data handling was done using the Perkin Elmer PECSS2 programme and temperature conversion programme TCONV3 (modified for our laboratory by Dr G.A. Leonard).

*Ultraviolet Thermal Denaturation of dCATATG.* A sample of HPLC purified desalted freeze dried dCATATG was dissolved in SSC buffer (2.5ml) in a matched cell with quartz windows to give an absorbance of about 0.5Au at room temperature. The cells were heated to 60°C, slowly cooled to 1.3°C and allowed to equilibrate until the absorbance had stabilised. A series of timedrives was then done at 254nm over the temperature range 1.3-56.8°C. The data was stored and the absorbance versus time curves converted to absorbance versus temperature (degrees absolute) curves. The first derivatives of the profiles were used to evaluate the melting temperature of the oligonucleotide.

*Ultraviolet Thermal Denaturation of dCA'TA'TG.* The melting profiles and hence the melting temperature was obtained by the same method as used for dCATATG.

*Ultraviolet Thermal Denaturation of dTA'TA'.* A sample of HPLC purified and desalted d(TA')<sub>2</sub> was dissolved in SSC buffer as for d(TA')<sub>3</sub> and attempts made to obtain the melting profiles as before. This was unsuccessful due to the low melting temperature of the oligonucleotide.

*Ultraviolet Thermal Denaturation of dTA'TA'TA'.* HPLC purified and desalted d(TA')<sub>3</sub> was dissolved in SSC buffer (2.5ml) in a matched cell with quartz windows to give an absorbance of about 1au at room temperature. The cells were heated to 60°C, slowly cooled to 0.0°C and left to equilibrate until the absorbance had stabilised. A series of timedrives at 260nm was made over the temperature range 0.0-54.5°C heating at the nominal rate of 1.0°C per minute. The data was stored and the absorbance versus time curves converted to absorbance versus temperature (degrees absolute) curves. The first derivatives of the melting profiles were used to evaluate T<sub>m</sub> for the oligonucleotide.

*Ultraviolet Thermal Denaturation of dTATATA.* A sample of HPLC purified and desalted d(TA)<sub>3</sub> was dissolved in SSC buffer as for d(TA')<sub>3</sub> and attempts were made to obtain the

melting profiles as before. This was unsuccessful due to the low melting temperature of the oligonucleotide.

*Ultraviolet Thermal Denaturation of dTATATATA.* HPLC purified and desalted d(TA)<sub>4</sub> was dissolved in a matched cell with quartz windows containing SSC buffer (2.5ml) so as to give an initial absorbance of about 1au against a blank containing only the buffer at room temperature. The cells were heated to 60°C, slowly cooled and allowed to equilibrate until the absorbance had stabilised. A series of timedrives at 260nm was then made over the temperature range 0.0-72.7°C heating at the nominal rate of 1.0°C per minute. The data was stored the absorbance versus time curves converted to absorbance versus temperature (degrees absolute) curves and the first derivatives taken to establish the melting temperature (T<sub>m</sub>). Once the melting temperature for the oligonucleotide in SSC had been established MgCl<sub>2</sub>.6H<sub>2</sub>O (0.127g 0.625mmol) was added to each cell to give a solution of 0.25M MgCl<sub>2</sub> in SSC. The timedrives were repeated over the same temperature range to give the melting temperature for the oligonucleotide in 0.25M MgCl<sub>2</sub>. Further MgCl<sub>2</sub>.6H<sub>2</sub>O was added to give solutions of 0.5M, 1.0M and 2.0M MgCl<sub>2</sub> and the timedrives repeated to give the melting temperature of the oligonucleotide at each of these salt concentrations.

*Ultraviolet Thermal Denaturation of dTA'TA'TA'TA'.* This was determined in the same way as for d(TA)<sub>4</sub> over the same temperature ranges and at the same salt concentrations.

#### *Synthesis of Trial Probes.*

HLP-9tet	dA'CTA'TCGA'C
HLP-10tet	dCA'CTA'TCGA'C
51P-10	dGA'A'GA'GA'GTT

These were prepared using a small scale 0.2umol cycle SSCE103a on an ABI 380B DNA synthesiser utilising standard phosphoramidite methods. Coupling efficiency was assayed by means of the released trityl cation and was found to be 98-99%.

*Evaluation of 2-Amino-2'-deoxyadenosine Extinction Coefficient at 260nm.* 2,6-diaminopurine (0.026mg 0.01mmol) was dissolved in H<sub>2</sub>O (25ml) and the UV spectrum recorded over the range 320-190nm. The absorbance at 260nm was used to calculate the extinction coefficient.  $E_{260} = 8200$ .

*Ultraviolet Thermal Denaturation of Trial Probes and their Native Equivalents.* The native versions of HLP-9tet, HLP-10tet and 51P-10 were synthesised and purified along with their native complimentary strands to give a set of nine related oligonucleotides

		$\Sigma_{260}$	$T_m(K)$
HLP-9tet	dA'CTA'TCGA'C	68.22	311.10
HLP-10tet	dCA'CTA'TCGA'C	74.79	320.13
51P-10	dGA'A'GA'GA'GTT	87.48	321.78
HLP-9Ntet	dACTATCGAC	81.72	303.65
HLP-10Ntet	dCACTATCGAC	74.79	313.62
51P-10N	dGAAGAGAGTT	113.40	307.57
HLP-9tetC	dGTCGATAGT	87.66	
HLP-10tetC	dGTCGATAGTG	100.17	
51P-10C	dAACTCTCTTC	85.68	

The extinction coefficient for each oligonucleotide was calculated using the values  $\Sigma_{260}^{289}$   $T = 8.8$ ,  $C = 7.3$ ,  $G = 11.7$ ,  $A = 15.4$  and  $A' = 8.2 \text{ cm}^2\text{umol}^{-1}$ . To allow for suppression due to single strand base stacking this value was multiplied by 0.9 to give the values in the table above.

A sample of each oligonucleotide was dissolved in H<sub>2</sub>O (1ml). An OD<sub>260</sub> reading was taken for each sample and their molarities calculated. An aliquot of each probe or its native equivalent (N) and its compliment (C) was taken and annealed in a UV cell containing SSC buffer (2.75ml) at 80°C to give a solution which was 0.36 umolar in duplex. The annealed solution was slowly cooled to 10°C and allowed to equilibrate until the temperature and absorbance had stabilised. A series of three timedrives at 260nm was carried out for each

duplex collecting 480 data points at 10 second intervals over the temperature range 10.0-83.2°C. The heating rate was nominally 1.0°C per minute. The absorbance versus time curves were stored and converted to absorbance versus temperature curves (degrees absolute). The  $T_m$  value for each probe duplex and its native analogue was obtained from the first derivative of the curve.

## 12.0 References

1. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737-738.
2. S. Furberg, *Acta. Chem. Scand.*, 1952, **6**, 634-640.
3. S. Zamenhof, G. Brawerman and E. Chargaff, *Biochim. Biophys. Acta*, 1952, **9**, 402-405.
4. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 964-967.
5. E. C. Freidberg, 1985, *DNA Repair*. W. H. Freeman and Co., New York
6. M. D. Topal and J. R. Fresco, *Nature*, 1976, **263**, 285-289.
7. F. H. C. Crick, *J. Mol. Biol.*, 1966, **19**, 548-555.
8. D. J. Patel, S. A. Koslowski, L. A. Marky, J. A. Rice, C. Broka, J. Dallas, K. Itakura and K. Breslauer, *Biochemistry*, 1982, **21**, 437-444.
9. D. J. Patel, S. A. Koslowski, S. Ikuta and K. Itakura, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 1984, **43**, 2663-2670.
10. T. Brown, O. Kennard, G. Kneale and D. Rabinovich, *Nature*, 1985, **315**, 604-606.
11. G. Kneale, T. Brown, O. Kennard and D. Rabinovich, *J. Mol. Biol.*, 1985, **186**, 805-814.
12. W. N. Hunter, G. Kneale, T. Brown, D. Rabinovich and O. Kennard, *J. Mol. Biol.*, 1986, **190**, 605-618.
13. P. S. Ho, C. A. Frederick, G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. H. -J. Wang and A. Rich, *The EMBO Journal*, 1985, **4**, 3617-3623.
14. W. N. Hunter, T. Brown, N. N. Anand and O. Kennard, *Nature*, 1986, **320**, 552-555.
15. W. N. Hunter, T. Brown and O. Kennard, *Nucleic Acids Res.*, 1987, **15**, 6589-6606.
16. X. Gao and D. J. Patel, *J. Biol. Chem.*, 1987, **262**, 16973-16984.
17. T. Brown, G. Kneale, W. N. Hunter and O. Kennard, *Nucleic Acids Res.*, 1986, **14**, 1801-1809.
18. M. Coll, D. Saal, C. A. Frederick, J. Aymani, A. Rich and A. H. -J. Wang, *Nucleic Acids Res.*, 1989, **17**, 911-923.
19. W. B. T. Cruse, J. Aymani, O. Kennard, T. Brown, A. G. C. Jack and G. A. Leonard, *Nucleic Acids Res.*, 1989, **17**, 55-72.
20. T. A. Steitz, L. Beese, P. S. Freemont, J. M. Freedman and M. R. Sanderson, *Cold Spring Harbor Symp. Quant. Biol.*, 1987, **52**, 465-471.
21. J. -P. Claverys, V. Mejan, A. -M. Gasc and A. M. Sicard, *Proc. Natl. Acad. Sci. USA.*, 1983, **80**, 5956-5960.

22. B. Kramer, W. Kramer and H. -J. Fritz, *Cell*, 1984, **38**, 879-887.
23. C. Dohet, R. Wagner and M. Radman, *Proc. Natl. Acad. Sci. USA.*, 1985, **82**, 503-505.
24. G. V. Fazakerley, E. Quignard, A. Woisard, W. Guschlbauer, G. A. van der Marel, J. H. van Boom, M. Jones and M. Radman, *The EMBO Journal*, 1986, **5**, 3697-3703.
25. S. -S. Su, R. S. Lahue, K. G. Au and P. Modrich, *J. Biol. Chem.*, 1988, **263**, 6829-6835.
26. R. E. Wagner and M. Meselson, *Proc. Natl. Acad. Sci. USA.*, 1976, **73**, 4135-4139.
27. P. J. Pukila, J. Peterson, G. Herman, P. Modrich and M. Meselson, *Genetics*, 1983, **104**, 571-582.
28. R. L. Letsinger and V. Mahadevan, *J. Am. Chem. Soc.*, 1965, **87**, 3526-3527.
29. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.*, 1981, **22**, 1859-1862.
30. V. P. Chuprina and V. I. Poltev, *Nucleic Acids Res.*, 1983, **11**, 5205-5222.
31. M. McCall, T. Brown, W. B. T. Cruise and O. Kennard, *Acta Crystallogr.*, 1984, **A40**, 6-46.
32. A. R. Fersht, J. -P. Shi and W. -C. Tsui, *J. Mol. Biol.*, 1983, **165**, 655-667.
33. S. -H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. -J. Wang, N. C. Seeman and A. Rich, *Science*, 1974, **185**, 435-440.
34. G. J. Quigley, N. C. Seeman, A. H. -J. Wang, F. L. Suddath and A. Rich, *Nucleic Acids Res.*, 1975, **2**, 2329-2341.
35. G. J. Quigley and A. Rich, *Science*, 1976, **194**, 776-806.
36. A. Rich, *Acc. Chem. Res.*, 1977, **10**, 388-402.
37. W. Traub and J. L. Sussman, *Nucleic Acids Res.*, 1982, **10**, 2701-2708.
38. D. J. Patel, L. Shapiro, S. A. Kozlowski, B. L. Gaffney and R. A. Jones, *J. Mol. Biol.*, 1986, **188**, 677-692.
39. L. -S. Kan, S. Chandrasegaran, S. M. Pulford and P. S. Miller, *Proc. Natl. Acad. Sci. USA.*, 1983, **80**, 4263-4265.
40. D. J. Patel, S. A. Kozlowski, S. Ikuta and K. Itakura, *Biochemistry*, 1984, **23**, 3207-3217.
41. O. Kennard, *J. Biomolec. Str. Dyn.*, 1985, **3**, 205-226.
42. T. Brown, W. N. Hunter, G. Kneale and O. Kennard, *Proc. Natl. Acad. Sci. USA.*, 1986, **83**, 2402-2406.
43. W. N. Hunter, T. Brown and O. Kennard, *J. Biomolec. Str. Dyn.*, 1986, **4**, 173-191.

44. G. G. Privé, U. Heinemann, S. Chandrasegaran, L. -S. Kan, M. L. Kopka and R. E. Dickerson, *Science*, 1987, **238**, 498-504.
45. X. Gao and D. J. Patel, *J. Am. Chem. Soc.*, 1988, **110**, 5178-5182.
46. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1980, **287**, 755-758.
47. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA.*, 1981, **78**, 2179-2183.
48. R. E. Dickerson and H. R. Drew, *J. Mol. Biol.*, 1981, **149**, 761-786.
49. H. R. Drew and R. E. Dickerson, *J. Mol. Biol.*, 1981, **151**, 535-556.
50. D. J. Patel, S. A. Koslowski, L. A. Marky, C. Broka, J. A. Rice, K. Itakura and K. S. Breslauer, *Biochemistry*, 1982, **21**, 428-436.
51. A. H. -J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature*, 1979, **282**, 680-686.
52. A. H. -J. Wang, G. Ughetto, G. J. Quigley, T. Hakoshima, G. A. van der Marel, J. H. van Boom and A. Rich, *Science*, 1984, **225**, 115-121.
53. T. Brown, G. A. Leonard, E. D. Booth and J. Chambers, *J. Mol. Biol.*, 1989, **207**, 455-457.
54. G. A. Leonard, E. D. Booth and T. Brown, *Nucleic Acids Res.*, 1990, **19**, 5617-5623.
55. T. A. Jones, *J. Applied Cryst.*, 1978, **11**, 268-272.
56. H. A. Sober, R. A. Harte and E. K. Sober, "Handbook of Biochemistry. Selected Data for Molecular Biology", Chemical Rubber Company, Cleveland, 1970, pp. G-3 to G-98
57. L. C. Sowers, G. V. Fazakerley, H. Kym, L. Dalton and M. F. Goodman, *Biochemistry*, 1986, **25**, 3983-3988.
58. T. Brown, G. A. Leonard, E. D. Booth and G. Kneale, *J. Mol. Biol.*, 1990, **212**, 437-440.
59. A. N. Lane, personal communication.
60. B. Pullman and A. Pullman, *Adv. Heterocycl. Chem.*, 1971, **13**, 77-159.
61. R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1984, **106**, 244-248
62. H. C. Nelson, J. T. Finch, B. F. Luisi and A. Klug, *Nature*, 1987, **330**, 221-226.
63. M. Coll, C. A. Frederick, A. H. -J. Wang and A. Rich, *Proc. Natl. Acad. Sci. USA.*, 1987, **84**, 8385-8389.
64. C. A. Frederick, A. H. -J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. A. van der Marel and A. Rich, *Nature*, 1979, **286**, 680-686.

65. F. H. C. Crick, *J. Mol. Biol.*, 1966, **19**, 548-555.
66. Y. Takahashi, K. Kato, Y. Hayashizaki, T. Wakabayashi, E. Ohtsuka, S. Matsuki, M. Ikehara and K. Matsubara, *Proc. Natl. Acad. Sci. USA.*, 1985, **82**, 1931-1935.
67. B. D. Davis, P. Anderson and P. F. Sparking, *J. Mol. Biol.*, 1973, **76**, 223-232.
68. E. Ohtsuka, S. Matsuki, M. Ikehara, Y. Takahashi and K. Matsubara, *J. Biol. Chem.*, 1985, **260**, 2605-2608.
69. F. H. Martin, M. M. Castro, F. Aboul-ela and I. Tinoco Jr, *Nucleic Acids Res.*, 1985, **13**, 8927-8938.
70. T. Lindahl, *Prog. Nucleic Acids Res. Mol. Biol.*, 1979, **22**, 135-192.
- 71a. P. Karran and T. Lindahl, *J. Biol. Chem.*, 1978, **253**, 5877-5879.
- 71b. P. Karran and T. Lindahl, *Biochemistry*, 1980, **19**, 6005-6011.
72. B. Myrnes, P-H. Guddal and H. Krokan, *Nucleic Acids Res.*, 1982, **10**, 3693-3701.
73. P. W. R. Corfield, W. N. Hunter, T. Brown, P. Robinson and O. Kennard, *Nucleic Acids Res.*, 1987, **15**, 7935-7949.
74. S. Uesugi, Y. Oda, M. Ikehara, Y. Kawase and E. Ohtsuka, *J. Biol. Chem.*, 1987, **262**, 6965-6968.
75. W. A. Hendrickson and J. H. Konnert, in "Biomolecular Structure, Conformation, Function and Evolution", ed. R. Srinivasan, Pergamon Press, Oxford, 1981.
76. T. Brown, E. D. Booth and G. A. Leonard, in "Molecular Mechanisms in Bioorganic Processes", eds. C. Bleasdale and B. T. Golding, The Royal Society of Chemistry, Cambridge, 1990, 115-132.
77. I. Kulakowska, M. Geller B. Lesyng and K. L. Wierzchowski, *Biochim. Biophys. Acta*, 1974, **361**, 119-130.
78. E. Freese, *J. Mol. Biol.*, 1959, **1**, 87-105.
79. T. R. Skopek and F. Hutchinson, *J. Mol. Biol.*, 1982, **159**, 19-33.
80. R. L. Hopkins and M. F. Goodman, *Proc. Natl. Acad. Sci. USA.*, 1980, **77**, 1801-1805.
81. R. S. Lasken and M. F. Goodman, *J. Biol. Chem.*, 1984, **259**, 11491-11495.
82. A. R. Katritzky and A. J. Waring, *J. Chem. Soc.*, 1962, 1540-1544.
83. H. Iwahashi and Y. Kyogoku, *J. Am. Chem. Soc.*, 1977, **99**, 7761-7765.
84. M. D. Kirnos, I. Y. Khudyakov, N. I. Alexandrushkina and B. F. Vanyushin, *Nature*, 1977, **270**, 369-370.
85. I. YA. Khudyakov, M. D. Kirnos, N. I. Alexandrushkina and B. F. Vanyushin, *Virology*, 1978, **88**, 8-18.

86. F. B. Howard, J. Frazier and H. T. Miles, *J. Biol. Chem.*, 1966, **241**, 4293-4295.
87. S. J. Stahl and M. D. Chamberlin, *J. Biol. Chem.*, 1978, **253**, 4951-4959.
88. A. Cerami, E. Reich, D. C. Ward and I. H. Goldberg, *Proc. Natl. Acad. Sci. USA.*, 1967, **57**, 1036-1042.
89. K. H. Scheit and H. -R. Rackwitz, *Nucleic Acids Res.*, 1982, **10**, 4059-4069.
90. B. L. Gaffney, L. A. Marky and R. A. Jones, *Nucleic Acids Res.*, 1982, **10**, 4351-4361.
91. B. L. Gaffney, L. A. Marky and R. A. Jones, *Tetrahedron*, 1984, **40**, 3-13.
92. F. B. Howard and H. T. Miles, *Biochemistry*, 1984, **23**, 6723-6732.
93. F. B. Howard and H. T. Miles, *Biopolymers*, 1983, **22**, 597-600.
94. J. -M. Neumann, W. Guschlbauer and S. Trinh-Dinh, *Eur. J. Biochem.*, 1979, **100**, 141-148.
95. A. Chollet, A. Chollet-Damerius and E. H. Kawashima, *Chemica Scripta*, 1986, **26**, 37-40.
96. A. Chollet and E. Kawashima, *Nucleic Acids Res.*, 1988, **16**, 305-317.
97. C. Cheongh, I. Tinoco Jr. and A. Chollet, *Nucleic Acids Res.*, 1988, **16**, 5115-5122.
98. A. D. DiGabriele, M. R. Sanderson and T. A. Steitz, *Proc. Natl. Acad. Sci. USA.*, 1989, **86**, 1816-1820.
99. M. Falk, K. A. Hartman Jr. and R. C. Lord, *J. Am. Chem. Soc.*, 1963, **85**, 391-394.
100. B. N. Conner, T. Takanu, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1982, **295**, 294-299.
101. S. Arnott and D. W. L. Huskins, *Biochem. Biophys. Res. Commun.*, 1972, **47**, 1504-1510.
102. E. Selsing, S. Arnott and R. L. Ratliff, *J. Mol. Biol.*, 1975, **98**, 243-248.
103. S. Arnott, R. Chandrasekaran, D. W. L. Huskins, P. J. C. Smith and L. Watts, *J. Mol. Biol.*, 1974, **88**, 523-533.
104. F. M. Pohl and T. M. Jovin, *J. Mol. Biol.*, 1972, **67**, 375-396.
105. D. J. Patel, L. L. Canuel and F. M. Pohl, *Proc. Natl. Acad. Sci. USA.*, 1979, **76**, 2508-2511.
106. H. Drew, T. Takano, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1980, **286**, 567-573.

107. E. M. Lafer, A. Moller, A. Nordheim, B. D. Stoller and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 3546-3550.
108. E. Hamori and T. M. Jovin, *Biophysical J.*, 1981, **33**, 67a.
109. M. Behe, S. Zimmerman and G. Felsenfeld, *Nature*, 1981, **293**, 233-235.
110. H. H. Chen, M. J. Behe and D. C. Rau, *Nucleic Acids Res.*, 1984, **12**, 2381-2389.
111. B. Borah, J. S. Cohen and A. D. Bax, *Biopolymers*, 1985, **24**, 747-765.
112. S. Devarajan and R. H. Schafer, *Nucleic Acids Res.*, 1986, **14**, 5099-5109.
113. A. Woisard, W. Guschlbauer and G. V. Fazakerley, *Nucleic Acids Res.*, 1986, **14**, 3515-3519.
114. A. Moller, A. Nordheim, S. R. Nichols and A. Rich, *Proc. Natl. Acad. Sci. USA.*, 1981, **78**, 4777-4781.
115. F. Quadrifoglio, G. Manzini, M. Vasser, K. Dinkelspiel and R. Crea, *Nucleic Acids Res.*, 1981, **9**, 2195-2206.
116. F. Quadrifoglio, G. Manzini, K. Dinkelspiel and R. Crea, *Nucleic Acids Res.*, 1982, **10**, 3759-3769.
117. H. Hamada and T. Kakunaga, *Nature*, 1982, **298**, 396-398.
118. L. P. McIntosh, I. Greiger, D. A. Zarlring, M. Robert-Nicoud, J. H. van de Sande, K. F. Jorgenson and F. Eckstein, *Nature*, 1983, **304**, 83-86.
119. T. M. Jovin, L. P. McIntosh, D. A. Zarlring, D. J. Arndt-Jovin, M. Robert-Nicoud and J. H. Van de Sande, *Jerusalem Symp. Quantum Chem. Biochem.*, 1983, **16**, 89-99.
120. T. M. Jovin, L. P. McIntosh, D. J. Arndt-Jovin, D. A. Zarlring, M. Robert-Nicoud, J. H. van de Sande, K. F. Jorgenson and F. Eckstein, *J. Biomolec. Str. Dyn.*, 1983, **1**, 21-55.
121. M. Vorlíčková, J. Kypr, S. Stokrova and J. Sponar, *Nucleic Acids Res.*, 1982, **10**, 1071-1080.
122. C. Zimmer, S. Tymen, C. Marck and W. Guschlbauer, *Nucleic Acids Res.*, 1982, **10**, 1081-1091.
123. S. Ikuta and Y. -S. Wang, *Nucleic Acids Res.*, 1989, **17**, 4131-4144.
124. A. H. -J. Wang, R. Gessner, G. A. van der Marel, J. H. van Boom and A. Rich, *Proc. Natl. Acad. Sci. USA.*, 1985, **82**, 3611-3615.
125. A. H. -J. Wang, T. Hakoshima, G. H. van der Marel, J. H. van Boom and A. Rich, *Cell*, 1984, **37**, 321-331.
126. J. Feigon, A. H. -J. Wang, G. A. van der Marel, J. H. van Boom and A. Rich, *Science*, 1985, **230**, 82-84.

127. J. A. Taboury, S. Adam, E. Taillandier, J.-M. Neumann, S. Tran-Dinh, T. Huynh-Dinh, B. Langlois d'Estaintot, M. Conti and J. Igolen, *Nucleic Acids Res.*, 1984, **15**, 6291-6305.
128. S. P. A. Fodor, P. A. Starr and T. G. Spiro, *Biopolymers*, 1985, **24**, 1493-1500.
129. M. Vorlíčková and J. Kypr, *J. Biomolec. Str. Dyn.*, 1985, **3**, 67-83.
130. M. Vorlíčková, P. Sedláček, J. Kypr and J. Sponar, *Nucleic Acids Res.*, 1982, **10**, 6969-6979.
131. M. Vorlíčková and J. Kypr, *Int. J. Biol. Macromol.*, 1984, **6**, 77-80.
132. B. Hartmann, J. Pilet, M. Ptak, J. Ramstein, B. Malfoy and M. Leng, *Nucleic Acids Res.*, 1982, **10**, 3261-3277.
- 133a. W. Saenger, W. N. Hunter and O. Kennard, *Nature*, 1986, **324**, 385-388.
- 133b. W. Saenger, W. N. Hunter and O. Kennard, *Nature*, 1988, **335**, 596.
134. S. N. Rao, T. Jovin and P. A. Kollman, *Proc. Gulf Shores Symposium*, 1988, 267-285.
135. D. R. Davies and R. L. Baldwin, *J. Mol. Biol.*, 1963, **6**, 251-255.
136. H. R. Drew and R. E. Dickerson, *The EMBO Journal*, 1982, **1**, 663-667.
137. I. E. Scheffler, E. L. Elson and R. L. Baldwin, *J. Mol. Biol.*, 1968, **36**, 291-304.
- 138a. M. A. Viswamitra, O. Kennard, Z. Shakked, P. G. Jones, G. M. Sheldrick, S. Salisbury and L. Falvello, *Curr. Sci.*, 1978, **47**, 289-292.
- 138b. M. A. Viswamitra, O. Kennard, P. G. Jones, G. M. Sheldrick, S. Salisbury, L. Falvello and Z. Shakked, *Nature*, 1978, **273**, 687-688.
139. A. Klug, M. A. Viswamitra, O. Kennard, Z. Shakked and T. A. Steitz, *J. Mol. Biol.*, 1979, **131**, 669-680.
140. M. Rajagopalen, G. Gupta and V. Sasisekharan, *FEBS Lett.*, 1983, **159**, 285-289.
141. C. Yoon, G. G. Privé, D. S. Goodsell and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA.*, 1988, **85**, 6332-6336.
142. H. Shindo, R. T. Simpson and J. S. Cohen, *J. Biol. Chem.*, 1979, **254**, 8125-8128.
143. J. Kypr, M. Vorlíčková, M. Budesinsky and V. Sklenar, *Biochem. Biophys. Res. Commun.*, 1981, **99**, 1257-1264.
144. J. S. Cohen, J. B. Wooten and C. L. Chatterjee, *Biochemistry*, 1981, **20**, 3049-3055.
145. H. Shindo, *Eur. J. Biochem.*, 1981, **120**, 309-312.
146. D. J. Patel, S. A. Kozlowski, J. W. Suggs and S. D. Cox, *Proc. Natl. Acad. Sci. USA.*, 1981, **78**, 4063-4067.

147. C-W. Chen and J. S. Cohen, *Biopolymers*, 1983, **22**, 879-893.
148. F. Eckstein and T. M. Jovin, *Biochemistry*, 1983, **22**, 4546-4550.
149. N. Assa-Munt and D. R. Kearns, *Biochemistry*, 1984, **23**, 791-796.
150. J. Kypr, V. Sklenar and M. Vorlíčková, *Gen. Physiol. Biophys.*, 1986, **5**, 77-87.
151. E. Suzuki, N. Pattabiraman, G. Zon and T. L. James, *Biochemistry*, 1986, **25**, 6854-6865.
152. M. Vorlíčková, J. Kypr, V. Kleinwachter and E. Palecek, *Nucleic Acids Res.*, 1980, **8**, 3965-3973.
153. M. Vorlíčková, J. Kypr and V. Sklenar, *J. Mol. Biol.*, 1983, **166**, 85-92.
154. D. S. Zarlenga, H. B. Halsall and R. A. Day, *Nucleic Acids Res.*, 1984, **12**, 6325-6335.
155. J. Kypr and M. Vorlíčková, *Gen. Physiol. Biophys.*, 1984, **3**, 223-229.
156. M. Vorlíčková, E. E. Minyat and J. Kypr, *Biopolymers*, 1984, **23**, 1-4.
157. J. Kypr and M. Vorlíčková, *Biochem. Biophys. Res. Commun.*, 1984, **123**, 831-835.
158. M. Vorlíčková, E. Birch-Hirschfeld, H. Lang and J. Kypr, *Biochem. Biophys. Res. Commun.*, 1986, **139**, 1158-1163.
159. J. Kypr and M. Vorlíčková in "Structure and Expression", eds. R. H. Sarma and M. H. Sarma, Adenine Press, London, 1988, vol. 2, p. 105-121
160. M. Atreyi, M. V. R. Rao and S. Kumar, *Ind. J. Biochem. Biophys.*, 1988, **25**, 385-388.
161. D. W. Gruenwedel and M. K. Cruikshank, *Nucleic Acids Res.*, 1989, **17**, 9075-9086.
162. V. P. Chuprina, *FEBS Lett.*, 1985, **186**, 98-102.
163. M. J. Tunis and J. E. Hearst, *Biopolymers*, 1968, **6**, 1345-1353.
164. H. W. Zimmerman in, "Organic Liquids ( Structure Dynamics and Chemical Properties )," A. D. Buckingham, E. Lippert and S. Bratos eds. John Willey and Sons Ltd, Town 1978, p1-15.
165. I. Tinoco Jr., C. Bustamante and M. F. Maestre, *Ann. Rev. Biophys. Bioeng.*, 1980, **9**, 107-141.
166. B. B. Johnson, K. S. Dahl, I. Jr. Tinoco, V. I. Ivanov and V. B. Zhurkin, *Biochemistry*, 1981, **20**, 73-78.
167. D. G. Gorenstein, *Ann. Rev. Biophys. Bioeng.*, 1981, **10**, 355-386.
168. F. B. Howard, C. Chen, J. S. Cohen and H. T. Miles, *Biochem. Biophys. Res. Commun.*, 1984, **118**, 848-853.

169. F. B. Howard and H. T. Miles, *Nucleic Acids: The Vectors of Life*. D. Reidel
170. B. Borah, J. S. Cohen, F. B. Howard and H. T. Miles, *Biochemistry*, 1985, **24**, 7456-7462.
171. L. J. Rinkel, J. -R. Mellema, G. A. van der Marel, J. H. van Boom and C. Altona, *Eur. J. Biochem.*, 1986, **154**, 259-265.
172. M. Vorlíčková, J. Sagi, A. Szabolcs, A. Szemzo, L. Otvos and J. Kypr, *Nucleic Acids Res.*, 1988, **16**, 279-289.
173. M. Vorlíčková, J. Sagi, A. Szabolcs, A. Szemzo, L. Otvos and J. Kypr, *J. Biomolec. Str. Dyn.*, 1988, **6**, 503-510.
174. B. Borah, F. B. Howard, H. T. Miles and J. S. Cohen, *Biochemistry*, 1986, **25**, 7464-7470.
175. J. C. Sutherland, K. P. Griffen, P. C. Keck and P. Z. Jaknes, *Proc. Natl. Acad. Sci. USA.*, 1981, **78**, 4801-4804.
176. V. N. Bartenev, E. I. Golovamov, K. A. Kapitonova, M. A. Mokulskii, L. I. Volkova and I. Y. Skuratovskii, *J. Mol. Biol.*, 1983, **169**, 217-234.
- T. J. Thomas and V. A. Bloomfield, *Biopolymers*, 1985, **24**, 2185-2194.
177. V. I. Ivanov, L. E. Minchenkova, E. E. Miyat, M. D. Frank-Kamenetskii and A. J. Schyolkina, *J. Mol. Biol.*, 1974, **87**, 817-833.
178. M. Vorlíčková, J. Kypr, J. Sagi, L. Otvos and V. Sklenar, *Int. J. Biol. Macromol.*, 1987, **9**, 131-136.
179. M. Coll, A. H. -J. Wang, G. A. van der Marel, J. H. van Boom and A. Rich, *J. Biomolec. Str. Dyn.*, 1986, **4**, 157-172.
180. R. L. Letsinger, J. L. Finnan, G. A. Heavner and W. B. Lunsford, *J. Am. Chem. Soc.*, 1975, **97**, 3278-3279.
181. G. Alvarado-Urbina, G. M. Sathe, W. C. Liu, M. F. Gillen, P. D. Duck, R. Bender and K. K. Ogilvie, *Science*, 1981, **214**, 270-274.
182. R. L. Letsinger and W. B. Lunsford, *J. Am. Chem. Soc.*, 1976, **98**, 3655-3661.
183. M. D. Matteucci and M. H. Caruthers, *Tetrahedron Lett.*, 1980, **21**, 719-722.
184. M. D. Matteucci and M. H. Caruthers, *J. Am. Chem. Soc.*, 1981, **103**, 3185-3191.
185. G. W. Daube and E. E. van Tamelen, *J. Am. Chem. Soc.*, 1977, **99**, 3526-3528.
186. S. P. Adams, K. S. Kavka, E. J. Wykes, S. B. Holder and G. R. Gallupi, *J. Am. Chem. Soc.*, 1983, **105**, 661-663.
187. L. J. McBride and M. H. Caruthers, *Tetrahedron Lett.*, 1983, **24**, 245-248.
188. J. L. Fourrey and J. Varenne, *Tetrahedron Lett.*, 1983, **24**, 1963-1966.

189. S. L. Beaucage, *Tetrahedron Lett.*, 1984, **25**, 375-378.
190. A. D. Barone, J. -Y. Tang and M. H. Caruthers, *Nucleic Acids Res.*, 1984, **12**, 4051-4061.
191. G. W. Daube and E. E. van Tamelen, *J. Am. Chem. Soc.*, 1977, **99**, 3526-3528.
192. M. D. M. Gray and D. J. H. Smith, *Tetrahedron Lett.*, 1980, **21**, 859-860.
193. D. J. H. Smith, K. K. Ogilvie and M. F. Gillen, *Tetrahedron Lett.*, 1980, **21**, 861-864.
194. T. Atkinson and M. Smith in, "Oligonucleotide Synthesis a Practical Approach," ed. M. J. Gait, IRL Press, Oxford, 1984, p69.
195. X. Gao, B. L. Gaffney, M. Senior, R. R. Riddle and R. A. Jones, *Nucleic Acids Res.*, 1985, **13**, 573-584.
196. L. J. McBride, J. S. Eadie, J. W. Efcavitch and W. A. Andrus, Proceedings of the Seventh International Round Table Nucleosides Nucleotides and their Biological Applications 1986.
197. T. Tanaka and R. L. Letsinger, *Nucleic Acids Res.*, 1982, **10**, 3249-3260.
198. K. K. Ogilvie, N. Y. Theriault, J. -M. Seifert, R. T. Pon and M. J. Nemer, *Can. J. Chem.*, 1980, **58**, 2686-2693.
199. N. Balgobin and J. Chattopadhyaya, *Acta Chem. Scand., Ser. B*, 1985, **39**, 883-887.
200. C. A. A. Claesen, R. P. A. M. Segers and G. I. Tesser, *Recl. Trav. Chim. Pays-Bas*, 1985, **104**, 119-122.
201. C. A. A. Claesen, R. P. A. M. Segers and G. I. Tesser, *Recl. Trav. Chim. Pays-Bas*, 1985, **104**, 209-214.
202. R. L. Letsinger, E. P. Groody and T. Tanaka, *J. Am. Chem. Soc.*, 1982, **104**, 6805-6806.
203. A. H. Beiter and W. Pfeleiderer, *Tetrahedron Lett.*, 1984, **25**, 1975-1978.
204. N. D. Sinha, J. Biernat and H. Köster, *Tetrahedron Lett.*, 1983, **24**, 5843-5846.
205. N. D. Sinha, J. Biernat, J. McManus and H. Köster, *Nucleic Acids Res.*, 1984, **12**, 4359-4557.
206. G. M. Tener, *J. Am. Chem. Soc.*, 1961, **83**, 159-168.
207. R. L. Letsinger, K. K. Ogilvie and P. S. Miller, *J. Am. Chem. Soc.*, 1969, **91**, 3360-3365.
208. G. Zon, K. A. Gullo, C. J. Samson, K. Shao, M. F. Summers and R. A. Byrd, *Nucleic Acids Res.*, 1985, **13**, 8181-8196.
209. H. Schaller, G. Weimann, B. Lerch and H. G. Khorana, *J. Am. Chem. Soc.*, 1963, **85**, 3821-3827.

210. H. Schaller and H. G. Khorana, *J. Am. Chem. Soc.*, 1963, **85**, 3828-3835.
- 211a. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. USA.*, 1977, **74**, 560-564.
- 211b. A. M. Maxam and W. Gilbert, *Methods in Enzymology*, 1980, **65**, 499-504.
212. J. Stawinski, T. Hozumi, S. A. Narang, C. P. Bahl and R. Wu, *Nucleic Acids Res.*, 1977, **4**, 353-371.
213. K. Miyoshi, T. Huang and K. Itakura, *Nucleic Acids Res.*, 1980, **8**, 5491-5505.
214. M. J. Gait, S. G. Popov, M. Singh and R. C. Timas, *Nucleic Acids Res. Symp. Ser.*, 1980, **7**, 243-257.
215. V. Kohli, H. Blöcker and H. Köster, *Tetrahedron Lett.*, 1980, **21**, 2683-2686.
216. M. D. Matteucci and M. H. Caruthers, *Tetrahedron Lett.*, 1980, **21**, 3243-3246.
217. T. Neilson and E. Werstuik, *J. Am. Chem. Soc.*, 1973, **96**, 2295-2297.
218. J. H. van Boom and P. M. J. Burgers, *Tetrahedron Lett.*, 1977, 4875-4878.
219. J. H. Chattopadhyaya, C. B. Reese and A. H. Todd, *J. Chem. Soc., Chem. Commun.*, 1979, 981-983.
220. E. J. Corey and A. Venkateswarlu, *J. Am. Chem. Soc.*, 1972, **94**, 6190-6191.
221. E. J. Corey, J-L. Gras and P. Ulrich, *Tetrahedron Lett.*, 1976, **11**, 809-812.
222. M. Sekine and T. Hata, *J. Org. Chem.*, 1983, **48**, 3011-3014.
223. S. S. Jones, C. B. Reese and S. Sibanda, *Tetrahedron Lett.*, 1981, **22**, 1933-1936.
224. R. L. Letsinger and J. L. Finnan, *J. Am. Chem. Soc.*, 1975, **97**, 7197-7198.
225. J. B. Chattopadhyaya and C. B. Reese, *J. Chem. Soc., Chem. Commun.*, 1978, 639-640.
226. User Bulletin Issue No 13 Revised, " The Evaluation and Purification of Synthetic Oligonucleotides ", Applied Biosystems, 1987.
227. H. Takaku, K. Marita and T. Sumiuchi, *Chemistry Letters*, 1983, 1661-1664.
228. M. Sekine, N. Masuda and T. Hata, *Tetrahedron*, 1985, **41**, 5445-5453.
- 229a. A. Kume, M. Sekine and T. Hata, *Tetrahedron Lett.*, 1982, **23**, 4365-4368.
- 229b. A. Kume, M. Sekine and T. Hata, *Chemistry Letters*, 1983, 1597-1600.
230. A. Kume, R. Iwase, M. Sekine and T. Hata, *Nucleic Acids Res.*, 1984, **12**, 8525-8538.
231. C. Morin, *Tetrahedron Lett.*, 1983, **24**, 53-56.
232. H. Klenow and S. Frederick, *Biochim. Biophys. Acta*, 1961, **52**, 384-386.

233. K. K. Ogilvie, M. J. Nemer, G. H. Hakiilahi, Z. A. Proba and M. Lucas, *Tetrahedron Lett.*, 1982, **23**, 2615-2618.
234. J. Heikkilä and J. Chattopadhyaya, *Acta Chem. Scand., Ser. B*, 1983, **37**, 263-265.
235. J. Heikkilä, N. Balgobin and J. Chattopadhyaya, *Acta Chem. Scand., Ser. B*, 1983, **37**, 857-864.
236. X.-X. Zhou, I. Ugi and J. Chattopadhyaya, *Acta Chem. Scand., Ser. B*, 1985, **39**, 761-765.
237. B. C. Froehler and M. D. Matteucci, *Nucleic Acids Res.*, 1983, **11**, 8031-8036.
238. J. Zemlička, S. Chladek, A. Holý and J. Smrt, *Collect. Czech. Chem. Commun.*, 1966, **31**, 3198-3211.
239. J. Zemlička and A. Holý, *Collect. Czech. Chem. Commun.*, 1967, **32**, 3159-3168.
240. L. J. McBride and M. H. Caruthers, *Tetrahedron Lett.*, 1983, **24**, 2953-2956.
241. G. S. Ti, B. L. Gaffney and R. A. Jones, *J. Am. Chem. Soc.*, 1982, **104**, 1316-1319.
242. J. A. Zoltewicz, D. F. Clark, T. W. Sharpless and G. Grahe, *J. Am. Chem. Soc.*, 1970, **92**, 1741-1750.
243. E. R. Gamett and P. J. Mehta, *J. Am. Chem. Soc.*, 1972, **94**, 8532-8541
244. J. L. York, *J. Org. Chem.*, 1981, **46**, 2171-2173.
245. M. Oivanen and H. Lönnberg, *Tetrahedron*, 1987, **43**, 1133-1140.
246. C. B. Reese, *Tetrahedron Lett.*, 1978, **34**, 3143-3179.
247. G. Remaud, X-X. Zhou, J. Chattopadhyaya, M. Oivanen and H. Lönnberg, *Tetrahedron*, 1987, **43**, 4453-4461.
248. N. C. Gonella and J. D. Roberts, *J. Am. Chem. Soc.*, 1982, **104**, 3162-3164.
249. A. Sygula and A. Buda, *J. Mol. Struct.*, 1983, **92**, 267-277.
250. Y. Maki, M. Suzuki, K. Kameyama and M. Sako, *J. Chem. Soc., Chem. Commun.*, 1981, 658-659.
251. T. Ueda, K. Miura and T. Kasai, *Chem. Pharm. Bull.*, 1978, **26**, 2122-2127.
252. J. Gralla and D. M. Crothers, *J. Mol. Biol.*, 1973, **78**, 301-319.
253. N. Tibanyenda, S. H. De Bruin, C. A. G. Haasnoot, G. A. van der Marel, J. H. van Boom and C. W. Hilbers, *Eur. J. Biochem.*, 1984, **139**, 19-27.
- 254a. W. L. Sung, *J. Chem. Soc., Chem. Commun.*, 1981, 1089.
- 254b. W. L. Sung, *Nucleic Acids Res.*, 1981, **9**, 6139-6151.

255. T. Huynh-Dinh, B. Langois d'Estaintot, P. Allard and J. Igolen, *Tetrahedron Lett.*, 1985, **26**, 431-434.
256. J. S. Eadie and D. S. Davidson, *Nucleic Acids Res.*, 1987, **15**, 8333-8349.
257. B. L. Gaffney and R. A. Jones, *Tetrahedron Lett.*, 1982, **23**, 2253-2256.
258. B. L. Gaffney and R. A. Jones, *Tetrahedron Lett.*, 1982, **23**, 2257-2260.
- 259a. P. K. Bridson, W. Markiewicz and C. B. Reese, *J. Chem. Soc., Chem. Commun.*, 1977, 447-448.
- 259b. P. K. Bridson, W. T. Markiewicz and C. B. Reese, *J. Chem. Soc., Chem. Commun.*, 1977, 791-792.
260. H. P. Daskalov, M. Sekine and T. Hata, *Tetrahedron Lett.*, 1980, **21**, 3899-3902.
261. C. B. Reese and A. Ubasawa, *Tetrahedron Lett.*, 1980, **21**, 2265-2268.
262. J. Kiburis and J. H. Lister, *J. Chem. Soc. ( C )*, 1971 3942-3947.
263. M. J. Robbins and G. L. Basom, *Can. J. Chem.*, 1973, **51**, 3161-3169.
- 264a. J. Nielsen, J. E. Marugg, J. H. van Boom, J. Honnens, M. Taagaard and O. Dahl, *J. Chem. Res. ( S )*, 1986, 26-27.
- 264b. J. Nielsen and O. Dahl, *Nucleic Acids Res.*, 1987, **15**, 3626.
265. J. Nielsen, M. Taagaard J. E. Marugg, J. H. van Boom and O. Dahl, *Nucleic Acids Res.*, 1986, **14**, 7391-7403.
266. T-D. Son, W. Guschlbauer and M. Guéron, *J. Am. Chem. Soc.*, 1972, **94**, 7903-7911.
267. A. Poustka, T. Pohl, D. P. Barlow, G. Zehetner, A. Craig, F. Michiels, E. Erich, A. M. Frischauf and H. Lehrach, *Cold Spring Harbor Symp. Quant. Biol.*, 1986, **51**, 131-139.
268. F. Michiels, A. G. Craig, G. Zehetner, G. P. Smith and H. Lehrach, *CABIOS*, 1987, **3**, 203-210.
269. C. R. Cantor and P. R. Schimmel, in "Biophysical Chemistry, The Behaviour of Macromolecules" W. H. Freeman, San Francisco 1980.
270. J. C. Schulhof, D. Molko and R. Teoule, *Nucleic Acids Res.*, 1987, **15**, 397-415.
271. H. Vu, C. McCollum, K. Jacobson, P. Theisen, R. Vinayak, E. Spiess and A. Andrus, *Tetrahedron Lett.*, 1990, **31**, 7269-7272.
272. Y. Okada, E. Terzaghi, G. Streisinger, J. Emrich, M. Inouye and A. Tsugita, *Proc. Natl. Acad. Sci. USA.*, 1966, **56**, 1692-1698.
273. L. S. Ripley, *Proc. Natl. Acad. Sci. USA.*, 1982, **79**, 4128-4132.
274. J. G. de Boer and L. S. Ripley, *Proc. Natl. Acad. Sci. USA.*, 1984, **81**, 5528-5531.

275. D. J. Patel, S. A. Kozlowski, L. A. Marky, J. A. Rice, C. Broka, K. Itakura and K. J. Breslauer, *Biochemistry*, 1982, **21**, 445-451.
276. D. Hare, L. Shapiro and D. J. Patel, *Biochemistry*, 1986, **25**, 7456-7464.
277. M. W. Kalnik, D. G. Norman, P. F. Swann and D. J. Patel, *J. Biol. Chem.*, 1989, **264**, 3702-3712.
278. D. H. Evans and A. R. Morgan, *J. Mol. Biol.*, 1982, **160**, 117-122.
279. D. H. Evans and A. R. Morgan, *Nucleic Acids Res.*, 1986, **14**, 4267-4280.
280. K. M. Morden, Y. G. Chu, F. H. Martin and I. Tinoco Jr., *Biochemistry*, 1983, **22**, 5557-5563.
281. M. W. Kalnik, D. G. Norman, M. G. Zagarski, P. F. Swann and D. J. Patel, *Biochemistry*, 1989, **28**, 294-303.
282. S. A. Woodson and D. M. Crothers, *Biochemistry*, 1987, **26**, 904-912.
283. Y. T. van den Hoogen, A. A. van Beuzekom, H. van den Elst, G. A. van der Marel, J. H. van Boom and C. Altona, *Nucleic Acids Res.*, 1988, **16**, 2971-2986.
284. Y. T. van den Hoogen, A. A. van Beuzekom, E. de Vroom, G. A. van der Marel, J. H. van Boom and C. Altona, *Nucleic Acids Res.*, 1988, **16**, 5013-5030.
285. M. W. Kalnik, D. G. Norman, B. F. li, P. F. Swann and D. J. Patel, *J. Biol. Chem.*, 1990, **265**, 636-647.
286. M. A. Saper, H. Eldar, K. Mizuuchi, J. Nickol, A. Appella and J. L. Sussman, *J. Mol. Biol.*, 1986, **188**, 111-113.
287. B. S. Sproat and M. J. Gait in, "Oligonucleotide Synthesis a Practical Approach," ed. M. J. Gait, IRL Press, Oxford, 1984, p91.
288. J. Igolen and C. Morin, *J. Org. Chem.*, 1980, **45**, 4802-4804.
289. B. S. Sproat and M. J. Gait in, "Oligonucleotide Synthesis a Practical Approach," ed. M. J. Gait, IRL Press, Oxford, 1984, p109.

## **Appendix: Publications**

**Crystal Structure and Stability of a DNA Duplex  
Containing A(*anti*) · G(*syn*) Base-pairs**

**Tom Brown, Gordon A. Leonard  
Ewan D. Booth and Jenny Chambers**

## Crystal Structure and Stability of a DNA Duplex Containing A(*anti*)·G(*syn*) Base-pairs

The synthetic dodecanucleotide d(CGCAAATTGGCG) has been analysed by single-crystal X-ray diffraction techniques and the structure refined to  $R = 0.16$  and  $2.25 \text{ \AA}$  resolution, with the location of 94 solvent molecules. The sequence crystallizes as a full turn of a B-DNA helix with ten Watson-Crick base-pairs and two adenine·guanine mismatches. The analysis clearly shows that the mismatches are of the form A(*anti*)·G(*syn*). Thermal denaturation studies indicate that the stability of the duplex is strongly pH dependent, with a maximum at pH 5.0, suggesting that the base-pair is stabilized by protonation. Three different arrangements have been observed for base-pairs between guanine and adenine and it is likely that A·G mismatch conformation is strongly influenced by dipole-dipole interactions with adjacent base-pairs.

Base-pair mismatches are occasionally formed in DNA during replication or genetic recombination. Proofreading and mismatch repair enzymes usually excise the incorrect base but, if the error is not detected, substitution mutations arise (Friedberg, 1985). Enzymic recognition of mismatches may depend on structural or thermodynamic factors, or a combination, or both (Steitz *et al.*, 1987) and in order to understand the mechanisms by which enzymes recognize mispaired bases it is important to study the properties of DNA duplexes containing mismatched base-pairs. There has recently been a great deal of progress in this field and X-ray crystallography (Kennard, 1987) and nuclear magnetic resonance spectroscopy (Patel *et al.*, 1987) have been used to elucidate the structure of several base-pair mismatches in a variety of base-stacking environments. The G·A base-pair is of particular interest, as experiments with DNA polymerase III have shown that about 10% of these mismatches avoid detection by proofreading, whereas only

0.5% of G·T mismatches escape (Fersht *et al.*, 1982). The G·A mismatch has been shown to be conformationally variable and both G(*anti*)·A(*syn*) and G(*anti*)·A(*anti*) base-pairs have been identified in different DNA sequences at neutral pH (Kan *et al.*, 1983; Patel *et al.*, 1984a,b; Hunter *et al.*, 1986; Prive *et al.*, 1987). Theoretical calculations suggest that both have similar stability when incorporated in a DNA duplex although the G(*anti*)·A(*anti*) base-pair causes greater distortion of the sugar-phosphate backbone (Chuprina & Polter, 1983; Keepers *et al.*, 1984). It has recently been shown by nuclear magnetic resonance that the G·A mismatches in the sequence d(CGGAATTACG) display conformational flexibility as a function of pH (Gao & Patel, 1988). The G(*anti*)·A(*anti*) base-pair is observed at neutral pH, whereas the G(*syn*)·A(*anti*) base-pair predominates in the pH range 4.0 to 5.5. In the latter case the adenine base is protonated.

We report the crystal structure of the synthetic

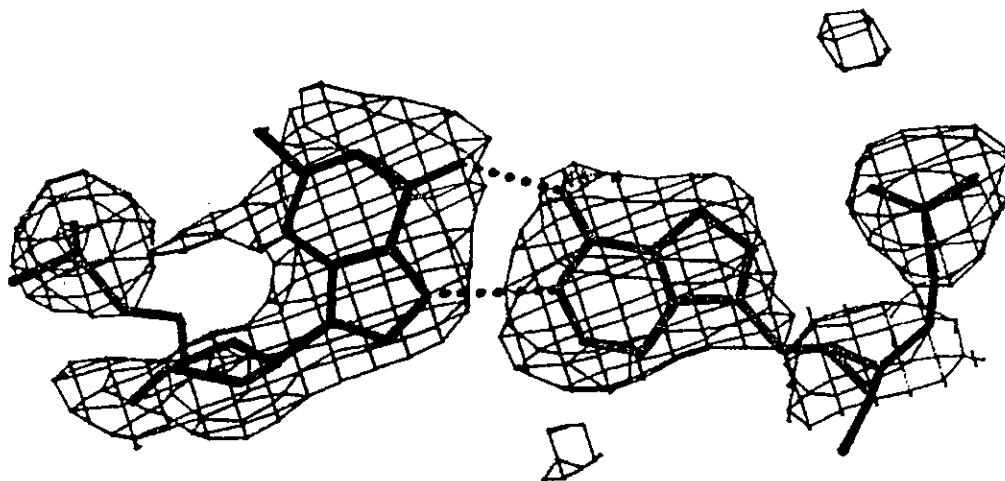


Figure 1. Fragment  $F_o - F_c$  electron density map of the G(*syn*)·A(*anti*) base-pair in d(CGCAAATTGGCG), calculated with all mismatch bases removed. Contour level  $0.35 \text{ e/\AA}^3$ ,  $1.75 \times \text{root-mean-square value}$ .

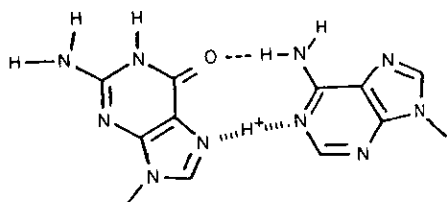


Figure 2. The  $G(\text{syn}) \cdot A(\text{anti})$  base-pair.

dodecanucleotide d(CGCAAATTGGCG) to a resolution of 2.25 Å (1 Å=0.1 nm) and show that in this sequence  $G(\text{syn}) \cdot A(\text{anti})$  base-pairs are present at near-neutral pH.

The self-complementary dodecamer was synthesized by the phosphoramidite method and purified by ion-exchange and reverse-phase high-pressure liquid chromatography. Crystals were grown at pH 6.6 and 4°C from a solution of DNA (0.5 mM) in sodium cacodylate buffer (50 mM), magnesium chloride (20 mM), spermine (1.0 mM) and 2-methyl-2,4-pentanediol (18%, v/v). A single crystal measuring 2.0 mm × 0.4 mm × 0.3 mm, mounted in a sealed capillary was used to collect the diffraction data. A unit cell  $a = 25.23$  Å,  $b = 41.16$  Å and  $c = 65.01$  Å, space group  $P2_12_12_1$  indicated quasi-isomorphism with the native dodecamer d(CGCGAATTCGCG) (Wing *et al.*, 1980). X-ray data to 2.25 Å were measured at 4°C on a Stoe-Siemens AED2 four-circle diffractometer equipped

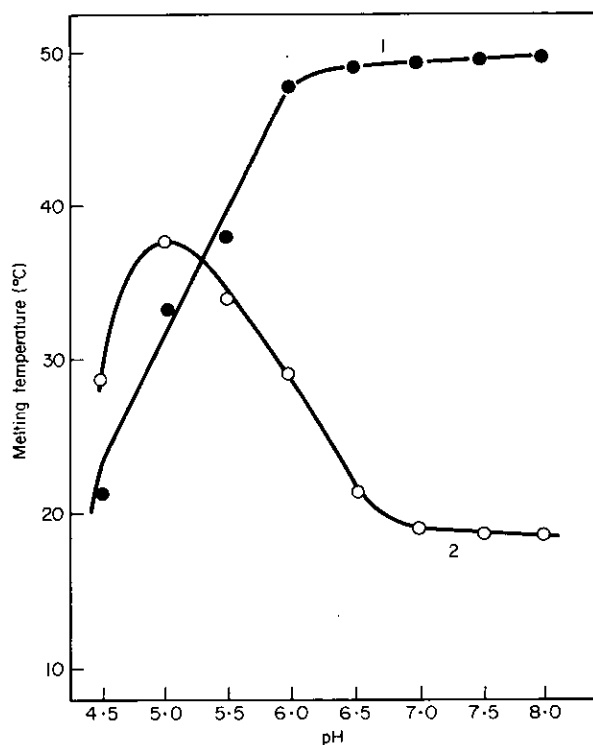


Figure 3. Ultraviolet melting-curves in 0.1 M-phosphate buffer. Curve 1: d(CGCGAATTCGCG), native dodecamer. Curve 2: d(CGCAAATTGGCG), A·G mismatch dodecamer.

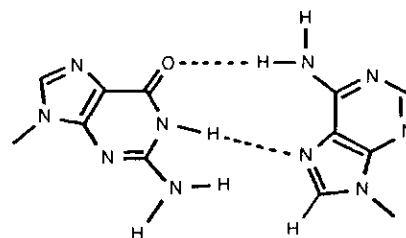


Figure 4. The  $G(\text{anti}) \cdot A(\text{syn})$  base-pair.

with a long arm and helium path and intensities were corrected for Lorentz and polarization factors, absorption and time-dependent decay. The data collection yielded a total of 2840 unique reflections with  $F \geq 1 \sigma(F)$  and 2262 of these in the range 8 Å to 2.25 Å, with  $F \geq 2 \sigma(F)$  were used in the refinement. Starting co-ordinates were those of the native dodecamer and the model was refined using restrained least-squares methods (Hendrickson & Konnert, 1981; Westhof *et al.*, 1985), with periodic examination of  $F_o - F_c$  maps on an Evans and Sutherland PS300 graphics system. During these calculations the mismatch bases were allowed to move freely relative to each other. It was clear from "fragment"  $F_o - F_c$  electron density maps calculated with the mismatched base-pairs omitted that  $G(\text{syn}) \cdot A(\text{anti})$  gave a significantly better electron density fit than either  $G(\text{anti}) \cdot A(\text{syn})$  or  $G(\text{anti}) \cdot A(\text{anti})$  (Fig. 1). The refinement converged at  $R = 0.16$ , with the location of 94 solvent molecules. Figure 2 shows the A·G base-pair schematically. The N-6 atom of adenine and the O-6 atom of guanine are 2.6 Å apart and the N-1 atom of adenine and the N-7 atom of guanine are separated by 2.8 Å. These distances are indicative of strong hydrogen bonds, but with both bases in their major tautomeric forms a proton is required to enable the  $N_{(1)}$  adenine- $N_{(7)}$  guanine hydrogen bond to form. There is good evidence from thermal denaturation studies that the mismatch base-pair is indeed protonated. Ultraviolet melting-studies in the pH range 8.0 to 4.5 on the A·G dodecamer d(CGCAAATTGGCG) and the native sequence d(CGCGAATTCGCG) show that the behaviour of the two duplexes is strikingly different (Fig. 3). The melting temperature ( $t_m$ ) of the native sequence decreases sharply below pH 6.5 whereas the  $t_m$  of the A·G dodecamer reaches a maximum at pH 5.0 before falling away at lower pH. These results suggest that the  $G(\text{syn}) \cdot A(\text{anti})$  base-pair is present

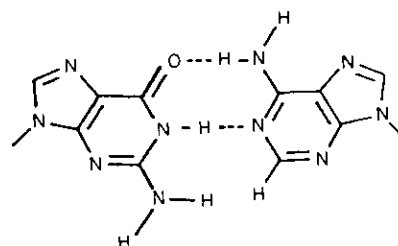


Figure 5. The  $G(\text{anti}) \cdot A(\text{anti})$  base-pair.

in solution over a wide pH range. In the G(*syn*)·A(*anti*) base-pair, the 2-amino group of guanine lies in the major groove, where it is free to hydrogen bond to two water molecules. This is not the case for G(*anti*)·A(*syn*) (Fig. 4) and G(*anti*)·A(*anti*) (Fig. 5) base-pairs in which the 2-amino group of guanine is sterically crowded in the minor groove and can only interact with a single water molecule. This will lead to a loss of hydrogen bonds when the fully hydrated single strands come together to form a duplex. The disposition of major and minor groove hetero-atoms in the A(*anti*)·G(*syn*) mismatch is completely different from that of other forms of the G·A mismatch, although in general shape it resembles the A(*syn*)·G(*anti*) mismatch. This may be important when considering recognition of mismatches by proofreading and repair enzymes. In the current study and that of Gao & Patel (1988), the guanine base of the G(*syn*)·A(*anti*) base-pair is flanked on the 5'-side by the guanine of a G·C base-pair. Guanine has a large dipole moment (7.5 D) (Pullman & Pullman, 1971) and a GpG step in B-DNA involving G(*anti*) bases may have an unfavourable electrostatic component, as the dipole moments of the two bases would be unfavourably aligned. However, with the guanine base of the G·A mismatch in the *syn*-conformation, the dipole-dipole interactions are likely to be much more favourable. In the sequences d(CGAGAATTCGCG) and d(CGCGAATTAGCG), in which the guanine bases of the G·A mismatches are *anti* (G(*anti*)·A(*anti*) and G(*anti*)·A(*syn*), respectively), the guanines are involved in CpG stacking interactions. The dipole moment of cytosine is almost the same as that of guanine (7.6 D) and in its normal orientation in B-DNA it has almost exactly the opposite direction. Thus, the CpG(*anti*) base-stacking step will produce favourable electrostatic interactions. The dipole moment of adenine is small (2.9 D), so its influence on G·A mismatch conformation will be less than that of guanine. Dipole moments may be particularly important as the two single strands come together and intra-strand stacking interactions can be optimized, locking the G·A mismatch into a specific conformation as the duplex is formed.

An interesting feature of the d(CGCAAA-TTGGCG) duplex is the network of bifurcated hydrogen bonds in the major groove of the AAATTG core, similar to that observed in the A<sub>n</sub>·T<sub>n</sub> region of other sequences (Nelson *et al.*, 1987; Coll *et al.*, 1987). In the present structure, mispaired bases are involved and it is likely that the additional interstrand hydrogen bonds stabilize the region of the duplex containing the mismatches.

We are grateful to Olga Kennard, Bill Hunter and Geoff Kneale for helpful discussions. This work was

funded by an SERC Molecular Recognition Grant and a Research Corporation Trust Award to T.B.

**Tom Brown**  
**Gordon A. Leonard**  
**Ewan D. Booth**  
**Jenny Chambers**

Department of Chemistry  
 University of Edinburgh  
 West Mains Road  
 Edinburgh EH9 3JJ, U.K.

Received 14 December 1988

### References

- Chuprina, V. P. & Poltev, V. I. (1983). *Nucl. Acids Res.* **11**, 5205–5222.
- Coll, M., Frederick, C. A., Wang, A. H.-J. & Rich, A. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8385–8389.
- Fersht, A. R., Knill-Jones, J. W. & Tsui, W. C. (1982). *J. Mol. Biol.* **156**, 37–51.
- Friedberg, E. C. (1985). *DNA Repair*. W. H. Freeman & Co., New York.
- Gao, X. & Patel, D. J. (1988). *J. Amer. Chem. Soc.* **110**, 5178–5182.
- Hendrickson, W. A. & Konnert, J. H. (1981). In *Biomolecular Structure, Conformation, Function and Evolution* (Srivanan, R., ed.), pp. 43–57, Plenum Press, Oxford.
- Hunter, W. N., Brown, T. & Kennard, O. (1986). *J. Biomol. Struct. Dynam.* **4**, 173–191.
- Kan, L.-S., Chandrasegaran, S., Pulford, S. M. & Miller, P. S. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 4263–4265.
- Keepers, J. W., Schmidt, P., James, T. L. & Kolman, P. A. (1984). *Biopolymers*, **23**, 2901–2929.
- Kennard, O. (1987). In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D. M., eds), vol. 1, pp. 25–52, Springer-Verlag, Berlin, Heidelberg, New York.
- Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). *Nature (London)*, **330**, 221–226.
- Patel, D. J., Koslowski, S. A., Ikuta, S. & Itakura, K. (1984a). *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* **43**, 2663–2670.
- Patel, D. J., Koslowski, S. A., Ikuta, S. & Itakura, K. (1984b). *Biochemistry*, **23**, 3207–3217.
- Patel, D. J., Shapiro, L. & Hare, D. (1987). In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D., eds), vol. 1, pp. 70–84, Springer-Verlag, Berlin, Heidelberg, New York.
- Prive, G. G., Heinemann, U., Kan, L.-S., Chandrasegaran, S. & Dickerson, R. E. (1987). *Science*, **238**, 498–504.
- Pullman, B. & Pullman, A. (1971). *Advan. Het. Chem.* **13**, 77–159.
- Steitz, T. A., Beese, L., Freemont, P. S., Friedman, J. M. & Sanderson, M. R. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 465–471.
- Westhof, E., Dumas, P. & Moras, D. (1985). *J. Mol. Biol.* **184**, 119–145.
- Wing, R., Drew, H. R., Takano, T., Broka, L., Takana, S., Itakura, K. & Dickerson, R. E. (1980). *Nature (London)*, **287**, 755–758.

## **Influence of pH on the Conformation and Stability of Mismatch Base-pairs in DNA**

**Tom Brown, Gordon A. Leonard, Ewan D. Booth  
and Geoff Kneale**

## Influence of pH on the Conformation and Stability of Mismatch Base-pairs in DNA

Tom Brown, Gordon A. Leonard, Ewan D. Booth

Department of Chemistry, University of Edinburgh  
West Mains Road, Edinburgh EH9 3JJ, U.K.

and Geoff Kneale

Biophysics Laboratories, School of Biological Sciences  
Portsmouth Polytechnic, White Swan Road, Portsmouth PO1 2DT, U.K.

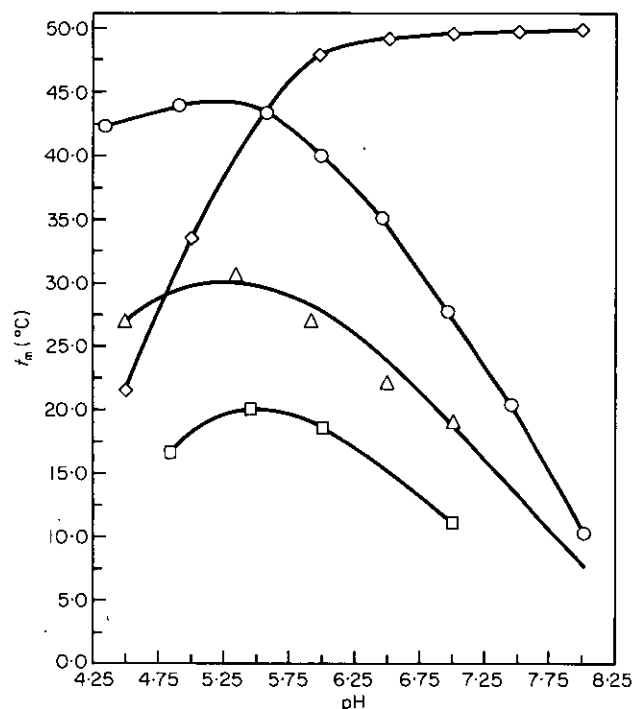
(Received 10 November 1989; accepted 11 December 1989)

A series of self-complementary dodecanucleotide duplexes containing two symmetrically disposed mismatches have been studied by pH-dependent, ultraviolet light melting techniques. The results indicate that A·C, and C·C mismatches are strongly stabilized by protonation and that the degree of stabilization of the A·C mismatch depends greatly on the flanking bases. In one case, a duplex containing two A·C mismatches is more stable than the native sequence below pH 5.5. The G·A mismatch displays conformational flexibility, with a protonated G(*syn*)·A(*anti*) base-pair occurring in certain base stacking environments but not in others. The A·A and T·C mismatches are not stabilized at low pH. These solution studies correlate well with predictions based on X-ray crystallographic data.

In recent years, physical methods such as nuclear magnetic resonance spectroscopy (Patel *et al.*, 1984*a,b*, 1987), X-ray crystallography (Kennard, 1987), and ultraviolet light melting studies (Martin *et al.*, 1985) have been employed to investigate the conformation and stability of base-pair mismatches in DNA duplexes. Very recently, nuclear magnetic resonance spectroscopy (Gao & Patel, 1988) and ultraviolet light melting studies (Brown *et al.*, 1989) have been used to probe the variability in the conformation and stability of the G·A mismatch as a function of pH. The N<sub>(1)</sub> atom of deoxyadenosine and the N<sub>(3)</sub> atom of deoxycytidine both have pK<sub>a</sub> values around 4 (Saenger, 1984); so, in principle, any mismatch containing an adenine or cytosine base can be stabilized by protonation at or near physiological pH. Moreover, the true pK<sub>a</sub> values for these atoms in base-pairs in DNA duplexes may be significantly higher. Accordingly, we undertook to study several mismatches of this type in sequences based on the Dickerson dodecamer, d(CGCGAATTCGCG), a self-complementary deoxyoligonucleotide that forms a B-DNA duplex (Wing *et al.*, 1980). These sequences were chosen because the structures of a number of them have been studied in detail by X-ray crystallography. We now report the results of detailed pH-dependent melting studies and interpret these in terms of possible conformations of specific base-pair mismatches.

An X-ray crystallographic analysis of the A·C

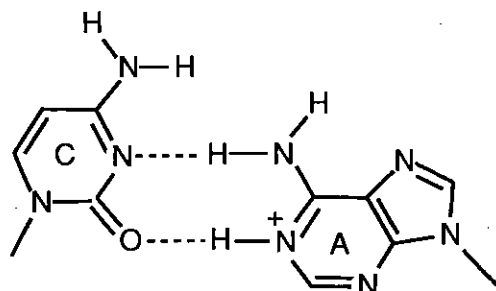
mismatch in the self-complementary B-DNA dodecamer d(CGCGAATTCGCG) at 2.5 Å resolution (1 Å = 0.1 nm) had shown the presence of a wobble base-pair (Hunter *et al.*, 1986*a*). At this resolution, it is not possible to locate hydrogen atoms, so the crystal structure is consistent with both minor tautomer and protonated major tautomer base-pairs. In order to resolve the situation, we measured the melting temperature of the A·C dodecamer as a function of pH (circles in Fig. 1). The result clearly shows a striking correlation between duplex stability and proton concentration, strongly supporting a protonated base-pair (Fig. 2). At pH 8.0, the duplex is very unstable, with a *t<sub>m</sub>* of 11°C and, in the absence of a high concentration of protons, it is possible that A·C base-pairs do not form. Interestingly, the presence of a protonated base does not seem to destabilize the duplex *per se* and at the highest point on the curve (pH 5.3), the A·C dodecamer, with a *t<sub>m</sub>* of 44°C, is more stable than the native Dickerson dodecamer (*t<sub>m</sub>* 40°C, diamonds in Fig. 1(b)). The correlation between pH and duplex stability of the sequence d(CGCCAATTAGCG), the C·A dodecamer was examined. This sequence was derived from the A·C mismatch by inverting the base-pairs at positions 4 and 9, thus changing the base-stacking environment. Again, the duplex becomes more stable as the pH is lowered until maximum stability is reached at pH 5.5 (squares in Fig. 1). Beyond this point, protonation of cytosine



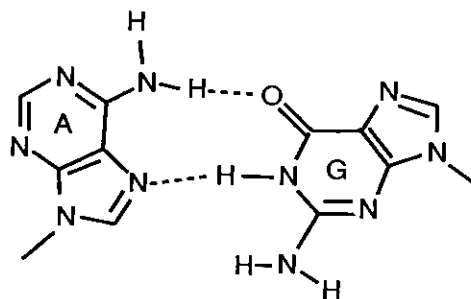
**Figure 1.** Plot of  $t_m$  versus pH for duplexes of the form d(CGCAATTGCG). Points are accurate to within 0.5 deg. C. All melting curves were measured at least 3 times in 0.1 M-sodium phosphate buffer at the appropriate pH. ( $\diamond$ ) X=G, Y=C; ( $\circ$ ) X=A, Y=C; ( $\triangle$ ) X=Y=C; ( $\square$ ) X=C, Y=A.

$N_{(3)}$ , and adenine  $N_{(1)}$  atoms destabilizes all Watson-Crick base-pairs in the usual way. The much lower melting temperature of the C·A dodecamer relative to the A·C dodecamer (a difference in  $t_m$  of 23 deg. C at pH 5.5) may be due to an unfavourable dipole moment component of the CpC mismatch base-stacking step in the former compared to a favourable CpG step in the latter. No X-ray crystallographic information is available on the C·A mismatch duplex.

The G·A mismatch has been shown to be conformationally variable (Kan *et al.*, 1983; Patel *et al.*, 1984a,b; Hunter *et al.*, 1986b; Prive *et al.*, 1987) and both G(*anti*)·A(*syn*) and G(*anti*)·A(*anti*) base-pairs have been identified in different DNA sequences at neutral pH (Figs 3 and 4). It has been shown by nuclear magnetic resonance spectroscopy

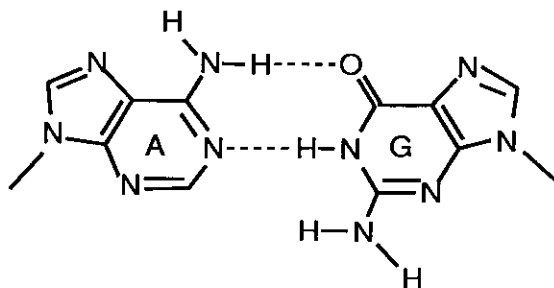


**Figure 2.** The C·A<sup>+</sup> mismatch.



**Figure 3.** The A(*syn*)·G(*anti*) mismatch.

that the G·A mismatch in the sequence d(CGG-GAATTCAG) displays conformational flexibility as a function of pH (Gao & Patel, 1988). The G(*anti*)·A(*anti*) base-pair is observed at neutral pH, whereas the G(*syn*)·A(*anti*) base-pair (Fig. 5) predominates below pH 5.5. In the latter case, there is evidence that the adenine base is protonated. Our earlier crystallographic analysis (Brown *et al.*, 1986) of a G·A mismatch in the sequence d(CGG-GAATTAGCG) showed the presence of a G(*anti*)·A(*syn*) pairing. We have solved the structure of the dodecamer d(CGCAAATTGGCG), which is related to the previous sequence by interchanging the adenine and guanine bases of the G·A mismatches (Brown *et al.*, 1989), thus placing the mismatches in a different base stacking environment. In this case, a G(*syn*)·A(*anti*) base-pair was identified, so it is clear that base-stacking interactions are influential in determining the form of conformationally flexible mismatches. In the G(*syn*)·A(*anti*) mismatch, with both bases in their major tautomeric forms, protonation is necessary to stabilize the adenine  $N_{(1)}$  to guanine  $N_{(7)}$  hydrogen bond. Ultraviolet light melting studies in the pH range 8.0 to 4.5 (squares in Fig. 6) indicate that the G(*syn*)·A(*anti*) base-pair is present in solution, as the stability of the A·G dodecamer reaches a maximum at pH 5.2 ( $t_m$  38.5°C). It can, however, be concluded that between pH 6.5 and pH 8.0, non pH-dependent forms of the G·A mismatch are present, presumably involving a G(*anti*) base, as in this range the melting temperature is constant at 19°C. In both dodecamers that have been shown to accommodate G(*syn*)·A(*anti*) base-pairs, d(CGCAAATTGGCG) and d(CGG-GAATTCAG), the guanine base of the mismatch is flanked by the guanine(s) of a G·C base-pair(s).



**Figure 4.** The A(*anti*)·G(*anti*) mismatch.

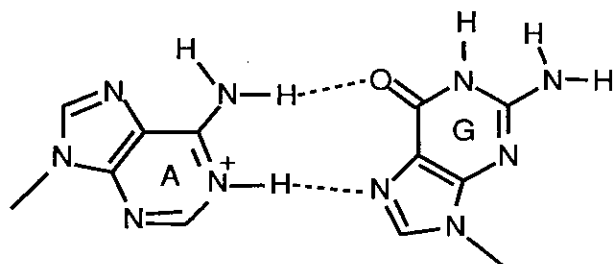


Figure 5. The A<sup>+</sup>(*anti*)·G(*syn*) mismatch.

Guanine has a large dipole moment (7.5 Debye units; Pullman & Pullman, 1971), and a GpG step in *B*-DNA involving G(*anti*) bases would be expected to have an unfavourable base-stacking interaction as the dipole moments of the two bases would be closely aligned, with a relative rotation of only about 33°. However, with the guanine base of the G·A mismatch in the *syn* conformation, the static dipole-dipole interactions are likely to be less unfavourable.

In the sequences d(CGAGAATTCGCG) (Patel *et al.*, 1984b) and d(CGCGAATTAGCG) (Brown *et al.*, 1986) in which the guanine bases of the G·A mismatches have been shown to be *anti* (Figs 4 and 3, respectively), the guanine bases are involved in CpG stacking interactions. The dipole moment of cytosine is almost identical to that of guanine (7.6 Debye units) and when on the 5' side of guanine in *B*-DNA it lies in almost exactly the opposite direction. This should result in a favourable CpG(*anti*) base-stacking interaction and, in order to further investigate this, we carried out thermal denaturation studies on the dodecamer d(CGAGAATTCGCG), in which the mismatched guanine is flanked by two cytosine bases (triangles in Fig. 6). The much higher  $t_m$  of this sequence at pH 7.0 (35°C) compared with 19°C for d(CGCAATTCGCG), demonstrates the greater stability of CpG and GpC steps relative to the GpG stacking interaction for the G·A mismatch. In this case, there is no indication of increased duplex stability with decreasing pH, indicating that the G(*syn*)·A(*anti*) mispair, which requires protonation, probably does not occur. Hence, it seems that the *anti* conformation is preferred for a mismatched guanine base stacked on a cytosine base. Dipole moments may be particularly important, as the two single strands interact to form a duplex and stacking forces can be optimized. Once the duplex is formed, any change in the conformation of the G·A mismatch will necessitate rotation of a purine base around its glycosidic bond, an operation that would require considerable disruption of the surrounding Watson-Crick base-pairs. It is known that poly[d(G·C)] forms a more stable duplex than poly(dG)·poly(dC) (Gotoh & Takashira, 1981), and that the CpG base-stacking step is the most stable in *B*-DNA (Ornstein *et al.*, 1978), so our results with mismatches are consistent with these observations. As all the DNA sequences

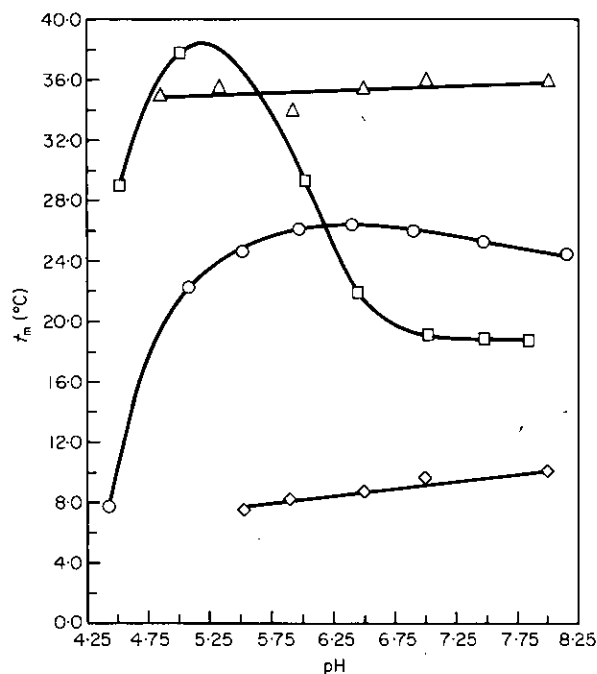


Figure 6. Plot of  $t_m$  versus pH for duplexes of the form d(CGCAATTCGCG), or (b) d(CGAGAATTCGCG). Points are accurate to within 0.5 deg. C. All melting curves were measured 3 times in 0.1 M-sodium phosphate buffer at the appropriate pH. ( $\Delta$ ) X = A, Y = C; ( $\circ$ ) X = T, Y = C; ( $\square$ ) X = A, Y = G; ( $\diamond$ ) X = Y = A.

discussed here are self-complementary, we determined their melting temperatures over a 20-fold range of DNA concentration, both at pH 5.5 and pH 7.0, to ascertain that duplexes, not single-stranded structures are formed. In only one of the cases studied, that of the sequence d(CGCGAATTAGCG), was  $t_m$  independent of DNA concentration, suggesting a single-stranded hairpin loop. Consequently, it was not possible to obtain duplex melting data on this particular sequence, even though it has been crystallized as *B*-DNA.

Thermal denaturation studies on the dodecamer d(CGCAATTCGCG) suggest that in this sequence the A·A mismatch is not stabilized by protonation (diamonds in Fig. 6). The duplex is very unstable and we are unable to propose a structure for the A·A base-pair. We are investigating the properties of the A·A mismatch in a variety of base-stacking environments.

So far, we have been unable to grow suitable crystals of a DNA duplex containing pyrimidine·pyrimidine base-pairs and it is likely that presence of such a mismatch in a DNA duplex leads to some distortion of the sugar-phosphate backbone. The stability of the C·C mismatch in the sequence d(CGCAATTCGCG) shows a striking pH dependence (triangles in Fig. 1), the  $t_m$  increasing from 19°C at pH 7.0 to 30°C at pH 5.3, suggesting the protonated base-pair shown in Figure 7. Other forms of the C<sup>+</sup>·C base-pair have been proposed for

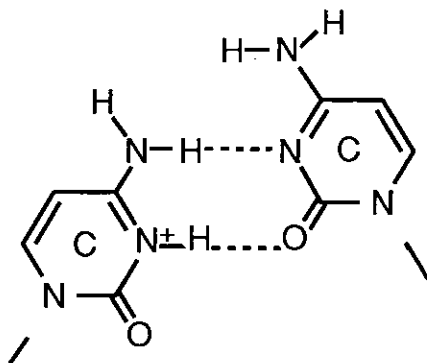


Figure 7. The C<sup>+</sup>·C base-pair.

homopolymers (Borah & Wood, 1976) but it is not clear if they could be accommodated in a normal antiparallel duplex. The  $t_m$  of the T·C mismatch (Fig. 6(d)) in the dodecamer d(CGCTAATTCGCG) shows very little variation between pH 5.3 and 8.25, and this is consistent with the base-pair stabilized by a metal ion, which was suggested by Patel *et al.* (1984a). However, the  $t_m$  is at a maximum around pH 6.3, so we cannot discount the participation of a protonated wobble base-pair.

Our results indicate that the formation of some mismatches involves protonation of an adenine or cytosine base and it is notable that in all such cases both hydrogen bond donors are situated on the same base. Throughout this communication, we have represented the bases in their most common tautomeric forms, but we cannot exclude the possibility that after mismatch formation a rearrangement occurs, resulting in minor tautomer base-pairs with one hydrogen bond donor on each base.

We are grateful to Professor Olga Kennard and Professor Albert Eschenmoser for helpful discussions. This work was funded by SERC Molecular Recognition grants, a Research Corporation Trust Award and a grant from the Nuffield Foundation to T.B.

## References

- Borah, B. & Wood, J. L. (1976). *J. Mol. Struct.* **30**, 13–30.
- Brown, T., Hunter, W. N., Kneale, G. G. & Kennard, O. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 2401–2406.
- Brown, T., Leonard, G. A., Booth, E. D. & Chambers, J. (1989). *J. Mol. Biol.* **207**, 455–457.
- Gao, X. & Patel, D. J. (1988). *J. Amer. Chem. Soc.* **110**, 5178–5182.
- Gotoh, O. & Takashira, Y. (1981). *Biopolymers*, **20**, 1033–1042.
- Hunter, W. N., Brown, T., Anand, N. N. & Kennard, O. (1986a). *Nature (London)*, **320**, 552–555.
- Hunter, W. N., Brown, T. & Kennard, O. (1986b). *Biomol. Struct. Dynam.* **4**, 173–191.
- Kan, L.-S., Chandrasegaran, S., Pulford, S. M. & Miller, P. S. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 4263–4265.
- Kennard, O. (1987). In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D. M., eds), vol. 1, pp. 25–52, Springer-Verlag, Berlin, Heidelberg, New York.
- Martin, F. H., Castro, M. M., Aboul-Ela, F. & Timoko, L., Jr (1985). *Nucl. Acids Res.* **13**, 8927–8938.
- Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. (1978). *Biopolymers*, **17**, 2341–2360.
- Patel, D. J., Koslowski, S. A., Ikuta, S. & Itakura, K. (1984a). *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* **43**, 2663–2670.
- Patel, D. J., Koslowski, S. A., Ikuta, S. & Itakura, K. (1984b). *Biochemistry*, **23**, 3207–3217.
- Patel, D. J., Shapiro, L. & Hare, D. (1987). In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D. M., eds), vol. 1, pp. 70–84, Springer-Verlag, Berlin, Heidelberg, New York.
- Prive, G. G., Heinemann, U., Kan, L.-S., Chandrasegaran, S. & Dickerson, R. E. (1987). *Science*, **238**, 498–504.
- Pullman, B. & Pullman, A. (1971). *Advan. Het. Chem.* **13**, 77–159.
- Saenger, W. (1984). In *Principles of Nucleic Acid Structure*, chapt. 5, Springer-Verlag, New York.
- Wing, R., Drew, H. R., Takano, T., Broka, L., Takana, S., Itakura, K. & Dickerson, R. E. (1980). *Nature (London)*, **287**, 755–758.

Edited by A. R. Fersht

Reproduced by permission from "Molecular Mechanisms in  
Bioorganic Processes", eds Christine Bleasdale and  
Bernard T. Golding, Royal Society of Chemistry, Cambridge, 1990.

Permission of co-authors obtained.

# THE INFLUENCE OF BASE-PAIR MISMATCHES AND MODIFIED BASES ON THE STRUCTURE AND STABILITY OF THE DNA DUPLEX

Tom Brown, Ewan D Booth, and Gordon A Leonard

*Department of Chemistry, University of Edinburgh,  
Edinburgh EH9 3JJ, UK*

## 1 INTRODUCTION

The accurate transmission of genetic information is essential in maintaining the integrity of living organisms. Point mutations can arise in DNA as a result of base mispairing during replication and a complex group of enzymes has evolved to recognise and remove these mistakes.<sup>1</sup> These proofreading and repair enzymes detect some base-pair mismatches more efficiently than others and certain modified bases give rise to incorrect pairs which are particularly evasive. Such mutagens, which are produced by irradiation or chemical damage to the four natural bases, may give rise to genetic defects and cancer. A number of theories have been proposed to explain the observed frequency of substitution mutations,<sup>2</sup> but until recently no detailed information was available on the precise conformations of mismatched base pairs in DNA. In the late 1970s the development of reliable methods for the synthesis and purification of deoxyoligonucleotides made it possible to contemplate structural studies on these materials by *X*-ray crystallography and high-field NMR methods. In this chapter we discuss some recent results obtained in this field in our laboratory.

## 2 THE A.C MISMATCH

An *X*-ray crystallographic analysis of the A.C mismatch in the self-complementary B-DNA dodecamer, d(CGCAAATTCGCG) at 2.5 Å resolution had previously shown the presence of a wobble base

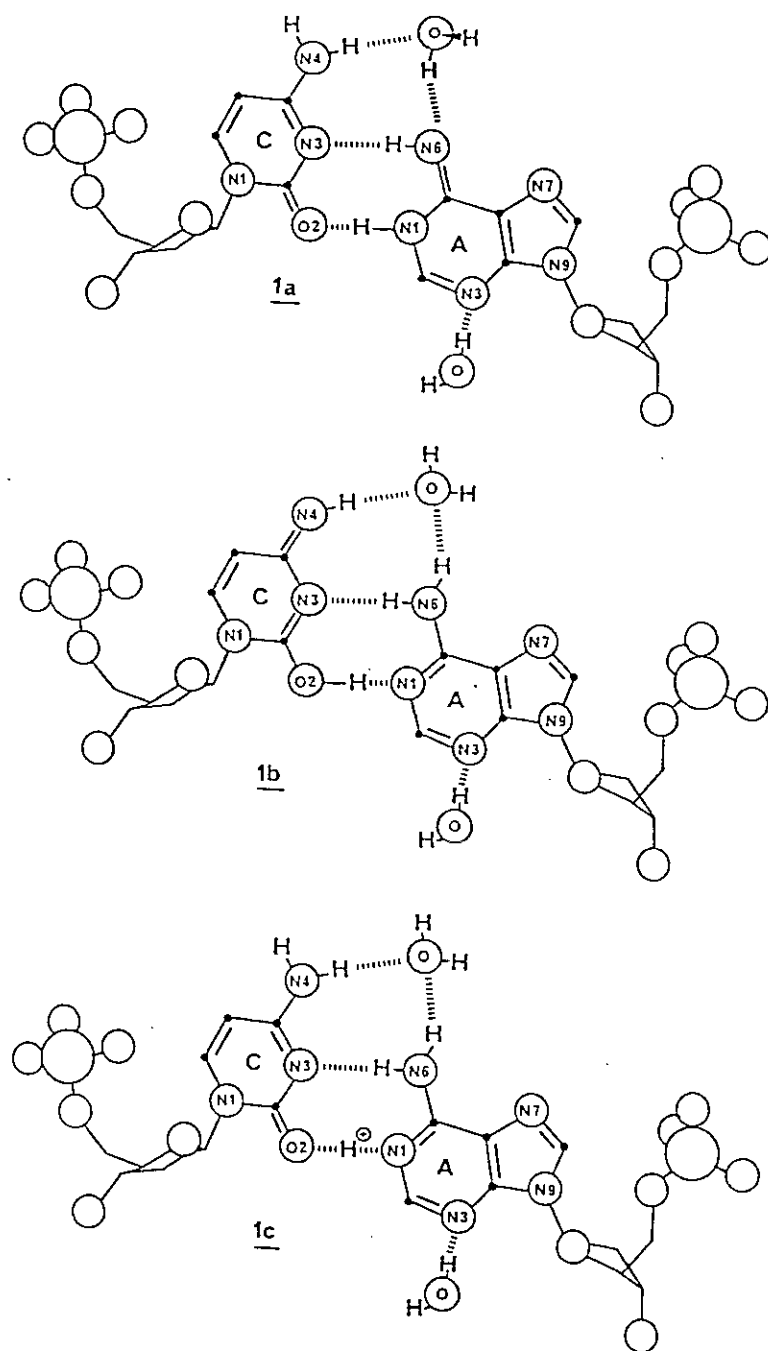


Figure 1: Possible forms of the A:C mismatch: (1a) C.A(imino); (1b) C(enol, imino).A; (1c) C.A<sup>+</sup>.

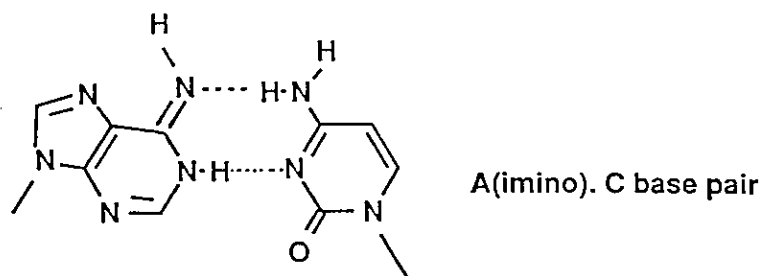


Figure 2: A minor tautomer C.A mismatch isomorphous with a Watson-Crick base pair.

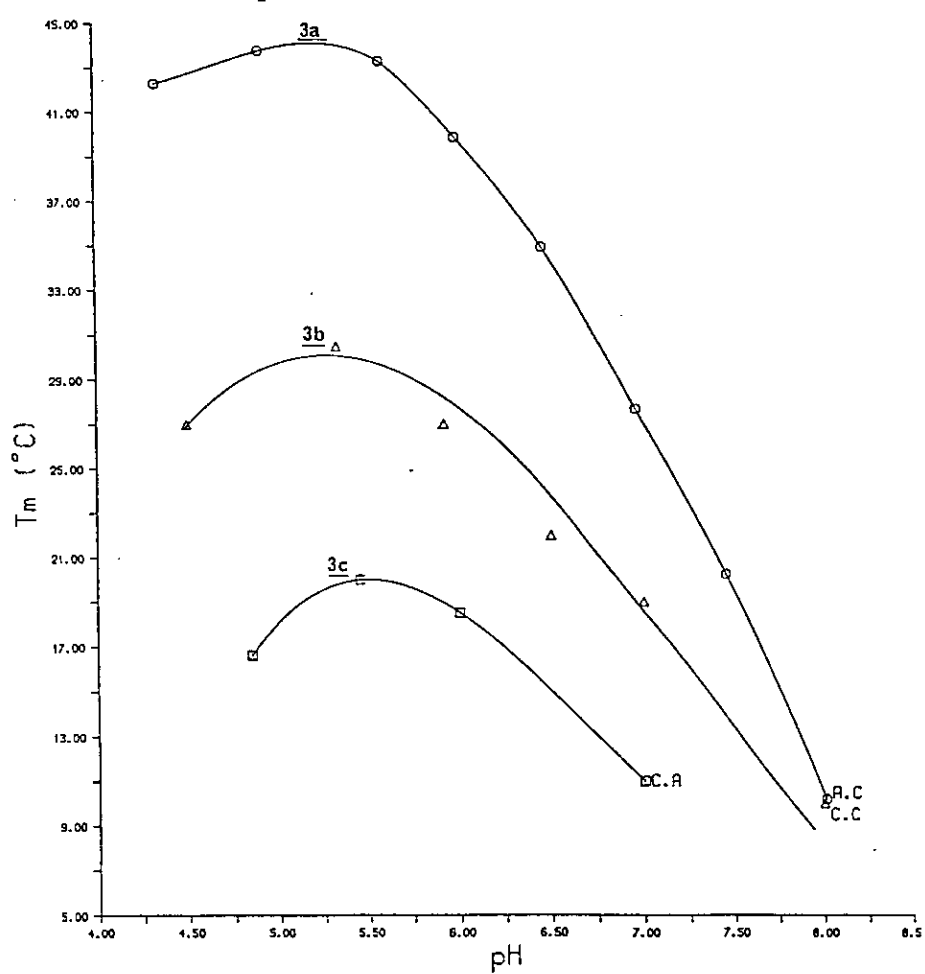


Figure 3: pH-dependence of melting temperature in 0.1M phosphate buffer: (3a) d(CGCAAATTCGCG); (3b) d(CGCCAATTCGCG); (3c) d(CGCCAATTAGCG).

pair.<sup>3</sup> At this resolution it is not possible to identify hydrogen atoms, so the crystal structure is consistent with any of the three A.C mismatches shown in Figure 1. The base pairs (1a) and (1b) seem unlikely as in each case one base is present as a rare tautomer. In particular, the occurrence of the imino tautomer of adenine (1a) might be expected to produce a Watson-Crick base pair (Figure 2) rather than a wobble base pair. The existence of the protonated base pair (1c) is also uncertain as the crystals were grown in sodium cacodylate buffer at pH 7, far from the  $pK_a$  for protonation of adenine N1 (*ca* 4.1). In order to resolve the situation we measured the melting temperature of the A.C dodecamer as a function of pH (Figure 3a). The results clearly show a correlation between duplex stability and hydrogen ion concentration, strongly supporting base pair (1c). At high pH the duplex is very unstable and in the absence of a high proton concentration it is likely that A.C base pairs do not form. Interestingly, the presence of a protonated base does not seem to destabilise the duplex and the A.C dodecamer is more stable than the native sequence around pH 5.

The correlation between pH and duplex stability for the sequence d(CGCCAATTAGCG) was also examined. This sequence relates to the previous one by inverting the A.C mismatches at positions 4 and 9, and this changes the base stacking environment. Again, the duplex becomes more stable as the pH is lowered until maximum stability is reached just above pH 5. Beyond this point protonation of cytosine N3 and adenine N1 atoms destabilises all Watson-Crick base pairs in the usual way (Figure 3c). The lower melting temperature of the C.A dodecamer relative to the A.C dodecamer might be due to the occurrence of an unfavourable dipole moment contribution to the CpC base stacking step in the former, compared to a favourable CpG step in the latter. This is speculative in the absence of accurate theoretical calculations.

### 3 THE G.A MISMATCH

The G.A mismatch is of particular interest as experiments with DNA polymerase III have shown that about 10% of these mismatches avoid detection, whereas only 0.5% of G.T mismatches escape.<sup>4</sup> The G.A

mismatch has been shown to be conformationally variable<sup>5-8</sup> and both *G(anti).A(syn)* and *G(anti).A(anti)* base pairs have been identified in different DNA sequences at neutral pH (Figure 4b, 4a). Theoretical calculations suggest that both have similar stability when incorporated in a DNA duplex although the *G(anti).A(anti)* base pair causes greater distortion to the sugar-phosphate backbone.<sup>9,10</sup> It has recently been shown by nuclear magnetic resonance that the G.A mismatch in the sequence d(CGGAATTCACG) displays conformational flexibility as a function of pH.<sup>11</sup> The *G(anti).A(anti)* base pair (4a) is observed at neutral pH, whereas the *G(syn).A(anti)* base pair (4c) predominates below pH 5.5. In the latter case the adenine base is protonated.

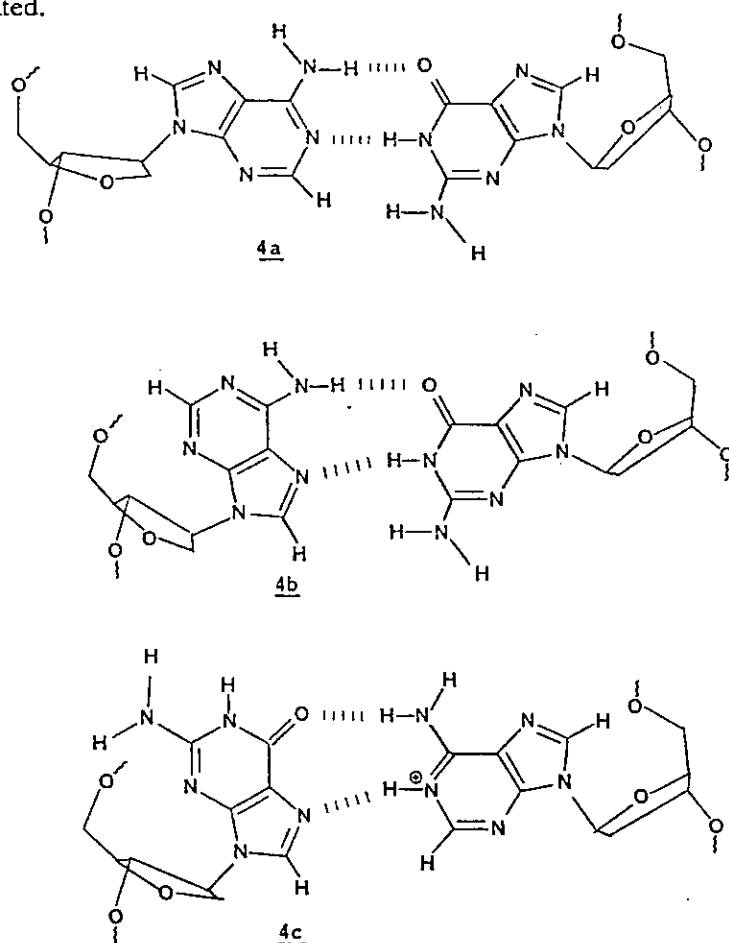


Figure 4: Possible forms of the G.A mismatch:  
 (4a) *A(anti).G(anti)*; (4b) *A(syn).G(anti)*; (4c)  $G^+(syn).A(anti)$ .

Our earlier crystallographic analysis of a G.A mismatch in the sequence d(CGCGAATTAGCG) showed the presence of a G(*anti*).A(*syn*) pairing (Figure 4b).<sup>7</sup> Very recently we have solved the structure of the dodecamer d(CGCAAATTGGCG) which is related to the previous sequence by interchanging the adenine and guanine bases of the G.A mismatches.<sup>12</sup> As a consequence of this change the mismatches are in a different base stacking environment. Clearly base stacking interactions must be influential in determining the form of conformationally flexible mismatches. Crystals were grown at pH 6.6 and X-ray data were measured at 4°C on a Stoe-Siemens AED2 four circle diffractometer equipped with a long arm and a helium path. Intensities were corrected for Lorenz and polarisation factors, absorption and time dependent decay. A total of 2262 unique reflections with  $F \geq 2$  sigma(F) in the range 8Å to 2.25Å were used in the refinement. Starting co-ordinates were those of the native dodecamer d(CGCGAATTCGCG) and the model was refined using the restrained least squares methods<sup>13,14</sup> with periodic examination of  $F_0-F_c$  maps on an Evans and Sutherland PS300 graphics system. During these calculations the mismatches were allowed to move freely relative to each other. It was clear from the fragment  $F_0-F_c$  electron density maps calculated with the mismatched base pairs omitted, that G(*syn*).A(*anti*) gave a much better electron density fit than either G(*anti*).A(*anti*) or G(*anti*).A(*syn*) [Figure 5].

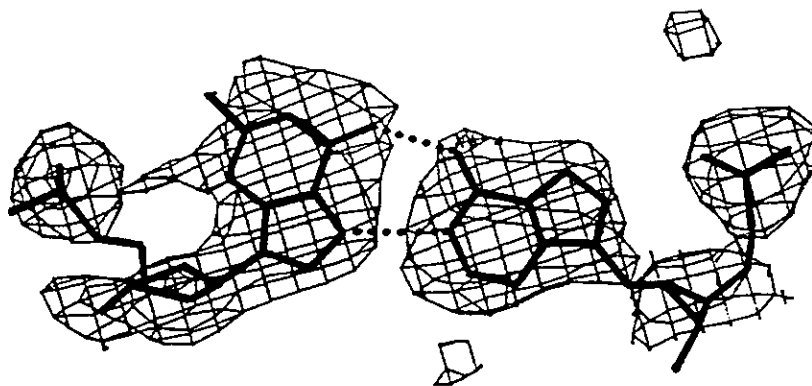


Figure 5: Fragment  $F_0-F_c$  electron density map of the G(*syn*).A(*anti*) base pair in the dodecanucleotide duplex d(GCGAAATTGGCG). All atoms have been omitted from the structure factor calculations to give an unbiased representation of the true electron density.

The refinement converged at  $R = 0.16$  with the location of 94 solvent molecules. Figure 4c shows the  $G(\text{syn}).A(\text{anti})$  base pair schematically. The N6 atom of adenine and the O6 atom of guanine are  $2.6\text{\AA}$  apart and the N1 atom of adenine and the N7 atom of guanine are separated by  $2.8\text{\AA}$ . These distances are indicative of strong hydrogen bonds, but with both bases in their major tautomeric forms a proton is required to enable the N1-adenine N7-guanine hydrogen bond to form. There is good evidence from thermal denaturation studies that the mismatch base pair is indeed protonated. Ultraviolet melting studies in the pH range 8.0 to 4.5 on the A.G dodecamer d(CGCAAATTGGCG) [Figure 6b] and the native control sequence d(CGCGAATTCGCG) [Figure 9a] show that the behaviour of the two duplexes in aqueous media is strikingly different (Figure 6). The melting temperature of the native sequence decreases sharply below pH 6.5, whereas that of the A.G dodecamer reaches a maximum around pH 5.0 before falling away at lower pH. These results suggest that the  $G(\text{syn}).A(\text{anti})$  base pair is present in solution over a wide pH range. It can, however, be concluded that at high pH other species are present as the melting temperature does not change between pH 7 and 8.

In the  $G(\text{syn}).A(\text{anti})$  base pair the 2-amino group of guanine lies in the major groove where it is free to hydrogen bond to two water molecules. This is not the case for  $G(\text{anti}).A(\text{syn})$  and  $G(\text{anti}).A(\text{anti})$  base pairs (Figure 4b and 4a, respectively) in which the 2-amino group is sterically crowded in the minor groove and can only interact with a single water molecule. This will lead to a loss of hydrogen bonding when the fully hydrated single strands come together to form a duplex. The disposition of major groove hydrogen bond acceptors and donors (heteroatoms) in the  $A(\text{anti}).G(\text{syn})$  base pair is completely different from that of other forms of the G.A mismatch although in general shape it resembles the  $A(\text{syn}).G(\text{anti})$  mismatch. These observations may be important when considering the recognition of mismatches by proofreading and repair enzymes. In the sequences d(CGCAAATTGGCG) and d(CGGAATTCACG), both of which have been shown to accommodate  $G(\text{syn}).A(\text{anti})$  base pairs, the guanine base of the mismatch is flanked by the guanine of a G.C base pair. Guanine has a large dipole moment ( $7.5D$ )<sup>15</sup> and a GpG step in B-DNA involving  $G(\text{anti})$  bases would be expected to have an

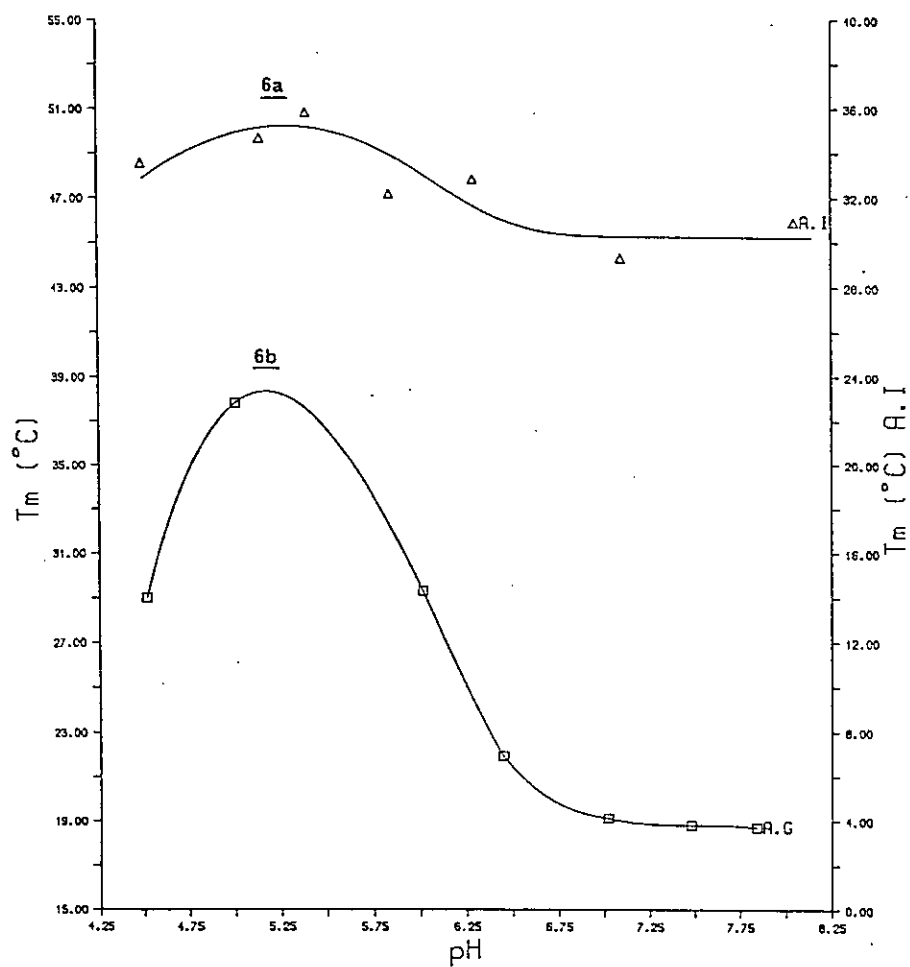


Figure 6: pH-dependence of duplex melting in 0.1M phosphate buffer.  
(6a): d(CGCAAATTIGCG), temperature scale on right hand side.  
(6b): d(CGCAAATTGGCG), temperature scale on left hand side.

unfavourable dipole-dipole contribution as the dipole moments of the two bases would be unfavourably aligned, with a relative rotation of only *ca* 33°. However, with the guanine base of the G.A mismatch in the *syn*-conformation, the static dipole-dipole interactions are likely to be much more favourable. In the sequences d(CGAGAATTCGCG) and d(CGCGAATTAGCG) in which the guanine bases of the G.A mismatches are *anti* [G(*anti*).A(*anti*) and

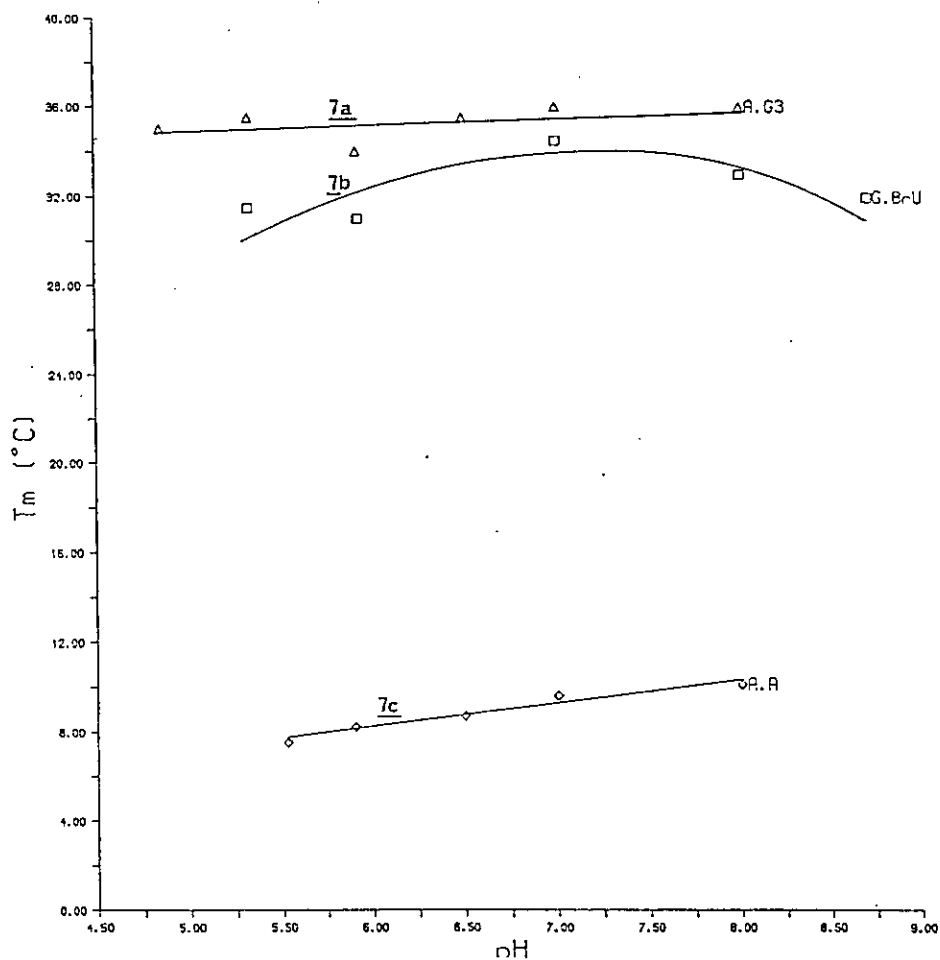
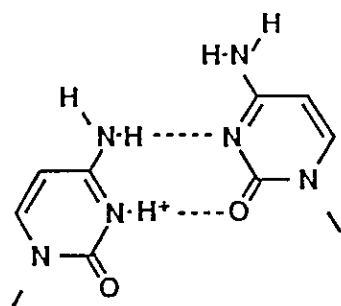
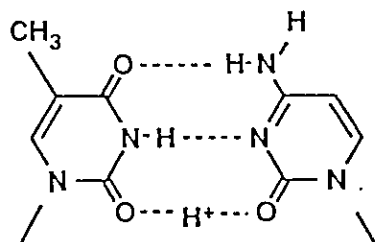


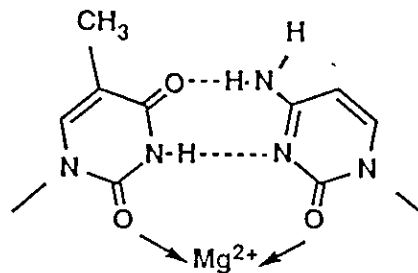
Figure 7: pH dependence of duplex melting in 0.1M phosphate buffer.  
 (7a) d(CGAGAATTCGCG); (7b) d(CGCGAATTBrUGCG);  
 (7c) d(CGCAAATTAGCG).



8a C+.C mismatch.



8c T.C mismatch?



8b T.C mismatch?

Figure 8: (8a): Putative C+.C mismatch; (8b) and (8c): Possible forms of the T.C mismatch.

G(*anti*).A(*syn*), respectively] the guanines are involved in CpG stacking interactions. The dipole moment of cytosine is almost identical to that of guanine (7.6D) and when on the 5'-side of guanine in B-DNA it lies in almost exactly the opposite direction. Thus, the CpG(*anti*) base stack will have a favourable static dipole contribution. The dipole moment of adenine is relatively small so its influence on the G.A mismatch will be less than that of guanine. Dipole moments may be particularly important in the two *single* strands as they come together and intrastrand stacking interactions can be optimised, locking the G.A mismatch into a specific conformation as the duplex is formed. Once the duplex is formed, any change in the conformation of the G.A mismatch will require rotation of a

purine base around its glycosidic bond, an operation that would considerably disrupt the surrounding Watson-Crick base pairs. It is well known that poly d(CG) forms a more stable duplex than poly dG.poly dC<sup>16</sup> and that the CpG base stacking step is the most stable in B-DNA,<sup>17</sup> so our results with mismatches are consistent with these observations. Thermal denaturation studies on the dodecamer d(CGAGAATTCGCG) in our laboratory show no indication of increased stability with decreasing pH (Figure 7a), suggestion that the guanine base of the G.A mismatch has no tendency to adopt the *syn*-conformation in this sequence. This observation is consistent with the prediction that the *anti*-conformation should be strongly preferred for the mismatched guanine base when it is on the 3'-side of a cytosine.

#### 4 OTHER MISMATCHES

Any mismatch containing either adenine or cytosine bases can in principle be stabilised by protonation near physiological pH as the N1-atom of adenine and the N3-atom of cytosine both have  $pK_a$  values around 4. Thermal denaturation studies on the dodecamer d(CGCAAATTAGCG) indicate that in this sequence the A.A. mismatch is not stabilised by protonation (Figure 7c). Due to a lack of high resolution crystallographic data we are unable to speculate on the precise form of the A.A mismatch. So far we have been unable to grow suitable crystals of a DNA duplex containing pyrimidine-pyrimidine base pairs and no crystallographic information is available. It is likely that these mismatches will not span the same distance across the duplex as Watson-Crick base pairs so their incorporation into B-DNA will lead to some distortion of the sugar-phosphate backbone. In general pyrimidine-pyrimidine base pairs substantially destabilise the duplex and the strong pH dependence of the stability of the C.C mismatch in the sequence d(CGCCAATTCGCG) (Figure 3b) is consistent with the protonated base pair in Figure 8a. The T.C mismatch in the dodecamer d(CGCTAATTCGCG) shows no such pH dependent stability (Figure 9b) so the base pair (8c) is unlikely to predominate. An alternative base pair that is not stabilised by protonation is illustrated in Figure 8b.

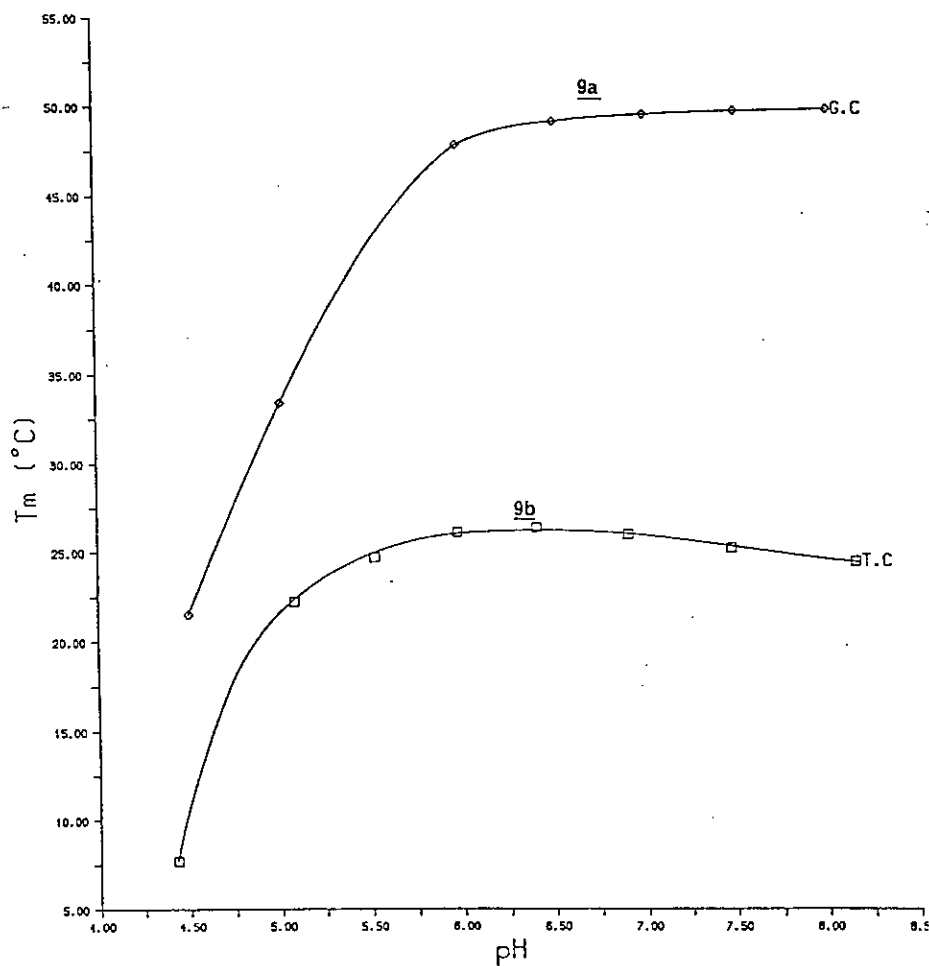


Figure 9: pH-dependence of duplex melting in 0.1M phosphate buffer: (9a) d(CGCGAATTCGCG); (9b) d(CGCTAATTCGCG).

### 5 5-BROMO-URACIL

5-Bromo-uracil is an analogue of thymine that has an enhanced tendency to form mismatches with guanine.<sup>18</sup> The stability of the G.BrU mismatch could be due to any of the following:

- (i) A stable wobble base pair (10a) with the bromine atom involved in favourable base stacking interactions.
- (ii) The formation of (10b), a Watson-Crick-like base pair with the deprotonated N3-atom of 5-bromo-uracil acting as a hydrogen bond acceptor.
- (iii) The occurrence of (10c), a Watson-Crick-like base pair involving a rare enol tautomer of 5-bromo-uracil.

Although (10b) and (10c) would be indistinguishable at the limit of resolution of a typical oligonucleotide crystal structure, (10c) is unlikely to occur, as there are no grounds for proposing a very high ratio of the enol form of 5-bromo-uracil. However, as the  $pK_a$  for deprotonation of the N3 position of 5-bromo-uracil is 8.5, compared to a  $pK_a$  of 10 for thymine, base pair (10b) might well occur. In order to attempt to identify this species we crystallised the hexamer  $d(\text{BrUGCGCG})$  at pH 8.5 where 5-bromo-uracil should be 50% ionised. The hexamer crystallised in the Z-form and the structure was solved to a resolution of  $2.2\text{\AA}$ .<sup>19</sup> The Z-DNA conformation was deliberately chosen because of the high diffracting power of crystals

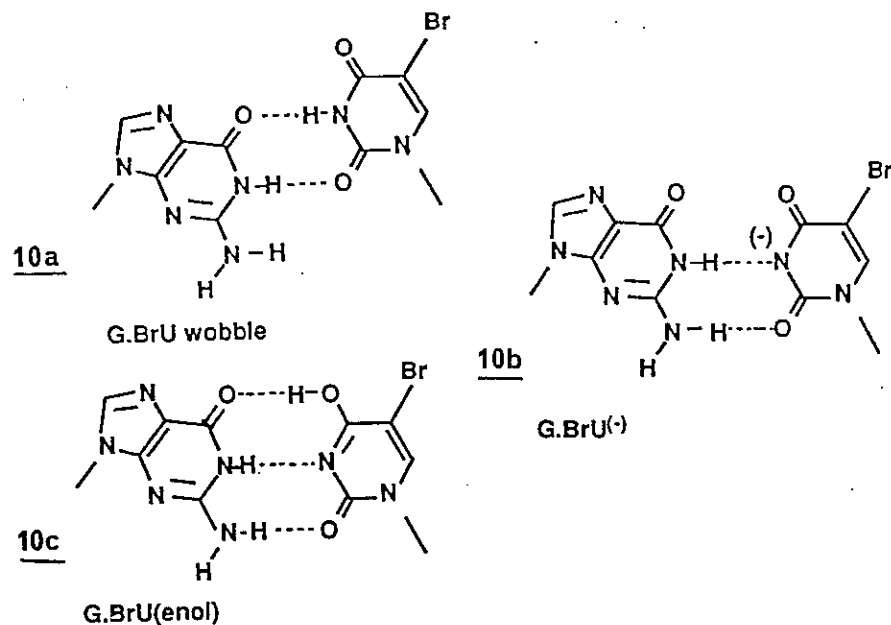


Figure 10: Possible G.BrU mismatches: (10a) G.BrU(wobble); (10b) G.BrU<sup>-</sup>; (10c) G(enol).BrU or G.BrU(enol).

of this form. Crystallisation of oligonucleotides is impeded at high pH so we were forced to work with small crystals. Analysis of the structure revealed that the wobble base pair (10a) was the only detectable species. Ultraviolet melting studies (Figure 7b) of the sequence d(CGCGAATTBrUGCG), a self-complementary dodecanucleotide containing two G.BrU mismatches were consistent with this finding. There was no significant change in the melting temperature of the duplex between pH 5.5 and 8.5 even though the formation of (10b) might be expected to be more favourable at high pH. If this base pair *were* to form it would be destabilised by the close contact between the O6 atom of guanine and the O4 atom of BrU. This unfavourable interaction could only be stabilised at *very low* pH by protonation of the carbonyl oxygen of either base. Therefore conditions suitable for the stabilisation of the guanine O6-bromo-uracil O4 hydrogen bond are totally unsuitable for the formation of the guanine N1-bromo-uracil N3 hydrogen bond. It has been shown by NMR that the presence of the electron withdrawing bromine atom in the 5-position of uracil strengthens the hydrogen bonding capacity of the O2-atom and weakens the hydrogen bonding capacity of the O4-atom.<sup>20</sup> Clearly this would stabilise the G.BrU wobble base pair relative to the A.BrU 'Watson-Crick' base pair and produce a mutagenic effect.

## 6 DEOXYINOSINE

Inosine is an analogue of guanosine which lacks the 2-amino group. It occurs in ribonucleic acids, especially in the wobble position in some transfer RNA anticodons, where it is known to base pair with A, C or U at the messenger RNA codon.<sup>21</sup> It is also able to occupy the middle position of the anticodon and pair with A.<sup>22</sup> The occurrence of deoxyinosine in genomic DNA is much less common, although it is occasionally produced by the deamination of deoxyadenosine. As it is potentially mutagenic<sup>23</sup> it is efficiently removed by the enzyme hypoxanthine DNA glycosylase.<sup>24-26</sup> Because of its tendency to form stable mismatches, deoxyinosine has been used as a universal base in synthetic hybridisation probes.<sup>27-29</sup>

In order to investigate the precise nature of the I.A pairing in a

DNA helix we carried out an *X*-ray structural analysis on the synthetic dodecamer d(CGCI AATTAGCG) to 2.5Å resolution.<sup>30</sup> This analysis established that an I(*anti*).A(*syn*) base pair (Figure 11c) can be accommodated in the B-DNA duplex without any major distortion of the local or global conformation. The structure is essentially isomorphous with that of d(CGCGAATTAGCG) which contains two G(*anti*).A(*syn*) mispairs (Figure 11d), but because inosine lacks a 2-amino group there is no possibility of a steric clash in the minor groove. This is likely to be the origin of the greater stability of I.A mismatches and it can be seen from Figure 6 that A.I mismatches are significantly more stable than G.A mismatches around neutral pH. We are currently crystallising the dodecamer d(CGCAAATTIGCG).

An analogous comparison was made between the I.T and G.T mismatches in the A-DNA octamers d(GGGGCTCC)<sup>31</sup> and d(GGIGCTCC),<sup>32</sup> respectively. A high resolution *X*-ray analysis<sup>32</sup> on the I.T octamer (1.7Å) showed the presence of wobble base pairs (Figure 11a) similar to those found previously (Figure 11b) for the G.T mismatches.<sup>31,33,34</sup> As the 2-amino group of guanine in the G.T mismatch is not involved in any unfavourable steric clashes it is not surprising that the G.T and I.T mismatches have almost identical stability.

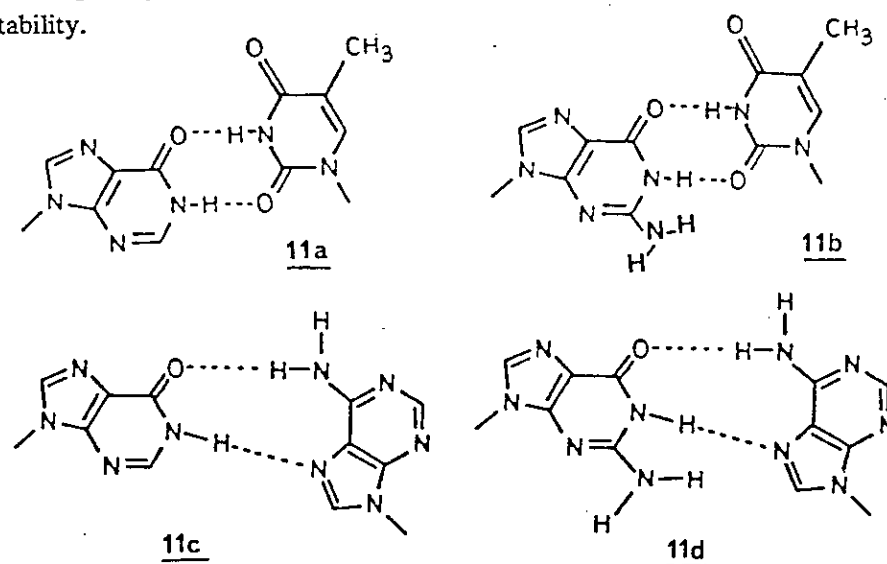


Figure 11: A comparison of mismatches of guanine and of hypoxanthine: (11a): I.T(wobble); (11b): G.T(wobble); (11c): I.A(*syn*); (11d): G.A(*syn*).

## 7 SUMMARY

A combination of X-ray crystallography and ultraviolet melting studies have been used to examine the structure and stability of DNA duplexes containing mismatched base pairs and modified bases. Base stacking interactions clearly have an influence on mismatch conformation and we are currently measuring thermodynamic parameters of a wide range of duplexes in order to gain a clearer insight into the factors influencing the stability of abnormal base pairs in DNA.

## ACKNOWLEDGEMENTS

We are grateful to Olga Kennard, Bill Hunter and Geoff Kneale for helpful discussions. This work was funded by SERC Molecular Recognition grants, and Awards from the Research Corporation Trust and the Nuffield Foundation to TB.

## REFERENCES

- 1 E C Friedberg, 'DNA Repair', W H Freeman and Co, New York, 1985.
- 2 M D Topal and J R Fresco, *Nature*, 1976, 263, 285.
- 3 W N Hunter, T Brown, N N Anand, and O Kennard, *Nature* 1986, 320, 552.
- 4 A R Fersht, J W Knill-Jones, and W C Tsui, *J Mol Biol*, 1982, 156, 37.
- 5 L-S Kan, S Chandrasegaran, S M Pulford, and P S Miller, *Proc Nat Acad Sci USA*, 1983, 80, 4263.
- 6 D J Patel, S A Koslowski, S Ikuta, and K Itakura, *Biochemistry*, 1984, 23, 3207
- 7 T Brown, W N Hunter, G Kneale and O Kennard, *Proc Nat Acad Sci USA*, 1986, 83, 2402.
- 8 G G Prive, U Heinemann, L-S Kan, S Chandrasegaran, and R E Dickerson, *Science*, 1987, 238, 498.
- 9 V P Chuprina and V I Poltev, *Nucleic Acids Res*, 1983, 11, 5205.

- 10 J W Keepers, P Schmidt, T L James, and P A Kolman, *Biopolymers*, 1984, 23, 2901.
- 11 X Gao and D J Patel, *J Am Chem Soc*, 1988, 110, 5178.
- 12 T Brown, G A Leonard, E D Booth, and J Chambers, *J Mol Biol*, 1989, 207, 455.
- 13 W A Hendrickson and J H Koonin, 'Biomolecular Structure, Conformation, Function and Evolution', Plenum Press, 1981, pp43-57.
- 14 E Westhof, P Dumas, and D Moras, *J Mol Biol*, 1985, 184, 119.
- 15 B Pullman and A Pullman, *Adv Het Chem*, 1971, 13, 77
- 16 O Gotoh and Y Takashira, *Biopolymers*, 1981, 20, 1033.
- 17 R L Ornstein, R Rein, D L Breen, and R D MacElroy, *Biopolymers*, 1978, 17, 2341.
- 18 E Freeze, *J Mol Biol*, 1959, 1, 87.
- 19 T Brown, G Neale, W N Hunter, and O Kennard, *Nucleic Acids Res*, 1986, 14, 1801.
- 20 H Iwahashi and Y Kyogoku, *J Am Chem Soc*, 1977, 99, 7761.
- 21 F H C Crick, *J Mol Biol*, 1966, 19, 548.
- 22 B D Davis, P Anderson, and P F Sparking, *J Mol Biol*, 1973, 76, 223.
- 23 T Lindahl, *Prog Nucleic Acids Res Mol Biol*, 1979, 22, 135.
- 24 P Karran and T Lindahl, *J Biol Chem*, 1978, 253, 5877.
- 25 B Myrnes, P-H Guddal, and H Krokan, *Nucleic Acids Res*, 1982, 10, 3693.
- 26 T Lindahl and P Karran, *Biochemistry*, 1980, 19, 6005.
- 27 Y Takahashi, K Kato, Y Hayashizaka, T Wakabayashi, E Ohtsuka, S Matsuki, I Ikehara, and K Matsubara, *Proc Nat Acad Sci USA*, 1985, 82, 1931.
- 28 E Ohtsuka, S Matsuki, M Ikehara, Y Takahashi, and K Matsubara, *J Biol Chem*, 1985, 260, 2605.
- 29 F H Martin, M M Castro, F Aboul-ela, and I Tinoko, *Nucleic Acids Res*, 1985, 13, 8927.
- 30 P W R Corfield, W N Hunter, T Brown, P Robinson, and O Kennard, *Nucleic Acids Res*, 1987, 15, 7935.
- 31 T Brown, O Kennard, G Kneale, and D Rabinovich, *Nature*, 1985, 315, 604.
- 32 W B T Cruse, W Aymani, O Kennard, T Brown, A G C Jack, and G A Leonard, *Nucleic Acids Res*, 1989, 17, 55.

- 33 G Kneale, T Brown, O Kennard, and D Rabinovich, *J Mol Biol*, 1985, 186, 805.
- 34 W N Hunter, T Brown, G Kneale, N N Anand, D Rabinovich, and O Kennard, *J Biol Chem*, 1987, 262, 9962.

# Structural and thermodynamic studies on the adenine.guanine mismatch in B-DNA

Gordon A. Leonard, Ewan D. Booth and Tom Brown

Department of Chemistry, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JJ, UK

Received August 8, 1990; Revised and Accepted September 7, 1990

## ABSTRACT

The structure of the synthetic dodecamer d(CGCAAATTGGCG) has been shown by single crystal X-ray diffraction methods to be that of a B-DNA helix containing two A(*anti*).G(*syn*) base pairs. The refinement, based on data to a resolution of 2.25 Å shows that the mismatch base pairs are held together by two hydrogen bonds. The *syn*-conformation of the guanine base of the mismatch is stabilised by hydrogen bonding to a network of solvent molecules in both the major and minor grooves. A pH-dependent ultraviolet melting study indicates that the duplex is stabilised by protonation, suggesting that the bases of the A.G mispair are present in their most common tautomeric forms and that the N(1)-atom of adenine is protonated. The structure refinement shows that there is some disorder in the sugar-phosphate backbone.

## INTRODUCTION

The removal of mispaired bases by proofreading and repair enzymes is essential to the maintenance of fidelity during replication (1) and the enzymic recognition of base pair mismatches in genomic DNA is likely to depend upon structural and thermodynamic factors (2). In order to gain insight into the molecular basis of these recognition processes the structure and stability of isolated base pair mismatches in synthetic oligonucleotide duplexes have been investigated by a variety of methods, such as NMR spectroscopy (3), X-ray crystallography (4) and ultraviolet duplex melting techniques (5–8). Recent studies have revealed that the properties of certain mispairs are strongly sequence-dependent (9). Hence, a comprehensive understanding of the properties of DNA duplexes containing mispaired bases can only be gained by thoroughly investigating mismatches in a variety of base stacking environments.

The guanine-adenine mismatch is of particular interest as proofreading enzymes remove this base pair relatively inefficiently (10). Furthermore, it has been shown that the G.A base pair displays significant conformational variability (11, 12), both as a function of pH and base-stacking environment. As part of a study on the nature of base pair mismatches in DNA duplexes we now report the X-ray structure and thermodynamic properties of the deoxydodecanucleotide d(CGCAAATTGGCG)<sub>2</sub>.

## EXPERIMENTAL METHODS

### Chemical synthesis

The self-complementary dodecanucleotide d(CGCAAATTGGCG) was synthesised on an ABI 380B DNA synthesiser by the phosphoramidite method (6 × 1 μmole preps) and purified by ion-exchange HPLC followed by reversed-phase HPLC and Sephadex gel filtration. This protocol yielded 10mgs of product which eluted as a single peak when injected on analytical reversed-phase HPLC. (Ion-exchange gradient; 0.04M to 0.67M potassium phosphate buffer pH 6.4, in 20% acetonitrile, 30 minutes. Reversed-phase gradient; 0.1M ammonium acetate buffer, 3% to 14% acetonitrile, 30 minutes).

### Ultraviolet melting studies

Melting curves were measured at 264nm on a Perkin-Elmer Lambda 15 ultraviolet spectrometer equipped with a Peltier block and controlled by an IBM PS2 microcomputer. A heating rate of 1.0 deg.C per minute was used throughout and all melting curves were measured in triplicate. The crude data was collected and processed using the PECSS2 software package. For the pH-dependent ultraviolet melting data, the oligonucleotides were dissolved in a buffer consisting of aqueous sodium dihydrogen orthophosphate (0.1M) and EDTA (1mM) which had been adjusted to the appropriate pH by the addition of sodium hydroxide. Thermodynamic parameters were determined from the concentration-dependence of ultraviolet melting by standard methods (5). For this purpose melting curves were recorded in aqueous sodium chloride (1M), sodium dihydrogen orthophosphate buffer (10mM) and EDTA (1mM) which had been adjusted to the desired pH by the addition of sodium hydroxide. Oligonucleotide concentrations were calculated from the ultraviolet absorbance (264nm) of digested samples dissolved in the UV melting buffer, using the published ε<sub>260</sub> values of the nucleosides dA, dG, dC and T as standards. Enzymic digestion was carried out using snake venom phosphodiesterase 1 and alkaline phosphatase (Sigma Chemical Co. Ltd.)

All but one of the DNA sequences discussed in this paper were shown by the concentration-dependence of UV melting to form duplexes, not single-stranded structures. However, the T<sub>m</sub> of the dodecanucleotide d(CGCGAATTAGCG) was found to be independent of DNA concentration, suggesting a single stranded hairpin loop. Consequently it was not possible to obtain duplex melting data on this particular sequence, even though it has been

previously crystallised as B-DNA (13). We are currently investigating this structure by high-field NMR techniques.

### X-ray crystallography

Crystals were grown at 4°C from a solution containing the oligonucleotide (0.5mM), sodium cacodylate buffer (10mM, pH 6.6), magnesium chloride (20mM), spermine tetrahydrochloride (1.0mM) and 2-methyl-2,4-pentanediol (18% vol/vol). The crystals were orthorhombic, space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with unit cell dimensions of a=25.23Å, b=41.16Å and c=65.01Å. A single crystal of dimensions 2.0mm×0.4mm×0.3mm was selected and mounted in a sealed Lindemann capillary with a small quantity of mother liquor. A CuK-α radiation source was used and X-ray data were collected at 4°C on a Stoe-Siemens AED2 four circle diffractometer equipped with a long arm and a helium path. Intensities were corrected for Lorentz and polarisation effects, absorption and time dependent decay. The data collection yielded 2840 unique reflections with  $F \geq \sigma(F)$  to a resolution of 2.25Å. The unit cell parameters indicated that the structure was quasi-isomorphous with the native B-DNA dodecanucleotide d(CGC-GAATTCGCG) (14) and the co-ordinates for this structure, idealised for the slightly different unit cell, were used as a starting model for the refinement.

Two approaches to the refinement were used. In the first method the model was refined using only restrained least-squares methods (15, 16), initially against data in the range 10Å to 7Å. Upon convergence the resolution was increased in five steps to 2.5Å. Throughout the calculations the atoms of the nucleotides G4, C9, G16 and C21 were given a very low occupancy and were thus effectively omitted from the structure factor calculations. At 2.5Å, the  $2F_o - F_c$  and  $F_o - F_c$  maps were

examined on an Evans and Sutherland PS300 graphics system using FRODO (17) and the necessary changes were made at the sites of the G.A mispairs. It was immediately apparent that G(*syn*).A(*anti*) gave a much better fit to the density than either G(*anti*).A(*syn*) or G(*anti*).A(*anti*). The atoms of A4.G21 and G9.A16 were given full occupancy in this conformation and the refinement was continued using data to 2.25Å, with the inclusion of isotropic thermal parameters and the progressive addition of water molecules. Solvent molecules were included on the criteria of good spherical density in the difference maps and acceptable hydrogen bonding distances and geometry. The refinement converged after the identification of 94 solvent molecules with  $R = 0.158$  for 2262 reflections with  $F \geq 2.0 \sigma(F)$  in the range 8Å to 2.25Å.

In the second method the starting model was refined as a rigid body using a modified version of SHELX (18). Initially data in the region 10.0Å to 6.0Å were used and the resolution was then increased in steps of 1.0Å up to 3.0Å. This part of the refinement converged at  $R=0.34$  for 1275 reflections. The nucleotides G4, C9, G16 and C21 were then removed from the structure factor calculations and the refinement was continued, extending the resolution first to 2.5Å then to 2.25Å using Konnert-Hendrickson techniques (15) with the program NUCLSQ (16) with tighter geometric constraints than in the first refinement. This part of the refinement converged at  $R=0.35$  for 2301 reflections in the range 10.0Å to 2.25Å with  $F \geq 2\sigma(F)$ . Electron density ( $2F_o - F_c$ ) and difference ( $F_o - F_c$ ) maps were examined on the graphics system and the refinement was continued in the manner described for the first method. A total of 62 solvent molecules were identified and the final crystallographic residual was  $R=0.167$  for 2212 reflections with  $F \geq 2\sigma(F)$  in the region 7.0Å to 2.25Å.

**Table 1.** Torsion angles (degrees) and distances (Å) between adjacent phosphorous atoms.

<i>Residue</i>	$\chi$	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$P_i - P_{i+1}$
C1	-117	-	-	170	142	-169	-124	-
G2	-94	-47	181	38	120	-204	-73	6.5
C3	-142	-96	182	81	91	-169	-98	6.6
A4	-101	-45	170	53	159	-139	-133	7.0
A5	-138	-74	136	75	134	-151	-132	7.2
A6	-105	-40	150	41	127	-172	-113	6.6
T7	-137	-53	155	59	85	-198	-63	6.6
T8	-113	-74	197	52	110	-198	-62	6.5
G9	80	-44	190	4*	147	-157	-111	6.4
G10	-88	-58	165	45	153	-118	-213	6.8
C11	-122	-41	123	42	133	-172	-90	6.7
G12	-96	-105	199	64	109	-	-	-
C13	-131	-	-	39	124	-139	-166	-
G14	-125	-10	114	42	111	-168	-126	6.7
C15	-126	-6	122	44	113	-165	-119	6.4
A16	-107	-76	165	77	143	-166	-128	7.2
A17	-110	-32	160	28	156	-155	-123	6.8
A18	-97	-56	178	30	141	-178	-101	6.6
T19	-149	-66	156	89	100	-160	-88	6.9
T20	-93	-17*	194	-2*	145	-159	-108	6.3
G21	86	-55	174	29	137	-161	-135	6.8
G22	-76	22*	137	-17*	163	-102	-198	6.5
C23	-93	-42	168	-7*	154	-161	-124	6.6
G24	-138	-73	156	47	80	-	-	-

The main chain torsion angles are defined by :  
 $P_i - \alpha - O5' - \beta - C5' - \gamma - C4' - \delta - C3' - \epsilon - O3' - \zeta - P_{i+1}$

Glycosyl torsion angles are given by:  
 $\chi = O4' - C1' - N1 - C2$  for pyrimidines  
 $\chi = O4' - C1' - N9 - C4$  for purines

\*Outside the normal range for B-DNA (see text).

## RESULTS AND DISCUSSION

### The sugar-phosphate backbone

The first refinement, based purely on the Konnert-Hendrickson technique, gave rise to a number of unacceptable backbone torsion angles. Some values of  $\alpha$  and  $\gamma$  in particular differed significantly from those usually found in B-DNA. Examination of the  $F_o - F_c$  fragment maps with backbone atoms removed indicated that all the atoms were in density and attempts to constrain all  $\alpha$  and  $\gamma$  angles to correct values resulted in an increase in the  $R$ -factor. The occurrence of non-standard torsion angles was surprising, as previous structures of DNA duplexes containing mismatches have shown that the mispairs are accommodated into the double helix with only small, highly localised changes relative to the native duplex. One exception is the structure of a deoxydecamer containing G(*anti*). A(*anti*) base pairs (19), but this is clearly a special case due to the abnormally large C1'-C1' separation in the mismatch base pairs. It is, however, interesting to note that the X-ray crystal structure of a B-DNA dodecamer containing N(6)-methyl adenine-thymine base pairs revealed some unusual  $\alpha$  and  $\gamma$  torsion angles (21)

In order to correct the abnormal torsion angles, we refined

**Table 2.** Average torsion angles in B-DNA dodecamer structures (estimated standard deviations in parenthesis)

Angle	G.C	G.T	I.A	A.T	GATATC	A.G
$\alpha$	-63(8)	-63(16)	-68(35)	-48.2	-85(40)	-32(59)
$\beta$	171(14)	-179(28)	175(31)	166.3	176(51)	169(57)
$\gamma$	54(8)	51(13)	49(19)	39.2	62(54)	25(88)
$\delta$	123(21)	119(24)	120(27)	136.5	135(35)	134(28)
$\epsilon$	-169(25)	-171(30)	-183(46)	-181.9	-129(79)	-168(29)
$\zeta$	-108(34)	-98(19)	-126(75)	-95	-95(54)	-121(55)
$\chi$	-117(14)	-109(19)	-111(18)	-108	-119(35)	-118(17)*

\* omits G9 and G21 bases which have *syn*-orientation

Torsion angles defined by:

P- $\alpha$ -O5'- $\beta$ -C5'- $\gamma$ -C4'- $\delta$ -C3'- $\epsilon$ -O3'- $\zeta$ -P

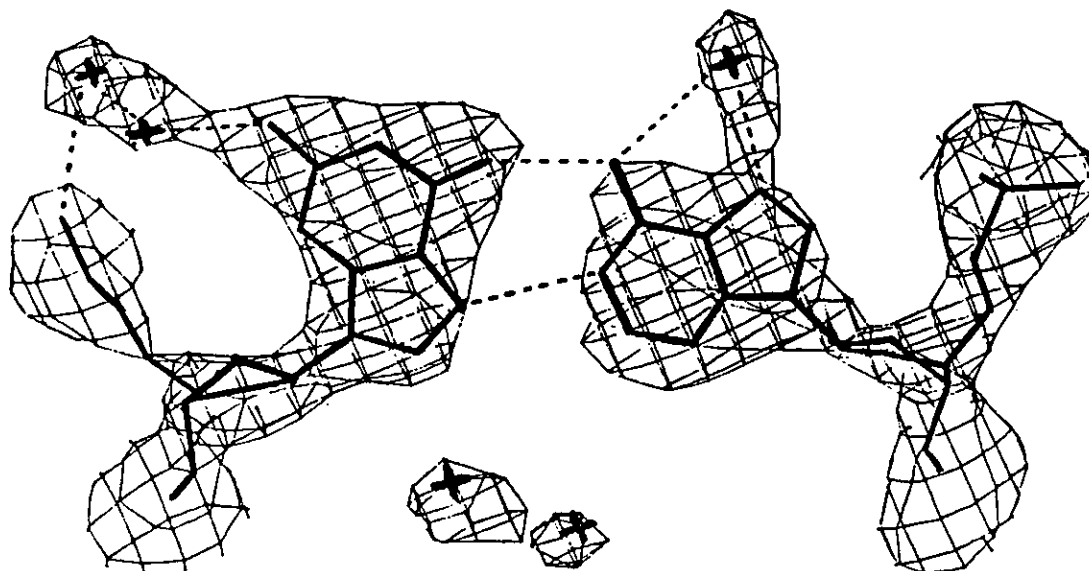
and  $\chi$ =O4'-C1'-N1-C2 for pyrimidines and  $\chi$ =O4'-C1'-N9-C4 for purines

the structure a second time using a stricter protocol. This refinement generally produced a more characteristic B-DNA backbone conformation (table 1, 2). However, the values of  $\alpha$  and  $\gamma$  for the nucleotides T20 and G22 were still outside the acceptable range. The structure was then examined by high field NMR methods and the details of this analysis will be described in full at a later date (22). The NMR study provided important information relating to the dynamic processes around the mismatch sites. In addition, it confirmed the presence of disorder in the backbone around the A.G base pairs. Hence, the abnormal torsion angles quoted in table 1 are probably not meaningful. The initial rigid body refinement and tighter geometric constraints applied in the second analysis were clearly helpful, but were not sufficient to produce acceptable values of  $\alpha$  and  $\gamma$  for T20 and G22. However, the overall structure produced from the second refinement is far more acceptable than that produced in the first. For this reason the remainder of the discussion is based on the results of refinement 2.

### The A.G mismatch base pair

The  $2F_o - F_c$  and  $F_o - F_c$  fragment maps clearly show the conformation of each A.G mispair within the B-DNA helix to be A(*anti*).G(*syn*). The geometry of both base pairs is essentially identical with two inter-base distances, N(6)-adenine to O(6)-guanine (2.6 Å) and N(1)-adenine to N(7)-guanine (2.8 Å) indicative of strong hydrogen bonds (figure 1). The average C1'-C1' separation (10.8 Å) and propeller twist (15.5°) of the mispaired bases are within the accepted range for Watson-Crick base pairs.

At the resolution of the structure determination it is not possible to locate hydrogen atoms so at least two tautomeric forms are consistent with the A.G base pair. With each mispaired base as its major tautomer (figure 2a), protonation is required to stabilise the hydrogen bond between the adenine-N(1) and guanine-N(7) atoms. It is reasonable to postulate such a structure, as the  $pK_a$  value of the N(1) atom of a free adenine base is *ca.* 4.5. An alternative base pair, with the adenine base in its minor imino-tautomeric form (23) can occur without protonation of either base



**Figure 1.**  $2F_o - F_c$  map of the G(9).A(16) base pair. Hydrogen bonds involving bases and major groove solvent molecules are shown in dotted lines.

(figure 2b), but minor tautomers are thought to occur with a very low frequency. Ultraviolet melting studies in the pH range 8.0 to 4.5 (9) indicate that the protonated  $AH^+$ (*anti*). $G$ (*syn*) base pair is present, as the stability of the A.G dodecamer reaches a maximum at pH 5.2 ( $T_m=38.5^\circ C$ ) and falls away at higher pH. Between pH 6.5 and pH 8.0 the melting temperature becomes constant at  $19^\circ C$ , indicative of non pH-dependent forms of the G.A mismatch ( $T_m$  values quoted at a single strand concentration of  $5 \mu\text{molar}$ ).

In the  $AH^+$ (*anti*). $G$ (*syn*) base pair, the 2-amino group of guanine protrudes into the major groove. This is in contrast to other forms of the G.A mismatch, namely  $G$ (*anti*). $A$ (*anti*) (19) and  $G$ (*anti*). $A$ (*syn*) (13) (figure 3, figure 4) in which the amino group lies in the minor groove. In these base pairs steric hindrance inhibits hydrogen bonding between one of the hydrogen atoms of the 2-amino group of guanine and surrounding water molecules. Thus there is a net loss of hydrogen bonding when the fully hydrated single strands interact to form a duplex. In general the disposition of hydrogen bond donors and acceptors in the  $A$ (*anti*). $G$ (*syn*) base pair is completely different from other forms of the G.A mismatch. These differences may be significant during enzymic recognition, as repair frequency will be influenced by the conformation of the base pair.

#### Base stacking and helical parameters

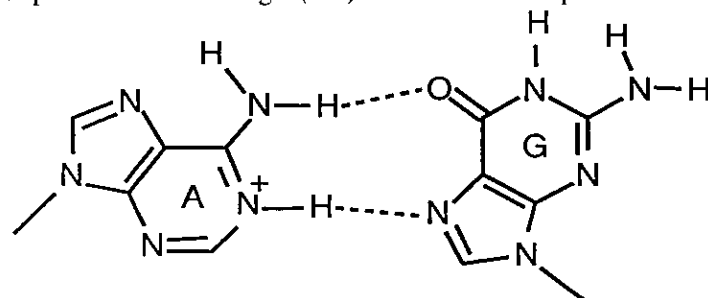
The global twist of the helix, which is  $36^\circ$  for the A.G mismatch compared with  $37^\circ$  for the native duplex, is essentially unaffected by the disorder in the backbone. The average rise per residue is 3.3A in both structures and other parameters, such as the separation between adjacent phosphorous atoms along each strand (average=6.7A) are also very similar.

Figure 5 compares base stacking interactions involving the 4.21 base pair for the native sequence  $d(\text{CGCGAATTCGCG})$  and the A.G dodecamer  $d(\text{CGCAAATTGGCG})$ . In general the base overlaps for the steps involving the  $A$ (*anti*). $G$ (*syn*) mismatch are similar to those for the  $G$ (*anti*). $A$ (*syn*) (13) and the  $I$ (*anti*). $A$ (*syn*) base pairs (24) in the same base stacking environment. In all these structures, the *syn*-purine at position 21 displays slightly enhanced overlap with thymine at position 20 compared to the native structure. Likewise, the presence of a purine-purine step (21-p-22) allows substantial overlap between the five membered rings of the *syn*-purine (G21 or A21) and the *anti*-purine (G22). Such an interaction is not possible in the native dodecamer and a smaller pyrimidine-purine overlap is observed. An increase in base-base overlap does not necessarily reflect greater stability, as the relative contributions and precise nature of the various forces involved in base stacking are not understood. Indeed, duplexes containing purine-purine mismatches are considerably less stable than the corresponding native duplexes (5-8).

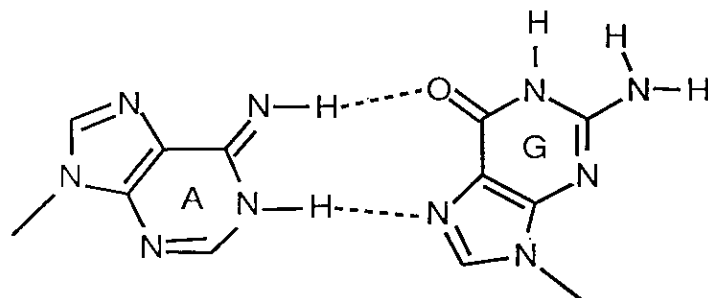
#### Sequence dependence of A.G mismatch conformation

It has recently been shown by nuclear magnetic resonance techniques that the G.A mismatch in the sequence  $d(\text{CGGGAATTCACG})$  displays conformational flexibility as a function of pH (11). The  $G$ (*anti*). $A$ (*anti*) base pair is observed at neutral pH whereas the  $G$ (*syn*). $AH^+$ (*anti*) base pair predominates below pH 5.5. Our earlier crystallographic analysis of a G.A mismatch in the sequence  $d(\text{CGCGAATTAGCG})$  (13) showed the presence of a  $G$ (*anti*). $A$ (*syn*) pairing. The current structure, that of the dodecamer  $d(\text{CGCAAATTGGCG})$  with  $G$ (*syn*). $A$ (*anti*) base pairs is related to the previous sequence by interchanging the adenine and guanine bases of the G.A

mismatches thus changing the base stacking environment. Hence it is clear that base stacking interactions are influential in determining the form of conformationally flexible mismatches. In both DNA duplexes that have been shown to accommodate  $G$ (*syn*). $A$ (*anti*) base pairs,  $d(\text{CGCAAATTGGCG})$  and  $d(\text{CGGGAATTCACG})$ , the guanine base of the mismatch is flanked by the guanine or guanines of G.C base pairs. Guanine has a large dipole moment (7.5D) (25) and an intrastrand GpG step in B-DNA involving  $G$ (*anti*) bases would be expected to have



2a



2b

Figure 2. 2a. The  $AH^+$ (*anti*). $G$ (*syn*) base pair. 2b. The  $A$ (*imino,anti*). $G$ (*syn*) base pair.

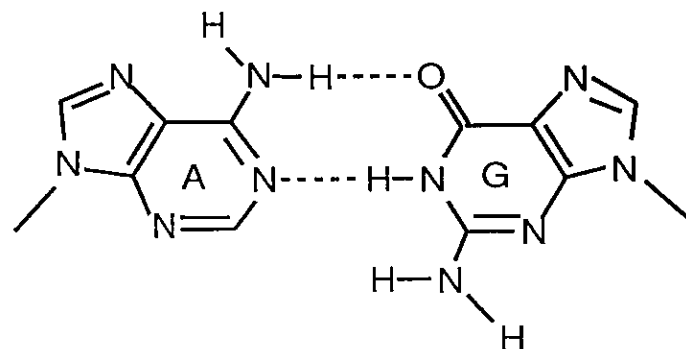
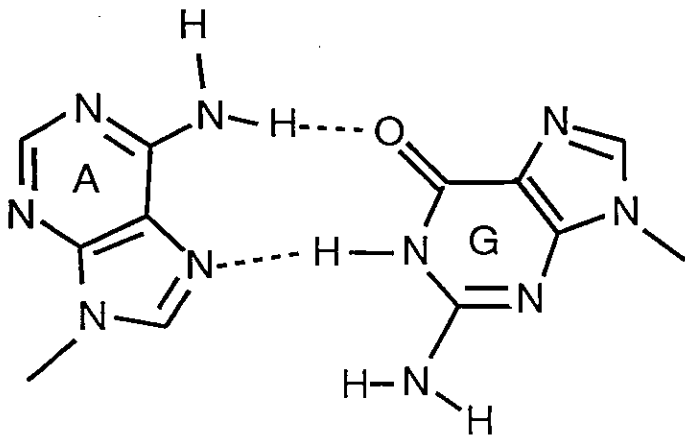
A(*anti*). $G$ (*anti*) mismatch

Figure 3. The  $A$ (*anti*). $G$ (*anti*) base pair.



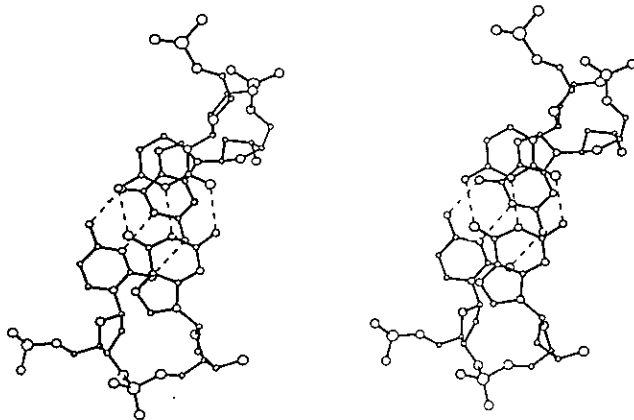
A(syn).G(anti) mismatch

Figure 4. The A(syn).G(anti) base pair.

an unfavourable base stacking interaction. This is because the dipole moments of the two bases would be closely aligned, with a relative rotation of only ca. 33°. However, with the guanine base of the G.A mismatch in the *syn*-conformation, the static dipole-dipole interactions are likely to be less unfavourable.

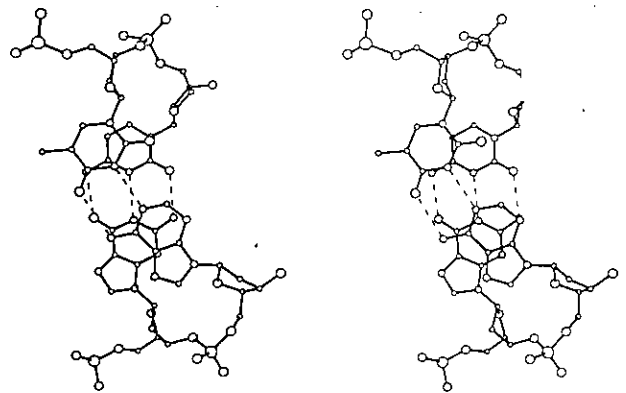
In the sequences d(CGAGAATTCGCG) (26,27) and d(CGCGAATTAGCG) (13), the guanine bases of the G.A mismatches have been shown to be *anti*, (figure 3, figure 4). In most cases the mispaired guanine bases are involved in CpG intrastrand stacking interactions. The dipole moment of cytosine is almost identical to that of guanine (7.6D) and it lies in almost exactly the opposite direction. Hence, this should produce a favourable CpG(*anti*) base stacking interaction. In the dodecamer d(CGAGAATTCGCG) the mismatched guanine is flanked by two cytosines, so the situation should be even more favourable. We have shown that the  $T_m$  of this sequence at pH 7.0 is 35°C, much higher than that of d(CGCAAATTCGCG), ( $T_m$ =19°C). In the former case there is no indication of increased duplex stability with decreasing pH, indicating that the G(*syn*).AH<sup>+</sup>(*anti*) mispair probably does not occur. Hence it seems that

A

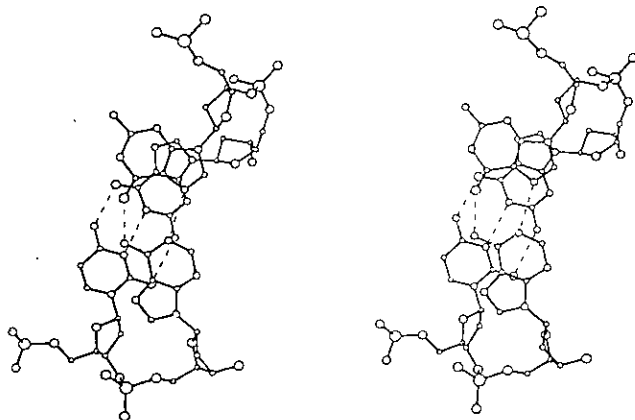


Native dodecamer. C3.G22 on G4.C21

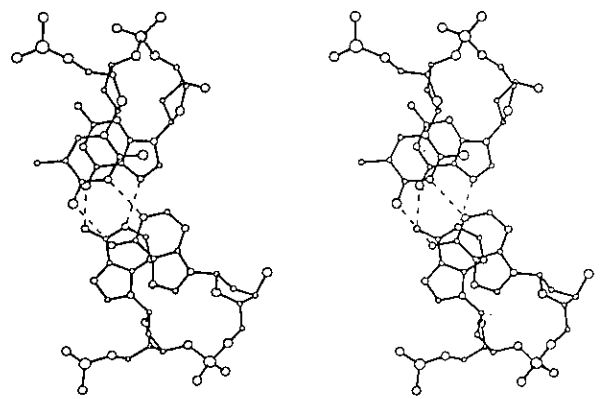
B



Native dodecamer. G4.C21 on A5.T20



A.G dodecamer. C3.G22 on A4.G21

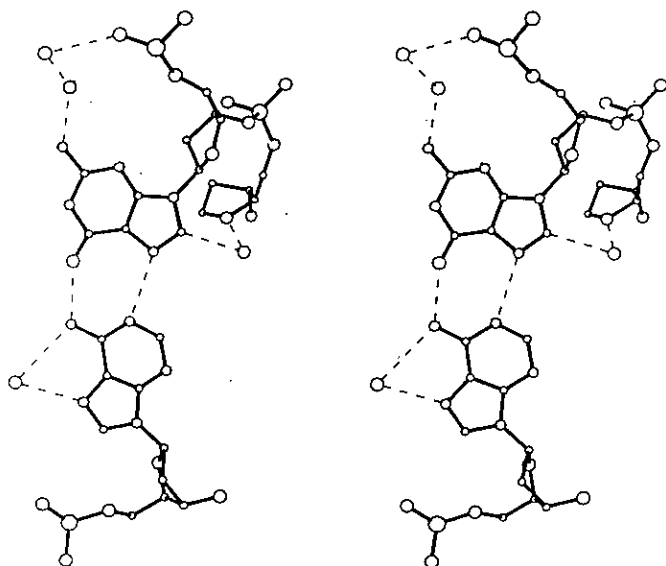


A.G dodecamer. A4.G21 on A5.T20

Figure 5. A comparison of stacking interactions around base pair 4.21 for the native dodecamer d(CGCGAATTCGCG)<sub>2</sub> and the A.G mismatch dodecamer, d(CGCAAATTCGCG)<sub>2</sub>.

**Table 3.** Propellor Twist in the base pairs in d(CGCAAATTGGCG)

Base pair	Propellor twist (degrees)
C1.G24	11.8
G2.C23	11.8
C3.G22	8.9
A4.G21	14.4
A5.T20	16.9
A6.T19	19.1
T7.A18	21.9
T8.A17	20.1
G9.A16	16.8
G10.C15	7.7
C11.G14	23.0
G12.C13	7.2

**Figure 6.** The network of water molecules stabilising the AH<sup>+</sup>(*anti*).G(*syn*) base pair.

the that the *anti*-conformation is preferred for a mismatched guanine base stacked on a cytosine. The high  $T_m$  of this duplex at neutral pH demonstrates the greater stabilising effect of CpG(*anti*) and G(*anti*)pC steps relative to the G(*anti*)pG step for the G.A mismatch and illustrates the dramatic effect of base stacking environment on the stability of mismatch-containing DNA duplexes.

Static dipole moments of free bases are likely to be important as the two single strands interact to initiate duplex formation, when intra-strand stacking forces can be optimised. Once the duplex is formed, any change in the conformation of the G.A mismatch will necessitate rotation of a purine base around its glycosidic bond, an operation that will require considerable disruption of the surrounding Watson-Crick base pairs. Hence, reformation of the mismatch base pair will again be influenced by the dipole moments of the unpaired bases in the vicinity of the mispair.

### Bifurcated hydrogen bonds

A network of bifurcated hydrogen bonds appears to run through the central part of the major groove of the d(CGCAAATTGGCG) duplex. In this respect the mismatch duplex behaves like the native

**Table 4a.** Thermodynamic parameters for d(CGCAATTGGCG)<sub>2</sub> duplexes at pH 7.0 (1M NaCl, 10mM phosphate, 1mM EDTA)

X.Y	$\Delta H^\circ$ Kcal/mole	$\Delta S^\circ$ cal/mole/deg K	$\Delta G^\circ$ Kcal/mole	$T_m$ degK (40 $\mu$ M)
G.C	-102.9	-278.4	-19.9	344.2
A.T	-95.4	-259.5	-18.1	341.6
A.C	-48.6	-139.4	-7.0	298.0
G.T	-79.2	-224.3	-12.5	324.6
A.G	-66.9	-189.9	-10.3	319.0

**Table 4b.** Thermodynamic Parameters for d(CGCAATTGGCG)<sub>2</sub> duplexes at pH 5.6 (1M NaCl, 10mM phosphate, 1mM EDTA)

X,Y	$\Delta H^\circ$ Kcal/mole	$\Delta S^\circ$ cal/mole/degK	$\Delta G^\circ$ Kcal/mole	$T_m$ degK (40 $\mu$ M)
A.C	-67.0	-191.5	-9.9	317.2
A.G	-51.7	-142.8	-9.2	318.1

**Table 4c.** Thermodynamic parameters for the d(CGCAAATTGGCG)<sub>2</sub> duplex (0.1M sodium phosphate buffer, 1mM EDTA)

pH	$\Delta H^\circ$ Kcal/mole	$\Delta S^\circ$ cal/mole/degK	$\Delta G^\circ$ Kcal/mole	$T_m$ degK (40 $\mu$ M)
7.0	-32.3	-85.3	-6.9	307.5
5.5	-54.9	-153.9	-9.1	316.2

A.T sequence, d(CGCAAATTGGCG) (28) and the related non-complementary duplex examined by Nelson *et al.* (29). These bifurcated hydrogen bonds can form because of the high propeller twist of the core base pairs (average = 18.5°), (table 3). This has the effect of tilting the N(6)-atom of an adenine base towards the O(4)-atom of the thymine base on the 3'-side (table 3). Similar close contacts are observed between the N(4)-amino group of each cytosine of base pairs C3.G22 and G10.C15 and the O(6)-atom of the guanine base of the A.G mismatch on its 3'-side. The putative bifurcated hydrogen bonds are rather long, averaging 3.17Å, and assuming that the amino groups of the heterocyclic bases have trigonal geometry, the N-H...O angles are generally much smaller than the ideal of 180°. The net result will be very weak hydrogen bonds which are unlikely to make a major contribution to duplex stability. Such interactions are probably best regarded as non-specific electrostatic forces in the major groove of the B-DNA duplex.

### The Solvent Environment

The B-values of the 62 solvent molecules located during the refinement range from 20Å<sup>2</sup> to 80Å<sup>2</sup>. The phosphate groups are individually hydrated as expected for B-DNA, and 17 solvent molecules interact directly with the sugar-phosphate backbone. A total of 10 solvent molecules are bound to the functional groups of the bases in the major groove and only 7 are bound directly to the minor groove. The minor groove spine of hydration, which is characteristic of the native dodecamer, was not found in the present structure and there is no discernible pattern of hydration in either groove. The remaining 28 solvent molecules are involved in solvent-solvent interactions and represent fragments of the second shell of hydration. It would have been possible to include many more solvent molecules during the refinement by relaxing the criteria for good hydrogen bonding geometry and distance,

but this would have resulted in an artificial lowering of the *R*-factor.

The most interesting feature of the hydration of the A.G mismatch duplex occurs in the vicinity of the AH<sup>+</sup> (*anti*).G(*syn*) base pairs. There is a well ordered water structure around the guanidine nucleotide which appears to stabilise the *syn*-conformation. The 2-amino group of guanine, which protrudes into the major groove, is linked to its 5'-phosphate group by a solvent bridge. In the minor groove, the guanine H(8)-atom forms a weak hydrogen bond to a water molecule which is also H-bonded to the O(4)'-atom of the deoxyribose sugar of the 3'-cytidine nucleotide (figure 6). A similar water structure was found around the *syn*-guanine in the crystal structure of Z-DNA (20).

### Duplex stability

The A.G mismatch duplex is much less stable than Watson-Crick duplexes in 1M NaCl at pH7.0 (table 4a), with a  $\Delta G^\phi$  value of  $-10.3$  kcal/mole compared with  $-19.9$  kcal/mole for the G.C dodecamer and  $-18.1$  kcal/mole for the A.T dodecamer. Hence, despite the presence of two hydrogen bonds, the A.G mismatch base pair has a significant destabilising effect on the B-DNA duplex. Under the above conditions the A.G duplex is less stable than the G.T mismatch duplex ( $\Delta G^\phi = -12.5$  kcal/mole) despite the fact that G.T mismatches are generally repaired more efficiently than A.G mismatches *in vivo*. The A.G mismatch duplex and the A.C mismatch duplex both contain protonated mismatch base pairs, and the A.C mismatch duplex is less stable than the A.G duplex in high salt conditions at pH 7.0 ( $\Delta G^\phi = -7.0$  kcal/mole).

The thermodynamic parameters for the A(*anti*).G(*syn*) base pair in 0.1M buffer (table 4b) confirm that there is a significant increase in stability when going from pH 7.0 to pH 5.5 ( $\Delta \Delta G^\phi = -2.2$  kcal/mole). This pH-dependent increase in stability disappears in 1.0M salt, so the A.G mismatch behaves differently from the A.C mismatch, which has increased stability at acidic pH in both low and high salt conditions. This difference in behaviour between the A.G mismatch and the A.C mismatch is presumably due to the fact that the latter has only one stable form and cannot exhibit conformational flexibility.

The  $T_m$  values in table 4 are those determined at a single strand concentration of 40  $\mu$ molar as part of a concentration-dependent study, whereas those quoted previously were determined during pH-dependent studies at a strand concentration of 5  $\mu$ molar (9).

### CONCLUSIONS

The A.G mismatch is conformationally variable and the relative populations of the various conformers depends on pH, base stacking environment and salt concentration. The AH<sup>+</sup> (*anti*).G(*syn*) base pair is present to a significant degree below pH 6.5 in sequences where the mispaired guanine is flanked by the guanine base of a G.C base pair. The base pair is stabilised by protonation of adenine-N(1) and the *syn*-conformation of the G-nucleotide is stabilised by surrounding solvent molecules. The disposition of hydrogen bond donor and acceptor groups in the major and minor groove at the A(*anti*).G(*syn*) site is completely different from that of the A(*anti*).G(*anti*) and the A(*syn*).G(*anti*) base pairs. This may be important when considering interactions with mismatch repair enzymes. Above all, the present work shows that the interaction

between a guanine and an adenine base in B-DNA is rather complex, and simplistic interpretations of the relationship between base pair structure and mismatch repair frequency are likely to be misleading.

Refined atomic co-ordinates have been deposited with the Cambridge Crystallographic Data Centre, University Chemical Laboratory, ????????? Road, Cambridge CB2 1EN, and are available on request.

### ACKNOWLEDGEMENTS

We wish to thank the SERC for providing funding under the molecular recognition initiative.

### REFERENCES

- Friedberg, E.C. (1985). *DNA Repair*. W.H.Freeman & Co., New York.
- Steitz, T.A., Beeze, L., Freemont, P.S., Friedman, J.M. & Sanderson, M.R. (1987). *Cold Spring Harbour Symp. Quant. Biol.* **52**, 465-471.
- Patel, D.J., Shapiro, L. & Hare, D. (1987). In *Nucleic Acids and Molecular Biology*. (Eckstein, F. & Lilley, D. eds.) Vol 1. pp70-84. Springer Verlag, Berlin, Heidelberg, New York.
- Kennard, O. *ibid.* pp25-52.
- Aboul-ela, F., Koh, D. & Tinoco, I. Jr. (1985). *Nucleic Acids Res.* **13**, 4811-4824.
- Tibanyenda, M., DeBruin, S.H., Hasnoot, C.A.G., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1984). *Eur. J. Biochem.* **139**, 19-27.
- Gralla, J., & Crothers, D.M. (1973) *J. Mol. Biol.* **78**, 301-319.
- Werntges, H., Steger, G., Riesner, D. & Fritz, H.J. (1986) *Nucleic Acids Res.* **14**, 3773-3790.
- Brown, T., Leonard, G.A., Booth, E.D. & Kneale, G. (1990). *J. Mol. Biol.* **212**, 437-440.
- Fersht, A.R., Knill-Jones, J.W. & Tsui, W.C. (1982). *J. Mol. Biol.* **156**, 37-51.
- Gao, X. & Patel, D.J. (1988). *J. Amer. Chem. Soc.* **110**, 5178-5182.
- Brown, T., Leonard, G.A., Booth, E.D. & Chambers, J. (1989). *J. Mol. Biol.* **207**, 455-457.
- Brown, T., Hunter, W.N., Kneale, G. & Kennard, O. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2402-2406.
- Wing, R., Drew, H.R., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1980). *Nature*. **287**, 755-758.
- Hendrickson, W.A. & Konnert, J.H. (1981). in *Biomolecular Structure, Conformation, Function and Evolution*, Ed: Srinivasan, R., Pergamon Press, Oxford. VOLL, 43-57.
- Westhof, E., Dumas, P. & Moras, D. J. (1985). *J. Mol. Biol.* **184**, 119-145.
- Jones, T.A. (1978). *J. Applied. Cryst.* **11**, 268-272.
- Sheldrick, G.M. (1976). Shelx 76 system of Computing Programs, Univ. Cambridge, UK.
- Prive, G.G., Heinemann, U., Kan, L-S., Chandrasegaran, S. & Dickerson, R.E. (1987). *Science*. **238**, 498-504.
- Frederick, C.A., Wang, A.H-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G.A. & Rich, A. (1979). *Nature*. **286**, 680-686. 17872-17879.
- Frederick, C.A., Quigley, G.J., van der Marel, G.J., van Boom, J.H., Wang, A.H-J. & Rich, A. (1988). *J. Biol. Chem.* **263**, 17872-17879.
- A.N. Lane, & T.Brown. manuscript in preparation.
- Topal, A.D. & Fresco, J.R. (1976). *Nature*. **263**, 285-289.
- Corfield, P.W., Hunter, W.N., Brown, T., Robinson, P. & Kennard, O. (1987). *Nucleic Acids Res.* **15**, 7935-7949.
- Pullman, B. & Pullman, A. (1971). *Advan. Het. Chem.* **13**, 77-159.
- Patel, D.J., Koslowski, S.A., Ikuta, S. & Itakura, K. (1984). *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* **43**, 2663-2670.
- Patel, D.J., Koslowski, S.A., Ikuta, S. & Itakura, K. (1984). *Biochemistry*. **23**, 3207-3217.
- Coll, M., Frederick, C.A., Wang, A.H-J. & Rich A. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8385-8389.
- Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987). *Nature*. **330**, 221-226.