Studies on Chemically Modified Oligonucleotides and on DNA Triplexes

by

Ana-Ariana Moraru Allen

A thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

1997
Declaration

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

Acknowledgements 6

List of Abbreviations 7

Abstract 11

Part 1. Studies on Chemically Modified Oligonucleotides:
Modified DNA as a Substrate for the DNA Repair Enzyme
O\(^6\)-Methylguanine DNA Methyl Transferase 12

1. Introduction 13

1.1 Introduction to DNA 13

1.2 DNA damage 15

1.3 Carcinogens 16

1.3.1 Physical carcinogens 16

1.3.2 Viral agents 16

1.3.3 Chemical carcinogens 17

1.4 Alkylating agents and their interaction with DNA 17

1.4.1 Sites of DNA alkylation 17

1.4.2 Categories of alkylating agents 19

1.4.3 The Activation \textit{in vivo} of secondary carcinogens 20

1.4.4 Products of DNA alkylation and their effects 23

1.5 O\(^6\)-Alkylguanine lesions 29

1.5.1 Biological significance 29

1.5.2 Consequences of O\(^6\)-guanine alkylation on DNA structure 29

1.5.3 Stability of DNA containing O\(^6\)-alkylguanine 32
1.6 DNA repair

1.6.1 Nucleotide excision repair

1.6.2 Base excision repair

1.6.3 Direct reversal of damage to bases

1.6.4 Post-replicative repair

1.7 O\textsuperscript{6}-methyl-G-transferase

1.7.1 Mechanism of O\textsuperscript{6}-methylguanine demethylation

1.7.2 Structure and putative mode of action of O\textsuperscript{6}-methyl-G-transferase

1.7.3 Regulation of O\textsuperscript{6}-methyl-G-transferase expression

1.8 Aims

2. Results and Discussion

2.1 Structure of a phosphoramidite

2.2 Synthesis of N\textsuperscript{2}-acetyl-O\textsuperscript{6}-alkyl-dG

2.3Synthesis of O\textsuperscript{6}-pentafluorophenyl-N\textsuperscript{2}-trifluoroacetyl-2'-deoxy-guanosine

2.4 Synthesis of phosphoramidites from the modified deoxyguanosine derivatives

2.5 Oligonucleotide synthesis

2.5.1 Short history

2.5.2 Solid support DNA synthesis by the phosphoramidite method

2.6 Deprotection and postsynthetic modification of oligonucleotides

2.7 Optimization of the deprotection time, purification, enzymatic digestion and CZE analysis of oligonucleotides

2.7.1 Deprotection and HPLC purification of the O\textsuperscript{6}-alkyl-G oligonucleotides

2.7.2 Enzymatic digestion of the O\textsuperscript{6}-alkyl-G oligonucleotides

2.7.3 CZE analysis of the O\textsuperscript{6}-alkyl-G oligonucleotides

2.7.4 Deprotection, HPLC purification, enzymatic digestion and CZE analysis of the PFPG / N\textsuperscript{6}-methyldiaminopurine oligonucleotides

2.8 UV melting (thermal denaturing)
3. Experimental

3.1 Chemicals and solvents
3.2 Thin layer chromatography (TLC)
3.3 Wet flash column chromatography
3.4 Instrumentation
3.5 Oligonucleotide synthesis
3.6 HPLC analysis and purification of oligonucleotides
3.7 Enzymatic digestion of oligonucleotides
3.8 Sephadex gel-filtration (desalting) of oligonucleotides
3.9 Capillary zone electrophoresis (CZE)
3.10 UV melting (thermal denaturing) studies
3.11 Synthetic chemistry

4. Conclusions

5. References

Part 2. Studies on DNA Triplexes

1. Introduction

1.1 Nucleic acid triple helices: history and structure
1.1.1 Discovery
1.1.2 Evidence for triple helix formation
1.1.3 Types of nucleic acid triplexes
1.1.4 Triplexes of the pyrimidine motif
1.1.5 Triplexes of the purine motif
1.1.6 Fine structure of triple helices. A-DNA versus B-DNA conformation
1.2 Types of DNA triplex structures found in vivo and their potential biological role
1.2.1 H-DNA
1.2.2 R-DNA
1.2.3 Detection of triplex DNA structures in vivo
1.2.4 Implication of triplexes in biological processes
1.3 Applications of DNA triplexes
1.3.1 Specificity of triple helix formation
1.3.2 The stabilisation of DNA triple helices
1.3.3 Therapeutic applications
1.3.4 DNA triplexes as a tool in molecular biology
1.4 Aims

2. Results and Discussion
2.1 Synthesis of foldback triplex forming oligonucleotides
2.1.1 Synthesis of the linker phosphoramidite monomers
2.1.2 Oligonucleotide synthesis
2.2 UV melting
2.2.1 Preliminary experiments
2.2.2 UV melting of the (T-A-T)₈ triplexes in the presence of NaCl
2.2.3 UV melting of the mixed sequence triplexes in the presence of NaCl
2.2.4 UV melting behaviour of T-A-T triplexes compared to that of their corresponding duplexes in the presence of MgCl₂
2.2.5 UV melting behaviour of the mixed sequence triplexes compared to that of their corresponding duplexes in the presence of MgCl₂
2.2.6 UV melting temperatures of DNA triplex-coralyne complexes
2.3 Attempts to crystallise DNA foldback triplexes
2.4 Results from CD and NMR studies carried out in collaboration
3. Experimental

3.1 Chemicals and solvents
3.2 Thin layer chromatography (TLC)
3.3 Wet flash column chromatography
3.4 Instrumentation
3.5 Oligonucleotide synthesis
3.6 HPLC analysis and purification of oligonucleotides
3.7 Sephadex gel-filtration (desalting) of oligonucleotides
3.8 Capillary zone electrophoresis (CZE)
3.9 Crystallisation experiments
3.10 UV melting (thermal denaturing) studies
3.11 Synthetic chemistry

4. Conclusions

5. References

Appendix - Publication
Acknowledgements

I am indebted to:

- my supervisor Prof. Tom Brown for his guidance and encouragement,
- the Wellcome Trust for the funding of my course,
- the Committee of Vice-Chancellors and Principals for the partial funding of my course through the Overseas Research Students' Award Scheme,
- the OSWEL DNA Service for the use of their chemicals and facilities,
- Edinburgh University for the use of their facilities, including CHN analysis and NMR and mass spectrometry services,
- my husband Keith for his support and advice regarding wording.
Abbreviations

A  adenine
Ac₂O  acetic anhydride
AcOH  acetic acid
Anal.  microanalysis
AP  apurinic/apyrimidinic (sites)
Arg  arginine
BCNU  1,3-Bis-(2-chloroethyl)-N-nitrosourea
bs  broad singlet
C  cytosine
CAA  chloroacetaldehyde
calcd.  calculated
ccod  cacodylate
CD  circular dichroism
CEO  2-chloroethylenoxide
Cys  cysteine
CZE  capillary zone electrophoresis
d  doublet
dA  2'-deoxyadenosine
DAP  2,6-diaminopurine
DBU  1,8-diazabicyclo[5.4.1]undec-7-ene
dC  2'-deoxycytidine
DCM  dichloromethane
dG  2'-deoxyguanosine
DIPEA  diisopropylethylamine
DMAP  4-dimethylaminopyridine
DMSO  dimethyl sulphoxide
DMTr  4,4'-dimethoxytrityl
DNA   deoxyribonucleic acid
dT    2'-deoxythymidine
εA    1,N6-ethenoadenine
εC    3,N4-ethenocytosine
εG    N2,3-ethenoguanine
EDTA  ethylenediaminetetraacetic acid
eq.   equivalents
EtOAc ethylacetate
EtOH  ethanol
FAB   fast atom bombardment
FTFO  foldback triplex forming oligonucleotide
G     guanine
G^Et  O6-ethylguanine
G^Me  O6-methylguanine
G^{NHMe}  N6-methyldiaminopurine
HEG   hexaethyleneglycol
His   histidine
HIV   human immunodeficiency virus
HPLC  high performance liquid chromatography
hr(s) hour(s)
HS    Hoogsteen
m     multiplet
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeCN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>MeOH</td>
<td>methanol</td>
</tr>
<tr>
<td>MGMT</td>
<td>O(^6)-methylguanine DNA methyltransferase</td>
</tr>
<tr>
<td>min(s)</td>
<td>minute(s)</td>
</tr>
<tr>
<td>MNNG</td>
<td>N-methyl-N'-nitro-N-nitrosoguanidine</td>
</tr>
<tr>
<td>MNU</td>
<td>N-methyl-N-nitrosourea</td>
</tr>
<tr>
<td>MPD</td>
<td>2-methyl-2,4-pentanediol</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MSCI</td>
<td>mesitylenesulphonyl chloride</td>
</tr>
<tr>
<td>N(^6)-MeDAP</td>
<td>N(^6)-methyldiaminopurine</td>
</tr>
<tr>
<td>NAP</td>
<td>nucleic acid purification</td>
</tr>
<tr>
<td>NEt(_3)</td>
<td>triethylamine</td>
</tr>
<tr>
<td>NH(_4)OAc</td>
<td>ammonium acetate</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>O(^6)-EtdG</td>
<td>O(^6)-ethyl-2'-deoxyguanosine</td>
</tr>
<tr>
<td>O(^6)-EtG</td>
<td>O(^6)-ethylguanine</td>
</tr>
<tr>
<td>O(^6)-MedG</td>
<td>O(^6)-methyl-2'-deoxyguanosine</td>
</tr>
<tr>
<td>O(^6)-MeG</td>
<td>O(^6)-methylguanine</td>
</tr>
<tr>
<td>O(^6)-PFPG</td>
<td>O(^6)-pentafluorophenylguanine</td>
</tr>
<tr>
<td>Ocd</td>
<td>1,8-octanediol</td>
</tr>
<tr>
<td>oxet-G</td>
<td>7-(2-oxoethyl)guanine</td>
</tr>
<tr>
<td>PFTF(_d)G</td>
<td>O(^6)-pentafluorophenyl-N(^2)-trifluoroacetyl-2'-deoxyguanosine</td>
</tr>
<tr>
<td>PFTFG</td>
<td>O(^6)-pentafluorophenyl-N(^2)-trifluoroacetylguanine</td>
</tr>
<tr>
<td>prep</td>
<td>preparative</td>
</tr>
<tr>
<td>Pro</td>
<td>proline</td>
</tr>
</tbody>
</table>
pur  purine
Py   pyridine
pyr  pyrimidine
R_f  relative to front
rHS  reverse Hoogsteen
RNA  ribonucleic acid
s    singlet
SAM  S-adenosylmethionine
t    triplet
T    thymine
TCA  trichloroacetic acid
TEA  triethylamine
TFA  trifluoroacetic anhydride
TFO  triplex forming oligonucleotide
THF  tetrahydrofuran
TLC  thin layer chromatography
T_m  UV melting temperature
tRNA transfer RNA
UV   ultraviolet
W/C  Watson/Crick
Abstract

Studies on chemically modified oligonucleotides: Modified DNA as a substrate for the DNA repair enzyme O^6^-methyl-G-transferase

The interactions between modified oligonucleotides and DNA binding proteins can provide important information on the mode of binding and the structure of the enzyme-substrate complex. In collaboration with Dr M. Moore at York University studies were carried out to determine the interaction of DNA with O^6^-methyl-G-transferase, a DNA repair enzyme which removes the alkyl group from guanine residues which are alkylated on the O^6^-position.

A series of modified oligonucleotides have been synthesised containing O^6^-methylguanine, O^6^-ethylguanine and N^6^-methyl-2,6-diaminopurine. The oligonucleotides were obtained in high purity. UV melting (thermal denaturing) experiments have been carried out on duplexes containing the modified bases paired with cytosine and thymine respectively. These oligonucleotides were prepared in order to study their interactions with a mutant O^6^-methyl-G-transferase enzyme lacking DNA repair activity. Gel shift experiments suggest that the above oligonucleotides are recognised by the enzyme. Attempts were made to crystallise the enzyme-DNA complex.

Studies on DNA Triplexes

Triplex DNA is formed by the specific binding of a DNA strand in the major groove of a preformed DNA duplex. It has been suggested that such structures might have a biological role, especially in gene regulation, and that there is potential for therapeutic applications in which gene expression is repressed by triplex formation.

One difficulty encountered in the studies of DNA triplexes is their low thermodynamic stability. In order to overcome this problem oligonucleotides with appropriate triplex forming sequences have been synthesised with the three strands linked by hexaethyleneglycol or 1,8-octanediol chains. These molecules are capable of folding back on themselves to give intramolecular triplexes with significantly increased stabilities compared to intermolecular triplexes.

UV melting studies have been carried out to determine the thermal stability of these triplexes, in the presence and in the absence of the DNA triplex binding drug coralyne. Circular dichroism and NMR studies were carried out in collaborarion with Dr A. Lane at the National Institute of Medical Research, London.
Part 1. Studies on Chemically Modified Oligonucleotides: Modified DNA as a Substrate for O6-Methylguanine DNA Methyl Transferase
1. Introduction

1.1 Introduction to DNA

The elucidation of the DNA secondary structure by Watson and Crick (1953) is a historical landmark in modern biochemistry. DNA has the shape of a double helix with antiparallel strands which consist of a phosphodiester backbone and four types of heterocyclic bases - adenine (A), cytosine (C), thymine (T) and guanine (G) - linked to the backbone by a sugar ring, which is β-D-2'-deoxyribose (Figure 1).

![Chemical Structure of the DNA bases](image)

**Figure 1.** Chemical Structure of the DNA bases
Adenine and guanine are purines, and cytosine and thymine are pyrimidines. The basic monomeric unit of DNA is called a nucleotide. The nucleotides are joined together by 3'→5' phosphodiester linkages, which gives the nucleic acid chains a directional nature.

The two strands of the DNA double helix are held together by the hydrogen bonds between the bases. Only two types of base pairs involving the major tautomeric forms of the bases can occur: guanine always pairs with cytosine and adenine with thymine (Figure 2), which explains the findings of Chargaff and co-workers (Zamenhof et al., 1952) that the guanine to cytosine and adenine to thymine ratios were always very close to 1:1, irrespective of the species the DNA was obtained from. As a consequence, the sequence of bases in one strand determines the sequence in the other - complementary - strand. A schematic representation of the DNA double helix is shown in Figure 3 below.

![Figure 2. GC and AT Watson-Crick base pairs](image)

![Figure 3. Schematic representation of the DNA double helix](image)
DNA is the carrier of genetic information for all cellular organisms and for some categories of viruses. The genome size varies between a few thousand base pairs in some viruses to 2.9 million in humans, and even more in some other organisms. It undergoes replication, in which each parent strand serves as a template for the synthesis of a complementary daughter strand by DNA polymerases, thus ensuring that following cell division the daughter cells will have the same amount of genetic material as the parent cell. The genetic information contained in DNA is expressed during protein synthesis, when the DNA template is transcribed to give mRNA with a complementary sequence, which is in turn translated with the help of tRNA and ribosomes to synthesise proteins.

### 1.2 DNA Damage

The DNA sequence is crucial for protein synthesis, as one single modification may lead to a wrong amino acid being incorporated into a protein or, in some cases, to premature termination of the amino acid chain, both of which could give rise to a reduction in the activity or to the inactivation of that particular protein, and which are often detrimental to the cell. Therefore, high fidelity of DNA replication is essential if it is to maintain its capability of encoding for ‘correct’ proteins.

However, DNA has limited chemical stability. Hydrolysis, oxidation and non-enzymatic methylation of DNA occur at a significant rate in vivo (Lindahl, 1993), and are counteracted by specific DNA repair processes. The spontaneous decay of DNA is likely to be a major factor in mutagenesis, carcinogenesis and ageing (Lindahl, 1993). In addition, DNA is subject to spontaneous mutations, which usually arise from the mispairing of bases during replication. This process, however, has a very small error rate; this ranges generally from one misincorporation in $10^8$ to one in $10^{11}$ replicated bases (Drake, 1969).

Exposure to environmental factors, such as radiation, various chemicals and even viruses, induces damage in the DNA, which might lead to cancer and ultimately to death.
1.3 Carcinogens

A mutation is defined as one heritable change. Agents that interact with DNA and result in damage to the molecule, are referred to as mutagens. If they cause mutations which alter normal cell function and give rise to cancer, they are called carcinogens.

Carcinogens are responsible for the change from a normal to a malignant cell pattern. They can be physical carcinogens, viral agents or chemical carcinogens. Since the discoveries of the first naturally-occurring chemical carcinogen (tobacco for snuff taking) around the middle of the 18th century and the first industrially-generated chemical carcinogen (a component of coal tar) more than sixty years ago, awareness of the human risk of cancer arising from exposure to environmental agents has increased dramatically.

1.3.1 Physical Carcinogens

Physical carcinogens include ionising radiation, X-rays and UV light, which lead to chemical modifications of DNA. Ionising radiation generates hydroxyl radicals, which convert thymine into hydroperoxides and cause imidazole ring opening in adenine and guanine. In addition, it leads to cleavage of the phosphodiester backbone with the formation of single strand breaks (SSB) and double strand breaks (DSB), which seem to have greater damaging effect than the reactions at the base residues (Blackburn and Kellard, 1986c). Exposure to UV light leads to the formation of pyrimidine photodimers, pyrimidine photohydration, crosslinking of DNA to protein and photochemical addition of chemicals to DNA bases.

Asbestos fibres and wood dust also belong to this category, and are thought to work via irritation or physical abrasion of the cell (Blackburn and Kellard, 1986a).

1.3.2 Viral Agents

Viruses such as Epstein-Barr virus (EBV) and Herpes Simplex Virus Type II (HSV-II) have been proven to cause certain types of cancer, and Human T-cell Leukaemia Virus (HTLV) is the cause of some types of leukaemia. It is likely that other viruses are also involved in carcinogenesis, the papilloma viruses being prime suspects (Blackburn and Kellard, 1986a).
1.3.3 Chemical Carcinogens

Chemicals are by far the greatest threat in causing cancer. Chemical carcinogens are compounds that interact directly or indirectly with the cellular target molecule, which is usually DNA, RNA or proteins. They can be alkylating agents, polycyclic aromatic hydrocarbons, N-aryl compounds, halocarbons or hormones (Blackburn and Kellard, 1986a,b,c). If they combine spontaneously with DNA, they are called primary or ultimate carcinogens; these are typically electrophilic alkylating agents. Most chemical carcinogens, however, require a structural change which can occur either spontaneously under physiological conditions, or through oxidative metabolic activation by enzymes involved in detoxification pathways (Mitra and Kaina, 1993); such compounds are called secondary carcinogens or precarcinogens.

In addition, there are chemicals which are not carcinogenic themselves, but they promote the carcinogenic activity of other compounds; these are called co-carcinogens.

Chemical carcinogens which affect nucleic acids bind to the bases and cause a change in the normal Watson-Crick (W/C) base pairing pattern, which, if not repaired, will result in mutations after replication. It has been proposed that mutations may also arise by chemical modification of the DNA precursors in the cell and their subsequent incorporation into DNA (Dodson et al., 1982; Eadie et al., 1984). In vitro experiments have shown that modified nucleotides are much more difficult to incorporate into DNA than unmodified ones. Consequently, this mechanism may not be as important as the direct modification of DNA itself; however, its relevance is still to be determined (Saffhill et al., 1985).

1.4 Alkylating Agents and Their Interactions with DNA

1.4.1 Sites of DNA Alkylation

DNA is considered the most important target of alkylating compounds, since its modification can lead to alterations in gene expression, replication and repair, ultimately resulting in permanent heritable changes - mutations - which may represent the initiation step of the carcinogenic process (Basu and Essigmann, 1988).
DNA offers a wide range of sites for attack by electrophiles: the base residues react at all the ring nitrogens (except the ones involved in the glycosyl bond) and at the oxygen atoms (Figure 4). Chemical (as opposed to biological) DNA alkylation at carbon centres has not been observed (Blackburn and Kellard, 1986a). In addition, alkylation at the oxygen atoms of the phosphodiester bonds by 'hard' electrophiles, occurs to a significant extent. The terms 'hard' and 'soft' are defined in Section 1.4.2.

\[ \begin{align*}
\text{G} & \quad \text{C} \\
\text{A} & \quad \text{T}
\end{align*} \]

\[ \text{dR} = 2'\text{-deoxyribose} \]

\[ \begin{array}{c}
\text{targets for 'hard' electrophiles} \\
\text{targets for 'soft' electrophiles}
\end{array} \]

**Figure 4.** Principal sites of alkylation of the DNA bases
1.4.2 Categories of Alkylating Agents

The majority of alkylating agents are direct electrophiles. Monofunctional alkylating agents (Figure 5a) such as dialkyl sulphate, alkyl alkane sulphonates and epoxides, can only react at one nucleophilic site. Bifunctional alkylating agents (Figure 5b) can alkylate twice, and may crosslink a nucleic acid base either with protein or with another base within the same or the alternate strand, or they may lead to the formation of cyclic DNA adducts. Typical examples are sulphur mustard, nitrogen mustards, bis-chloromethylether and cis-diamminedichloroplumate(II) (cisplatin).

![Diagrams of Monofunctional and Bifunctional Alkylating Agents]

In addition, Pearson and Songstad's (1967) concept of 'hard and soft' acids (electrophilic species) and bases (nucleophilic species) can be applied. This states that hard acids react preferentially with hard bases, and soft acids react preferentially with soft bases. According to this theory, alkylating species have intermediate acidity and the ring nitrogens and exocyclic oxygens of the nucleic acid bases have intermediate basicity, which is why alkylating agents react preferentially at these sites (Saffhill et al., 1985). Moreover, within the intermediate classification, carbonium ions are harder acids than compounds such as alkanesulphonates; therefore alkylating agents are considered 'hard' if their reactions have more $S_N1$ character and 'soft' if they have more $S_N2$ character. The oxygen atoms of both the nucleic acid bases and the phosphodiester backbone are considered harder bases than the ring nitrogens, therefore...
'hard' electrophiles will react more at the oxygens and 'soft' electrophiles at the ring nitrogens (Saffhill et al., 1985).

1.4.3 The Activation in Vivo of Secondary Carcinogens

Simple alkylation agents are generated from secondary carcinogens in vivo either spontaneously or in metabolic processes catalysed by enzymes of the cytochrome P-450 class. Typical examples of secondary carcinogens are represented in Figure 6 below.

![Figure 6. Secondary carcinogens](image)

N-nitroso compounds, which are part of this group, are very widely spread in the environment, both from natural and pollutant sources. They include nitrosamines, nitrosoureas and nitrosoguanidines. Some, such as 1,3-bis-(2-chloroethyl)-N-nitrosourea (BCNU), are used in cancer chemotherapy, but they can themselves give rise to secondary tumors. In addition, nitrosamines can be formed in the stomach by nitrosation of secondary amines, ideally at a pH around 4 to 5. These amines are usually protonated at the normal pH value of 3 in the stomach. However, processes such as ulcer treatment may result in a higher than normal stomach pH when the amines are no longer protonated and can therefore react with the nitrous
acid and nitrites (Figure 7) generated through bacterial degradation of nitrates present in food (Blackburn and Kellard, 1986a; Bartsch and Montesano, 1984). It has been suggested that alcohol consumption may increase the risk of cancer from nitrosamines in primates by suppressing or reducing the detoxification of these compounds by the liver (Anderson et al., 1996), whereas certain organosulphur compounds present in garlic and cruciferous vegetables (Lin et al., 1994; Schaffer et al., 1996), as well as vitamin C (Bartsch and Montesano, 1984; Dyke et al., 1994), may reduce the risk.

![Figure 7. Formation of nitrosamines in the stomach](image)

The activation of nitrosoureas, nitrosoguanidines and methyl phenyltriazole occurs through spontaneous hydrolysis and generates alkyl diazohydroxides. Other compounds require one (Figure 8a) or several (Figure 8b) oxidations by cytochrome P-450 to give the primary carcinogen, which is also alkyl diazohydroxide. The latter compound rearranges to positively charged alkylating species, which is usually a carbocation (Singer and Kusmierek, 1982) (Figure 9).
Figure 8. Metabolic activation of (a) dimethylnitrosamine and (b) N,N'-dimethylhydrazine by cytochrome P-450

Figure 9. Spontaneous rearrangement of methyldiazoiodrooxide
Vinyl chloride, another secondary carcinogen, is a major industrial chemical produced in large quantities around the world. It is also metabolised by enzymes of the cytochrome P-450 class to 2-chloroethylenoxide (CEO), which spontaneously rearranges to chloroacetaldehyde (CAA) (Guengerich et al., 1979), which is less hazardous and is detoxified by reaction with glutathione (Figure 10). Both CEO and CAA are bifunctional alkylating agents.

![Figure 10. Metabolic activation of vinyl chloride](image)

1.4.4 Products of DNA Alkylation and Their Effects

Most alkylation reactions of nucleic acids are methylations. The reactions of methylating and ethylating agents with DNA have been studied extensively both in vitro and in vivo. The mutagenic properties of alkylating agents are thought to arise mainly as a consequence of the modification of the Watson-Crick base pairing pattern (Saffhill et al., 1985), with some sequences being more prone to mutation than others (Basu and Essigmann, 1988). The most common nucleic acid base modifications due to alkylation are illustrated in Figure 11.
Figure 11. The most common DNA base alkylation products
N-Alkylation

N-alkylpurines appear not to be directly mutagenic. However, they can lead to mutagenesis indirectly because their removal, either in spontaneous chemical reactions or during excision repair, leads to apurinic sites which normally create replication blocks, but under some circumstances they can lead to mutations (Mitra and Kaina, 1993).

It should be noted that DNA damage is caused not only by chemicals taken up by the cells from the environment, but also by some endogenous small molecules, such as S-adenosylmethionine (SAM), which serves as methyl donor in most cellular enzymatic transmethylation reactions (Lindahl, 1993). SAM is a weak alkylating agent, and it can methylate proteins and DNA nonenzymatically. It mainly targets ring nitrogens of purine residues, generating mostly two of the modified bases discussed below: 7-methylguanine and 3-methyladenine.

The major alkylation product in DNA is 7-methylguanine. It does not cause mispairing and does not appear to be mutagenic; however it has been implicated in the ageing process on the basis of circumstancial evidence (Mitra and Kaina, 1993). It is poorly repaired and may tend to accumulate in cells, although the lability of the N-glycosyl bond leads to a significant amount of depurination (Lindahl, 1993). However, cross-linking of of two neighbouring guanines at N\(^7\) by a bifunctional alkylating agent such as nitrogen mustard is a lesion which can lead to cell death (Blackburn, 1990).

3-Methyladenine is also generated to a significant extent. Although the methyl group is located in the minor groove where it would not hinder normal base pairing, this lesion blocks DNA replication and is cytotoxic (Lindahl, 1993).

1-Alkyladenine, 3-alkylcytosine and 3-alkylthymine are minor products of N-alkylation of the DNA bases. They were not found to be mutagenic, but their alkyl groups are located within the base pairing region and as a result disrupt hydrogen bonding. They are inhibitors of DNA polymerase and may therefore play a role in determining the toxicity of alkylating agents (Saffhill et al., 1985).

Chemical alkylation of the 1-position of guanine and of the exocyclic amino groups have been found to occur, although to a very limited extent, but they have only been detected in \textit{in vitro} experiments (Singer, 1996).
The metabolites of vinyl chloride can lead to the formation of exocyclic DNA adducts as well as to cross links either with protein or with another base. CEO is more reactive towards DNA \textit{in vivo} than CAA, and the two compounds are thought to give rise to different patterns of DNA modifications (Barbin and Bartsch, 1986).

The major lesion generated by CEO is 7-(2-oxoethyl)guanine (oxet-G). This has been postulated to lead, after formation of a hemiacetal (Scherer \textit{et al.}, 1981), to a modification at the O$^6$-position of guanine (Figure 12). It was found, however, that the hemiacetel structure is highly unlikely under physiological conditions (Politzer \textit{et al.}, 1986), and misincorporation of bases tends to occur opposite cytosine rather than opposite the modified guanine residues (Barbin and Bartsch, 1986). This may be due to the formation of intrastrand links between oxet-G and an adjacent cytosine, which changes the base pairing properties of the latter, leading to misincorporation mainly of adenine (Scherer \textit{et al.}, 1986).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure12}
\caption{Formation of O$^6$,7-(1'-hydroxyethano)guanine from oxet-G}
\end{figure}

Exposure to vinyl chloride also leads to the formation of the cyclic adducts 1,N$^6$-etheno-adenine (εA), 3,N$^4$-ethenocytosine (εC) (Eberle \textit{et al.}, 1989) and N$^2$,3-ethenoguanine (εG) (Laib, 1986) in DNA (Figure 13a), but sites involved in base pairing in double-stranded DNA are generally protected from reaction. Heavily transcribed sequences and single-stranded regions of DNA arising due to 'breathing' or the formation of unusual DNA structures, are thought to be most at risk (Palejwala \textit{et al.}, 1991).

\textit{In vitro} experiments have shown εA and εC to be non-instructional lesions, as their hydrogen bonding properties are severely disrupted by the formation of the etheno bridge, εA having low
but εC high mutagenic activity (Jacobsen et al., 1989). They create kinetic pause sites rather than absolute replication blocks (Jacobsen and Humayun, 1990).

εG can form two hydrogen bonds with thymine and with cytosine (Oesch and Doerjer, 1982) (Figure 13b) and is mutagenic (Singer et al., 1987). Although in vitro CAA reacts with guanine to give mainly 1, N2-ethenoguanine (Sattsangi et al., 1977), this has not been detected in vivo, possibly because of efficient repair (Khazanchi et al., 1993).

![Figure 13. (a) Cyclic DNA adducts of vinyl chloride metabolites and (b) εG-T and εG-C base pairs](image)

### O-Alkylation

The attack of DNA by ‘hard’ electrophiles generates a significant amount of phosphotriesters, but the base residues are still preferentially alkylated on nitrogen. However, a larger proportion of O-alkylated products is formed (Blackburn and Kellard, 1986a), primarily at G-O6, but also at T-O2 and to a lesser extent at T-O4 and C-O2. The nature of the alkyl group
influences the distribution of adducts at various sites as well; ethylation generates a higher proportion of modification at oxygens than methylation (Singer, 1996; Mitra and Kaina, 1993).

The major O-alkylation product of DNA bases is O⁶-alkylguanine. This modified base directs the misincorporation of thymine to a significant extent during nucleic acid synthesis which, if left unrepaired, would lead to G:C→A:T transition mutations upon further replication (see Section 1.5 below). The existence of minor miscoding pathways leading to incorporation of guanine and adenine opposite O⁶-alkyl-G has also been reported (Saffhill, 1986).

Another miscoding lesion is O⁴-alkylthymine. As in the case of O⁶-alkylguanine, O⁴-alkylthymine has a tautomeric form different from that of the unmodified base, and consequently its hydrogen bonding properties are changed. It leads preferentially to misincorporation of guanine, resulting in A:T→G:C transitions (Saffhill et al., 1985). O⁴-Alkylthymine is repaired \textit{in vivo} by the same enzyme as O⁶-alkylguanine, but at a much slower rate (Basu and Essigmann, 1988).

O²-Alkylthymine and O²-alkylcytosine have low mutagenic efficiency, but they also inhibit DNA synthesis to a significant extent. Modelling studies have shown that the O²-alkyl group interferes sterically with the DNA sugar-phosphate backbone, and this interference may lead to distortions of the DNA structure (Saffhill et al., 1985).

Phosphotriesters could lead to changes in the binding of proteins to DNA due to neutralisation of negative charge along the phosphodiester backbone (Saffhill et al., 1985). They may hydrolyse with the cleavage of a nucleoside-phosphate link resulting in single strand breaks (Blackburn and Kellard, 1986a). Their role in carcinogenesis is not clear.
1.5 O⁶-Alkylguanine Lesions

1.5.1 Biological Significance

Alkylation of DNA at the O⁶-position of guanine is regarded as one of the most critical events leading to mutations and cancers. O⁶-alkyl-G directs the incorporation of either thymine or cytosine residues without blocking DNA replication (Lindahl et al., 1988). Where O⁶-alkyl-G pairs with thymine, a G·C→A·T transition mutation will occur. It has been suggested that the potency of carcinogens may depend on their ability to form O⁶-alkylguanine (Loveless, 1969).

O⁶-alkyl-G in DNA has also been linked with increased chromosome instability. This appears to be caused, at least in part, by futile cycles of mismatch repair of the newly synthesised strand which leave persistent single-stranded regions in the DNA (Rehakrantz and Nonay, 1995). Genetic experiments indicated the active removal by mismatch repair of both cytosine and thymine when paired with an O⁶-methylguanine residue in newly replicated DNA. It is assumed, therefore, that it is a matter of chance which pyrimidine will escape this correction system (Lindahl et al., 1988).

The mutagenic potential of O⁶-alkyl-G, especially in the case of O⁶-methyl-G and O⁶-ethyl-G, has been proved both in vitro, in templates copied by RNA and DNA polymerases (Mehta and Ludlum, 1978; Abbott and Saffhill, 1979), and in vivo, through site-directed mutagenesis (Chambers et al., 1985; Essigmann et al., 1986). DNA precursor pool alkylation is not thought to be relevant to mutagenesis in this case (Basu and Essigmann, 1988).

1.5.2 Consequences of O⁶-Guanine Alkylation on DNA Structure

NMR studies on double-stranded oligonucleotides of B-DNA have shown that O⁶-methylguanine causes a small helical distortion and altered hydrogen bonding when opposite either cytosine or thymine; with thymine it forms a Watson-Crick pair, and with cytosine it forms a wobble pair (Patel et al., 1986a,b). The O⁶-methyl-G·C wobble pair distorts the DNA helix more than the pair with thymine does. The hydrogen bonding possibilities O⁶-methyl-G with T and C are illustrated in Figure 14a, in comparison with the normal G·C and the wobble G·T base pairs (Figure 14b).
Figure 14. Hydrogen bonding scheme (a) of \( \text{O}^6\)-methyl-G with C and T and (b) of G with C and T.
The alkyl group of O^6-alkyl-G can adopt one of two conformations in the plane of the base: either proximal, pointing away from the hydrogen bonding region, or distal, pointing towards it; these conformations are also called \textit{anti} and \textit{syn} (relative to N^1 of the base) respectively.

NMR studies indicated that in solution the preferred conformation is distal (Kalnik \textit{et al.}, 1989a,b). In contrast, the X-ray structure of a self-complementary DNA duplex containing O^6-methyl-G paired with thymine, showed that the methyl group was in the proximal orientation (Leonard \textit{et al.}, 1990; Vojtechovsky \textit{et al.}, 1995). It is possible that the difference between the structures determined by X-rays and NMR respectively reflects the differences between the solution and crystal environments (Vojtechovsky \textit{et al.}, 1995).

The X-ray structure of a short Z-DNA duplex revealed that O^6-methyl-G paired with cytosine adopted a Watson-Crick configuration, with the methyl group in the proximal orientation, but the hydrogen bonds could not be defined. It is unclear, therefore, if in this structure cytosine exists as a minor tautomer, if it is protonated or if there is no hydrogen bond at all between N^1 of O^6-methyl-G and N^3 of C (Ginell \textit{et al.}, 1990).

Molecular modelling and computational studies have indicated that when O^6-methyl-G is paired with thymine, the energy difference between the two conformations of the methyl group is likely to be very small (Dosanjh \textit{et al.}, 1993), with the result that the methyl group stays in the conformation in the plane of the base, either distal or proximal, which was adopted prior to base pairing (Cruzeiro-Hansson \textit{et al.}, 1995).

The local conformation of the base pair is more disrupted when the alkyl group adopts the distal conformation. The latter conformation affects the base pairing with thymine to a greater extent than that with cytosine; the steric clashes between the O^6-alkyl group of guanine and the 4-amino group of cytosine are relieved through wobble pairing (Kalnik \textit{et al.}, 1989b), whereas in the case of thymine the base pair spreads apart slightly in the major groove, with the result that N^1 of O^6-methyl-G and N^3 of T are no longer within hydrogen bonding distance - they may at best form a very weak hydrogen bond, leaving one single intact hydrogen bond between the two bases (Kalnik \textit{et al.}, 1989a).

Support for the existence of the two conformations comes also from data obtained from \textit{in vitro} experiments, which show that in some cases O^6-methyl-G is replicated quickly, and in others replication is virtually blocked; the latter was presumed to arise due to the methyl group in the distal conformation interfering with replication to a significant extent, explaining thus the observation that O^6-methyl-G lesions can be lethal as well as mutagenic (Dosanjh \textit{et al.}, 1993).
NMR (Kalnik et al., 1989a,b) and computer modelling studies (Sayle and Goodfellow, 1996) showed that although the ethyl group at the O⁶-position of guanine introduces a slightly bigger local distortion of the phosphodiester backbone than methyl, the base pair containing the lesion was well contained within the double helix.

The O⁶-alkyl-G·C base pair has distorted phosphodiester links both 3’ and 5’ to the cytosine (Kalnik et al., 1989b), whereas the O⁶-alkyl-G·T base pair is accommodated within the DNA helix with minimal distortion (Kalnik et al., 1989a). The slow incorporation of C opposite O⁶-methyl-G, and of the next correct nucleotide following the incorporation of C, can be ascribed to the stereochemical problems encountered when these distorted phosphodiester links are formed (Tan et al., 1994).

The flanking sequence of the modified base pairs can influence the extent of the DNA structure perturbations, because the final dynamic structure of the helix will be influenced by the adjustment of stacking interactions, as the more hydrophobic O⁶-alkyl-G will tend to stack towards the centre of the helix (Wong et al., 1992). In addition, the alkyl group is likely to disrupt the spine of hydration in the major groove (Ginell et al., 1990).

1.5.3 Stability of DNA Containing O⁶-Alkylguanine

UV melting (thermal denaturing) studies of DNA oligomers containing O⁶-methylguanine paired in turn with A, C, T and G, showed that the duplexes were significantly destabilised compared to the one containing a normal G·C base pair and even compared to the G·T base pair, the order of stability being C > A > G > T, but the difference in the stability of the modified duplexes was very small (Gaffney et al., 1984). The thermal destabilisation of DNA brought about by alkylation at the O⁶-position of guanine shows little sequence dependence (Gaffney and Jones, 1989).

The fact that the O⁶-alkyl-G·T base pair is thermodynamically the least stable, combined with structural information from NMR studies - which showed that it is stabilised by one hydrogen bond and possibly another weak one while maintaining a Watson-Crick configuration of the bases - indicate that certain structural features, such as the alignment of the bases at the active site of the polymerase during replication, play a role in the preferential pairing of O⁶-alkylguanine with thymine, rather than the number of hydrogen bonds that can be formed between the two bases (Kalnik et al., 1989a).
The stability profile of a DNA oligomer containing O\(^{6}\)-methylguanine paired with cytosine studied over a wide pH range showed that at lower pH the duplex is stabilised, probably because of the protonation of the cytosine N\(^3\)-position, which enables the formation of a Watson-Crick pair between the bases, as represented in Figure 15 (Leonard et al., 1990).

![Figure 15. Protonated O\(^{6}\)-alkyl-G\(\cdot\)C\(^+\) base pair](image)

The O\(^{6}\)-alkyl-G\(\cdot\)C wobble base pair, which predominates at physiological pH, is likely to be removed by normal proofreading and repair enzymes, whereas O\(^{6}\)-alkyl-G\(\cdot\)C\(^+\), which is probably present to some extent at physiological pH, may escape detection (Brown and Kennard, 1992).

1.6 DNA Repair

Another factor regarding the biological relevance of DNA adducts is the mechanism of repair, because repair processes can also introduce errors into the DNA. There are three main types of repair mechanisms in vivo, which are widely, if not universally, distributed, and which maintain the integrity of the genome as long as repair occurs before replication: nucleotide excision repair, base excision repair and the direct reversal of damage to bases (Lindahl, 1982). The first two are so-called ‘short patch’ repair systems and are generally error-free.

There are ‘long patch’ repair systems as well. These are switched on after replication and are error-prone; their mode of action is best understood in bacteria (Blackburn, 1990). Mismatch
repair systems also occur universally in cells; they are very complex and involve a multitude of proteins (Lindahl, 1982).

1.6.1 Nucleotide Excision Repair

The incision-excision mechanism of nuclease cleaves phosphodiester bonds resulting in single strand breaks (Lindahl, 1982). After removal of the modified nucleotide, exonucleases enlarge the gap to 30-60 nucleotides long, and the DNA sequence is reconstructed by polymerases and DNA ligase (Blackburn, 1990). It is mainly bulky DNA lesions which cause major helix distortions that are removed by this mechanism, such as those caused by polycyclic aromatic hydrocarbons, aromatic amines and amides, and some photoadducts.

1.6.2 Base Excision Repair

A second mechanism requires the action of DNA N-glycosylases which are highly substrate specific and generate apurinic/apyrimidinic (AP) sites. These in turn are recognised by AP endonucleases which produce strand breaks usually on the 5'-end of the base-free sites, then excised by a phosphodiesterase, and the gap is filled by a polymerase and DNA ligase (Lindahl, 1993).

N-alkylpurines are generally repaired by this mechanism. 3-Methyladenine, which is a potentially lethal lesion because it blocks DNA synthesis, is removed in *Escherichia coli* by two different glycosylases, one of which is constitutively expressed and only recognises 3-methyladenine, and one which is inducible and - unusually for a N-glycosylase - cleaves other alkylated bases as well, such as 3-methylguanine, 7-methylguanine and O²-methylpyrimidines (Mitra and Kaina, 1993). In mammals, 3-methyladenine appears to be released by only one glycosylase, which is similar to the inducible *E.coli* enzyme in its broader substrate specificity (Mitra and Kaina, 1993).

7-Methylguanine is susceptible to imidazole ring-opening, and the resulting secondary lesion is removed by formamidopyrimidine-DNA glycosylase. This secondary lesion is formed to a very limited extent, but the reaction could be significant because 7-methylguanine is the main alkylation product of DNA bases (Lindahl, 1982).
Another important lesion repaired by this mechanism is uracil generated by hydrolytic deamination of cytosine. The enzyme responsible is uracil-DNA glycosylase. There is also a glycosylase specific for hypoxanthine formed by adenine deamination.

The release of exocyclic etheno base adducts by N-glycosylases has first been observed a decade ago in a rat brain tumour cell line (Oesch et al., 1986), but it was only recently shown in vitro that these lesions are repaired by glycosylases in human cell extracts (Dosanjh et al., 1994).

1.6.3 Direct Reversal of Damage to Bases

Direct reversal of damage to bases has been observed for O⁶-alkylguanine modifications, especially for small alkyl groups - such as methyl and ethyl - and for O⁴-alkylthymine, but to a lesser extent (Lindahl, 1982). The enzyme responsible for this particular repair mechanism is O⁶-methylguanine DNA methyl transferase (MGMT) or O⁶-methyl-G-transferase. In E.coli, MGMT is also called Ada-protein, as it was found to be the product of the ada gene, which is responsible for the adaptive response of the bacterium.

Photoreactivation is another example of this type of repair mechanism. Photolyase cleaves pyrimidine photodimers to their constituent pyrimidines. Light with a wavelength in the range of 340-400nm is required for this process (Lindahl, 1982).

1.6.4 Post-Replicative Repair

When the DNA polymerase encounters a base modification that inhibits it, a long gap is left opposite the lesion before replication starts again. This gap is repaired by the induction of the SOS response or by recombination-repair, both of which are error-prone. The filling of this gap is determined by several factors, such as the type of lesion, the template sequence, the type of polymerase and its 3'→5' proofreading activity (Blackburn, 1990). The latter is usually diminished in this type of repair in order to let the polymerase bypass the lesion.
1.7 O⁶-Methyl-G-Transferase

An inducible DNA repair pathway that protects specifically against the damaging effects especially of 'hard' alkylating agents, such as MNU and MNNG, was discovered in *E. coli* by Samson and Cairns (1977), and it was termed 'adaptive response'. The major mutagenic lesion generated by these chemicals is O⁶-methylguanine, and the repair of this lesion is induced as part of the adaptive response, leading to increased resistance to alkylation mutagenesis (Lindahl *et al*., 1988).

1.7.1 Mechanism of O⁶-Methylguanine Demethylation

O⁶-Alkylguanine is repaired *in vivo* by direct transfer of the methyl (or ethyl) group from the alkylated DNA to a protein, which is O⁶-methyl-G-transferase. This repair mechanism is unusual because it represents an enzymatic methyl group transfer to a protein without the involvement of the cofactor S-adenosylmethionine (Lindahl, 1982). The methyl group is accepted by one of the protein's cysteine residues, and an unmodified guanine is simultaneously regenerated in DNA (Olsson and Lindahl, 1980; Demple *et al*., 1982). Bulkier alkyl groups are also removed, but more slowly, and with the increase in bulk, other DNA repair systems become involved in their removal; however, the extent to which an adduct is repaired by the alkyltransferase rather than by excision repair depends on the organism (Pegg and Byers, 1992). The rates of reaction of MGMT with a given alkyl group are different in different organisms (Mitra and Kaina, 1993).

O⁶-methyl-G-transferase acts as a suicidal enzyme, i.e. it undergoes irreversible inactivation in its reaction with an O⁶-methylguanine residue, because no mechanism appears to exist to demethylate the resulting S-methylcysteine moiety, which is chemically very stable (Lindahl *et al*., 1988). As a consequence, cells become sensitive to the mutagenic effects of alkylating agents when the number of O⁶-alkylguanine lesions generated in their DNA exceeds the number of MGMT molecules they contain (Lindahl *et al*., 1988).

The preferred substrate for MGMT is double-stranded DNA. MGMT also reacts with single-stranded DNA, albeit at a slower rate (Lindahl *et al*., 1988). This suggests that the conformation of the phosphodiester backbone may play a role in the recognition of the substrate by the protein (Pegg and Byers, 1992). The free O⁶-alkylguanine base and the free nucleoside are very poor substrates for the enzyme, but they are still recognised.
It has been proposed that the methyl transfer reaction follows an $S_{N}2$ mechanism, the active site cysteine in its thiolate form acting as the nucleophile and guanine as the leaving group (Spratt and de los Santos, 1992) (Figure 16). Studies on the interactions of MGMT with O$^6$-methyl-guanine analogues containing sulfur or selenium instead of the O$^6$-atom, or with the ring nitrogens replaced, as well as with oligonucleotides containing these modified bases, suggested that O$^6$-methyl-G binds to the active site of the protein with hydrogen bonds to O$^6$ and N$^1$, and that the methyl group is then displaced from the guanine as a proton is transferred to the oxygen, neutralising the charge on the leaving group (Spratt and de los Santos, 1992; Spratt and Campbell, 1994).

![Figure 16. Proposed mechanism for O$^6$-methylguanine demethylation](image)

1.7.2 Structure and Putative Mode of Action of O$^6$-Methyl-G-transferase

The O$^6$-methyl-G-transferase of E.coli or Ada-protein, is the most extensively studied MGMT. It is a 39 kDa protein with two distinct domains: a 19 kDa C-terminal domain responsible for the repair of O$^6$-alkylguanine and O$^4$-alkythymine, and a 20 kDa N-terminal domain which demethylates one of the two diastereoisomers of the methylphosphotriesters generated by the action of 'hard' alkylating agents on DNA (Sedgwick et al., 1988).

Cys-321 has been identified as the methyl acceptor in the reaction with O$^6$-methyl-G (Demple et al., 1985). The active site Pro-Cys-His-Arg sequence is conserved in all known MGMT proteins (Sedgwick et al., 1988) and seems to be essential for the alkyl transfer function, as site directed mutagenesis studies have shown (Ihara et al., 1994). Cys-69 has been found to be the methyl acceptor in the reaction with methylphosphotriesters (Sedgwick et al., 1988).
The human MGMT has a molecular weight of 24 kDa and has only retained the O6-alkyl-G-transferase activity, but not the capacity to repair the methylated phosphodiester backbone, and is more similar to the C-terminal domain of Ada rather than to the whole Ada-protein (Lindahl et al., 1988).

Mammalian MGMTs repair bulkier adducts more readily than bacterial ones. They also have the ability to repair O6-benzylguanine, which the bacterial enzymes cannot. This fact has been exploited in cancer treatment, as many tumours are resistant to alkylating agents used in chemotherapy because of their high MGMT levels (Pegg and Byers, 1992). Resistance is acquired not only towards methylation agents, but also towards chloroethylating agents, such as BCNU. The main cytotoxic lesion caused by chloroethylating agents is an intrastrand cross-link from the N1 of guanine to N3 of cytosine in the opposite strand. Intermediates in the formation of this cross-link are O6-chloroethylguanine and 1,O6-ethanoguanine (Figure 17), and both these adducts are substrates for MGMT (Ludlum, 1990). Depletion of the alkyltransferase activity with O6-benzylguanine renders the tumour cells more susceptible to the action of the drugs.

Figure 17. Formation of interstrand cross-links by chloroethylating agents
The ability of human O⁶-methyl-G-transferase to generate guanine from O⁶-benzylguanaine is significantly enhanced in the presence of DNA, which suggests that it undergoes a conformational change upon DNA binding; this may facilitate the reaction of the O⁶-position of the modified guanine with the protein's acceptor site cysteine (Goodtzova et al., 1994). This also indicates that MGMT binds DNA in a non-specific manner, regardless of its O⁶-methyl-G content, in addition to the specific binding for dealkylation.

The tertiary structure of the 19 kDa C-terminal domain of the Ada-protein has recently been elucidated by X-ray crystallography (Moore et al., 1994). Surprisingly, the active site cysteine is buried. As a consequence, it has been proposed that part of the protein would swivel to expose a DNA binding site and provide access to the active site (Moore et al., 1994).

The crystal structure of a mammalian MGMT has not been obtained yet; however, a homology model of the three-dimensional structure of human MGMT based on the crystal structure of the C-terminal domain of the MGMT of E.coli (Ada-protein) and on the sequences of the two proteins has been proposed (Wibley et al., 1995). This and other studies with O⁶-methyl-G analogues containing different alkyl groups (Arris et al., 1994) suggest that there is a hydrophobic pocket with strict constraints at the active site of the human MGMT which, in addition to the stabilisation of the negative charge by the aromatic ring, facilitates reaction with O⁶-benzylguanaine rather than that with O⁶-methylguanaine; this is why the former compound is used as an aid in chemotherapy to deplete tumour cells of MGMT.

The DNA may undergo a conformational change as well during repair of a O⁶-alkylguanaine residue. The proposal that N¹ of the modified guanine is involved in hydrogen bonding with the enzyme during the alkyltransfer suggests that the double helix must open up for the dealkylation reaction to take place. This has been shown to happen in case of Hhal methyltransferase, which is part of a restriction-modification system from Haemophilus haemolyticus (Klimasauskas et al., 1994). This enzyme catalyses the transfer of a methyl group from S-adenosylmethionine to the 5-position of cytosine within a certain DNA sequence. The crystal structure of the Hhal bound to an oligonucleotide containing fluorocytosine (C⁵-F) shows a disrupted G·C⁵-F base pair, with the modified cytosine flipped out of the double helix and situated in the active site, which itself has undergone a large conformational change (Klimasauskas et al., 1994).

A similar suitable substrate-bound complex of MGMT would be required in order to elucidate the protein's mode of action, but such stable complexes are difficult to obtain because in the Sₜ₂ alkyl transfer reaction there is a single transition state and no chemical intermediates (Moore et al., 1994).
1.7.3 Regulation of O\textsuperscript{6}-Methyl-G-transferase Expression

The adaptive response of \textit{E.coli} protects the cells against the toxic and mutagenic effects of certain alkylating agents. The Ada-protein itself is regulating this response, acting as a strong transcriptional activator of the \textit{ada} gene when methylated at its Cys-69 residue, and as an inhibitor in its unmethylated state (Saget and Walker, 1994).

In mammals, the level of expression of O\textsuperscript{6}-methyl-G-transferase varies between organs and even between types of cells within a particular organ (Pegg and Byers, 1992). Some mammalian cell lines do not express the protein at all; it is thought that this is due to switching off of the corresponding gene.

It is not clear how the expression of the gene for MGMT is regulated in mammals. Induction of the enzyme has been observed in mammals in some circumstances, but its mechanism is likely to be quite different from the adaptive response in \textit{E.coli}, and the increase in MGMT levels are more modest (Lindahl \textit{et al.}, 1988). However, the number of transferase molecules in unadapted bacterial cells is much lower than that in mammalian cells: about 20 compared to 30,000 - 200,000 (Lindahl, 1982; Mitra and Kaina, 1993).
1.8 Aims

It is difficult to obtain a DNA-MGMT complex for X-ray studies, because the rate of reaction of O\(^6\)-methyl-G-transferase with O\(^6\)-methylguanine is extremely high. In addition, substrate analogues that bind without methyl transfer have not been identified yet.

In collaboration with Dr M. Moore from York University, this study proposes to synthesise an analogue of O\(^6\)-methylguanine with a nitrogen atom instead of the oxygen at the 6-position and incorporate it in oligonucleotides, in order to test it for MGMT binding. Analogues containing sulfur and selenium at the 6-position have been found to undergo demethylation (Spratt and de los Santos, 1992; Spratt and Campbell, 1994). Bacteria generate 6-methyladenine enzymatically for the modulation of gene expression (Lindahl, 1993), and this base is not recognised by O\(^6\)-methyl-G-transferases. However, N\(^6\)-methyldiaminopurine (N\(^6\)-MeDAP) has similarities to both 6-methyladenine and O\(^6\)-methylguanine, and could therefore perhaps be a substrate for MGMT. If it is recognised by the enzyme and if the methyl group is not cleaved, oligonucleotides containing this modification could be used to obtain the enzyme-DNA complex for crystallisation studies.

In addition, oligonucleotides containing O\(^6\)-methyl and O\(^6\)-ethylguaninate have been synthesised to be used as substrates for a mutant MGMT containing an alanine instead of cysteine at the active site, to be cloned and isolated by the research group of Dr M. Moore. This protein should retain its DNA binding properties but not its dealkylation function.

It is hoped that at least one of these two approaches will give stable DNA-enzyme complexes which could then be investigated by NMR and X-ray crystallography to provide information on the structural change the protein and/or the DNA must undergo upon binding.
2. Results and Discussion

Modified oligonucleotides can be used in conjunction with DNA binding proteins to provide useful information on binding and structure of the enzymes. In collaboration with Dr M. Moore at York University, studies were carried out to determine the interaction of DNA with O\(^6\)-methyl-G-transferase (MGMT), a suicidal DNA repair enzyme which removes the alkyl group from guanine residues methylated or ethylated on the O\(^6\)-position, and attaches it irreversibly to one of its cysteine residues.

A series of modified oligonucleotides have been synthesised containing O\(^6\)-alkylguanine (Figure 18) and a similar base - a purine having a nitrogen instead of an oxygen atom in the 6-position (N\(^6\)-methyl-2,6-diaminopurine) (Figure 19). The O\(^6\)-alkyl-G oligonucleotides are to be tested for the binding of a mutant MGMT containing an alanine residue instead of the cysteine in the reactive site, and the oligonucleotides containing N\(^6\)-methyl-2,6-diaminopurine (N\(^6\)-DAP) are to be used in conjunction with the wild type enzyme.

The procedures followed for the synthesis of suitably modified deoxyguanosine derivatives and for their transformation into phosphoramidite monomers for automated solid support DNA synthesis are outlined below. Optimisation of the deprotection time, purification and analysis of the oligonucleotides synthesised are also discussed. Comparative UV melting (thermal denaturing) studies were carried out on some of the sequences containing the above modified nucleic acid bases.
2.1 Structure of a Phosphoramidite

Phosphoramidites are chemically modified nucleosides which are used in DNA synthesis, and they are added one at a time to the support bound nucleotide chain. The reactive groups on these nucleosides are blocked or protected to prevent unwanted side reactions. The characteristics of cyanoethyl phosphoramidites (Figure 20) are:

- a trivalent phosphorus moiety on the 3' end, which will form the phosphite internucleotide linkage with the 5' hydroxyl of the support bound oligonucleotide,
- a diisopropylamino group on the 3' phosphorus moiety on which it confers stability; it also acts as a leaving group when protonated,
- a β-cyanoethyl protecting group on the 3' phosphorus moiety, which prevents side reactions and improves the solubility of the compound,
- a DMTr protecting group on the 5' end, to prevent self-polymerisation,
- protecting groups on the exocyclic amines of dA, dC and dG (dT does not have any exocyclic amino groups), to prevent or minimise side-reactions.
2.2 Synthesis of N\textsuperscript{2}-Acetyl-O\textsuperscript{6}-alkyl-dG

When synthesising a phosphoramidite with an alkyl group in the O\textsuperscript{6}-position of deoxyguanosine, a few difficulties must be overcome. Methods involving direct alkylation are unsuitable, mainly due to the fact that alkylation agents give significant N-alkylation.

Gaffney & Jones (1982) were the first to devise a practicable route for the synthesis of N\textsuperscript{2}-isobutyryl-O\textsuperscript{6}-alkyl-dG that did not involve direct alkylation. This method exploited the findings of Bridson et al. (1977) and Daskalov et al. (1981) that the O\textsuperscript{6}-position of guanosine and deoxyguanosine is susceptible to acylation, sulphonylation and phosphorylation. The problem with this method is that the isobutyryl group used for protection of the 2-amino group

---

**Figure 20.** Standard phosphoramidites of (a) dA, (b) dG, (c) dT and (d) dC
of the $O^6$-substituted guanine requires prolonged exposure of the oligonucleotide to concentrated ammonia at high temperatures ($65\,^\circ\mathrm{C}$ for 72 hours). Even so, the final product may be contaminated by incompletely deprotected material as well as by the formation of 2,6-diaminopurine (DAP) (Figure 21) through displacement of the $O^6$-alkyl function through nucleophilic attack by ammonia during the prolonged exposure at high temperature (Fowler et al., 1982; Kuzmich et al., 1983).

![Figure 21. Formation of DAP by prolonged exposure of $O^6$-alkyl-G to conc. NH$_3$](image)

The presence of a DAP residue in oligonucleotides has been shown to increase the stability of duplexes when it is paired with thymine, as it can form three hydrogen bonds with this DNA base (Howard et al., 1966; Scheit & Rackwitz, 1982). For this study, DAP is an undesirable side product, since it reduces the amount of $O^6$-alkyl-G present in oligonucleotides and would affect their UV melting temperatures.

The synthesis of $N^2$-acyl-$O^6$-alkyl-2'-deoxyguanosine employed here (Figure 22) generally follows the method of Li & Swann (1989), which was itself adapted from the method of Gaffney and Jones (1982), but the acyl group used for the protection of the 2-amino group of $O^6$-alkyl-G is acetyl (Gaffney et al., 1984), rather than phenylacetyl (Li & Swann, 1989; Smith et al., 1990), as this is more convenient. Deoxyguanosine triacetate is relatively easy to prepare, and the acetyl group is quite readily removed from the 2-amino group of the substituted guanine residues in the oligonucleotides. In addition, the number of steps in the synthesis is reduced when using acetyl as a protecting group rather than phenylacetyl, and the overall yields for 3',5'-acetyl-$N^2$-acyl-2'-deoxyguanosine are comparable for the two methods - 78% for acetyl versus 83% for phenylacetyl.
Deoxyguanosine (1) is reacted with a large excess of acetic anhydride (Ac₂O) at reflux in pyridine (Py). Various purification methods of the resulting deoxyguanosine triacetate (2) were tried. Recrystallisation from EtOH / H₂O (1:9) gave a deep pink solid in a fairly low yield (53%). Boiling with a small amount of charcoal did not improve the yield by very much, and not enough impurities were removed. Boiling with a larger amount of charcoal gave much clearer and purer crystals, but unfortunately the yield dropped dramatically, as the bulk of the compound crystallised within the charcoal layer during filtration. Flash column chromatography was considered, but the compound did not dissolve well in dichloromethane. However, upon filtration, a very small amount of product was isolated. Recrystallisation from acetone, followed by two more recrystallisations of the supernatant, gave off-white, well defined crystals, raising the yield to 70 - 80%.

The deoxyguanosine triacetate (2) is first sulphonylated (3) with mesitylenesulphonyl chloride, then displaced by a sterically hindered tertiary amine - in this case N-methylpyrrolidine - to give the N-methylpyrrolidinium derivative (4), which is in turn displaced by the appropriate alcohol in the presence of DBU, to give O⁶-alkyl-deoxyguanosine triacetate (5) in a one-pot reaction. After work up of the reaction mixture, the product is purified by wet flash column chromatography.

The next step is the removal of the acetyl groups from the 3' and 5' oxygen atoms of the sugar ring by adding NaOH to compound 5 dissolved in pyridine and methanol, to obtain N²-acetyl-O⁶-alkyldeoxyguanosine (6). The reaction time and NaOH concentration were optimised so that there was minimal deprotection of the N²-acetyl. The methanol is required in order to obtain a homogenous reaction mixture, otherwise the NaOH solution and the compound 5 in pyridine solution do not mix. The mixture is then neutralised with pyridinium-Dowex. This procedure has the benefit of removing the sodium ions while generating pyridine, which can then be removed in vacuo.

After protection of the 5'-OH group and subsequent transformation into a phosphoramidite, N²-acetyl-O⁶-alkyldeoxyguanosine can be used in solid support automated DNA synthesis.
Figure 22. Synthesis of N\textsuperscript{2}-acetyl-O\textsuperscript{6}-alkyldeoxyguanosine
2.3 Synthesis of O\textsuperscript{6}-pentafluorophenyl-N\textsuperscript{2}-trifluoroacetyl-2'-deoxyguanosine

In order to obtain oligonucleotides with a N-methyl group rather than the usual oxygen at the 6-position of a guanine residue, the nucleoside to be incorporated during DNA synthesis has to be chosen carefully. As is the case for O\textsuperscript{6}-substituted guanines, the N\textsuperscript{2}-acyl protecting group should be relatively labile, otherwise - as is the case for isobutyryl - deprotection is extremely slow and the end product may be contaminated with incompletely deprotected material.

In addition, the 6-methylamino group can pose a few problems. If it is inserted in the phosphoramidite, it will undergo side reactions during DNA synthesis, notably during coupling, when it would be phosphitylated. It could be protected with an acyl group, but the presence of two N-acyl bonds in the purine moiety makes the glycosidic bond between deoxyribose and the nitrogen base very sensitive to acid. Protic acids lead to depurination within minutes, and even with Lewis acids such as ZnBr\textsubscript{2}, depurination has been observed after about one hour (Gaffney \textit{et al.}, 1984).

Postsynthetic modification of a functional group at the 6-position of a deoxyguanosine derivative incorporated in oligonucleotides appears to be the best option. The research group of Gaffney and Jones (The State University of New Jersey) found that O\textsuperscript{6}-pentafluorophenyl-N\textsuperscript{2}-trifluoroacetyl-2'-deoxyguanosine (PFTFdG) was a useful synthon for the postsynthetic modification of oligonucleotides (Gao \textit{et al.} 1992). Their method for its synthesis (Fathi \textit{et al.}, 1990) was used here (Figure 23).

Deoxyguanosine (1) in pyridine was treated with trifluoroacetic anhydride (TFA) and then with a pentafluorophenol (PFP) solution in pyridine to give O\textsuperscript{6}-pentafluorophenyl-N\textsuperscript{2}-trifluoroacetyl-2'-deoxyguanosine (9) in a one-pot reaction. The mechanism of the synthesis has not been completely elucidated, as the authors were not able to isolate any reaction intermediates. It is presumed, however, that the reaction leads first to exhaustive trifluoroacetylation of the starting material, giving compound 7, with the O\textsuperscript{6}-trifluoroacetyl group being then displaced by pyridine, yielding the 6-pyridyl derivative 8, which is a fluorescent polar material (TLC) (Fathi \textit{et al.}, 1990). PFP then displaces the 6-pyridyl group, to give PFTFdG (9).
2.4 Synthesis of Phosphoramidites from the Modified Deoxyguanosine Derivatives

In order to incorporate the modified deoxyguanosine derivatives in oligonucleotides, they have to be first protected at the 5'-OH with a 4,4'-dimethoxytrityl (DMTr) group, and then phosphitylated to obtain the phosphoramidites (Figure 24).

The standard methods of dimethoxytritylation (Jones, 1984) and phosphitylation (Sinha et al., 1984) were used. The deoxyguanosine derivative (6, 9) is treated with DMTrCl in pyridine in the presence of triethylamine to give the 5'-dimethoxytrityl compound (10 a-c), which is then reacted with the phosphitylating agent 2-cyanoethyl-N,N-(diisopropylamino)chlorophosphite in anhydrous tetrahydrofuran (THF) in the presence of diisopropylethylamine (DIPEA), the end product being the desired phosphoramidite (11 a-c).
The phosphitylating agent was synthesised according to the following method (Sinha et al., 1983) (Figure 25): Phosphorus trichloride is treated with 3-hydroxypropionitrile to give 2-cyanoethyl-dichlorophosphite. This is then reacted with diisopropylamine to obtain 2-cyanoethyl-N,N-(diisopropylamino)chlorophosphite. P(III) compounds are very sensitive to moisture and oxygen, therefore the synthesis of the phosphitylating agent as well as the
phosphitylation of nucleosides must be carried out in anhydrous conditions under argon or nitrogen, with the exclusion of air.

Figure 25. Synthesis of phosphitylating agent

2.5 Oligonucleotide synthesis

After suitably modified phosphoramidite monomers were obtained, the synthesis of the oligonucleotides involves straightforward solid support automated DNA synthesis by the phosphoramidite method, which is the method of choice for most laboratories because of efficient and rapid coupling and the stability of the starting materials (Beaucage and Caruthers, 1981).

2.5.1 Short History

The first synthesis of a dinucleotide containing a 3'→5' phosphate internucleotide linkage identical to the one in natural DNA was reported in 1955 by Michelson and Todd. Khorana and co-workers developed the protection of the 5'-hydroxyl group with trityl derivatives and the protection of the exocyclic amino groups of adenine and guanine with benzoyl and isobutyryl groups, respectively, during their pioneering work carried out in the late nineteen-fifties and in the nineteen-sixties (Beaucage and Iyer, 1992; Blackburn and Gait, 1990).

The introduction of the ‘phosphite triester’ method (Letsinger et al., 1975) revolutionised the synthesis of oligonucleotides, which up to then was carried out mostly by the ‘phosphotriester’ method (Letsinger and Mahadevan, 1965 and 1966). However, the deoxynucleoside chlorophosphite intermediates were unstable to moisture and difficult to handle because of
their high reactivity. A new class of intermediates for oligonucleotide synthesis - the phosphoramidites (Figure 26) - was introduced by Beaucage and Caruthers (1981). These compounds could be isolated and stored as stable powders.

\[
\text{DMTrO} \quad \text{B} \\
\text{O} \\
\text{O} \\
\text{H}_3\text{C} \quad \text{N} \quad \text{P} \quad \text{OCH}_3 \\
\text{CH}_3
\]

B is a suitably protected nucleic acid base

**Figure 26. Phosphoramidite (Beaucage and Caruthers, 1981)**

The concept of solid phase synthesis was developed by Merrifield (1963) for peptide synthesis. The use of organic - mainly polystyrene based (Letsinger and Mahadevan, 1965) - and inorganic - silica based (Köster, 1972) - solid supports for the synthesis of nucleic acids was then investigated. After the introduction of the phosphite triester method of oligonucleotide synthesis, the interest in suitable solid supports was revived (Matteucci and Caruthers, 1981). Currently, the most widely used solid support is silica based and is controlled pore glass (CPG) (Beaucage and Iyer, 1992).

Reliable automation of DNA synthesis became possible after Beaucage and Caruthers (1981) introduced phosphoramidites as a new class of monomers for DNA synthesis, which were more stable under normal laboratory conditions than the nucleoside chlorophosphites or the corresponding tetrazolides commonly used at the time. A number of DNA synthesisers employing the phosphoramidite approach are currently available.

The N,N-dimethylamino group was subsequently replaced with N,N-diisopropylamino (McBride and Caruthers, 1983; Adams et al., 1983), which facilitated the purification of the phosphoramidite monomers and increased their stability in solution.

The removal of the methyl group on the phosphorus moiety required treatment with thiophenol at the end of the synthesis. The latter is an unpleasant reagent, so methyl protection on the phosphoramidites has been superseded by β-cyanoethyl protection (Sinha et al., 1983 and
1984), since this group can be easily removed under the basic conditions required to cleave the oligonucleotides from the solid support, using the same reagent - concentrated ammonia - employed in base deprotection.

2.5.2 Solid Support DNA Synthesis by the Phosphoramidite Method

The synthesis is performed with the growing DNA chain attached to the solid support trapped on the synthesis column between two filters. The chemical reactions are driven to completion by the use of an excess of reagents relative to the support bound oligonucleotide, and the unreacted compounds, which are in the liquid phase, can easily be removed by washing with an appropriate solvent (Brown and Brown, 1991; Gait, 1984). Therefore, no purification steps are required between cycles.

Each cycle of base addition consists of four steps (Figure 27):

(i) Detritylation - occurs in the presence of a protic acid, usually trichloroacetic acid (TCA) in anhydrous dichloromethane (DCM). The removal of the acid-labile dimethoxytrityl (DMTr) protecting group on the 5'-end of the support-bound nucleoside enables the coupling of the 5'-hydroxyl group with a phosphoramidite during the following step.

(ii) Coupling - the phosphoramidite is activated by tetrazole and then forms an internucleotide phosphite with the support-bound 5'-hydroxyl.

(iii) Capping - the oligonucleotide chains that failed to undergo addition in the previous step are capped by acetylation with Ac₂O in the presence of 1-methylimidazole, to prevent their propagation in subsequent cycles, which will make them easier to separate from the main product.

(iv) Oxidation - the phosphite linkage, being susceptible to acid and base cleavage, is then oxidised by iodine in a solution of water, pyridine and THF, and the cycle can start again.
B is a suitably protected nucleic acid base

**Figure 27.** Solid support DNA synthesis cycle

---

54
When the synthesis is complete, the DMTr protection on the last nucleoside can be left on or removed. For this study it was removed, as the oligonucleotides were quite short. It is mainly in case of long oligonucleotides that reverse phase HPLC purification with the DMTr on presents a distinct advantage, as this hydrophobic group makes the oligomer to which it is attached elute later from the column, thus enabling a better separation, after which the DMTr group is removed by treatment with 80% acetic acid. Short oligonucleotides are usually quite easy to separate, even without an attached DMTr group. In addition, 'trityl off' purification avoids the additional exposure of the modified oligomers to aqueous protic acid, which could lead to some depurination.

### 2.6 Deprotection and Postsynthetic Modification of Oligonucleotides

Cleavage from the solid support and deprotection of the phosphate groups are usually done automatically, in concentrated ammonia. The crude oligonucleotide synthesis mix is then placed at 55°C to remove the base protecting groups. The O⁶-alkylguanine oligonucleotides were cleaved and deprotected in this way.

The O⁶-pentafluorophenyl-N²-trifluoroacetyl guanine (PFTFG) oligonucleotides were treated manually with a 40% aqueous solution of methylamine at 55°C during the final step, which is cleavage from the solid support and deprotection. If it were to be used on the DNA synthesizer, methylamine could attack the plastic tubing and the seals. At the same time the O⁶-pentafluorophenyl group of the modified guanines is displaced by methylamine to give N⁶-methyldiaminopurine (N⁶-MeDAP) (Figure 28). The deprotection time of these oligomers seemed to be a crucial factor in respect of the ease of their HPLC purification and had to be optimized (see Section 2.7.4).
The O\(^6\)-methyl-G oligonucleotides could also have been obtained by postsynthetic modification. Treatment of the PFTFG containing oligomers with DBU in methanol leads to the replacement of the pentafluorophenyl group of the modified guanine to give O\(^6\)-methyl-G. However, the reaction is much slower in case of the oligonucleotides compared to that of the monomer; Gao et al. (1992) reported that the reaction was less than two-thirds complete even after five days. Consequently, the synthesis of an O\(^6\)-methyl-G phosphoramidite which would then be used in DNA synthesis was deemed a more suitable method for obtaining oligonucleotides containing this particular modified base.
2.7 Optimization of the Deprotection Time, Purification, Enzymatic Digestion and CZE Analysis of the Oligonucleotides

The purification of the oligonucleotides consists of preparative reverse phase HPLC, desalting on Sephadex G-10 and freeze-drying.

2.7.1 Deprotection and HPLC Purification of the O\textsuperscript{6}-Alkyl-G Oligonucleotides

The unmodified bases making up the oligonucleotides have standard protection on their amino groups (see Section 2.1) and the O\textsuperscript{6}-alkyl-G bases are protected by an acetyl group. To determine the optimum deprotection time of the O\textsuperscript{6}-alkyl-G oligonucleotides, one of these (CGC G\textsuperscript{Me}AA TTC GCG) was treated with concentrated ammonia at 55°C for over five days, and analytical HPLC traces of the crude mixture were done after 4.5, 16, 19, 36 and 126 hours.

The chromatogram obtained after 4.5 hours (Figure 28a) indicates a two main overlapping peaks, the one on the right being the bigger one. After 16 and 19 hours (Figure 28b and c, respectively) the peak on the left has increased significantly in size, the one on the right appearing as a shoulder to the former. After 36 hours (Figure 28d) there is just one main peak, and after 126 hours (Figure 28e) there appears to be no significant change. It was therefore assumed that deprotection was complete at 36 hours. Consequently, judging by the size of the shoulder to the main peak after 19 hours and trying to avoid unnecessary prolonged exposure to concentrated ammonia at 55°C, the O\textsuperscript{6}-methyl-G and O\textsuperscript{6}-ethyl-G oligonucleotides were deprotected for 24 hours and then purified.
Figure 29. Analytical HPLC traces of CGC GMeAA TTC GCG after (a) 4.5, (b) 16, (c) 19, (d) 36 and (e) 126 hrs deprotection in conc. NH$_3$ at 55°C
Prior to purification by preparative HPLC, analytical HPLC of the oligonucleotides was carried out. Figure 30 contains examples of the HPLC traces obtained.

![HPLC traces](image)

**Figure 30.** Prep HPLC trace of (a) GCA GGMeG AAC and (b) GCA GGGEt AAC

2.7.2 Enzymatic Digestion of the O^6^-Alkyl-G Oligonucleotides

In order to make sure that O^6^-methyl-G and O^6^-ethyl-G, respectively, have been incorporated in oligonucleotides, these were submitted to enzymatic digestion with phosphodiesterase from snake venom and alkaline phosphatase from bovine intestinal mucosa. This generates a mixture of free nucleosides in the proportions that they were incorporated in the oligonucleotides.

The enzyme digests were analysed by reverse HPLC on their own, as well as coinjected with an HPLC standard, i.e. the appropriate modified nucleoside. It was found that an extra peak due to the modified nucleoside appeared on the chromatograms of the enzyme digests. This had a longer retention time than the unmodified (standard) nucleosides. When coinjected with the HPLC standard, the size of the peak attributed to the modified nucleosides had increased significantly. This proves that the modified nucleosides were successfully incorporated in DNA oligomers. Examples of the HPLC chromatograms obtained are shown in Figure 31.
The order of elution of the peaks is dC, dG, dA and O\textsuperscript{6}-alkyl-dG. No peaks corresponding to DAP, which elutes after dG and before dA, were detected.

**Figure 31.** Enzyme digest of GCA GGG\textsuperscript{Me} AAC (a) on its own and (b) coinjected with O\textsuperscript{6}-MedG

**Figure 32.** Enzyme digest of GCA GG\textsuperscript{Et}G AAC (a) on its own and (b) coinjected with O\textsuperscript{6}-EtdG
2.7.3 CZE Analysis of the O\textsuperscript{6}-Alkyl-G Oligonucleotides

CZE was used to check the purity of the oligonucleotides after they had been purified, desalted and freeze dried, as it is a very sensitive method. The CZE traces of the O\textsuperscript{6}-alkyl-G oligonucleotides all showed one single sharp main peak. Examples of the obtained CZE traces are shown in the appropriate experimental section.

2.7.4 Deprotection, HPLC Purification, Enzymatic Digestion and CZE Analysis of the PFPG / N\textsuperscript{6}-Methyldiaminopurine Oligonucleotides

To determine the optimum deprotection time of the PFPG / N\textsuperscript{6}-DAP containing oligonucleotides, one of these - GCA GG\textsuperscript{NHMe}G AAC - was treated with 40% aqueous methylamine at 55°C for 12 days. The crude oligonucleotide mixture was analysed by HPLC after 1, 3, 5, 7 and 12 days (Figure 33).
Figure 33. Analytical HPLC traces of GCA GGNHMeG AAC after (a) 1, (b) 3, (c) 5, (d) 7 and (e) 12 days treatment with methylamine at 55°C
The chromatogram obtained after 1 day (Figure 33a) indicates a main peak with a shoulder on the right. After 3 days (Figure 33b) the shoulder on the right has increased in size. After 5 and 7 days (Figure 33c and d, respectively) there appear to be two main peaks running very close together and partially overlapping, the one on the left being the bigger one, and after 12 days (Figure 33e) the one on the right is bigger. At first it was assumed that the postsynthetic modification of PFPG oligonucleotides may be quite slow, especially when they contain three guanines in a row (one of them being modified), which generates an environment that is quite hydrophobic and may have the tendency to prevent the access of solvent (water) and of the deprotecting reagent. Based on that assumption, the peak on the left would be due to the as yet unmodified oligomers.

When the oligonucleotides were submitted to preparative HPLC purification, the chromatograms showed either two overlapping main peaks in case of GCA GGNHMeG AAC (Figure 34a) or a broad main peak with a few significant shoulders for GCA GGGNHMe AAC (Figure 34b). This made purification very difficult.

![Figure 34](image)

Figure 34. Prep HPLC of (a) GCA GGNHMeG AAC and (b) GCA GGGNHMe AAC

The two main peaks of GCA GGNHMeG AAC were collected separately, submitted to enzymatic digestion and analysed by Capillary Zone Electrophoresis. HPLC analysis of the enzyme digests showed that, surprisingly, both peaks contained N6-methylaminopurine nucleoside. CZE of the first HPLC peak showed one single main peak, sometimes with a broad base, whereas the second HPLC peak seemed to block the capillary. In most cases the
second HPLC peak showed nothing at all on CZE, but occasionally a distortion of the base line in the shape of a bulge, a small peak with a broad base or a great number of single narrow peaks would appear. It was concluded that the second HPLC peak was due to secondary structure of the oligonucleotides, which would explain why it might block the capillary. It may also be possible that this DNA secondary structure would sometimes get sheared in the capillary, which is consistent with a large number of peaks appearing occasionally during CZE analysis.

The broad HPLC peak of GCA GGG\textsuperscript{NHMe} AAC didn't block the capillary during CZE analysis, but appeared to be impure.

No such problems were encountered with the oligonucleotide GCA G\textsuperscript{PP}GG AAC, where the analytical HPLC traces after 2, 3 and 4 days treatment with methylamine were virtually identical, showing one sharp main peak (Figure 35). This indicated that the postsynthetic modification was complete.

![Figure 35. Analytical HPLC traces of GCA G\textsuperscript{NHMe}GG AAC after (a) 2 and (b) 4 days treatment with methylamine at 55°C](image)

The prep HPLC showed a single main peak (Figure 36), which appeared to contain no impurities when analysed by CZE.
Gao et al. (1992) had reported that postsynthetic modification of the PFPG residues with ammonia to 2,6-diaminopurine in oligonucleotides takes about 60 hours. Methylamine is more nucleophilic than ammonia, so it was considered reasonable to assume that its reaction with the PFPG oligomers would take less time. The oligonucleotides were resynthesised and their deprotection time shortened. Considering the analytical HPLC traces in Figure 33 and the fact that apparently the size of the shoulder/peak on the right of the main peak should be minimised to ease purification, newly synthesised PFPG oligonucleotides were treated with 40% aqueous methylamine for 10 - 12 hours and then purified by reverse phase HPLC. The prep HPLC traces obtained show only a small shoulder on the right of the main peak (Figure 37b).

Figure 37. (a) Analytical and (b) prep HPLC of GCA GG$^{\text{NHMeG}}$ AAC after 12 hours treatment with methylamine at 55°C
After desalting and freeze drying, each N⁶-methyl-DAP oligonucleotide was submitted to enzymatic digestion to check whether this modification was indeed present. An extra peak due to the modified nucleoside appeared on the chromatograms of the enzyme digests. This had a longer retention time than the unmodified (standard) nucleosides. When coinjected with N⁶-methyl-DAP nucleoside, the size of the peak attributed to the modified nucleoside had increased significantly. This proves that PFTFdG was successfully incorporated in DNA oligomers and then modified to N⁶-methyl-DAP. Examples of the HPLC chromatograms obtained are shown below. The order of elution of the peaks is dC, dG, dA and N⁶-methyl-DAP nucleoside.

![Figure 38. Enzyme digest of GCA GGNHMeG AAC (a) on its own and (b) coinjected with N⁶-MeDAP nucleoside](image)

CZE analysis of the oligonucleotides after 10 - 12 hours treatment with methylamine, reverse phase HPLC purification, desalting on Sephadex G-10 and freeze drying, showed one single sharp main peak. Examples of the CZE traces obtained are shown in the appropriate experimental section.
2.8 UV Melting (Thermal Denaturing)

UV melting studies were carried out on the duplexes formed from some of the modified oligonucleotides and a complementary strand with C and then with T opposite the modified base. The generic sequence of these duplexes is represented in Figure 39. The melting temperatures of the modified duplexes (Table 1) were compared with those of the duplex formed by the unmodified oligonucleotide, where each of the relevant guanines were paired in turn both with C and with T (Table 2).

\[
5' \text{GCA G}^{*}_{1}\text{G}^{*}_{2}\text{G}^{*}_{3} \text{AAC 3'}
\]

\[
3' \text{CGT C C C TTG 5'}
\]

Figure 39. Generic sequence of the G-modified duplexes submitted to UV melting

Table 1. Melting temperatures of the G-modified duplexes

<table>
<thead>
<tr>
<th>Tm (°C)</th>
<th>G_1 NHMe</th>
<th>G_2 NHMe</th>
<th>G_3 NHMe</th>
<th>G_1 Me</th>
<th>G_2 Me</th>
<th>G_3 Me</th>
<th>G_1 Et</th>
<th>G_2 Et</th>
<th>G_3 Et</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>38.5</td>
<td>38.9</td>
<td>37.3</td>
<td>22.1</td>
<td>24.0</td>
<td>22.8</td>
<td>21.2</td>
<td>23.3</td>
<td>21.8</td>
</tr>
<tr>
<td>C</td>
<td>25.7</td>
<td>29.3</td>
<td>29.1</td>
<td>27.9</td>
<td>30.4</td>
<td>28.1</td>
<td>27.0</td>
<td>30.8</td>
<td>27.3</td>
</tr>
</tbody>
</table>

Table 2. Melting temperatures of the unmodified duplexes

<table>
<thead>
<tr>
<th>Tm (°C)</th>
<th>G_1</th>
<th>G_2</th>
<th>G_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>31.1</td>
<td>29.4</td>
<td>27.4</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>43.8</td>
</tr>
</tbody>
</table>

When paired with a C, there is hardly any difference between the melting temperatures of the modified duplexes, but they all are significantly destabilised compared to the natural,
unmodified duplex. Those duplexes where the second in the row of three guanines (G₂) is substituted, appear to be somewhat more stable than their counterparts with G₁ or G₃ modified, the difference in melting temperatures ranging from 0.2 - 3.5°C. The structures of the natural G - C and of the modified base pairs are represented in Figures 40 and 41 respectively.

![Figure 40. Guanine - cytosine base pair](image)

![Figure 41. (a) O⁶-methylguanine - cytosine and (b) N⁶-methyldiaminopurine - cytosine wobble base pairs with the methyl group in the distal orientation](image)

When paired with a T, the situation is changed. The melting temperatures of the N⁶-methyldiaminopurine duplexes are quite high, almost as high as the one of the natural duplex, and significantly higher than that of the unmodified mismatched duplexes (with each G paired in turn with a T). The O⁶-alkyl duplexes, however, are even more destabilised than they were in the previous case. These results indicate that there is favourable base pairing, i.e.
there are three hydrogen bonds between \( G^{\text{Me}} \) and \( T \) (Figure 43b) with the methyl group probably adopting the proximal orientation, but, as discussed in Section 1.5.2, at best only two hydrogen bonds between \( G^{\text{Me}} \) and \( T \) (Figure 43a) and two between \( G \) and \( T \) (Figure 42).

![Figure 42. Guanine - thymine base mispair](image)

![Figure 43. (a) \( O^6 \)-methylguanine - thymine and (b) \( N^6 \)-methyldiaminopurine - thymine base pairs with the methyl group in proximal orientation](image)
3. Experimental

3.1 Chemicals and Solvents

All chemicals were obtained from Aldrich, Fluka, Sigma or Fisons. Silica gel 60 (80-430 mesh) was obtained from Fluka. All solvents were of analytical or HPLC grade. Anhydrous acetonitrile was purchased from Applied Biosystems as DNA synthesis grade.

Anhydrous pyridine was distilled over CaH$_2$. Anhydrous tetrahydrofuran (THF) and diethyl ether (ether) were distilled over sodium and benzophenone. Triethylamine (TEA), diisopropylamine and diisopropylethylamine (DIPEA) were dried by storage over CaH$_2$. Methanol and ethanol were dried by distillation from magnesium turnings and iodine. Toluene was dried by storage over sodium wire. Pyrogen-free, reverse osmosis purified water was used throughout.

3.2 Thin Layer Chromatography (TLC)

TLC was carried out on aluminium sheets coated with a 0.2mm layer silica gel 60 F$_{254}$, supplied by Merck, using the following solvent systems (all v/v):

I. dichloromethane : methanol (10 : 1)

II. dichloromethane : methanol (9.5 : 0.5)

III. ethylacetate : acetonitrile (1 : 1)

IV. ethylacetate : triethylamine (9.9 : 0.1)

V. ethylacetate : methanol : concentrated ammonia (5 : 1 : 1)

VI. dichloromethane : methanol : acetic acid (10 : 2 : 1)

VII. n-butanol : acetic acid : water (3 : 1 : 1)

Before carrying out TLC of dimethoxytritylated compounds, the plates were pretreated with triethylamine to prevent detritylation due to the acidic silica gel.

The purity of all compounds synthesised was verified by one spot on TLC.
For prep TLC, 20cm x 20cm glass plates were coated in the lab with silica gel GF$_{254}$ obtained from Fluka, by the following procedure: A mixture of 130g silica gel and 225g deionised water - vigorously shaken for 30 seconds - was poured in the spreader set at a distance of 1.0mm from the plates. The spreader was then pulled quickly in one single movement from one end to the other of the row of six glass plates clamped in a special holder. The plates were left to stand for 20 minutes and then placed in the drying oven overnight.

Compounds were visualised under UV light at 254nm. To see the presence of sugar groups, the TLC plates were sprayed with an anisaldehyde solution of the composition anisaldehyde (para-methoxybenzaldehyde) : AcOH : conc. H$_2$SO$_4$ : EtOH (5 : 1 : 1 : 100 v/v/v/v) and then heated on a hotplate. A positive test was indicated by a dark blue, green or brown coloration on heating. Dimethoxytrityl groups were visualised by fuming HCl as bright orange colours.

3.3 Wet Flash Column Chromatography

This was carried out under slight nitrogen or argon pressure (0.5 bar) using silica gel 60. For compounds containing the cyanoethyl-N$_2$N-diisopropyl-phosphoramidite group the silica was dried overnight at 120°C and then cooled in a dessicator over P$_2$O$_5$ prior to being used in order to remove any moisture present. For compounds containing the DMTr group, the silica was preequilibrated with solvent containing 1% TEA.

3.4 Instrumentation

$^1$H-NMR spectra were recorded on a Brucker WP-200 (200.13 MHz) Fourier transform spectrometer. $^{13}$C-NMR spectra were recorded on a Brucker WP-200 (50.32 MHz) Fourier transform spectrometer. Tetramethylsilane was used as internal reference and cited chemical shifts are given in ppm downfield of this standard. $^{31}$P-NMR spectra were recorded on a Jeol FX90Q (90 MHz) Fourier transform spectrometer. H$_3$PO$_4$ 85% was used as external reference and cited chemical shifts are given in ppm downfield of this standard. UV spectra were recorded on a Perkin Elmer Lambda 2 ultraviolet-visible spectrophotometer. Positive ion fast atom bombardment (FAB) mass spectra were recorded on a Kratos MS 50 TC spectrometer using either a thioglycerol or a 3-nitrobenzyl alcohol (3-NOBA) matrix. C, H and N microanalysis was performed on a Perkin Elmer 240 elemental analyser.
3.5 Oligonucleotide Synthesis

Oligonucleotide synthesis was performed using solid phase 2-cyanoethyl-N,N-diisopropylphosphoramidite chemistry (Beaucage and Caruthers 1981; Beaucage and Iyer, 1992; Brown and Brown 1991) on an Applied Biosystems 394 automated DNA synthesiser equipped with on-line long wave detector (498nm). All DNA synthesis reagents and standard phosphoramidite monomers were supplied by Applied Biosystems.

The modified phosphoramidites synthesised for this study were dried in vacuo, stored with P$_2$O$_5$ overnight and used as a 0.11 - 0.12M solution in anhydrous acetonitrile. Coupling efficiencies were in general > 98%. 'Trityl-off' syntheses were carried out. The O$_6$-alkyl-G and the unmodified oligonucleotides were cleaved automatically from the solid support with concentrated ammonia. Deprotection of O$_6$-alkyl-G and unmodified oligonucleotides was carried out by incubation in sealed vials with teflon caps at 55°C in concentrated ammonia for 24 hours and 12 hours, respectively. The PFPG oligonucleotides were cleaved manually and deprotected in 40% aqueous methylamine at 55°C for 10-12 hours.

The following DNA sequences were synthesised:

<table>
<thead>
<tr>
<th>GCA G$^{PFPG}$ AAC</th>
<th>GCA G$^{MeGG}$ AAC</th>
<th>GCA G$^{EtGG}$ AAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCA GG$^{PFPG}$ AAC</td>
<td>GCA GG$^{MeG}$ AAC</td>
<td>GCA GG$^{EtG}$ AAC</td>
</tr>
<tr>
<td>GCA GGG$^{PF}$ AAC</td>
<td>GCA GGG$^{Me}$ AAC</td>
<td>GCA GGG$^{Et}$ AAC</td>
</tr>
<tr>
<td>CGC G$^{PFPG}$AG CTC GCG</td>
<td>CGC G$^{MeAA}$ TTC GCG</td>
<td>GTC TCC TCG</td>
</tr>
<tr>
<td>CGC GAG CTC GCG</td>
<td>GCA GGG AAC</td>
<td>GTC TCT TCG</td>
</tr>
<tr>
<td>CGC G$^{PFPG}$AG CTT GCG</td>
<td>GTC CCC TCG</td>
<td>GTC CCT TCG</td>
</tr>
<tr>
<td>CGC GAG CTT GCG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

72
3.6 HPLC Analysis and Purification of Oligonucleotides

Oligonucleotide analysis and purification was carried out on a Gilson Model 306 equipped with UV detector, an Aquapore Octyl reverse phase column (10mm x 250mm) and an octadecyl pellicular guard column, controlled by a computer running Gilson 712 software.

The eluting buffers used were:

- buffer A - 0.1M NH₄OAc
- buffer B - 0.1M NH₄OAc plus 20% MeCN

At a flow rate of 3ml/min, the following gradient was used:

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>% Buffer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>3.00</td>
<td>0.0</td>
</tr>
<tr>
<td>4.00</td>
<td>15.0</td>
</tr>
<tr>
<td>24.00</td>
<td>70.0</td>
</tr>
<tr>
<td>26.00</td>
<td>100.0</td>
</tr>
<tr>
<td>28.00</td>
<td>100.0</td>
</tr>
<tr>
<td>29.00</td>
<td>0.0</td>
</tr>
<tr>
<td>30.00</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Analytical runs were monitored at 260nm and 0.2 absorbance units full scale (AUFS). Preparative scale runs were monitored at 293-298nm and 1 AUFS.

3.7 Enzymatic Digestion of Oligonucleotides

Enzymatic digestion of oligonucleotides allows the identification of their nucleoside composition. The enzymes used in this procedure are phosphodiesterase I (type VII from Crotalus atrox venom) and alkaline phosphatase (type VII from bovine intestinal mucosa). The former enzyme cleaves the oligonucleotides into its constituent nucleoside-5'-phosphates
(nucleotides), and the latter one hydrolyses the phosphate groups from the nucleotides to give the corresponding nucleosides.

One unit of phosphodiesterase I will hydrolyse 1.0 mM of bis(p-nitrophenyl)phosphate per minute at pH 8.8 and 37°C. One unit of alkaline phosphatase will hydrolyse 1 mM of p-nitrophenylphosphate at pH 10.4 and 37°C.

Oligonucleotides (2 OD260 units) dissolved in 500μl enzyme digest buffer (1M sodium chloride and 50mM Tris at pH 8.8) were incubated with 0.10 units phosphodiesterase I and 0.30 units alkaline phosphatase at 37°C for 24 hours. The digest was then analysed by reverse phase HPLC at 260nm and 0.2 AUFS using buffer A: 0.1M NH4Ac and buffer B: 0.1M NH4Ac plus 20% MeCN, at a flow rate of 3ml/min. One of the following two gradients was used:

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>% Buffer B</th>
<th>Time (min)</th>
<th>% Buffer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.0</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>2.00</td>
<td>0.0</td>
<td>2.00</td>
<td>0.0</td>
</tr>
<tr>
<td>3.00</td>
<td>15.0</td>
<td>3.00</td>
<td>15.0</td>
</tr>
<tr>
<td>26.00</td>
<td>100.0</td>
<td>26.00</td>
<td>100.0</td>
</tr>
<tr>
<td>29.00</td>
<td>100.0</td>
<td>29.00</td>
<td>0.0</td>
</tr>
<tr>
<td>29.50</td>
<td>0.0</td>
<td>30.00</td>
<td>0.0</td>
</tr>
<tr>
<td>30.00</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.8 Sephadex Gel-Filtration (Desalting) of Oligonucleotides

During the reverse phase HPLC purification step, the oligonucleotides were eluted with buffers containing salt and possibly other small molecules. In order to remove these small molecules, the oligonucleotides up to 10 bases in length were submitted to gel filtration on a Pharmacia Sephadex G-10 column (2.6cm x 40cm) connected to an UV detector, at a flow rate of
2 ml/min. The eluent which passed through the UV detector, was monitored at 260 nm. The DNA peak was collected and lyophilised to give a fluffy white powder.

The sequences of more than 10 bases in length were desalted on Pharmacia NAP-10 (nucleic acid purification) cartridges containing Sephadex G-25, according to the manufacturer's instructions.

3.9 Capillary Zone Electrophoresis (CZE)

The purity of all oligonucleotides synthesised was verified by one single sharp peak on CZE. This was carried out on an Applied Biosystems Model 270A Capillary Electrophoresis system with Microgel capillaries using 75 mM Tris-phosphate with 10% methanol at pH 7.6 elution buffer. The on-line detector was set at 272 nm with the loading voltage at -5 kV, the running voltage at -15 kV and the sample uptake time at 1 to 5 seconds. Sample runs were recorded on a chart recorder set at chart speed 1 cm/min, attenuation 16, offset 5 and peak threshold 10 000. Examples of the CZE traces obtained are shown below.

![CZE traces](image)
3.10 UV Melting (Thermal Denaturing) Studies

UV melting was carried out on a Lambda 15 UV-visible spectrophotometer controlled by a computer running a PECSS (Perkin Elmer Computerised Spectroscopy Software) package. The temperature of the sample was controlled by a Perkin Elmer Peltier block with a heating rate of 1°C/min. The absorbance of the sample at 260nm was measured at intervals of ten seconds to obtain absorbance versus time melting curves. The data was then converted to absorbance versus temperature curves and the melting temperature (which corresponds to the
point of inflection of the curves) was obtained by calculating the first derivative of the curve and determining its maximum point. The melting temperature of a nucleic acid is that where half of it is double stranded and half single stranded. All melting curves were determined in triplicate.

The micromolar extinction coefficients of the single stranded oligonucleotides were calculated using the following formula taken from Brown and Brown (1991):

\[ \varepsilon_{254} = [(8.8 \times nT) + (7.3 \times nC) + (11.7 \times nG) + (15.4 \times nT)] \times 0.9 \]

The factor of 0.9 is due to the fact that base stacking interactions in the single strand suppress the absorbance of DNA relative to the value calculated from the extinction coefficients of the individual nucleosides.

The values for the micromolar extinction coefficients \( \varepsilon(G^*) \) of the modified nucleosides were determined experimentally from the UV spectra of the compounds. The following values were found:

\[ \varepsilon_{254}(dG^Me) = 8.71 \]
\[ \varepsilon_{254}(dG^Et) = 3.49 \]
\[ \varepsilon_{254}(dG^NHMe) = 6.75 \]

Single-stranded oligonucleotides were dissolved in UV melting buffer (0.1M NaCl, 0.01M NaH_2PO_4, 20mM cacodylic acid, 1mM EDTA, adjusted to pH 7.0 with NaOH). Solutions were then mixed to give 5.0 \( \mu \)M solutions of double-stranded DNA when made up to 2.7ml with UV melting buffer in a cuvette. Baseline corrections were carried out against a reference cell containing UV melting buffer. Below 15°C, nitrogen was continuously flushed through the sample compartment to prevent condensate formation.
3.11 Synthetic Chemistry

3', 5'-Bis(O-acetyl)-N²-acetyl-2'-deoxyguanosine [Gaffney et al., 1984]

Deoxyguanosine monohydrate (5.03g, 17.5mmol) was dried by the evaporation of anhydrous pyridine (3 x 15ml), suspended in dry pyridine (100ml), and acetic anhydride (25ml, 265mmol, 15eq. [d=1.082]) was added. The reaction mixture was stirred at reflux for 4.5 hours, when TLC showed the absence of both starting material and 3', 5'-bisacetylated intermediate (R_f 0.0 and 0.15, respectively in solvent system I). The solvents were then removed in vacuo. Small amounts of toluene were added to help remove the pyridine. The solid residue was dissolved in acetone at reflux and left to cool. Crystalline deoxyguanosine triacetate precipitated and was filtered. The supernatant was concentrated on the rotary evaporator until crystals started to form again and was then filtered. These operations were repeated with the new filtrate. A cream coloured crystalline solid was obtained. Yield 78% (5.36g). R_f 0.30 (system I); 0.15 (system II); 0.74 (system VI). [FAB] m/z (M+H)^+: found 394.13768, calculated 394.13627. ¹H-NMR: δH (d₆-DMSO) 2.02, 2.08, 2.17 (3s, 9, 3'-N²-Ac: CH₃), 2.49-2.58 (m, 1, H₂'), 2.91-3.02 (m, 1, H₂''), 4.16-4.27 (m, 3, H₄'5'5''), 5.30-5.32 (m, 1, H₃'), 6.17-6.20 (t, 1, H₁'), 8.20 (s, 1, H₈), 11.70, 12.03 (2s, 2, 1-NH, N²-H). ¹³C-NMR: δC (d₆-DMSO) 20.70 CH₃, 20.96 CH₃, 23.94 CH₃, 38.93 CH₂, 63.71 CH₂, 74.55 CH, 81.90 CH, 83.13 CH, 120.49 CH, 137.54 C, 148.19 C, 148.71 C, 154.94 C, 170.21 C, 170.36 C, 173.71 C. Anal. calcd. for C₁₆H₁₉N₅O₇·3/4H₂O: C 47.23, H 5.08, N 17.21; found C 47.20, H 5.05, N 17.27. UV spectrum (MeOH): λₘₚₓ₁ = 202nm, λₘₚₓ₂ = 259nm, λₘₚₓ₃ = 279nm.
3', 5'-Bis(O-acetyl)-N2-acetyl-O6-methyl-2'-deoxyguanosine

3', 5'-Bis(O-acetyl)-N2-acetyl-2'-deoxyguanosine (5.07g, 12.9mmol) was suspended in dry dichloromethane (DCM) (80ml). Triethylamine (TEA) (7.2ml, 51.6mmol, 4eq. [d=0.726]), mesitylenesulphonyl chloride (MSCl) (5.64g, 25.8mmol, 2eq.) and 4-(N, N-dimethylamino)-pyridine (DMAP) (60mg, 0.5mmol) were added and the reaction was stirred at room temperature for 1.5 hrs, when TLC showed total conversion to the higher running O6-sulphonylated compound (Rf 0.64 in system I). The reaction mixture was then cooled in an ice bath and N-methylpyrrolidine (13.4ml, 129mmol, 10eq. [d=0.819]) was added. After 30 minutes, conversion to the 6-N-methylpyrrolidinium compound was complete (Rf 0.0 in system I). After warming to room temperature, the mixture was stirred with MeOH (3.6ml, 90.3mmol, 7eq. [d=0.791]) and DBU (5.8ml, 38.7mmol, 3eq. [d=1.018]) for 1.5 hrs, when a compound with a higher Rf was formed. The reaction was quenched with aqueous KH2PO4 (1M, 100ml, pH 6.5). The organic layer was drawn off, dried over Na2SO4, filtered and concentrated in vacuo. The residue was dissolved in a small amount of DCM and purified by wet flash column chromatography with a gradient of 0 - 3% MeOH in DCM. The appropriate fractions were combined and evaporated in vacuo obtaining the title compound as a pale yellow foam. Yield 62% (3.26g). Rf 0.48 (system I); 0.25 (system II); 0.82 (system VI).

[FAB] m/z (M+H)+: found 408.14951, calculated 408.15192. 1H-NMR: δ (d6-DMSO) 1.99, 2.01, 2.09 (3s, 9, 3'-5'-N2-Ac: CH3), 2.48-2.57 (m, 1, H2'), 3.16-3.30 (m, 1, H2''), 4.06 (s, 3, O6-CH3), 4.21-4.34 (m, 3, H4'5'5''), 5.42-5.45 (m, 1, H3'), 6.32-6.38 (t, 1, H1'), 8.40 (s, 1, H8), 10.46 (s, 1, N2-H). 13C-NMR: δ (d6-DMSO) 20.70 CH3, 20.99 CH3, 24.90 CH3, 35.18 CH2, 54.27 CH3, 63.84 CH2, 74.67 CH, 81.97 CH, 83.91 CH, 117.77 C, 141.65 CH, 152.25 C, 152.51 C, 160.61 C, 169.16 C, 170.19 C, 170.31 C. Anal. calcd. for C17H21N5O7: C 50.12, H 5.20, N 17.19; found C 50.47, H 4.70, N 16.39. UV spectrum (MeOH): λ1=202nm, λ2=259nm, λ3=279nm (shoulder).
3', 5'-Bis(O-acetyl)-N²-acetyl-O⁶-ethyl-2'-deoxyguanosine

The procedure above was followed up to the conversion of the O⁶-sulphonylated compound to the quaternary pyrrolidinium salt. After warming to room temperature, the mixture was stirred with EtOH (5.3ml, 90.3mmol, 7eq. [d=0.785]) and DBU (5.8ml, 38.7mmol, 3eq.) for 2.5 hrs, when a compound with a higher Rₚ was formed. The reaction was quenched with aqueous KH₂PO₄ (1M, 100ml, pH 6.5). The organic layer was drawn off, dried over Na₂SO₄, filtered and concentrated in vacuo. The organic was dissolved in a small amount of DCM and purified by wet flash column chromatography with a gradient of 0 - 3% MeOH in DCM. The appropriate fractions were combined and evaporated in vacuo to a pale yellow foam. Yield 52% (2.82g). Rₚ 0.50 (system I); 0.26 (system II); 0.83 (system VI). [FAB] m/z (M+H)+: found 422.16765, calculated 422.16757. ¹H-NMR: δH (CDCl₃) 1.42-1.47 (t, 3, O⁶-Et: CH₃), 2.04, 2.10 (2s, 6, 3'- 5'-Ac: CH₃), 2.50 (s, 3, N²-Ac: CH₃), 2.52-2.60 (m, 1, H2"), 2.89-2.98 (m, 1, H2"), 4.27-4.42 (m, 3, H4'(5'')), 4.50-4.59 (q, 2, O⁶-CH₂), 5.37-5.41 (m, 1, H3'), 6.29-6.32 (m, 1, H1'), 7.93 (s, 1, H8), 8.14 (s, 1, N²-H). ¹³C-NMR: δC (CDCl₃) 14.19 CH₃, 20.58 CH₃, 20.74 CH₃, 24.98 CH₃, 37.00 CH₂, 63.45 CH₂, 74.16 CH, 82.27 CH, 84.42 CH, 118.30 C, 135.96 C, 139.32 CH, 151.99 C, 160.84 C, 170.09 C, 170.31 C, 170.56 C. Anal. calcd. for C₁₈H₂₃N₅O₇.1/2H₂O: C 50.23, H 5.62, N 16.27; found C 50.54, H 5.68, N 16.05. UV spectrum (MeOH): λ₁=201nm, λ₂=219nm, λ₃=270nm.
3', 5'-Bis(O-acetyl)-N^2-acetyl-O^6-methyl-2'-deoxyguanosine (2.05g, 5mmol) was dissolved in pyridine (30ml) and methanol (25ml). NaOH solution (0.5M, 21ml, 2.1eq.) was added and the reaction mixture was vigorously shaken for 20 seconds, then it was poured over Dowex 50-X8 pyridinium form. The mixture was then filtered, and the filtrate was concentrated in vacuo to a pale yellow solid. Yield 96% (1.55g). The product was not soluble in solvent system I, and was therefore not purified by flash column chromatography. R_f 0.09 (system I); 0.22 (system VI); 0.64 (system VII). [FAB] m/z (M+H)^+: found 324.13084, calculated 324.13079.

^1H-NMR: δH(d_6-DMSO) 2.26 (s, 3, N^2-Ac: CH_3), 2.28-2.36 (m, 1, H2'), 2.68-2.78 (m, 1, H2''), 3.50-3.75 (m, 3, H5', 5'-OH), 3.82-3.89 (m, 1, H4'), 4.08 (s, 3, O^6-CH_3), 4.43-4.48 (m, 1, H3'), 5.33 (bs, 1, 3'-OH), 6.31-6.37 (t, 1, H1'), 8.45 (s, 1, H8), 10.44 (N^2-H). ^13C-NMR: δC (d_6-DMSO) 24.89 CH_3, 39.38 CH_2, 54.21 CH_2, 61.73 CH_2, 70.76 CH, 88.00 CH, 137.90 C, 141.27 CH, 152.23 C, 152.58 C, 160.48 C, 169.35 C. Anal. calcd. for C_{13}H_{17}N_5O_5.H_2O: C 45.75, H 5.61, N 20.52; found C 45.89, H 5.51, N 20.65. UV spectrum (MeOH): λ_1=202nm, λ_2=219nm, λ_3=270nm.

N^2-acetyl-O^6-ethyl-2'-deoxyguanosine
3', 5'-Bis(O-acetyl)-N²-acetyl-O⁶-ethyl-2'-deoxyguanosine (2.10g, 5mmol) was dissolved in pyridine (30ml) and methanol (25ml). A 0.5M NaOH solution (21ml, 2.1eq.) was added and the reaction mixture was vigorously shaken for 20 seconds, then it was poured over Dowex 50-X8 pyridinium form. The mixture was then filtered, and the filtrate was concentrated in vacuo to a pale yellow solid. Yield 97% (1.63g). Rₐ 0.17 (system I); 0.25 (system VI); 0.66 (system VII). [FAB] m/z (M+H)⁺: found 338.14694, calculated 338.14644. ¹H-NMR: δH (d₆-DMSO) 1.36-1.42 (t, 3, O⁶-Et: CH₃), 2.23 (s, 3, N²-Ac: CH₃), 2.27-2.32 (m, 1, H₂'), 2.66-2.77 (m, 1, H₂''), 3.48-3.72 (m, 3, H₅'₅'', 5'-OH), 3.83-3.88 (m, 1, H₄'), 4.41-4.45 (m, 1, H₃''), 4.51-4.59 (q, 2, O⁶-CH₂,), 5.05 (bs, 1, 3'-OH), 6.29-6.35 (t, 1, H₁''), 8.42 (s, 1, H₈). ¹⁳C-NMR: δC (d₆-DMSO) 14.53 CH₃, 24.84 CH₃, 39.42 CH₂, 61.76 CH₂, 62.83 CH₂, 70.80 CH, 83.42 CH, 87.98 CH, 117.49 C, 141.12 CH, 152.21 C, 152.63 C, 160.11 C, 169.29 C. Anal. calcd. for C₁₄H₁₉N₅O₅.H₂O: C 47.32, H 5.96, N 19.71; found C 47.46, H 5.79, N 19.98. UV spectrum (MeOH): λ₁=201nm, λ₂=220nm, λ₃=270nm.

O⁶-methyl-2'-deoxyguanosine [HPLC standard]

3', 5'-Bis(O-acetyl)-N²-acetyl-O⁶-methyl-2'-deoxyguanosine (203mg, 0.5mmol) was dissolved in pyridine (3ml) and methanol (3ml). NaOH solution (2M, 0.6ml, 2.4eq.) was added and the reaction was stirred overnight. The mixture was then poured over pyridinium-Dowex for neutralisation, filtrated, and the filtrate was concentrated in vacuo to a pale yellow solid. Yield 98% (136mg). Rₐ 0.14 (system I); 0.24 (system VI); 0.64 (system VII). [FAB] m/z (M+H)⁺: found 282.12000, calculated 282.12023. ¹H-NMR: δH (d₆-DMSO) 2.17-2.26 (m, 1, H₂'), 2.51-2.63 (m, 1, H₂''), 3.45-3.61 (m, 3, H₅'₅'', 5'-OH), 3.81-3.85 (m, 1, H₄'), 3.95 (s, 3, O⁶-CH₃), 4.34-4.37 (m, 1, H₃''), 5.15 (bs, 1, 3'-OH), 6.21-6.25 (t, 1, H₁''), 6.44 (s, 2, NH₂), 8.08 (s, 1, H₈). ¹³C-NMR: δC (d₆-DMSO) 39.29 CH₂, 53.29 CH₃, 61.88 CH₂, 70.92 CH, 82.94 CH, 87.75 CH, 114.11 C, 137.87 CH, 153.87 C, 159.91 C, 160.80 C. Anal. calcd. for
C_{11}H_{15}N_{5}O_{4}: C 46.97, H 5.38, N 24.90; found C 46.73, H 5.79, N 24.39. UV spectrum (MeOH): \( \lambda_1 = 211 \text{nm}, \lambda_2 = 248 \text{nm}, \lambda_3 = 282 \text{nm} \).

O^6\text{-ethyl-2'-deoxyguanosine [HPLC standard]}

3', 5'-Bis(O-acetyl)-N^2-acetyl-O^6-ethyl-2'-deoxyguanosine (210mg, 0.5mmol) was dissolved in pyridine (3ml) and methanol (3ml). A 2M NaOH solution (0.6ml, 2.4eq.) was added and the reaction was stirred overnight. The mixture was then poured over pyridinium-Dowex for neutralisation, filtrated, and the filtrate was concentrated in vacuo to a pale yellow solid. Yield 97% (143mg). R_f 0.22 (system I); 0.28 (system VI); 0.66 (system VII). [FAB] m/z (M+H)^+:

- found 296.13705, calculated 296.13588.

\(^1\)H-NMR: \( \delta_H (\text{d}_6\text{-DMSO}) 1.37-1.42 \text{ (t, 3, O^6-Et: CH}_3), 2.17-2.25 \text{ (m, 1, H2''), 2.53-2.65 \text{ (m, 1, H2''')}, 3.47-3.63 \text{ (m, 3, H5''', 5'-OH)}, 3.82-3.86 \text{ (m, 1, H4''), 4.35-4.39 \text{ (m, 1, H3'''), 4.48-4.56 \text{ (q, 2, O^6-CH}_2'), 5.22 \text{ (bs, 1, 3'-OH)}, 6.25-6.31 \text{ (t, 1, H1''), 6.42 \text{ (s, 2, NH}_2}), 8.10 \text{ (s, 1, H8)}. \(^{13}\)C-NMR: \( \delta_C (\text{d}_6\text{-DMSO}) 13.56 \text{ CH}_3, 39.27 \text{ CH}_2, 57.63 \text{ CH}_3, 61.82 \text{ CH}_2, 70.87 \text{ CH}, 83.06 \text{ CH}, 87.83 \text{ CH}, 115.29 \text{ C}, 138.17 \text{ CH}, 153.44 \text{ C}, 159.12 \text{ C}, 161.40 \text{ C}. \) Anal. caled. for C_{12}H_{17}N_{5}O_{4}.1/4H_2O: C 48.08, H 5.88, N 23.36; found C 48.35, H 5.89, N 23.44. UV spectrum (MeOH): \( \lambda_1 = 201, \lambda_2 = 247 \text{nm}, \lambda_3 = 281 \text{nm}. \)
N^2-acetyl-O^6-methyl-2'-deoxyguanosine (2.00g, 6.2mmol) was coevaporated from anhydrous pyridine (3 x 7ml) and dissolved in anhydrous pyridine (20ml) and TEA (950μl, 6.8mmol, 1.1eq. [d=0.726]). 4,4'-Dimethoxytrityl chloride (DMTrCl) (2.3g, 6.8mmol, 1.1eq.) and DMAP (150mg, 1.2mmol, 0.2eq.) were added and the mixture was stirred at room temperature for 2 hours. Methanol (20ml) was added to quench the reaction, and the solvents were removed in vacuo. The residue was dissolved in ethyl acetate (EtOAc) (200ml) and was washed three times with a saturated KCl solution (100ml). The organic layer was dried over Na\textsubscript{2}SO\textsubscript{4}, filtered and concentrated in vacuo to give the product as a gum, which was then purified by wet flash column chromatography, eluting with a gradient of 0 - 2.5% MeOH in DCM on a column with the silica preequilibrated with 1%TEA in DCM. The appropriate fractions were combined, and the solvents were removed in vacuo. A cream-yellow foam was obtained. Yield 64% (2.48g). R\textsubscript{f} 0.41 (system I); 0.27 (system IV). [FAB] m/z (M+H)\textsuperscript{+}: found 626.25918, calculated 626.26147. 1H-NMR: \(\delta\) (CDCl\textsubscript{3}) 2.46-2.58 (m, 1, H2"'), 2.62-2.73 (m, 1, H2'"), 3.25-3.44 (m, 2, H5'5'"), 3.72 (s, 6, Ar-OCH\textsubscript{3}), 4.08 (s, 3, O^6-CH\textsubscript{3}), 4.17-4.21 (m, 1, H4'), 4.65-4.68 (m, 1, H3'), 5.37 (bs, 1, 3'-OH), 6.41-6.47 (t, 1, H1'), 6.73-7.38 (m, 13, Ar), 7.96 (s, 1, H8), 8.29 (s, 1, N^2-H). 13C-NMR: \(\delta\) (CDCl\textsubscript{3}) 25.00 CH\textsubscript{3}, 40.37 CH\textsubscript{2}, 54.32 CH\textsubscript{3}, 55.01 CH\textsubscript{3}, 63.88 CH\textsubscript{2}, 72.21 CH, 84.16 CH, 86.31 CH, 112.96 CH, 118.09 C, 126.72 CH, 127.68 CH, 127.89 CH, 129.81 CH, 135.47 C, 139.95 CH, 144.33 C, 151.75 C, 152.07 C, 158.30 C, 160.97 C, 171.03 C. Anal. calcd. for C\textsubscript{34}H\textsubscript{35}N\textsubscript{5}O\textsubscript{7}: C 65.27, H 5.64, N 11.19; found C 64.44, H 5.62, N 11.12. UV spectrum (MeOH): \(\lambda_1=203\text{nm}, \lambda_2=236\text{nm}, \lambda_3=271\text{nm}\).
5'-O-(4,4'-Dimethoxytrityl)-N2-acetyl-O6-ethyl-2'-deoxyguanosine

N2-acetyl-O6-ethyl-2'-deoxyguanosine (1.38g, 4.1mmol), coevaporated from anhydrous pyridine (3 x 6ml) and dissolved in anhydrous pyridine (15ml) and TEA (630μl, 4.5mmol, 1.1eq.) was stirred with DMTrCl (1.5g, 4.5mmol, 1.1eq.) and DMAP (110mg, 0.9mmol, 0.2eq.) at room temperature for 2.5 hours. Methanol (15ml) was added to quench the reaction, and the solvents were removed in vacuo. The residue was dissolved in EtOAc (150ml) and was washed with a saturated KCl solution (3 x 75ml). The organic layer was dried over Na2SO4, filtered and concentrated in vacuo. The resulting gum was then purified by wet flash column chromatography, eluting with a gradient of 0-2.5% MeOH in DCM on a column with the silica preequilibrated with 1% TEA in DCM. The appropriate fractions were combined, and the solvents were removed in vacuo, obtaining a cream-yellow foam. Yield 58% (1.52g). Rf 0.43 (system I); 0.31 (system IV). [FAB] m/z (M+H)+: found 640.27696, calculated 640.27712. 1H-NMR: δH (CDCl3) 1.42-1.49 (t, 3, O6-Et: CH3), 2.46-2.78 (m, 5, H2'2' ), N2-Ac: CH3), 3.24-3.51 (m, 2, Ar-OCH3), 4.09-4.19 (m, 1, H4'), 4.49-4.60 (q, 2, O6-CH2-), 4.65 (m, 1, H3'), 5.41 (bs, 1, 3'-OH), 6.40-6.47 (t, 1, H1'), 6.73-7.44 (m, 13, Ar), 7.90 (s, 1, H8), 8.18 (s, 1, N2-H). 13C-NMR: δC (CDCl3) 14.25 CH3, 24.98 CH3, 40.40 CH2, 55.03 CH3, 63.37 CH2, 63.90 CH2, 72.19 CH, 84.11 CH, 86.32 CH, 112.97 CH, 118.09 C, 126.73 CH, 127.70 CH, 127.91 CH, 129.83 CH, 135.48 C, 139.79 CH, 144.36 C, 151.78 C, 152.13 C, 158.32 C, 160.72 C, 170.96 C. Anal. calcd. for C35H37N5O7: C 65.72, H 5.83, N 10.95; found C 65.52, H 5.76, N 11.02. UV spectrum (MeOH): λ1=202nm, λ2=235nm, λ3=271nm.
2-Cyanoethyl-N,N-(diisopropylamino)chlorophosphite [Sinha et al., 1983]

3-Hydroxypropionitrile (3.4ml, 50mmol, 0.14eq.) dissolved in anhydrous acetonitrile (20ml) under argon, was added dropwise, via a double ended needle, to a stirring solution of PCl₃ (30.6ml, 350mmol, 1eq.) in anhydrous acetonitrile (20ml) under argon. A further addition of solvent (5ml) was made to rinse the delivery flask and needle. After stirring for 15 minutes at room temperature the reaction mixture went cloudy, and the solvent was removed \textit{in vacuo} to leave an oil which was distilled under reduced pressure (0.5mm Hg, 125-130°C) in a Kugelrohr to give 2-cyanoethyl phosphodichlorite. This was dissolved in anhydrous ether (40ml) stirring, under argon, at 0°C (in an ice bath), and a solution of N,N-diisopropylamine (10.5ml, 74.9mmol, 2eq.) dissolved in anhydrous ether (40ml) was added dropwise via a double ended needle. A further addition of solvent (5ml) was made to rinse the delivery flask and needle. A white salt precipitated. After 5 minutes, the reaction mixture was filtered under argon and concentrated \textit{in vacuo} to give an oil which was purified by Kugelrohr distillation (0.5mm Hg, 160°C). A clear oil was collected. Yield (calculated in respect of 3-hydroxypropionitrile) 37% (4.38g). $^{31}\text{P-NMR}$: $\delta_p$ (CDCl$_3$) 180.25.
$5'-(4,4'-\text{Dimethoxytrityl})-N^2\text{-acetyl-O}^6\text{-methyl-2'-deoxyguanosine-3'}-(2\text{cyanoethyl-N,N-diisopropyl})\text{-phosphoramidite}$

$5'-(4,4'-\text{Dimethoxytrityl})-N^2\text{-acetyl-O}^6\text{-methyl-2'-deoxyguanosine (1.0g, 1.6mmol)}$ was dried by the evaporation of anhydrous pyridine ($2 \times 6\text{ml}$) and of anhydrous THF ($2 \times 7\text{ml}$), and dissolved in dry THF ($10\text{ml}$). Diisopropylethylamine (DIPEA) ($111\text{3}\mu\text{l, 6.4mmol, 4eq. [d=0.742]}$) and then 2-cyano-ethyl-N,N-(diisopropylamino)chlorophosphite ($393\mu\text{l, 1.76mmol, 1.1eq. [d=1.061]}$) was added dropwise with a syringe. A solid was seen to precipitate. After stirring the reaction mixture for 20 minutes, EtOAc ($50\text{ml, stored over Na}_2\text{SO}_4$) was added and the organic layer was extracted once with saturated KCl, then quickly dried over Na$_2$SO$_4$, filtered and the solvents removed in vacuo. The product was purified by wet flash column chromatography (silica pre-equilibrated with 1% TEA in EtOAc), eluting with EtOAc. The appropriate fractions were combined, and the solvents were removed in vacuo. The high vacuum pump was used as well in order to remove any remaining traces of TEA, which, if present, reduces the coupling efficiency of the phosphoramidite. A cream coloured foam was obtained. Yield 72% (950mg). $R_f$ 0.61 (system I); 0.47 (system IV). $^{31}$P-NMR: $\delta_p$ (CDCl$_3$) 149.02 (d).
5'-{(4,4'-Dimethoxytrityl)-N²-acetyl-O⁶-ethyl-2'-deoxyguanosine-3'-{(2-cyanoethyl-N,N-diisopropyl)-phosphoramidite

5'-{(4,4'-Dimethoxytrityl)-N²-acetyl-O⁶-ethyl-2'-deoxyguanosine (702mg, 1.1mmol) was dried by the evaporation of anhydrous pyridine (2 x 5ml) and of anhydrous THF (2 x 6ml), and dissolved in dry THF (10ml). DIPEA (765μl, 4.4mmol, 4eq.) and then 2-cyanoethyl-N,N-(diisopropylamino)chlorophosphite (268μl, 1.2mmol, 1.1eq.) was added dropwise with a syringe. A solid was seen to precipitate. After 20 minutes' stirring, EtOAc (40ml, stored over Na₂SO₄) was added to reaction mixture, and the organic layer was extracted once with saturated KCl, then quickly dried over Na₂SO₄, filtered and the solvents removed in vacuo.

The product was purified by wet flash column chromatography (silica preequilibrated with 1% TEA in EtOAc), eluting with EtOAc. The appropriate fractions were combined, and the solvents were removed in vacuo. A cream coloured foam was obtained. Yield 80% (738mg). Rₗ 0.63 (system I); 0.52 (system IV). ³¹P-NMR: δₚ (CDCl₃) 149.02 (d).
Deoxyguanosine monohydrate (1.14g, 4mmol) was coevaporated from anhydrous pyridine (3 x 5ml), suspended in dry pyridine (60ml), and trifluoroacetic anhydride (3.4ml, 24mmol, 6eq.) was added dropwise with a syringe, while cooling the reaction mixture in an ice bath. After 15 minutes, TLC showed the formation of a fluorescent compound with Rf 0. A solution of pentafluorophenol (9.6g, 52mmol) in dry pyridine (200ml) was added via a double ended needle. After a further 24 hours, the reaction mixture was concentrated in vacuo and poured into 500ml of water and extracted with EtOAc (4 x 200ml). The combined organic layers were washed with water (3 x 50ml), dried with Na₂SO₄, filtered and concentrated on the rotary evaporator. The resulting gum was redissolved in EtOAc (10ml) and added dropwise under rapid stirring to petrol ether 40-60 (700ml). A brown precipitate was formed and filtered. The precipitate was submitted to wet flash column chromatography eluting with 10% MeCN in EtOAc. The appropriate fractions were combined and dried in vacuo to a cream yellow foam. Yield 83% (1.75g). Rf 0.22 (system I); 0.47 (system III); 0.62 (system V); 0.70 (system VI). [FAB] m/z (M+H)+: found 530.07349, calculated 530.07107. 

1H-NMR: δH (d6-DMSO) 2.33-2.39 (m, 1, H2'), 2.75-2.82 (m, 1, H2"), 3.50-3.64 (m, 3, H5'S', 5'-OH), 3.87-3.91 (m, 1, H4'), 4.44-4.48 (m, 1, H3'), 5.31 (bs, 1, 3'-OH), 6.39-6.43 (t, 1, H1'), 8.79 (s, 1, H8), 12.08 (s, 1, N2-H). 13C-NMR: δC (d6-DMSO) 39.38 CH₂, 61.52 CH₂, 70.55 CH, 84.08 CH, 88.23 CH, 112.53 C, 118.00 CH, 135.12 C, 139.11 C, 140.15 C, 144.06 C, 149.60 C, 153.74 C, 154.34 C, 156.58 C, 159.50 C, 183.47 C. Anal. calcd. for C18H1N5O5F8: C 40.85, H 2.09, N 13.23; found C 40.54, H 1.86, N 12.81. UV spectrum (MeOH): λ₁=202nm, λ₂=222nm, λ₃=269nm.
O<sup>6</sup>-pentafluorophenyl-N<sup>2</sup>-trifluoroacetyl-2'-deoxyguanosine (1.70g, 3.2mmol) was dried by evaporation of anhydrous pyridine (3 x 6ml) and dissolved in dry pyridine (30ml). TEA (490µl, 3.5mmol, 1.1eq.) and DMTrCl (1.3g, 3.8mmol) were added. After 4 hours, methanol (10ml) was added to quench the reaction. The reaction mixture was then concentrated in vacuo, dissolved in ether (90ml) and washed with water (3 x 60ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and the solvent removed on the rotary evaporator. The residue was dissolved in a small amount of EtOAc and submitted to wet flash column chromatography (silica gel pre-equilibrated with 1% TEA in DCM), eluting with 0 - 2.5% MeOH in DCM. The appropriate fractions were combined and dried in vacuo, to give the title compound as a cream-yellow foam. Yield 65% (1.73g). R<sub>f</sub> 0.58 (system I); 0.71 (system III); 0.46 (system IV). [FAB] <i>m/z</i> (M+H)<sup>+</sup>: found 832.20796, calculated 832.20175. <sup>1</sup>H-NMR: <i>δ</i> (CDCl<sub>3</sub>) 2.41-2.48 (m, 1, H2'), 2.77-2.83 (m, 1, H2''), 3.36-3.38 (m, 2, H5'5''), 3.75 (s, 6, Ar-OCH<sub>3</sub>), 4.13-4.15 (m, 1, H4'), 4.64-4.67 (m, 1, H3'), 5.28 (bs, 1, 3'-OH), 6.28-6.33 (t, 1, H1'), 6.77-7.41 (m, 13, Ar), 7.83 (s, 1, H8), 12.10 (s, 1, N<sup>2</sup>-H). <sup>13</sup>C-NMR: <i>δ</i> (CDCl<sub>3</sub>) 39.53 CH<sub>2</sub>, 55.05 CH<sub>3</sub>, 63.58 CH<sub>2</sub>, 72.31 CH, 84.03 CH, 85.98 CH, 113.03 CH, 114.91 C, 126.81 CH, 127.75 CH, 127.92 CH, 129.89 CH, 135.39 C, 135.51 C, 137.20 C, 139.43 CH, 139.66 C, 141.20 C, 143.60C, 144.34C, 155.01C, 157.81 C, 158.41 C, 158.49 C. Anal. calcd. for C<sub>39</sub>H<sub>29</sub>N<sub>5</sub>O<sub>7</sub>F<sub>8</sub>·1/2C<sub>5</sub>H<sub>5</sub>N: C 57.15, H 3.76, N 8.83; found C 56.91, H 4.03, N 8.97. UV spectrum (MeOH): <i>λ</i><sub>1</sub>=202nm, <i>λ</i><sub>2</sub>=236nm, <i>λ</i><sub>3</sub>=285nm.
5'-(4,4'-Dimethoxytrityl)-O6-pentafluorophenyl-N2-trifluoroacetyl-2'-deoxyguanosine-3'-(2-cyanoethyl-N,N-diisopropyl)-phosphoramidite

5'-(4,4'-Dimethoxytrityl)-O6-pentafluorophenyl-N2-trifluoroacetyl-2'-deoxyguanosine (900mg, 1.08mmol) was coevaporated from dry pyridine (2 x 6ml) and dry THF (2 x 4ml), and dissolved in dry THF (10ml). DIPEA (750μl, 4.32mmol, 4eq.) and then 2-cyanoethyl-N,N-(diisopropylamino)chlorophosphite (265μl, 1.2mmol, 1.1eq.) was added dropwise with a syringe. A solid was seen to precipitate. After stirring the reaction for 20 minutes, EtOAc (50ml) was added, and the organic layer extracted once with saturated KCl, before being quickly dried over Na2SO4, filtered and the solvents removed on the rotary evaporator. The product was purified by wet flash column chromatography (silica preequilibrated with 1% TEA in EtOAc), eluting with EtOAc. The appropriate fractions were combined, and the solvents were removed in vacuo to obtain a cream coloured foam. Yield 94% (1045mg).

Rf 0.80 (system I); 0.71 (system IV). 31P-NMR: δ (CDCl3) 149.02 (d).
O_{6}^-\text{-pentafluorophenyl-N}^2^-\text{-trifluoroacetyl}-2'-deoxyguanosine (48mg, 0.09mmol) was dissolved in a small amount of methanol (0.5ml) and reacted with aqueous methylamine 40% (3ml) in a sealed vial for 24 hours at 60°C. The mixture was then concentrated \textit{in vacuo}. The residue was dissolved in methanol, loaded onto prep TLC plates and eluted with 10% MeOH in DCM. The band corresponding to the title compound was scraped of the plates, extracted with methanol and filtered. \( R_f \) 0.08 (system I); 0.03 (system III); 0.35 (system V). [FAB] \( m/z \) (M+H)^+: found 281.13559, calculated 281.13621. \( ^1\text{H-NMR: } \delta_\text{H} (d_6^{-\text{DMSO}}) 2.28-2.36 \) (m, 1, H2'), \( 2.49-2.64 \) (m, 1, H2''), \( 3.02 \) (s, 3, N^2^-CH_3), \( 3.38-3.53 \) (m, 3, H5', H5'', 5'-'OH), \( 3.84-3.89 \) (m, 1, H4'), \( 4.32-4.35 \) (m, 1, H3'), \( 5.28 \) (bs, 1, 3'-OH), \( 6.16-6.20 \) (t, 1, H1'), \( 6.41 \) (s, 2, NH_2), \( 7.86 \) (s, 1, N^6^-H), \( 8.15 \) (s, 1, H8). UV spectrum (MeOH): \( \lambda_1 = 201\text{nm}, \lambda_2 = 225\text{nm}, \lambda_3 = 249\text{nm} \).
4. Conclusions

Oligonucleotides containing the required chemical modifications were obtained in high purity, as shown by enzyme digestion studies and Capillary Zone Electrophoresis analyses. UV duplex melting (thermal denaturing) experiments have been carried out on duplexes containing the modified bases both paired with cytosine and thymine.

Currently the above oligonucleotides are being tested for the binding of wild type and mutant O\textsuperscript{6}-methyltransferase. Preliminary experiments suggest that they are recognised by the enzyme, which indicates that this approach to study the protein may be the right one. If the oligonucleotides will be bound by the enzyme, attempts will be made to obtain crystals of the enzyme-DNA complexes. The free oligonucleotides are also being crystallised on their own.

N\textsuperscript{6}-methyldiaminopurine may turn out to be a DNA base that mimics O\textsuperscript{6}-methylguanine, which would make it a useful tool in the study of O\textsuperscript{6}-methyltransferase, and which may help to elucidate the mechanism of DNA repair by this particular enzyme.
5. References


Part 2. Studies on DNA Triplexes
1. Introduction

1.1 Nucleic Acid Triple Helices: History and Structure

1.1.1 Discovery

The first observation of the formation of a nucleic acid triplex was made nearly four decades ago between two poly(rU) strands and one poly(rA) strand (Felsenfeld et al., 1957). Later it was shown that triple helical structures could be formed by two poly(rC) strands with guanine mononucleotides (Howard et al., 1964) and oligoribonucleotides (Lipsett, 1963 and 1964; Howard et al., 1964) when the pH was lowered, and it was postulated that the second poly(rC) strand was protonated. Morgan and Wells (1968) then reported the formation of an alternating sequence triplex between the DNA duplex d(CT)ₙ·d(GA)ₙ and an r(C⁺U)ₙ RNA third strand.

On the basis of an anti conformation for the bases, it was suggested that the third strand of a nucleic acid triplex is bound in the major groove parallel to the purine strand of the duplex by Hoogsteen (Hoogsteen, 1963) hydrogen bonds (Morgan and Wells, 1968). This hypothesis was supported by the X-ray fibre-diffraction work carried out by Arnott and co-workers (1974 and 1976), who were the first to propose a model for the molecular structure of the DNA triple helix d(T)ₙ·d(A)ₙ·d(T)ₙ, where "·" indicates Watson-Crick (W/C) and "*" Hoogsteen (HS) base pairing. A representation of a triplex is illustrated in Figure 1.

Figure 1. Representation of a triplex
1.1.2 Evidence for Triple Helix Formation

Moser and Dervan (1987) targeted duplex DNA with homopyrimidine oligonucleotides with an EDTA-Fe(II) moiety attached which could in principle bind the probe strand in parallel or antiparallel orientation. EDTA-Fe(II) cleaves the DNA backbone by oxidation of the deoxyribose with a short-lived diffusible hydroxyl radical, and is a nonspecific cleaving species. The resulting cleavage pattern proved that the homopyrimidine oligonucleotide (third or Hoogsteen strand) binds antiparallel to the pyrimidine Watson-Crick (first) strand and parallel to the purine (second) strand, confirming the binding model proposed earlier (see Section 1.1.1).

Rajagopal and Feigon (1989a,b) provided the first NMR data on the formation of a DNA triple helix, from d(GA)$_4$ and d(TC)$_4$ oligonucleotides, and observed the imino protons from protonated cytosines directly, proving their involvement in Hoogsteen hydrogen bonds. Thymine N$^3$ imino protons were shown to be implicated in both Watson-Crick and Hoogsteen base pairing in a T·A·T type triplex (Pilch et al., 1990b).

Direct evidence for the Hoogsten hydrogen bonding to the N$^7$ atoms of the purines came from NMR studies with oligodeoxynucleotides containing adenine and guanine labeled with $^{15}$N at these positions (Gaffney et al., 1995).

1.1.3 Types of Nucleic Acid Triplexes

For all triplexes, a third strand associates with a duplex tract that is purine-rich in one strand and therefore pyrimidine-rich in its complement. The pyrimidine strand of the Watson-Crick duplex is referred to as the first strand and the purine strand as the second strand. The third or Hoogsteen strand may be either pyrimidine-rich (the pyrimidine motif) or purine-rich (the purine motif).

Triplexes can be formed both by DNA and RNA as well as by their combinations. However, it has been suggested that RNA does not participate in triplexes of the purine motif (Maher, 1992b). Different research groups have obtained different results regarding the stability of mixed DNA-RNA triple helices made of the various possible combinations of strands. This is probably due to the fact that sequence and experimental conditions are likely to have an influence on relative stabilities (Sun and Hélène, 1993). However, these studies consistently
indicate that the triplexes are more stable when DNA represents the central homopurine strand than when RNA does (Frank-Kamenetskii and Mirkin, 1995).

Two major types of DNA triple helices have been described - intermolecular and intramolecular. In an intermolecular triplex, a separate third strand associates with a target duplex DNA. In an intramolecular triplex, the third strand is a portion of one of the strands of the duplex that has folded back to combine with the purine-pyrimidine tract. Intramolecular triplexes can be found either in supercoiled plasmids containing mirror repeats of purine-pyrimidine tracts or in single DNA strands that can fold back on themselves (Gee and Miller, 1992).

Sklenár and Feigon (1990) demonstrated the formation of an intramolecular foldback triplex containing both T-A*T and C-G*C base triplets where the strands were held together by loops consisting of pyrimidine nucleotides. Häner and Dervan (1990) also synthesised intramolecular foldback triplexes of the pyrimidine motif, but of opposite polarities, and showed that the triplex with the purine tract at the 5' end was more stable than the equivalent triplex with the purines located at the 3' end. Durand and co-workers (1992) used hexaethyleneglycol to link the triplex strands, and the use of other linkers, e.g. based on a terephtalamide derivative (Salunkhe et al., 1992), has also been reported.

It has been suggested that some intramolecular triplexes can form dimers (Borisova et al., 1995), but structures for such intermolecular associates have not been proposed yet.

1.1.4 Triplexes of the Pyrimidine Motif

In the pyrimidine motif, thymine residues in the third strand recognise adenines in the second strand, i.e. T-A base pairs, and cytosine recognises guanines, i.e. C-G base pairs. The latter requires cytosine protonation to enable the formation of two hydrogen bonds, and as a result the stability of the resulting base triplet increases as the pH decreases.

Some studies have suggested that the cytosine residues in the third strand of DNA triple helices are protonated at pH 7, far above their intrinsic pKa (Lavelle and Fresco, 1995), but most others, including the present study, indicated that triplexes containing C-G*C+ are significantly destabilised at neutral pH.
The T-A*T and C-G*C base triplets or triads can form in either the Hoogsteen (HS) or reverse Hoogsteen (rHS) configuration (Figure 2a and b, respectively). They are isomorphous in the Hoogsteen configuration, therefore the third strand of the triplex binds parallel to the purine strand and antiparallel to its chemically homologous pyrimidine strand. ⊕ and ⊗ indicate the relative orientations of the nucleic acid strands.

![Diagram of T-A*T and C-G*C base triplets in the (a) Hoogsteen and (b) reverse Hoogsteen configuration.](image)

**Figure 2.** T-A*T and C-G*C base triplets in the (a) Hoogsteen and (b) reverse Hoogsteen configuration
1.1.5 Triplexes of the Purine Motif

Triplex formation in the purine motif was first suggested by Hogan and co-workers (Cooney et al., 1988), whose experiments showed that in the presence of Mg$^{2+}$ a guanine-rich DNA oligonucleotide bound to a dG•dC rich sequence upstream of the human c-myc gene without the requirement of an acidic pH. Subsequent studies demonstrated that the third strand is antiparallel to the purine strand of the duplex DNA target (Beal and Dervan, 1991). This type of triplex is based on C•G•G and T•A•A base triplets in the reverse Hoogsteen configuration (Figure 3). The C•G•G and T•A•A triads within a triple helix are not isomorphous; therefore, backbone distortion occurs at each step between these triplets (Sun and Hélène, 1993).

![Figure 3. C•G•G and T•A•A in the reverse Hoogsteen configuration](image)

Triplexes of the purine motif are more versatile than the ones of the pyrimidine motif. It is possible to incorporate thymine residues in the purine-rich third strand which would then form T•A•T base triplets (Figure 4) in an otherwise pyr•pur•pur triple helix. In addition, in acidic conditions an adenine residue in the third strand may also oppose a guanine in the duplex, forming a protonated C•G•A$^+$ triad (Figure 4) (Malkov et al., 1993).
Both C-G*G and T-A*T base triplets can adopt either HS or rHS configurations. In both cases, the two triplets are not isomorphous and the HS configuration is energetically preferred, but the distortion of the third strand backbone would be more pronounced and the associated energy penalty greater for HS than for rHS configuration (Sun and Hélène, 1993).

A combined NMR-molecular dynamics study (Radhakrishnan and Patel, 1993) defined the solution structure of an intramolecular pyr-pur*pur triplex and identified the structural transitions necessary for accommodating thymines in a guanine rich third strand: an increase in axial rise and a decrease in axial twist at T-G steps.

*In vitro* experiments have produced evidence for the existence of a parallel pyr-pur*pur triplex structure - with the third purine-rich strand parallel to the purine strand involved in the Watson-Crick duplex, but it was found to be much less stable than its antiparallel counterpart (Shchyolkina *et al.*, 1995).
1.1.6 Fine Structure of Triple Helices. A-DNA versus B-DNA Conformation

Right-handed double helical DNA can adopt two distinct types of conformation: A-DNA is favoured at low humidity and B-DNA at high humidity. In both A-DNA and B-DNA the glycosidic bond is in the anti conformation.

A-DNA is characterised by a deep and narrow major groove and a shallow and wide minor groove, and it has 11 base pairs per turn with a vertical rise of 2.56Å per base pair. The bases are displaced by 4.5Å from the helix axis, creating a hollow core down the axis, and have a positive tilt angle of 20°C. The sugar rings have C3'-endo, also called N-type, puckers.

B-DNA has 10 base pairs per turn with a vertical rise of 3.3 - 3.4Å per base pair. The major and minor grooves are of similar depth. The displacement of the bases from the helix axis is much smaller than, and in the opposite direction to, that of A-DNA (-0.2 to -1.8Å), and as a result the bases straddle the helix axis. The base tilt is also smaller than that of A-DNA and is negative (-6°). The sugars have C2'-endo or S-type puckers.

The first model of a DNA triplex structure was proposed by Arnott and co-workers (1974 and 1976), based on an A-DNA like conformation of all three strands with C3'-endo sugar puckers, despite some short contacts which are generally known to be outside the allowed range. Some NMR studies appeared to support the A-DNA model (Umemoto et al., 1990).

IR data (Howard et al., 1992) and molecular modelling studies (Raghunathan et al., 1993), however, suggested a B-DNA conformation with C2'-endo sugar puckers, which had the advantage of no disallowed nonbonded distances present in the earlier A-DNA type model. NMR studies of intramolecular DNA triplexes selectively enriched with 13C in position C' on the third strand also suggest that the triple helix adopts a B-type conformation rather than an A-type one (Bornet and Lancelot, 1995). Other NMR studies of a DNA triplex containing a modified base in the third strand lend support to this conclusion by indicating that the duplex portion of the triple helix has a structure more similar to B-DNA than A-DNA (Wang et al., 1996).

It appears, however, that the situation is not as clear cut, and it is likely that the sequence of the triplex as well as the experimental conditions play a role. As a result, various studies can lead to somewhat different conclusions, but most of them conclude that DNA triplexes have a conformation that is neither completely A-form nor B-form.
While the structure of d(T-A*T)ₙ triple helices is thought to be very close to the B-form, having C²-endo sugar puckers and an axial rise per residue of ~3.3Å, a model for d(C-G*C⁺)ₙ triplexes based on the results of IR and Raman spectroscopy experiments proposed that only the guanine strand has C²-endo sugar puckers, the cytosine strands having puckers in the C³-endo region, with an axial rise per residue of ~3.9Å due to electrostatic repulsion between neighbouring positively charged base triplets (Ouali et al., 1993). For mixed sequence triplexes, these parameters will vary depending on the sequence.

Substitution of cytosine with 5-methylcytosine in either strand of a pyrimidine motif triplex does not change the conformation of the helix; IR studies with the cytosines partly or totally replaced by 5-methylcytosine have shown a relative proportion of two S-type sugars to one N-type (Fang et al., 1995a). Similar experiments have shown that the loop length of intramolecular foldback triplexes does not significantly influence their conformation (Fang et al., 1995b).

Computer modelling with a d(C-G*G)₇ and a d(T-A*T)₇ triplex suggested that the backbone geometry of the B-DNA target duplex has to be adjusted into an A-DNA like form with a deep major groove in order to accommodate the third strand in triplex formation; the sugar pucker, however, displayed a more rich variation (Cheng and Pettitt, 1995). Molecular dynamics simulations on a d(C-G*G)₇ triplex indicated that the the conformation of strands I and III is strictly neither A-form nor B-form, whereas the conformation of strand II is near the A-form (Weerasinghe et al., 1995).

The formation of a pyrimidine motif triple helix between a DNA duplex and an RNA third strand appears to be associated with a conversion of the sugars of the polypurine strand from S-type to N-type (Dagneaux et al., 1995). Within mixed DNA-RNA triplexes, the RNA strands maintain their A-type structure, as they have less conformational freedom than DNA.
1.2 Types of DNA Triplex Structures Found in Vivo and Their Potential Biological Role

Triplex DNA is formed through the binding with high sequence specificity of a third DNA strand in the major groove of a DNA duplex. They have been discovered in segments with predominantly purines in one strand (pur-pyr sequences). These sequences are overrepresented in eukaryotic DNA and have been mapped near genes and recombination hot spots (Wells et al., 1988). Several types of so-called ‘unorthodox’ structures can be adopted by this type of sequence, such as Z-DNA, cruciforms and triplexes (van Holde and Zlatanova, 1994). The formation of sequence specific triple helices can block DNA replication and transcription by inhibiting DNA binding proteins directed at overlapping sites in vitro. However, the biological role of the triplexes is as yet unclear, but it is presumed that they may be involved in DNA replication, transcription and recombination.

1.2.1 H-DNA

Sequences biased in favour of strings of contiguous purine or pyrimidine residues were found in a variety of eukaryotic organisms. They occur several times more often than would be expected from a statistical point of view, and they have some unusual properties, such as nuclease S1 hypersensitivity.

A substantial number of these sequences contain mirror-repeats; this is consistent with a proposed intramolecular triplex model for such tracts, called H-DNA (Mirkin et al., 1987). Interruptions in the regular pur-pyr sequences can occur only at the apex of the triplex without disturbing the conformation of the rest of the tract. The term ‘H form’ was used because the structural transition from B-DNA to this novel conformation occurred at acidic pH under low torsional tension and was clearly a protonated structure. A model for H-DNA is represented in Figure 5.
The H-DNA form consists of an intramolecular triple helix formed by the pyrimidine strand and half of the purine strand, leaving the other half of the purine strand - usually the 5' half - single-stranded (Frank-Kamenetskii and Mirkin, 1995).

Intramolecular triplexes in the purine motif can adopt a similar structure, called *H-DNA (or H'-DNA), where half of the pyrimidine strand remains single-stranded. As is the case for other triplexes of the purine motif, adenine residues can be replaced with thymine, and at low pH guanine can be replaced with adenine. Consequently, sequences adopting the *H form are not necessarily mirror-repeated, and they need not be homopurine-homopyrimidine either. This structure is stabilised by bivalent cations. However, different sequences have different cation requirements, and this is not well understood (Frank-Kamenetskii and Mirkin, 1995).

### 1.2.2 R-DNA

It was proposed that the triple-helical intermediate during RecA mediated homologous recombination is a novel form of structure designated as R-form or recombination form of DNA (Hsieh et al., 1990). The fundamental differences between R-DNA and the ‘classical’ triplexes are the facts that the chemically homologous strands are parallel rather than antiparallel and that any sequence can adopt an R-DNA structure. In addition, the proposed
base pairing schemes (Figure 6), arrived at through energy minimisation studies, suggest that the bases in the third strand are hydrogen-bonded to both bases of the Watson-Crick pair (Shchylkina et al., 1994; Kim et al., 1995).

Figure 6. Base pairing schemes for base triplets in R-DNA
Evidence to support the existence of R-DNA comes from in vitro experiments involving mainly chemical probing, but also other methods, such as enzymatic degradation, UV melting and binding of fluorescent dyes, on foldback triplexes (Shchyolkina et al., 1994 and 1995) as well as on plasmid DNA (Kim et al., 1995). However, the structure of the recombination intermediate is far from being understood.

1.2.3 Detection of Triplex DNA Structures in Vivo

Direct detection of H-DNA in vivo is difficult because of the complexity of genomic DNA, and as a consequence most studies used E.coli cells with recombinant plasmids containing triplex-forming inserts as a convenient model system. Certain chemicals give characteristic patterns of H- or *H-DNA modification, and as a result treatment of cultured cells with these chemicals followed by plasmid isolation and localisation of the modified bases proved to be a useful approach for the detection of these structures in vivo (Frank-Kamenetskii and Mirkin, 1995).

The presence of triplexes in vivo was probed with monoclonal antibodies raised against triple helical DNA. The antibodies were found to bind to eukaryotic chromosomes (Lee et al., 1987).

1.2.4 Implication of Triplexes in Biological Processes

Transcription

It was reported that triple helix formation significantly reduced transcription rates by RNA polymerases in vitro (Morgan and Wells, 1968). One study found that a purine-pyrimidine tract acted virtually like a transcriptional diode, i.e. transcription was effectively blocked in the direction that produces a purine-rich RNA due to triple helix formation which trapped the RNA polymerase, while transcription in the other direction was unrestricted (Grabczyk and Fishman, 1995).

In vitro experiments indicate that triple-helical complexes at homopurine-homopyrimidine sequences can block transcription initiation more efficiently than transcription elongation.
(Maher, 1992a). This suggests that oligonucleotide-directed triple helix formation can conceal the recognition sequences of DNA-binding proteins, and that helicase activities associated with elongating RNA polymerases can unwind the third strand of a triplex from the underlying duplex (Maher, 1996).

The idea that helicases may be able to disrupt triple helices is supported by experiments with Simian virus 40 (SV40) large T-antigen (T-ag) helicase. These have shown at first that DNA unwinding was significantly inhibited in regions susceptible to triple helix formation (Peleg et al., 1995). However, further studies indicated that this enzyme can unwind the third strand of DNA triplexes formed by the addition of a TFO to duplex DNA, but only if the enzyme binds to a single-stranded 3' flanking tract of the oligonucleotide first and then migrates in a 3'→5' direction, causing the release of the third strand in a process driven by ATP hydrolysis (Kopel et al., 1996). Unwinding of DNA triplexes may be required for processes such as DNA replication, transcription, recombination and repair.

The in vivo expression of genes was shown to be repressed when a poly(dG)-poly(dC) sequence capable of adopting a triplex structure under superhelical torsion in vitro, was placed 5' to a promoter, and this effect was dependent on the length of the insert (Kohwi and Kohwi-Shigematsu, 1991). These findings support the existence of triple helices in vivo.

It has been suggested that the mechanism of transcription repression by triple helix formation involves both interference with the binding of transcription factors to their specific recognition sequences and by blocking promoter assembly into initiation complexes (van Holde and Zlatanova, 1994). It has also been proposed that nucleic acid triplexes may play a role in regulating transcription elongation through a feedback mechanism: The transcription process generates levels of negative supercoiling high enough to drive the formation of unusual structures, such as triple helices, in the wake of elongating polymerases. If transcriptional activity in a given gene becomes very high, then the buildup of negative supercoils behind the polymerase could drive the formation of triplex structures within the gene, and these in turn would cause polymerases to pause until the supercoils have been relaxed by topoisomerases (van Holde and Zlatanova, 1994).
Replication

Triplexes have been suggested to play a role in DNA replication as well. This was based on the observation that DNA synthesis is prevented \textit{in vitro} by the formation of triple helical structures. Supercoiled templates containing *H-DNA lead to the premature termination of DNA synthesis at locations that correlate with the triplex border precisely (Dayn et al., 1992).

Triple helical structures can form during DNA polymerisation and efficiently block it. It was suggested that this may occur either through folding back of the template strand which will then form a triplex with the newly synthesised DNA, or through folding back of the displaced nontemplate strand to form a triplex downstream of the replication fork at an appropriate sequence (Frank-Kamenetskii and Mirkin, 1995).

Studies on the ability of DNA polymerases to extend an oligo(dA) primer through a poly(dT) tract showed that the polymerases terminate in the centre of the latter. However, replication arrest was abrogated by the addition of a single-stranded DNA binding protein or by substitution of 7-deaza-dATP for dATP (Mikhailov and Bogenhagen, 1996). These findings are consistent with the formation of a T·A+T DNA triplex between the primer stem and the folded template tract, and suggest that single-stranded DNA binding proteins may enable polymerases to replicate successfully through sequences which may be prone to intramolecular triplex formation.
1.3 Applications of DNA Triplexes

It has been suggested that there is potential for therapeutic applications in which gene expression is repressed by triplex formation - the so-called 'antigene' strategy - and that triplexes can be used as a tool in molecular biology (Gee and Miller, 1992). The formation of such sequence specific triple helices can block DNA replication and transcription by inhibiting DNA binding proteins directed at overlapping sites. As a consequence, oligonucleotide-directed triple helix formation could lead to the development of novel therapeutic agents to be used in the treatment of cancer, viral infections including AIDS and immunological diseases such as arthritis (Chubb and Hogan, 1992).

1.3.1 Specificity of Triple Helix Formation

The recognition of a purine base from the major groove forms the basis for sequence-specificity. As a consequence, base recognition of this type is limited to purine-rich tracts (Maher, 1992b).

It is possible, however, to target double helical DNA built of clusters of pyrimidines and purines by alternate strand triplex formation. The third strand may consist of adjacent homopyrimidine and homopurine blocks connected by natural 3'→5' phosphodiester linkages, forming HS and rHS hydrogen bonds, respectively, with purines on alternate strands of the target duplex (Beal and Dervan, 1992). This will result in two neighbouring triplexes, one in the pyrimidine motif and the other in the purine motif. Strand switching involving the formation of triplexes in a single structural motif would require a 3'-3' or a 5'-5' linkage between the two parts of the TFO, depending on the sequence requirements (Sun and Hélène, 1993).

The use of a polymeric linker to join two oligonucleotides which form triple helices at distant sites on a long duplex tract has also been reported (Kessler et al., 1993).

The major limitation in the use of TFOs is matching high sequence selectivity with binding that is sufficiently strong to interfere with genetic processes. Under physiological conditions, TFOs bind relatively weakly to their targets, which fact by itself favours a high sequence selectivity. However, in order to significantly influence genetic processes, the TFOs must be rather long, which limits the number of potential targets, as such long
Homopurine-homopyrimidine stretches are likely to be infrequent within a particular target gene (Frank-Kamenetskii and Mirkin, 1995). As a consequence, ways are being sought to lower the stringency of triple helix formation, if possible without compromising selectivity.

Triplexes can accommodate a few types of mismatched base triplets, but they are invariably destabilised, especially if the mismatches are in the centre of the helix rather than at its ends. Increasing the number of mismatches abrogates triple helix formation altogether.

Interruptions in the purine strand of the Watson-Crick duplex, i.e. A-T and G-C base pairs rather than the usual T-A and C-G within a pyrimidine motif triplex, can be recognised by G and T respectively, to form A-T\(\ast\)G and G-C\(\ast\)T base triplets (Yoon et al., 1992). Because the pyrimidine base in strand II can only provide one site for Hoogsteen hydrogen bonding, destabilisation of the triplex occurs.

Studies on the relative stabilities of individual mismatched base triads can give somewhat different results, probably as a consequence of different sequence contexts being studied, since stacking interactions within a triple helix are crucial for its stability. For example, the A-T\(\ast\)G triplet was shown to be destabilised when its flanking T-A\(\ast\)T triplet on the 5' side of guanine was replaced with C-G\(\ast\)C\(^+\) (Kiessling et al., 1992).

The formation of triplexes is roughly two orders of magnitude slower than duplex formation (Maher, 1992b), and it is an all-or-none process (Pilch et al., 1990a), as is duplex formation. Kinetic experiments indicate that triple helices are formed according to a nucleation-zipping model, where three to five base triplets are involved in the nucleation step, the latter being the rate-limiting step (Rouée et al., 1992). A base triplet mismatch in the centre of the target sequence disrupts the cooperativity of interactions between neighbouring base triplets, thereby destabilising the triplex structure (Mergny et al., 1991).

### 1.3.2 The Stabilisation of DNA Triple Helices

The presence of Mg\(^{2+}\) and Na\(^+\) and lowering of the pH promote triple helix formation. Three-stranded complexes are stabilised relative to their double-stranded counterparts by increasing Na\(^+\) or Mg\(^{2+}\) concentration as a result of increased electrostatic work needed to bring together three negatively charged phosphodiester chains (Riley et al., 1966).
Since pyrimidine motif triplexes require the protonation of the cytosine residues in the third (Hoogsteen) strand, they are not favoured at physiological pH, and triplexes of the purine motif are destabilised by physiological concentrations of $K^+$. However, polyamines - which are present in the cell and in greater concentrations in its nucleus - have been shown to facilitate triple helix formation at physiological pH (Hampel et al., 1991). Therefore the stability of triple helices in vivo probably depends on the interplay of a variety of factors, such as the ones mentioned above, in addition to the sequence and length of the triplexes.

The relatively low stability of the nucleic acid triplexes creates difficulties for their study, which makes the use of a high concentration of oligonucleotide and/or salt, polyamine, etc. necessary. Several approaches to stabilise the triple helices have been described, especially for the pyr-pur-pyr ones. These include the use of intercalating ligands covalently attached to one end of the Hoogsteen strand, covalent cross-linking of the triplex and incorporation in the third strand of modified bases that bind either by intercalating into the Watson-Crick duplex or by hydrogen bonding, mimicking other bases. Backbone modifications have also been investigated, especially with regard to improving oligonucleotide resistance to nucleases present in the cell. Combinations of these approaches have also been used.

However, increased stability of the triplexes entails decreased selectivity, as the TFOs, if they have very strong affinity for their target, could also bind to another site containing a mismatch (Frank-Kamenetski and Mirkin, 1995).

**Cross-linking**

Cross-linking between a TFO and its duplex DNA target blocks the action of RNA polymerases during transcription in vitro. Light-sensitive cross-linking agents, such as psoralene, when attached to one end of a TFO, lead to photoadduct formation with the target duplex at the triplex-duplex junction, therefore crosslinking the third DNA strand to the underlying duplex (Bates et al., 1995).

The photoadduct is repaired in vivo, which limits the usefulness of this approach for therapeutic purposes. However, the persistence of both the psoralen cross-link and of the corresponding purine-motif triple helix for at least 72 hours was demonstrated in vivo (Musso et al., 1996). Unrepaired cross-links may cause chromosome-specific rather than gene-specific toxicity due to inhibition of DNA replication (Maher, 1996).
Cross-linking can also be achieved through alkylation, especially of guanine residues in the target duplex, by a TFO joined to an electrophilic moiety (Povsic et al., 1990; Shaw et al., 1991). The formation of disulfide cross-links between the strands of a DNA triplex has also been reported (Goodwin et al., 1994).

**The Use of Ligands**

Minor groove binding ligands such as netropsin (Park and Breslauer, 1992) and Hoechst 33258 (Durand et al., 1994) bind to (dT)$_n$-(dA)$_n$•(dT)$_n$ type triplexes, but they destabilise the triplex while stabilising the duplex.

A series of heterocyclic compounds with several fused aromatic rings, some of them drugs used mainly in chemotherapy, have been shown to bind to DNA triplexes by intercalation due to strong stacking interactions. For example, coralyne and related compounds have a clear preference for triplexes over duplexes. Coralyne has four fused aromatic rings and binds triplexes much better than ethidium which has only three (Lee et al., 1993). The positive charge of these compounds helps them overcome the high charge density due to the three negatively charged phosphodiester chains.

Lee et al. (1993) reported that coralyne does not have a sequence preference, but experiments carried out for the present study seem to indicate otherwise. Derivatives of coralyne have been shown to intercalate preferentially in (dT)$_n$-(dA)$_n$•(dT)$_n$ rather than in d(TC)$_n$•d(GA)$_n$•d(C$^+$T)$_n$ triplexes (Latimer et al., 1995), which is probably due to the repulsion between the positive charges of the intercalating drug and of the C$^+$G$^+$C$^+$ base triads. Consequently, coralyne could be expected to behave in a similar fashion.

Two benzopyridoindole derivatives, also with four fused aromatic rings, have been shown to bind triplexes more tightly than duplexes (Mergny et al., 1992; Pilch et al., 1993). The stabilising effect of these ligands is stronger on triplexes containing T$^+$A$^+$T stretches.

A series of intercalators consisting of unfused aromatic cations have also been investigated and found to stabilise DNA triplexes (Wilson et al., 1993). The reasoning behind their design was that torsional flexibility within the ligands could be well accommodated within the triple helix, as there is a propeller twist between the bases within a triad.
DNase I footprinting studies have indicated that triplex-binding ligands promote the formation of triplexes even if the latter contain central triplet mismatches (Chandler et al., 1995). The triplex-duplex junction is favoured as an intercalation site for various ligands, which is probably due to special structural and conformational features of this particular site.

**Modified Bases**

The substitution of cytosine with 5-methylcytosine extends slightly the pH range over which a pyrimidine motif triplex is stable. Additional stabilisation is brought about by hydrophobic interactions between the methyl groups of thymine and those of 5-methylcytosine in the major groove (Sun and Hélène, 1993).

The stability of triplexes of the pyrimidine motif under physiological conditions would be strengthened if cytosines could be replaced by heterocycles that do not require protonation to form two hydrogen bonds with guanine residues in the Watson-Crick purine strand. Two such examples are pseudoisocytosine (Ono et al., 1991) and 8-oxo-adenine derivatives (Miller et al., 1992). The latter adopt a syn conformation about the glycosidic bond.

Interruptions in polypurine tracts have been targeted with TFOs containing non-natural bases that can span the major groove and form hydrogen bonds with both W/C strands (Miller, 1996), and the resulting triplexes are stable.

A non-natural base that binds preferentially at pur-pyr base pairs within a pyrimidine motif triplex, by intercalating on the 3' side with respect to the purine strand rather than by hydrogen bonding, has also been synthesised (Griffin et al., 1992). However, it does not discriminate between A-T and G-C base pairs, and the stability of the triplets it forms depends on nearest neighbour interactions.

Third strands containing an abasic site were tested over various interruptions in a polypurine-polypyrimidine tract, and it was found that they had significantly lower affinities for the duplex than some of those containing standard bases at the interruption sites (Horne and Dervan, 1991). This was attributed to the loss of base stacking interactions.
Backbone modifications

Modifications to the backbone of oligonucleotides have been used to increase nuclease resistance. Two such modifications, phosphorothioate (Kim et al., 1992) and methylphosphonate linkages (Callahan et al., 1991), appear to destabilise the triple helices.

Oligonucleotides of the (dT)$_m$ type containing $\alpha$-nucleosides have been shown to form triple helices only slightly less stable than their naturally occurring $\beta$-counterparts with (dA)$_n$ (dT)$_n$ duplexes (Le Doan et al., 1987). The orientation of the oligonucleotide third strand was found to be parallel to the purine strand of the W/C duplex, thus forming reverse Hoogsteen hydrogen bonds. If the TFO contained cytosine as well, it formed Hoogsteen hydrogen bonds, the chain orientation being antiparallel to the purine strand. This is due to the isomorphism of the T-A$\ast$T and C-G$\ast$C$^+$ base triplets formed by $\alpha$-nucleosides in the HS configuration, although the rHS configuration is energetically more stable for individual T-A$\ast$T and C-G$\ast$C$^+$ base triads (Sun and Hélène, 1993).

$\alpha$-Oligonucleotides with a cross-linking agent attached at one end and an intercalating agent at the other end have also been used. These formed very stable triplexes, as affinity cleavage reactions showed (Praseuth et al., 1988).

Polypyrmidine oligonucleotides with peptide backbones (PNA), instead of forming triplexes with a DNA duplex, have been shown to open the duplex and displace their chemically equivalent strand. Two PNA oligonucleotides then form a remarkably stable triplex with the complementary homopurine strand (Demidov et al., 1993). This type of structure is called P-loop.

1.3.3 Therapeutic Applications

Currently, most active drugs are inhibitors of proteins. It has been proposed that TFOs could form the basis for gene-targeted (antigene) drugs that might repress transcription from undesired genes in living cells. A number of studies employed TFOs to selectively modulate gene expression by altering DNA-protein interactions, both in vitro and in cultured cells.

Transcription of individual genes in cell culture was inhibited by sequence-specific triplex formation in regulatory enhancer sequences, which are crucial for binding of nuclear factors that regulate transcription (Lavrovsky et al., 1996). Oligonucleotide-dependent inhibition of
RNA polymerase II transcription initiation from the c-myc gene was observed at first in a cell-
free transcription system (Cooney et al., 1988). After exposure to the oligomer, a reduction of
the c-myc mRNA level resulted in cultured cells (Postel et al., 1991). TFOs with an attached
DNA intercalator (acridine) or cross-linking moiety (psoralen) have also been shown to act as
transcriptional repressors (Grigoriev et al., 1992 and 1993).

The action of HIV integrase was inhibited in vitro by targeting the binding sites of this enzyme
in the viral genome, which consist of a purine tract adjacent to a pyrimidine tract, with an
oligonucleotide capable of forming a continuous alternate strand triple helix; this prevented the
insertion of the viral genome into its target DNA (Bouziane et al., 1996). It is clear, therefore,
that blocking the recognition sites of DNA-binding proteins represents a useful approach to
therapy, with a wider scope than just the repression of transcription.

However, currently the use of TFOs in therapy suffers from important constraints regarding
requirements for stabilising binding conditions, restrictions on permitted target sequences, and
inefficient delivery of oligonucleotides to the cell nucleus (Maher, 1996). In addition, it has
been suggested that triplex formation cannot be targeted at DNA that is tightly associated with
nucleosome cores (Brown and Fox, 1996). If solutions to these problems are found,
oligonucleotide-directed triplex formation could become a viable approach to cancer therapy.

1.3.4 DNA Triples as a Tool in Molecular Biology

The formation of DNA triples can also be used as a tool in molecular biology. Site-specific
cleavage of large DNA fragments has been achieved by targeting homopyrimidine
oligonucleotides with an attached EDTA-Fe(II) moiety to a region of duplex DNA with a
contiguous run of purines in one strand (Moser and Dervan, 1987). In case of large DNA
fragments, the selectivity of restriction endonucleases may not be sufficient, as their target
sites are usually of 4-8 base pairs in size, but the TFOs are longer than that and therefore more
specific. However, the cleavage efficiency of this procedure was quite low (15-25%) and
needed refining in order to increase the yield.

Two approaches to solve this problem have been reported. One approach involves the
attachment of DNA cleaving enzymes to oligonucleotides. An oligonucleotide-staphylococcal
nuclease adduct cleaved plasmid DNA in 75% yield (Pei et al., 1990), but multiple cleavage
patterns were observed at the binding site, probably due to the conformational flexibility of the attached enzyme (Dervan, 1992).

The second approach is termed ‘Achilles' heel cleavage’ and involves transient site-specific protection by triple helix formation, from enzymatic methylation, followed by disruption of the triplex and consequent cleavage by a restriction enzyme (Strobel and Dervan, 1991). A 340 kilobase yeast chromosome was cleaved in this way at a unique site in near quantitative yield (Strobel and Dervan, 1991), and a 200 megabase human chromosome was cleaved in 80-90% yield (Strobel et al., 1991).

Sequence-specific double-strand alkylation mediated by triple helix formation can also lead to site-specific cleavage of DNA (Povsic et al., 1992). Attachment of an alkylating agent to the 5' end of a homopyrimidine oligonucleotide gave alkylation of a guanine to the 5' side of the local triple helix. Two such oligonucleotide-electrophile conjugates bound adjacent inverted purine tracts on duplex DNA by triple helix formation and alkylated single guanine residues on opposite strands of the duplex. Depurination at the alkylated sites gave a double-strand break with specific overhangs, suitable for ligation by restriction enzymes.

One limitation of triplex-mediated cleavage of DNA is that it can only be applied to mostly-purine tracts or to tracts containing consecutive runs of purines and pyrimidines (Dervan, 1992).

Formation of DNA triplexes at sequences that overlap protein binding sites, inhibits DNA binding by these proteins. By determining the 3' and 5' triplex interference boundaries, the core DNA-binding sequence for topoisomerase II has been determined (Spitzner et al., 1995). This approach may be useful for the characterisation of DNA-binding sites of other proteins as well.

Another application is the purification of double-stranded DNA by triplex-mediated affinity capture (Ji and Smith, 1993). The target DNA from bacterial cell lysates formed triple helices with biotinylated oligonucleotides immobilised on magnetic beads. Elution from the beads gave the desired DNA plasmids in high purity. This procedure can speed up DNA purification significantly and, in addition, it could be automated, which would be extremely useful, especially for large scale projects, such as the sequencing of the human genome (Jiu and Smith, 1993).
1.4 Aims

The main problem when studying DNA triplexes is their low stability. In order to overcome it, high concentrations of oligonucleotide and salt are usually required. The aim of this study was to synthesise oligonucleotides capable of folding back on themselves twice and thus form intramolecular triplexes, which are much more stable than triplexes made from single strands. The conformations of such intramolecular triple helices have been shown to be identical to the conformations of triplexes made up of separate (unlinked) strands, provided that the sequences of the two types of triplexes are identical (Durand et al., 1992).

The triplexes were to be investigated by UV melting to determine how various factors, such as their sequence and length, the nature of the linkers, pH (where appropriate), cation concentration and the addition of a triplex binding drug, affect their stability.

Another aim was to carry out NMR studies and crystallisation experiments with the triplexes. DNA triple helices made of separate strands tend to give fiber-type rather than normal X-ray diffraction patterns, even if single crystals are obtained (Liu et al., 1996). It was hoped that the crystallisation of intramolecular triplexes would provide a solution to this problem. The usefulness of the triple helix binding drug coralyne in crystallisation was assessed in these experiments as well, because drugs are often helpful in obtaining DNA crystals.
2. Results and Discussion

Triplex-forming oligonucleotides with appropriate sequences were synthesised, linked by hexaethyleneglycol (HEG) or hydrocarbon chains. Such oligonucleotides are capable of folding back on themselves to give triplexes which have identical conformations to those of the unlinked ones, but with significantly increased stabilities; they are called foldback triplex forming oligonucleotides (FTFOs). The strands of the triple helical structures are linked with hexaethyleneglycol (HEG) or two 1,8-octanediol (Ocd) moieties. HEG and Ocd were chemically modified so that they could be incorporated in oligonucleotides by the phosphoramidite method.

These oligonucleotides were synthesised in larger quantities for crystallisation experiments and they were to be studied by UV melting. For comparison purposes, the duplexes corresponding to the intramolecular triplexes were synthesised as well and included in the UV melting studies; the duplex strands were also linked by hexaethyleneglycol or octanediol chains.

In addition, CD and NMR studies were conducted in collaboration with Dr. A. Lane from the National Institute of Medical Research, London.

Durand et al. (1990) were the first to report the use of HEG to link the strands of a DNA duplex and then of a triplex (1992) to be studied thermodynamically and by circular dichroism (CD). HEG was chosen because it allows a large mobility of the oligonucleotide parts of the molecule and because the presence of the oxygen atoms allows a good solubility of the compounds.

For the purpose of crystallisation the HEG linker might make the oligonucleotides too soluble, so a linker of similar length was chosen but with a different structure (Figure 7); it is made of two 1,8-Octanediol molecules linked by a phosphodiester bond which is formed during automated DNA synthesis.

\[
\begin{align*}
\text{(a)} & \quad \text{HEG} & \text{(b)} & \quad \text{Ocd} \\
\text{O} & \quad \text{O} & \quad \text{O} \\
-\text{O-(CH}_2\text{-CH}_2\text{-O)}_6- & \quad \text{O-(CH}_2\text{)}_8\text{-O-P-O-(CH}_2\text{)}_8\text{-O} & \quad \text{O}\text{.}
\end{align*}
\]

**Figure 7.** HEG (a) and Ocd (b) linkers
The Ocd linker is only slightly longer than the HEG linker - 22 bonds compared to 20, and has fewer oxygen atoms and should therefore be more hydrophobic and help the triplexes come out of solution more easily during the crystallisation process.

2.1 Synthesis of Foldback Triplex Forming Oligonucleotides

The triplex sequences chosen for study were of the pyrimidine-purine-pyrimidine type: T·A·T on the one hand, and a mixed sequence alternating T·A·T and C·G·C+ base triplets on the other hand.

The two linkers used were hexaethyleneglycol (HEG) and a linker made of two 1,8-Octanediol (Ocd) units respectively, connected with the DNA bases via phosphodiester bonds.

2.1.1 Synthesis of the Linker Phosphoramidite Monomers

The synthesis of the linker phosphoramidite monomers suitable for incorporation into DNA was based on the method described by Durand et al. (1990) for HEG, and comprises two steps (Figure 8): dimethoxytritylation followed by phosphitylation.

Since both HEG and Ocd have two identical -OH groups in their molecules which can be dimethoxytritylated, an excess of HO-X-OH to DMTr-Cl of 5 : 1 was used. The monodimethoxytritylated compounds can then be separated from the reaction mixture by wet flash column chromatography on silica gel.
2.1.2 Oligonucleotide Synthesis

After the transformation of HEG and Ocd in suitable phosphoramidite monomers, the synthesis of the oligonucleotides involves straightforward solid support automated DNA synthesis by the phosphoramidite method.

The order of the individual DNA strands in the FTFOs is important, as this determines whether a triplex would form or not. There are two possible orientations for the third (pyrimidine) strand within the triple helix: parallel or antiparallel to the purine strand, corresponding respectively to Hoogsteen and reverse Hoogsteen binding of the bases as discussed in Section 1.1.4.

As both T·A·T and C·G·C⁺ base triplets favour the Hoogsteen configuration, the third DNA (pyrimidine) strand must be parallel to the purine strand and antiparallel to the other pyrimidine strand involved in duplex formation, in order for the triplex to be able to form. As a consequence, FTFOs of the following sequences were synthesised (Figures 9 and 10):

- \( dA_n-x-dT_n-x-dT_n \)
- \( dT_n-x-dT_n-x-dA_n \)
• $d(GA)_3 \times d(TC)_3 \times d(CT)_3$

• $d(TC)_3 \times d(CT)_3 \times d(AG)_3$

Note: In the above, $n$ is 6, 8 or 10 and $x$ is HEG or $(Ocd)_2$.

In addition, sequences with the purine strand in the middle of the oligonucleotide were synthesised for comparison in the UV melting studies.

The oligonucleotides were deprotected in concentrated ammonia, purified by reverse phase HPLC, desalted on NAP columns containing Sephadex G-25 and freeze-dried. As CZE analyses showed, the oligonucleotides were obtained in high purity. Examples of the CZE traces obtained are shown in the appropriate experimental section.
2.2 UV melting

2.2.1 Preliminary Experiments

It is well known that the presence of Na$^+$ and/or Mg$^{2+}$ favours DNA triplex formation because the electrostatic repulsion between three negatively charged phosphodiester chains is greater than that between the two chains in a duplex and would prevent triple helix formation in the absence of cations. Polyvalent cations are more effective than monovalent ones at promoting triplex formation.

Preliminary UV melting experiments were carried out on the sequence A$_8$-T$_8$-T$_8$ (HEG) [dA$_8$-HEG-dT$_8$-HEG-dT$_8$] in order to determine the optimal conditions for triplex formation and observation. The results are shown in Table 1.

Table 1. UV melting temperatures of A$_8$-T$_8$-T$_8$ (HEG) under various conditions

<table>
<thead>
<tr>
<th>Na conc (mM)</th>
<th>Salt</th>
<th>Salt conc (M)</th>
<th>pH</th>
<th>EDTA conc (mM)</th>
<th>$\lambda$(nm)</th>
<th>$T_m$(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>MgCl$_2$</td>
<td>0.01</td>
<td>6.0</td>
<td>-</td>
<td>260</td>
<td>57.6</td>
</tr>
<tr>
<td>50</td>
<td>MgCl$_2$</td>
<td>0.01</td>
<td>6.0</td>
<td>-</td>
<td>284</td>
<td>52.5</td>
</tr>
<tr>
<td>50</td>
<td>MgCl$_2$</td>
<td>0.002</td>
<td>6.0</td>
<td>-</td>
<td>260</td>
<td>51.7</td>
</tr>
<tr>
<td>10</td>
<td>MgCl$_2$</td>
<td>0.001</td>
<td>7.0</td>
<td>-</td>
<td>260</td>
<td>52.6</td>
</tr>
<tr>
<td>50</td>
<td>NaCl</td>
<td>0.1</td>
<td>6.0</td>
<td>0.2</td>
<td>260</td>
<td>39.8</td>
</tr>
<tr>
<td>50</td>
<td>NaCl</td>
<td>0.1</td>
<td>6.0</td>
<td>0.2</td>
<td>284</td>
<td>35.1</td>
</tr>
<tr>
<td>10</td>
<td>NaCl</td>
<td>0.1</td>
<td>7.0</td>
<td>0.2</td>
<td>260</td>
<td>36.5</td>
</tr>
<tr>
<td>10</td>
<td>NaCl</td>
<td>0.05</td>
<td>7.0</td>
<td>0.2</td>
<td>260</td>
<td>29.7</td>
</tr>
<tr>
<td>10</td>
<td>NaCl</td>
<td>0.02</td>
<td>7.0</td>
<td>0.2</td>
<td>260</td>
<td>20.9</td>
</tr>
</tbody>
</table>
The UV melting buffers containing NaCl allowed the observation of both the triplex-to-duplex and duplex-to-coil transitions when the Na cacodylate (ccod) concentration was kept relatively low (10mM as opposed to 50mM), whereas the ones with MgCl₂ did not, even at low salt and Na ccod concentrations. This is probably due to the fact that Mg²⁺ ions bind much more strongly to DNA than Na⁺. The addition of EDTA to the NaCl buffers is necessary to cancel out the action of any Mg²⁺ that may be present. The UV melting curves of A₈T₈T₈ (HEG) at three different NaCl concentrations in 10mM Na ccod buffer are shown in Figure 11.

![UV melting profiles of A₈T₈T₈ (HEG) in 0.02, 0.05 and 0.1M NaCl UV melting buffers](image)

Figure 11. UV melting profiles of A₈T₈T₈ (HEG) in 0.02, 0.05 and 0.1M NaCl UV melting buffers

The variation of the melting temperatures $T_{m1}$ (triplex-to-duplex) and $T_{m2}$ (duplex-to-coil) of the FTFO with varying NaCl concentration is represented in Figure 12.
Figure 12. Variation of the UV melting temperatures of $A_8T_8T_8$ (HEG) with increasing NaCl concentration

As the graph shows, $T_{m1}$ increases more rapidly than $T_{m2}$ when the NaCl concentration rises, which proves that, as expected, the stability of the triplex is influenced by salt concentration to a greater extent than that of the duplex.

For further experiments the following buffer was chosen: 10mM Na cacodylate / 0.05M NaCl / 0.2mM EDTA. This permits the observation of both transitions, which are, in addition, well spaced.

Previous studies on triple-helix formation of the T:A*T type (Riley et al., 1966; Cassani and Bollum, 1969) showed that the melting of the third strand from the underlying duplex - the Hoogsteen (HS) transition - is accompanied by an increase in absorbance at 284nm as well as at 260nm. The melting of the duplex - the Watson-Crick (WC) transition - is accompanied by an absorbance increase at 260nm, but not at 284nm. However, as can be seen from Table 1, the melting temperatures obtained for the HS transition at 284nm appear to be significantly lower than the ones obtained at 260nm. Also, the amplitude of the absorbance change for the HS transition at 284nm was much smaller. Consequently, all observations were carried out at 260nm.
Durand et al. (1992) have shown that for this type of FTFO the melting temperature is independent of the DNA concentration, thus proving the formation of intramolecular triplexes and ruling out the possible existence of intermolecular ones.

### 2.2.2 UV Melting of the (T·A·T)₈ Triplexes in the Presence of NaCl

The UV melting behaviour of the (T·A·T)₈ type oligonucleotides is detailed in Table 2. The buffer used was: 10mM Na cacodylate / 0.05M NaCl / 0.2mM EDTA (pH 7.0).

#### Table 2. UV melting temperatures of (T·A·T)₈ triplexes

<table>
<thead>
<tr>
<th>Sequence</th>
<th>HEG Tₘ₁(°C)</th>
<th>HEG Tₘ₂(°C)</th>
<th>(Ocd)₂ Tₘ₁(°C)</th>
<th>(Ocd)₂ Tₘ₂(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₈T₈T₈</td>
<td>28.2</td>
<td>45.3</td>
<td>30.3</td>
<td>44.2</td>
</tr>
<tr>
<td>T₈A₈T₈</td>
<td>-</td>
<td>48.0</td>
<td>-</td>
<td>46.3</td>
</tr>
<tr>
<td>T₈T₈A₈</td>
<td>19.1</td>
<td>48.0</td>
<td>26.0</td>
<td>46.2</td>
</tr>
</tbody>
</table>

As expected, the sequence with the purine strand in the middle does not form a triplex; it just shows a duplex-to-coil transition.

Surprisingly, the A₈T₈T₈ and T₈T₈A₈ sequences do not have identical melting temperatures. The triplex-to-duplex transition of A₈T₈T₈ occurs at a higher temperature whereas the duplex-to-coil transition at a lower temperature than the ones of T₈T₈A₈, for both linkers. This may be due to the sterical influence of the linkers themselves.

However, the duplex-to-coil transition of the FTFOs occurs at a temperature close to the one of the T₈A₈T₈, which only forms a double-stranded structure with a dangling end x-T₈ or T₈-x. In case of the triplex-to-duplex transition, the linker seems to make a greater difference.
2.2.3 UV Melting of the Mixed Sequence Triplexes in the Presence of NaCl

The mixed sequence triplexes consist of alternating T:A*T and C-G:C* base triplets. Since the C-G:C* triplet requires a protonated cytosine, the stability of a triplex containing such base triplets is pH dependent. The UV melting profiles change significantly as the pH changes, which is illustrated for the oligonucleotide GA.TC.CT (HEG) in Figure 13.

The UV melting temperatures of the mixed sequence triplexes as a function of pH, are detailed in Table 3. The buffer used was: 10mM Na ccod / 0.05M NaCl / 0.2mM EDTA.

Table 3. UV melting temperatures of mixed sequence triplexes as a function of pH

<table>
<thead>
<tr>
<th>Sequence</th>
<th>pH</th>
<th>HEG Tm1(°C)</th>
<th>HEG Tm2(°C)</th>
<th>(Ocd)2 Tm1(°C)</th>
<th>(Ocd)2 Tm2(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA.TC.CT [d(GA)3-x-d(TC)3-x-d(CT)3]</td>
<td>7.5</td>
<td>9.6</td>
<td>56.9</td>
<td>11.9</td>
<td>53.8</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>18.8</td>
<td>57.1</td>
<td>22.9</td>
<td>54.0</td>
</tr>
<tr>
<td></td>
<td>6.5</td>
<td>23.8</td>
<td>57.6</td>
<td>31.7</td>
<td>53.6</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>38.0</td>
<td>55.9</td>
<td></td>
<td>45.0</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>50.8</td>
<td></td>
<td>52.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>56.8</td>
<td></td>
<td>58.2</td>
<td></td>
</tr>
<tr>
<td>TC.GA.TC [d(TC)3-x-d(GA)3-x-d(TC)3]</td>
<td>6.0</td>
<td>13.7</td>
<td>59.3</td>
<td>16.5</td>
<td>56.4</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>19.0</td>
<td>59.6</td>
<td>-</td>
<td>52.7</td>
</tr>
<tr>
<td>TC.CT.AG [d(TC)3-x-d(CT)3-x-d(AG)3]</td>
<td>7.5</td>
<td>-</td>
<td>-</td>
<td>9.9</td>
<td>58.5</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>10.1</td>
<td>63.1</td>
<td>17.5</td>
<td>59.5</td>
</tr>
<tr>
<td></td>
<td>6.5</td>
<td>18.1</td>
<td>62.4</td>
<td>27.3</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>32.4</td>
<td>62.1</td>
<td>42.83</td>
<td>55.1</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>46.5</td>
<td>58.6</td>
<td>53.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>57.4</td>
<td></td>
<td>60.0</td>
<td></td>
</tr>
</tbody>
</table>
Figure 13. UV melting curves of GA.TC.CT (HEG) at pH (a) 7.5, (b) 7.0, (c) 6.5, (d) 6.0, (e) 5.5 and (f) 5.0
As expected, the stability of the triplexes increases significantly as the pH decreases. The melting temperature of the duplex-to-coil transition decreases slightly at lower pHs, as the more extensive protonation of the cytosines starts to affect the Watson-Crick hydrogen bonds as well.

In case of the Ocd linked oligonucleotides, the triplex-to-duplex transitions occur consistently at higher temperatures, and the duplex-to-coil transitions at lower temperatures, than the equivalent transitions of the HEG linked ones. For all mixed sequence FTFOs, as the pH is lowered, the two melting temperatures are getting closer, i.e. the temperature difference between the two transitions decreases. At low pH (5.0, 5.5 and even 6.0 in one case) only one transition was observed. At pH 5.5 the observed transition was broader than that at pH 5.0, which suggests that the triplex-to-duplex and the duplex-to-coil transitions occur close together and are unresolved.

The dependence of the melting temperatures $T_{m1}$ and $T_{m2}$ of GA.TC.CT (HEG) on pH is illustrated in Figure 14.

![Figure 14](image-url)
The apparently lower $T_m$ of the duplex at pH 5.5 is probably an artefact due to the fact that the two transitions are close together but unresolved.

The sequence GA.TC.CT and its reverse - TC.CT.AG - behave differently. For both linkers the triplex formed by GA.TC.CT seems to be more stable, whereas the duplex is less stable than that of TC.CT.AG. As the pH is lowered, the difference in triplex melting temperatures decreases, and at low pH values (5.5 and 5.0), where the two melting transitions are unresolved, TC.CT.AG appears to be more stable. However, the latter observation could arise as a consequence of the fact that the determined single melting temperature has a value inbetween the ones of the HS and the W/C transitions, and for TC.CT.AG the latter occurs at a higher temperature.

Surprisingly, the sequence with the purine strand in the middle does appear to form a triplex at acidic pH, but the corresponding melting temperatures are quite low; this may due to the formation of intermolecular triplexes, or to intramolecular triplexes with antiparallel reverse Hoogsteen binding of one pyrimidine to the purine strand, which are less stable.

2.2.4 UV Melting Behaviour of T-A•T Triplexes Compared to That of Their Corresponding Duplexes in the Presence of MgCl₂

This set of experiments was intended to compare the UV melting behaviour of the T-A•T FTFOs with that of their corresponding foldback duplex forming oligonucleotides, the strands of which are being held together by the same linkers.

When carrying out experiments on the sequence $A_6.T_6.T_6$ (Ocd), it became apparent that the two transitions for this FTFO could not be resolved, even at low salt concentrations. This is probably due to the fact that the sequence is quite short. The UV melting temperatures obtained for this FTFO under different conditions, are detailed in Table 4. The melting temperature of the duplex $A_6.T_6$ (Ocd) at one of the salt concentrations used is shown for comparison purposes. All UV melting buffers used for the study of T-A•T triplexes had pH 7.0.
Table 4. UV melting of $A_6.T_6.T_6(Ocd)$ at various salt concentrations

<table>
<thead>
<tr>
<th>Na ccod conc (mM)</th>
<th>EDTA conc (mM)</th>
<th>Salt</th>
<th>Salt conc (M)</th>
<th>$T_m(°C)$ duplex</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.2</td>
<td>NaCl</td>
<td>0.05</td>
<td>32.8</td>
</tr>
<tr>
<td>10</td>
<td>0.2</td>
<td>NaCl</td>
<td>0.02</td>
<td>27.6</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>18.9</td>
</tr>
</tbody>
</table>

It was then tried to increase the salt concentration in such a way as to make the HS and W/C transitions of the T-$A\cdot T$ triplexes occur simultaneously, thus obtaining monophasic melting curves, and then to determine whether the triplexes would be more stable than their respective duplex oligonucleotides.

A couple of triplex sequences and their corresponding duplexes have been used in order to establish optimal conditions for the proposed experiments. The results are shown in Table 5.

Table 5. UV melting temperatures of some T-$A\cdot T$ triplexes under different conditions

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Na ccod conc (mM)</th>
<th>Salt</th>
<th>Salt conc (M)</th>
<th>EDTA conc (mM)</th>
<th>$T_m(°C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_8.T_8.T_8(Ocd)$</td>
<td>10</td>
<td>NaCl</td>
<td>0.25</td>
<td>1</td>
<td>53.3</td>
</tr>
<tr>
<td>$A_8.T_8(Ocd)$</td>
<td>10</td>
<td>NaCl</td>
<td>0.25</td>
<td>1</td>
<td>58.2</td>
</tr>
<tr>
<td>$A_8.T_8.T_8(Ocd)$</td>
<td>10</td>
<td>NaCl</td>
<td>0.5</td>
<td>1</td>
<td>58.8</td>
</tr>
<tr>
<td>$A_8.T_8(Ocd)$</td>
<td>10</td>
<td>NaCl</td>
<td>0.5</td>
<td>1</td>
<td>61.7</td>
</tr>
<tr>
<td>$A_6.T_6.T_6(Ocd)$</td>
<td>10</td>
<td>NaCl</td>
<td>0.5</td>
<td>1</td>
<td>48.9</td>
</tr>
<tr>
<td>$A_6.T_6(Ocd)$</td>
<td>10</td>
<td>NaCl</td>
<td>0.5</td>
<td>1</td>
<td>50.3</td>
</tr>
<tr>
<td>$A_6.T_6.T_6(Ocd)$</td>
<td>10</td>
<td>MgCl$_2$</td>
<td>0.005</td>
<td>-</td>
<td>48.0</td>
</tr>
<tr>
<td>$A_6.T_6(Ocd)$</td>
<td>10</td>
<td>MgCl$_2$</td>
<td>0.005</td>
<td>-</td>
<td>47.4</td>
</tr>
<tr>
<td>$A_6.T_6.T_6(Ocd)$</td>
<td>10</td>
<td>MgCl$_2$</td>
<td>0.01</td>
<td>-</td>
<td>49.5</td>
</tr>
<tr>
<td>$A_6.T_6(Ocd)$</td>
<td>10</td>
<td>MgCl$_2$</td>
<td>0.01</td>
<td>-</td>
<td>48.5</td>
</tr>
</tbody>
</table>
A buffer containing 10mM Na acetate / 10mM MgCl₂ (pH 7.0) was chosen for the subsequent experiments. MgCl₂ has a much stronger effect on the melting temperatures of DNA triplexes than NaCl, and much smaller quantities of salt are required. At the concentrations used, saturation of the triplexes with Mg²⁺ seems to have been achieved, as only a small rise in melting temperature has occurred after doubling the MgCl₂ concentration from 5mM to 10mM. In addition, the melting of the FTFOs occurred at approximately the same or even at a slightly higher temperature than that of their respective duplex oligonucleotides.

The UV melting temperatures of all T·A·T FTFOs and of their corresponding duplex oligonucleotides in the 10mM MgCl₂ buffer are detailed in Table 6.

**Table 6.** UV melting temperatures of T·A·T triplexes and of their corresponding duplexes

<table>
<thead>
<tr>
<th>Sequence</th>
<th>HEG</th>
<th>(Ocd)₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₆·T₆·T₆ (dA₆-x-dT₆-x-dT₆)</td>
<td>49.7</td>
<td>49.5</td>
</tr>
<tr>
<td>A₆·T₆ (dA₆-x-dT₆)</td>
<td>49.7</td>
<td>48.5</td>
</tr>
<tr>
<td>T₆·T₆·A₆ (dT₆-x-dT₆-x-dA₆)</td>
<td>49.4</td>
<td>52.8</td>
</tr>
<tr>
<td>T₆·A₆ (dT₆-x-dA₆)</td>
<td>49.1</td>
<td>48.7</td>
</tr>
<tr>
<td>A₈·T₈·T₈ (dA₈-x-dT₈-x-dT₈)</td>
<td>59.5</td>
<td>59.7</td>
</tr>
<tr>
<td>A₈·T₈ (dA₈-x-dT₈)</td>
<td>59.5</td>
<td>59.3</td>
</tr>
<tr>
<td>T₈·T₈·A₈ (dT₈-x-dT₈-x-dA₈)</td>
<td>60.0</td>
<td>61.1</td>
</tr>
<tr>
<td>T₈·A₈ (dT₈-x-dA₈)</td>
<td>60.0</td>
<td>59.7</td>
</tr>
<tr>
<td>A₁₀·T₁₀·T₁₀* (dA₁₀-x-dT₁₀-x-dT₁₀)</td>
<td>65.0</td>
<td>65.1</td>
</tr>
<tr>
<td>T₁₀·T₁₀·A₁₀* (dT₁₀-x-dT₁₀-x-dA₁₀)</td>
<td>64.9</td>
<td>66.0</td>
</tr>
<tr>
<td>A₄·T₄·T₄ (dA₄-x-dT₄-x-dT₄)</td>
<td>-</td>
<td>28.9</td>
</tr>
<tr>
<td>A₄·T₄ (dA₄-x-dT₄)</td>
<td>-</td>
<td>**</td>
</tr>
</tbody>
</table>

Notes:  * No duplexes were available for comparison.
** Melting did not occur in the temperature range studied
As can be seen from the table above, the melting of the entire set of FTFOs occurred at the same or at slightly higher temperatures than that of the respective duplexes. The temperature difference between the melting of the triplex and that of the corresponding duplex is generally extremely small and could arise as a consequence of the calculations carried out to obtain the melting temperatures. There is, however, the possibility that the triplexes are slightly more stable than the duplexes. The nature of the linker seems to make relatively little difference to the thermal stability of the triplexes or of the duplexes.

The oligonucleotides $A_4T_4T_4$ (Ocd) and $A_4T_4$ (Ocd) were included in the study to see what impact the presence of the linker has on a very short sequence. Unlinked strands of $dA_4$ and $dT_4$ cannot form a stable duplex or a triplex. Because of their shortness (end effects) and the fact that only two W/C hydrogen bonds can form between dA and dT, the strands would come apart as soon as they had bound. The presence of the linker in the $A_4T_4$ (Ocd) duplex oligonucleotide is not enough to stabilise the double helix, and as a consequence no melting was observed. However, the triplex $A_4T_4T_4$ (Ocd) is quite stable, as its melting temperature of 28.9 °C shows. This clearly shows the stabilising effect the presence of the linkers in the molecule brings about.

The dependence of the melting temperatures of the (A.T.T)$_n$ FTFOs on their length, i.e., on the number of base triplets ‘n’ in the sequence, is shown in Figure 15. The melting profiles of these sequences are shown in Figure 16.
Figure 15. Variation of the UV melting temperature of (A.T.T)$_n$ triplexes with their length

Figure 16. UV melting curves of (A.T.T)$_n$ triplexes
It is clear that the longer the triplex, the higher its UV melting temperature is. However, the increase in melting temperature gets smaller as the number of base triplets in the oligonucleotides ‘n’ increases.

2.2.5 UV Melting Behaviour of the Mixed Sequence Triples Compared to That of Their Corresponding Duplexes in the Presence of MgCl₂

As was the case for the T·A·T triples, the UV melting temperatures of the mixed sequence triples were compared to those of their corresponding foldback duplex forming oligonucleotides, the strands of which are being held together by the same linkers (Table 7). Buffers of the same salt concentration were used - 10mM Na ccod / 10mM MgCl₂ - at a pH of 5.5 and then at 7.0.

Table 7. UV melting temperatures of mixed sequence triples and of their corresponding duplexes

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Tₘ(°C) / pH=5.5</th>
<th>Tₘ(°C) / pH=7.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HEG</td>
<td>(Ocd)₂</td>
</tr>
<tr>
<td>GA.TC.CT [d(GA)₃-x-d(TC)₃-x-d(CT)₃]</td>
<td>57.5</td>
<td>57.8</td>
</tr>
<tr>
<td>GA.TC [d(GA)₃-x-d(TC)₃]</td>
<td>61.6</td>
<td>60.1</td>
</tr>
<tr>
<td>TC.CT.AG [d(TC)₃-x-d(CT)₃-x-d(AG)₃]</td>
<td>52.8</td>
<td>65.1</td>
</tr>
<tr>
<td>CT.AG [d(CT)₃-x-d(AG)₃]</td>
<td>66.0</td>
<td>63.8</td>
</tr>
</tbody>
</table>

As expected, at pH 7.0 the triplex-to-duplex transition takes place at relatively low temperatures. In addition, the duplex-to-coil transition occurs at temperatures up to 2°C higher than that of the corresponding duplex oligonucleotide.

At pH 5.5, the melting of the duplex oligonucleotides takes place at higher temperatures than the equivalent transitions of the FTFOs. This probably means that the two transitions for the
FTFOs under these conditions do not occur simultaneously and are unresolved; only one transition can be seen, the melting temperature of which has a value in-between those of the HS and W/C transitions. In case of TC.CT.AG (HEG), two distinct transitions are visible.

The latter FTFO was then submitted to UV melting at pH 5.5 at higher MgCl$_2$ concentrations (the concentration of Na ccod was kept constant at 10mM), to see if the gap between the temperatures of the two transitions could be narrowed. The results are detailed in Table 8. For comparison purposes, some results at identical salt concentrations but at pH 7.0 were included.

**Table 8. UV melting temperatures of TC.CT.AG (HEG) at increasing MgCl$_2$ concentrations**

<table>
<thead>
<tr>
<th>pH</th>
<th>10mM</th>
<th>20mM</th>
<th>50mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>52.8</td>
<td>65.1</td>
<td>48.2</td>
</tr>
<tr>
<td>7.0</td>
<td>12.8</td>
<td>66.6</td>
<td>-</td>
</tr>
</tbody>
</table>

At pH 5.5 the temperature of the HS transition decreases as the MgCl$_2$ concentration increases. This destabilisation of the triplex structure may be due to electrostatic repulsion between Mg$^{2+}$ and the positively charged C-G*C$^+$ base triplets.

**2.2.6 UV Melting Temperatures of DNA Triplex-Coralyne Complexes**

Coralyne (Figure 1) is a DNA binding drug that possesses significant antitumor activity (Zee-Cheng *et al.*, 1974). It has been shown to intercalate into both triplex (Lee *et al.*, 1993) and duplex DNA, and it also stacks outside the helix, particularly at low ionic strengths (Wilson *et al.*, 1976). However, its affinity for triplexes seems to be greater than that for duplexes due to improved stacking interactions, and in its presence the UV melting temperature of the triplex has been shown to be higher than that of the duplex (Lee *et al.*, 1993). It may well be that the drug's antitumor activity is due to its binding to triplex rather than to duplex DNA, as the former structure has been found *in vivo*, and it is assumed to fulfil
a potentially important biological role. According to the work carried out by Lee et al. (1993), coralyne intercalates at every other base triplet in T·A*T type triplexes, and at every base triplet in triplexes where C·G·C+ and T·A*T alternate, and it shows little sequence preference. However, in view of the fact that it is positively charged, coralyne could be expected to favour consecutive T·A*T base triplets rather than intercalate at alternating T·A*T - C·G·C+ steps, due to electrostatic repulsion.

![Coralyne chloride](image)

**Figure 17. Coralyne chloride**

UV melting experiments have been carried out to assess the effect of coralyne on the melting temperatures of the T·A*T triplex hexamers (Table 9) and of the mixed sequence triplexes (Table 10), compared to the effects of the drug on the melting temperatures of their corresponding duplex oligonucleotides. These experiments were carried out in a 10mM Na ccod / 10mM MgCl₂ buffer, at pH 7.0 for the (T·A*T)₆ triplexes and at pH 5.5 for the mixed sequence ones. In case of the GA.TC.CT triplexes, the experiments were carried out at pH 7.0 as well. The amount of coralyne added was calculated so that one aliquot should provide 3 coralyne molecules per DNA molecule, i.e. one drug molecule for every other base triplet, with the exception of A₄ T₄ T₄ (Ocd) and its respective duplex, where it should provide one drug molecule per DNA molecule. The latter triplex was included in the study for comparison purposes. The DNA quantity was 10 nmoles in all samples.
Table 9. UV melting temperatures of the (T-A*T)_6 and A_4 T_4 (Ocd) triplex-coralyne complexes

<table>
<thead>
<tr>
<th>Sequence</th>
<th>T_m (°C)</th>
<th>1 aliquot (3 : 1)</th>
<th>2 aliquots (6 : 1)</th>
<th>3 aliquots (9 : 1)</th>
<th>4 aliquots (12 : 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_6.T_6 (HEG)</td>
<td>49.7</td>
<td>52.8</td>
<td>53.6</td>
<td>55.1</td>
<td>***</td>
</tr>
<tr>
<td>A_6.T_6 (HEG)</td>
<td>49.7</td>
<td>49.2</td>
<td>***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T_6.T_6.A_6 (HEG)</td>
<td>49.4</td>
<td>53.8</td>
<td>55.7</td>
<td>56.6</td>
<td>***</td>
</tr>
<tr>
<td>T_6.A_6 (HEG)</td>
<td>49.1</td>
<td>51.0</td>
<td>***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A_6.T_6.T_6 (Ocd)</td>
<td>49.5</td>
<td>55.2</td>
<td>57.3</td>
<td>58.7</td>
<td>58.8</td>
</tr>
<tr>
<td>A_6.T_6 (Ocd)</td>
<td>48.5</td>
<td>49.0</td>
<td>49.2</td>
<td>***</td>
<td>-</td>
</tr>
<tr>
<td>T_6.T_6.A_6 (Ocd)</td>
<td>52.8</td>
<td>59.3</td>
<td>61.2</td>
<td>62.4</td>
<td>***</td>
</tr>
<tr>
<td>T_6.A_6 (Ocd)</td>
<td>48.7</td>
<td>48.8</td>
<td>47.3</td>
<td>***</td>
<td>-</td>
</tr>
<tr>
<td>T_6.A_6.T_6 (Ocd)</td>
<td>14.0</td>
<td>51.7</td>
<td>34.4</td>
<td>51.6</td>
<td>38.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>T_m (°C)</th>
<th>1 aliquot (1 : 1)**</th>
<th>2 aliquots (2 : 1)**</th>
<th>3 aliquots (3 : 1)**</th>
<th>4 aliquots (4 : 1)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_4 T_4 (Ocd)</td>
<td>28.9</td>
<td>32.3</td>
<td>34.4</td>
<td>35.5</td>
<td>36.5</td>
</tr>
<tr>
<td>A_4 T_4 (Ocd)</td>
<td>****</td>
<td>****</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes:  
* 1 aliquot coralyne = 30 nmoles  
** 1 aliquot coralyne = 10 nmoles  
*** Spectrum gave an erratic curve and could not be used.  
**** Melting did not occur in the temperature range studied.
### Table 10. UV melting temperatures of the mixed sequence triplex-coralyne complexes

<table>
<thead>
<tr>
<th>Sequence</th>
<th>pH</th>
<th>-</th>
<th>1 aliquot (3:1)</th>
<th>2 aliquots (6:1)</th>
<th>3 aliquots (9:1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA.TC.CT (Ocd)</td>
<td>5.5</td>
<td>57.8</td>
<td>58.9</td>
<td>60.3</td>
<td>**</td>
</tr>
<tr>
<td>GA.TC (Ocd)</td>
<td>5.5</td>
<td>60.1</td>
<td>61.6</td>
<td>61.9</td>
<td>**</td>
</tr>
<tr>
<td>TC.CT.AG (Ocd)</td>
<td>5.5</td>
<td>59.2</td>
<td>59.1</td>
<td>60.1</td>
<td>**</td>
</tr>
<tr>
<td>CT.AG (Ocd)</td>
<td>5.5</td>
<td>63.8</td>
<td>64.1</td>
<td>64.2</td>
<td>**</td>
</tr>
<tr>
<td>GA.TC.CT (HEG)</td>
<td>5.5</td>
<td>57.5</td>
<td>50.7</td>
<td>58.3</td>
<td>48.2</td>
</tr>
<tr>
<td>GA.TC (HEG)</td>
<td>5.5</td>
<td>61.6</td>
<td>62.7</td>
<td>62.3</td>
<td>**</td>
</tr>
<tr>
<td>TC.CT.AG (HEG)</td>
<td>5.5</td>
<td>52.8</td>
<td>65.1</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>CT.AG (HEG)</td>
<td>5.5</td>
<td>66.0</td>
<td>66.2</td>
<td>66.5</td>
<td>**</td>
</tr>
<tr>
<td>GA.TC.CT (Ocd)</td>
<td>7.0</td>
<td>14.4</td>
<td>63.2</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>GA.TC.CT (HEG)</td>
<td>7.0</td>
<td>14.2</td>
<td>63.5</td>
<td>13.5</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**  
* 1 aliquot coralyne = 30 nmoles  
** Spectrum gave an erratic curve and could not be used.

As can be seen from Table 9, coralyne increases the melting temperatures of the T·A·T triplexes, and has a minimal effect on the melting of their respective duplexes. The effect on the Ocd linked oligonucleotides is greater than that on the HEG ones.

Upon addition of the first coralyne aliquot, the melting temperature rises by up to 6.5°C for the Ocd and 4.4°C for the HEG FTFOs. ΔT_m is greater in case of T₆·T₆·A₆ for both linkers. When the second addition of coralyne is made, another increase in melting temperatures of around 2°C occurs. A third aliquot brings about a further rise of 0.8 - 1.5°C. If a fourth aliquot of coralyne is added, the spectra become erratic curves which cannot be used for the calculation of the melting temperature, with the exception of A₆·T₆·T₆ (Ocd), where T_m increases by a further 0.1°C. The latter increase is, however, negligible compared to the
previous ones. Therefore it can be assumed that any further increase in drug concentration will not lead to a rise of the triplex melting temperature. $A_4\cdot T_4\cdot T_4$ (Ocd) is stabilised by coralyne as well.

The melting temperatures of the duplex oligonucleotides show mainly variations of less than 1°C, some positive and some negative. Their spectra become erratic after the addition of three or even two drug aliquots. $A_4\cdot T_4$ (Ocd) does not appear to form a duplex with or without coralyne.

The sequence $T_6\cdot A_6\cdot T_6$ (Ocd) appears to form a fairly weak triplex in the absence of coralyne. This is due either to the formation of an intermolecular triplex, or of an intramolecular one with parallel pyrimidine strands, the third strand of which binds to the purine strand by reverse Hoogsteen hydrogen bonds. Upon the addition of one coralyne aliquot, the temperature of the triplex-to-duplex transition increases significantly by 20.4°C. A second addition raises the melting temperature further by 3.6°C. The temperature of the duplex-to-coil transition remains virtually unchanged.

In contrast with the T*A*T triplexes, the Ocd linked mixed sequence triplexes (Table 10) are stabilised very little by coralyne: their melting temperatures at pH 5.5 are increased only by 2.5°C and 0.9°C after the addition of two drug aliquots. If more drug is added, the spectra become erratic.

The HEG linked mixed sequence triplexes are destabilised by coralyne at pH 5.5. As described in section 2.2.4, the two melting transitions of the mixed sequence FTFOs probably do not occur simultaneously and are unresolved under the same conditions as the UV melting experiments carried out with coralyne. Upon addition of the drug, the temperature of the triplex-to-duplex transition decreases, while that of the duplex-to-coil transition rises very slightly. The latter could be an apparent increase only; as the gap between the two transitions widens, it may be that the temperature of the W/C transition can be calculated more accurately.

The melting temperatures of the duplex oligonucleotides show very slight increases, the largest of which - 1.8°C after the addition of two coralyne aliquots - occurs for the GA.TC (Ocd) duplex.

At pH 7, the effect of coralyne on the melting temperatures of the mixed sequence triplexes is relatively insignificant.
These results suggest that under the conditions investigated coralyne has a sequence preference, in contrast to the findings of Lee et al. (1993). It binds to the T-A*T triplexes and stabilises them, the Ocd more so than the HEG linked ones, while the effect it has on the mixed sequence triplexes is much smaller: a slightly stabilising one in case of the Ocd linked ones and a destabilising one in case of the HEG linked FTFOs. This is consistent with the idea that there is electrostatic repulsion between the positively charged coralyne and the C-G•C+ base triplets in the mixed sequence triplexes, which may result in incomplete intercalation. The stability of the duplexes seems to be virtually unaffected by the presence of the drug.

It may be that coralyne would stabilise these types of triplexes and maybe even the duplexes to a greater extent, but this was not apparent because of the great stabilising effect of the linkers.

2.3 Attempts to Crystallise DNA Foldback Triplexes

The most powerful technique for determining the three-dimensional structure of biological macromolecules is single-crystal X-ray crystallography. Single-crystal analysis of DNA and RNA oligonucleotides and their complexes with drugs has provided a great deal of information about the structure and function of these molecules. However, the crystallisation of nucleic acids is currently still carried out in a trial-and-error manner, and the obtaining of diffraction quality crystals does not have a reliable success rate (Wang and Gao, 1990).

No normal X-ray diffraction pattern from crystals of a DNA triplex has yet been reported. Attempts to obtain crystals of triplexes from mixtures of single strands yield fibre patterns, even from well-formed single crystals (Liu et al., 1996). It was suggested that this diffraction behaviour may be an intrinsic property of triple helices, reflecting the lack of lateral ordering, as there is no strong interaction between the cylindrical triplex molecules in the lateral direction; this may arise because triple helices are much more rigid than Watson-Crick duplexes (Liu et al., 1996). Crystallography of DNA duplexes shows deviations from canonical structures, such as distortions which enable the double helices to interact laterally; for example, a portion of one duplex can interact with the major groove of another. In a triplex the major groove is filled with the third strand, preventing this kind of lateral interaction (Liu et al., 1996). In addition, the higher charge density of a triplex may also disfavour such interactions.
Arnott and co-workers (Arnott and Selsing, 1974; Arnott et al., 1976) were the first to propose a three-dimensional structure for a poly(dT)•poly(dA)•poly(dT) triplex on the basis of X-ray fiber diffraction data.

Van Meervelt et al. (1995) have described the structure of the C·G•G base triplet as part of an extended DNA crystal structure to a resolution of 2.05Å. The formation of this base triplet occurs as a result of interaction between adjacent overlapping duplexes. The same research group then obtained both parallel and antiparallel (C·G•G)\(_2\) fragments in a crystal structure of a DNA duplex with staggered ends (Vlieghe et al., 1996). Other triplet structures characterised by X-ray diffraction were in protein-DNA complexes studied at about 3Å (Luisi et al., 1991).

Most DNA duplex sequences that have been crystallised successfully are 4 to 15 nucleotides long; these molecules are nearly globular, which enables them to pack more effectively against other molecules of this shape (Wang and Gao, 1990). It was reasoned that the triplexes would have a good chance of crystallising if they were relatively short, 6 - 8 nucleotides in length. DNA triple helices of this length are stable at very low temperatures only. However, the foldback DNA triplexes studied are very stable at room temperature, so they were synthesised in multi-milligram quantities, which were then used in crystallisation experiments. The quantities of DNA synthesised for this purpose are listed in Table 11.

### Table 11. Quantities of oligonucleotides synthesised for crystallisation

<table>
<thead>
<tr>
<th>Sequence</th>
<th>(Ocd)(_2)</th>
<th>HEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>A(_6)T(_5)T(_5)</td>
<td>32.1</td>
<td>17.3</td>
</tr>
<tr>
<td>T(_5)T(_5)A(_6)</td>
<td>15.6</td>
<td>5.9</td>
</tr>
<tr>
<td>A(_8)T(_8)T(_8)</td>
<td>20.5</td>
<td>11.0</td>
</tr>
<tr>
<td>T(_8)T(_8)A(_8)</td>
<td>6.5</td>
<td>3.7</td>
</tr>
<tr>
<td>GA.TC.CT</td>
<td>34.2</td>
<td>9.8</td>
</tr>
<tr>
<td>TC.CT.AG</td>
<td>8.5</td>
<td>11.0</td>
</tr>
</tbody>
</table>
The purity of the oligonucleotides needs to be very high to increase the chance of obtaining good quality crystals. The synthesised FTFO-s were purified until CZE analyses showed one single peak in each case. Examples of the CZE traces are illustrated in the appropriate experimental section.

Crystallisations of the triplexes by the vapour diffusion and by the slow evaporation method have been tried. However, so far the experiments have been unsuccessful. For details of the procedures and solutions used, see the appropriate experimental section.

It may be that the linkers used to stabilise the DNA triplexes are too long and therefore have too many degrees of freedom, which reduces the ability of the molecules containing them to pack effectively in a lattice.

Coralyne was used in some of the crystallisation experiments, but this lead to the formation of a microcrystalline precipitate, which control experiments (without DNA) showed to be the drug on its own. It was concluded that coralyne does not represent a useful reagent in the crystallisation of nucleic acids, as it has a very poor solubility in the buffers and solutions normally used for this purpose.

2.4 Results from CD and NMR Studies Carried out in Collaboration

Circular dichroism and NMR studies were carried out on the $A_6.T_6.T_6$ and GA.TC.CT triplexes with and without coralyne. The nature of the linker - HEG or Ocd - did not make any difference to the CD and NMR spectra.

The CD spectrum of $A_6.T_6.T_6$ at 15°C did change significantly when coralyne was added (Figure 18), while the spectrum of GA.TC.CT showed a very small change (Figure 19). These results indicate that under the conditions used (8μM triplex and 4μM coralyne) $A_6.T_6.T_6$ binds coralyne tightly, probably intercalatively, whereas GA.TC.CT does not. It seems therefore that coralyne binds preferentially at consecutive T·A*·T base triplets, rather than at alternating T·A*·T/GA·TC*·CT sequences.
Figure 18. CD spectrum of $A_6$.T$_6$.T$_6$ with and without coralyne and difference spectrum.

Figure 19. CD spectrum of GA.TC.CT with and without coralyne.
$^1$H-NMR spectra in H$_2$O at 10°C of both triplexes showed large changes consistent with intercalation upon addition of coralyne. At the high DNA and drug concentrations (1.5mM and 0.5mM, respectively) used for the recording of the NMR spectra coralyne binds well not only to A$_6$.T$_6$.T$_6$, but also to GA.TC.CT, as the changes in the spectra (Figures 20 and 21) indicate.

**Figure 20.** $^1$H-NMR spectrum of A$_6$.T$_6$.T$_6$ (a) without and (b) with coralyne
Figure 21. $^1$H-NMR spectrum of GA.TC.CT (a) without and (b) with coralyne
3. Experimental

3.1 Chemicals and Solvents

All chemicals were obtained from Aldrich, Fluka or Fisons. Silica gel 60 (80 - 430 mesh) was obtained from Fluka. All solvents were of analytical or HPLC grade. Anhydrous acetonitrile was purchased from Applied Biosystems as DNA synthesis grade.

Anhydrous pyridine was distilled over CaH$_2$. Anhydrous THF and diethyl ether were distilled over sodium and benzophenone. TEA and DIPEA were dried by storage over CaH$_2$. Pyrogen-free, reverse osmosis purified water was used throughout.

3.2 Thin Layer Chromatography (TLC)

TLC was carried out on aluminium sheets coated with a 0.2mm layer silica gel 60 F$_{254}$, supplied by Merck, using the following solvent systems (all v/v):

I. dichloromethane : methanol (10 : 1)
II. ethylacetate : triethylamine (99 : 1)

Before carrying out TLC of dimethoxytritylated compounds, the plates were pretreated with triethylamine, to prevent detritylation due to the acidic silica gel. For compounds with R$_f$ values close to that of pyridine, the R$_f$ value was distorted. The R$_f$ values given below (section 3.11) correspond to that of the purified compounds.

Compounds were visualised under UV light at 254nm. Dimethoxytrityl groups were visualised by fuming HCl as bright orange colours. The purity of all compounds synthesised was verified by one spot on TLC.

3.3 Wet Flash Column Chromatography

This was carried out under slight nitrogen or argon pressure (0.5 bar) using silica gel 60. For compounds containing the cyanoethyl-N,N-diisopropyl-phosphoramidite group, in order to
remove any moisture present in the silica, this was dried overnight at 120°C and then cooled in a dessicator over P₂O₅ prior to being used.

3.4 Instrumentation

¹H-NMR spectra were recorded on a Brucker WP-200 (200.13 MHz) Fourier transform spectrometer. ¹³C-NMR spectra were recorded on a Brucker WP-200 (62.90 MHz) Fourier transform spectrometer. Tetramethyldisilane was used as internal reference and cited chemical shifts are given in ppm downfield of this standard. ³¹P-NMR spectra were recorded on a Jeol FX90Q (90 MHz) Fourier transform spectrometer. H₃PO₄ 85% was used as external reference and cited chemical shifts are given in ppm downfield of this standard. FAB mass spectra were recorded on a Kratos MS 50 TC spectrometer using a 3-nitrobenzyl alcohol (3-NOBA) matrix. C, H and N microanalysis was performed on a Perkin Elmer 240 elemental analyser.

3.5 Oligonucleotide Synthesis

Oligonucleotide synthesis was performed using solid phase cyanoethyl-N,N-diisopropylphosphoramidite chemistry (Beaucage and Caruthers, 1981; Brown and Brown, 1991) on an Applied Biosystems 394 automated DNA synthesiser equipped with on-line long wave detector (498nm). All DNA synthesis reagents and standard phosphoramidite monomers were supplied by Applied Biosystems.

The modified phosphoramidites synthesised for this study were dried in vacuo, stored with P₂O₅ overnight and used as a 0.11 - 0.12M solution in anhydrous acetonitrile. Coupling efficiencies were in general >98%. 'Trityl-off' syntheses were carried out. The oligonucleotides were cleaved from the solid support with concentrated ammonia and deprotected by incubation in sealed vials with teflon caps at 55°C in concentrated ammonia for 8 - 10 hours.
3.6 HPLC Analysis and Purification of Oligonucleotides

Oligonucleotide analysis and purification was carried out on a Gilson Model 306 equipped with an UV detector, an Aquapore Octyl reverse phase column (10mm x 250mm) and an octadecyl pelicular guard column, controlled by a computer running Gilson 712 software.

The eluting buffers used were:

buffer A - 0.1M NH₄OAc
buffer B - (a) 0.1M NH₄OAc plus 20% MeCN - for the HEG oligonucleotides
    (b) 0.1M NH₄OAc plus 35% MeCN - for the Ocd oligonucleotides, as the octanediol linker makes them more hydrophobic.

At a flow rate of 3ml/min, the following gradient was used:

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>% Buffer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>3.00</td>
<td>0.0</td>
</tr>
<tr>
<td>4.00</td>
<td>15.0</td>
</tr>
<tr>
<td>24.00</td>
<td>70.0</td>
</tr>
<tr>
<td>26.00</td>
<td>100.0</td>
</tr>
<tr>
<td>28.00</td>
<td>100.0</td>
</tr>
<tr>
<td>29.00</td>
<td>0.0</td>
</tr>
<tr>
<td>30.00</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Analytical runs were monitored at 260nm and 0.2 AUFS. Preparative scale runs were monitored at 293-295nm and 1 AUFS. Examples of the chromatograms obtained are given in Figures 22 and 23 below.
3.7 Sephadex Gel-Filtration (Desalting) of Oligonucleotides

In order to remove the salt accumulated in the sample during the reverse phase HPLC purification step, the oligonucleotides were submitted to gel filtration on Pharmacia NAP-10 (Sephadex G-25) cartridges. The DNA was lyophilised to give a fluffy white powder.
3.8 Capillary Zone Electrophoresis (CZE)

The purity of all oligonucleotides synthesised was verified by one single sharp peak on CZE. This was carried out on an Applied Biosystems Model 270A Capillary Electrophoresis system with Microgel capillaries using 75mM Tris-phosphate with 10% methanol at pH 7.6 elution buffer. The on-line detector was set at 272nm with the loading voltage at -5kV, the running voltage at -15kV and the sample uptake time between 1 and 5 seconds. Sample runs were recorded on a chart recorder set at chart speed 1cm/min, attenuation 16, offset 5 and peak threshold 10 000. Examples of the CZE traces obtained are shown below.

(a) ![CZE Trace 1](image1)
(b) ![CZE Trace 2](image2)
3.9 Crystallisation Experiments

Glass spot plates (Corning 7220) were used for the crystallisation of the foldback DNA triplex forming oligonucleotides. The plates were extensively cleaned and treated with silanising agent (5% chlorotrimethyl silane in DCM) to produce a water-repelling surface. This supports the crystallising solution in a round beaded shape, which allows the concentrations of all components in solution to increase uniformly as the droplet becomes smaller; otherwise it would collapse into a flat pool with an increased surface area, causing it to dry rapidly during equilibration, which could lead to the precipitation of the oligonucleotide (Wang and Gao, 1990).
The following aqueous solutions were used:

A: Na cacodylate buffer (100mM), pH=6.8 for T:A•T type triplexes or pH=5.5 for mixed sequence triplexes

B: MgCl\(_2\) (100mM, 200mM or 1M)

C: Coralyne.HCl (5mM)

D: MPD (50%)

E: Spermine.4HCl (10mM)

Each FTFO (1mg) was dissolved in the Na cacod buffer A (35\(\mu\)l in case of the vapour diffusion method, or 50\(\mu\)l in case of slow evaporation), and 5\(\mu\)l of this solution were placed in the top three and the bottom three wells of a nine wells crystallising plate in case of the vapour diffusion method, or in each well in case of slow evaporation. Then water was added to the wells, if needed, in such quantities to ensure that the volumes of the droplets on any one plate were identical after all solutions were added.

Specific quantities of the following combinations of solutions were then added sequentially (in the order indicated below), ensuring that the oligonucleotide did not precipitate:

- B - GA.TC.CT (HEG) and (Ocd)\(_2\)
- B and C - A\(_6\).T\(_6\).T\(_6\) (Ocd)\(_2\)
- B and D - GA.TC.CT and TC.CT.AG, (HEG) and (Ocd)\(_2\)
- B and E - TC.CT.AG (HEG)
- B, C and D - A\(_6\).T\(_6\).T\(_6\) and GA.TC.CT, (HEG) and (Ocd)\(_2\)
- B, D and E - all sequences

This protocol was then applied to each of the remaining wells containing DNA, the amount of solutions B,C,D and E being varied slightly in each well, to give a wide range of conditions. In case of the vapour diffusion method, 200\(\mu\)l or 100\(\mu\)l solution D was added to the three middle wells of the crystallising plate. The wells were then covered with glass microscope slides with vacuum grease applied to ensure an efficient seal. The majority of the plates were kept at room temperature, the remainder were kept at 9°C. They were examined daily under a microscope to follow the process of crystallisation.
3.10 UV Melting (Thermal Denaturing) Studies

UV melting was carried out on a Lambda 2 UV-visible spectrophotometer controlled by a computer running a PECSS (Perkin Elmer Computerised Spectroscopy Software) package. The temperature of the sample was controlled by a PTP-1 Peltier block with a heating rate of 1°C/min. The absorbance of the sample at 260nm was measured at intervals of ten seconds to obtain absorbance versus time melting curves. The data was then converted to absorbance versus temperature curves, and the melting temperatures (which correspond to the points of inflection of the curves) were obtained by calculating the first derivative of the curve and determining its maximum points. For the buffers used, see section 2.2.

The melting of the FTFOs synthesised for this study was either biphasic, in which case the melting curves had two inflection points, or monophasic, when the melting curve had just one inflection point. In case of the biphasic melting, the first inflection point, corresponding to a lower temperature, represents the triplex-to-duplex transition; the second one, corresponding to a higher melting temperature, represents the duplex-to-coil transition. In case of the monophasic melting, the transitions occur directly from triplex to single stranded DNA. All melting curves were determined in triplicate.

3.11 Synthetic Chemistry

1-O-(4,4'-Dimethoxytrityl)-hexaethyleneglycol [Durand et al., 1990]

![Chemical Structure]

HEG (12.50g, 45mmol, 11ml, 5eq.) and DMAP (215mg, 1.8mmol, 0.2eq.) were coevaporated from anhydrous pyridine (3 x 7ml) and dissolved in anhydrous pyridine (50ml). DMTTrCl (3g,
9mmol, 1eq.) was added portionwise over 30 minutes, and the reaction was stirred at room temperature for 5 hours. EtOAc (250ml) was added, and the mixture was washed with a 0.5M NaHCO₃ solution (125ml) and then with water (125ml). The organic layer was dried over Na₂SO₄, filtered and evaporated down, to give the product as an oil. This was dissolved in a small amount of EtOAc and purified by wet flash column chromatography, eluting with a gradient of 0 - 2% MeOH in DCM on a column with the silica preequilibrated with 1%TEA in DCM. The appropriate fractions were combined, and the solvents were removed in vacuo. An orange coloured oil was obtained. Yield (calculated in respect to DMTrCl) 79% (4.15g).

Rₜ 0.55 (system I); 0.35 (system II). [FAB] m/z M⁺: found 584.29856, calculated 584.29853. ¹H-NMR: δ (CDCl₃) 3.18-3.23 (t, 2, CH₂-OH), 3.54-3.69 (m, 23, CH₂, OH), 3.75 (s, 6, Ar-CH₃) 6.78-7.46 (m, 13, Ar). ¹³C-NMR: δ (CDCl₃) 45.85 CH₂, 54.97 CH₃, 61.46 CH₂, 62.92 CH₂, 70.06 CH₂, 70.31 CH₂, 70.36 CH₂, 70.46 CH₂, 70.51 CH₂, 72.36 CH₂, 85.68 C, 112.81 CH, 126.41 CH, 127.51 CH, 127.98 CH, 129.84 CH, 136.12 C, 144.87 C, 158.14 C. Anal. calcd. for C₃₃H₄₄O₉: C 67.79, H 7.58; found C 67.21, H 7.18.

1-O-(4,4'-Dimethoxytrityl)-hexaethyleneglycol-(2-cyanoethyl-N,N-diisopropyl)-phosphoramidite [Durand et al., 1990]

1-O-(4,4'-Dimethoxytrityl)-hexaethyleneglycol (993mg, 1.7mmol) was dried by the evaporation of anhydrous pyridine (2 x 5ml) and of anhydrous THF (5ml), and dissolved in dry THF (15ml). DIPEA (1.18ml, 6.8mmol, 4eq.) and then 2-cyanoethyl-N,N-(diisopropyl-amino)chlorophosphite (417µl, 1.9mmol, 1.1eq.) was added dropwise with a syringe. A white solid precipitated. After stirring the reaction for 20 minutes, EtOAc (50ml, stored over Na₂SO₄) was added to reaction mixture, and the organic layer was extracted once with
saturated KCI, then quickly dried over Na$_2$SO$_4$, filtered and the solvents removed in vacuo. The product was purified by wet flash column chromatography (silica preequilibrated with 1% TEA in EtOAc), eluting with EtOAc. The appropriate fractions were combined, and the solvents were removed in vacuo to obtain an orange coloured oil. Yield 76% (1013mg). R$_f$ 0.84 (system I); 0.65 (system II). $^{31}$P-NMR: $\delta_p$ (CDCl$_3$) 149.18.

1-O-(4,4'-Dimethoxytrityl)-1,8-octanediol

![Diagram of 1-O-(4,4'-Dimethoxytrityl)-1,8-octanediol]

1,8-Octanediol (6.75g, 45mmol, 5eq.) and DMAP (215mg, 1.8mmol, 0.2eq.) were dried by the evaporation of anhydrous pyridine (3 x 7ml) and dissolved in dry pyridine (50ml). DMTrCl (3g, 9mmol, 1eq.) was added portionwise over 30 minutes, and the reaction was stirred at room temperature for 5 hours. EtOAc (250ml) was added, and the mixture was washed with a 0.5M NaHCO$_3$ solution (125ml) and then with water (125ml). The organic layer was dried over Na$_2$SO$_4$, filtered and evaporated down. At this stage unreacted starting material precipitated. EtOAc was added and the mixture was filtered, then the solvent was removed in vacuo. The residue was dissolved in a small amount of EtOAc and submitted to wet flash column chromatography, eluting with a gradient of 0 - 2% MeOH in DCM on a column with the silica preequilibrated with 1%TEA in DCM. The appropriate fractions were combined, and the solvents were removed in vacuo, to give an orange coloured oil. Yield 70% (2.81g). R$_f$ 0.65 (system I); 0.54 (system II). [FAB] m/z $M^+$: found 448.25941, calculated 448.26136. $^1$H-NMR: $\delta$ (CDCl$_3$) 1.29-1.45 (m, 8, CH$_2$), 1.49-1.64 (m, 5, CH$_2$, OH), 3.01-3.06 (m, 2, CH$_2$-OH), 3.59-3.64 (t, 2, CH$_2$-OAr), 3.78 (s, 6, CH$_3$), 6.79-7.47 (m, 13, Ar). $^{13}$C-NMR: $\delta$ (CDCl$_3$) 25.53 CH$_2$, 26.06 CH$_2$, 29.19 CH$_2$, 29.88 CH$_2$, 32.60 CH$_2$, 55.02 CH$_3$, 62.81 CH$_2$, 162
63.25 CH₂, 85.44 C, 112.77 CH, 126.36 CH, 127.51 CH, 128.02 CH, 129.85 CH, 145.29 C, 158.09 C. Anal. calcd. for C₂₉H₃₆O₄: C 77.65, H 8.09; found C 77.32, H 8.03.

1-O-(4,4'-Dimethoxytrityl)-1,8-octanediol-8-(2-cyanoethyl-N,N-diisopropyl)-phosphoramide
e

1-O-(4,4'-Dimethoxytrityl)-1,8-octanediol (985mg, 2.2mmol) was coevaporated from anhydrous pyridine (2 x 5ml) and of anhydrous THF (5ml), and dissolved in dry THF (10ml). DIPEA (1.53ml, 8.8mmol, 4eq.) and then 2-cyanoethyl-N,N-(diisopropylamino)chlorophosphite (537μl, 2.4mmol, 1.1eq.) was added dropwise with a syringe. A solid was seen to precipitate. After stirring the reaction for 20 minutes, EtOAc (40ml) was added to reaction mixture, and the organic layer was extracted once with saturated KCl, then quickly dried over Na₂SO₄, filtered and the solvents removed in vacuo. The product was purified by wet flash column chromatography (silica preequilibrated with 1% TEA in EtOAc), eluting with EtOAc. The appropriate fractions were combined, and the solvents were removed in vacuo. An orange coloured oil was obtained. Yield 84% (1198mg). Rₚ 0.84 (system I); 0.69 (system II). ¹³¹P-NMR: δₚ (CDCl₃) 147.54.
4. Conclusions

T·A*T and mixed sequence foldback DNA triplexes were obtained in multi-milligram quantities in a high purity state, as CZE analyses showed. Their thermal stabilities were investigated by UV melting in a variety of conditions, in the presence and in the absence of the triplex binding drug coralyne. Attempts to crystallise the FTFOs have been unsuccessful. From these studies it can be concluded that:

1.) The two linkers used - HEG and (Ocd)₂ - stabilise the FTFOs to a similar extent.

2.) Very short triplexes, which under normal circumstances would not form (A₄·T₄·T₄), are stabilised by the presence in their molecules of flexible linkers which hold the three strands together, while the equivalent duplex is not sufficiently stabilised and does not form.

3.) FTFOs of a given sequence have UV melting temperatures different from those of the ones of reversed sequence, even though they form triplexes of identical sequences.

4.) As expected, the stability of the mixed sequence triplexes increases with the lowering of the pH, while the temperatures of the duplex-to-coil transition is much less affected.

5.) Coralyne stabilises the T·A*T type triplexes, while its effect on the mixed sequence ones is significantly smaller. This seems to indicate that the drug intercalates preferentially between consecutive T·A*T rather than between alternating T·A*T and C·G·C* base triplets, in contrast to the results of Lee et al. (1993), which indicated that coralyne has little sequence preference. This conclusion is consistent with the assumption that there is repulsion between the positively charged coralyne and the C·G·C* base triplets, which may result in incomplete intercalation of the drug.

6.) The effect of coralyne on the duplex-to-coil transitions of all DNA sequences studied was minimal, therefore confirming that coralyne binds preferentially to triplex rather than to duplex DNA.

7.) The addition of coralyne to the triplexes to be crystallised led to the formation of a microcrystalline precipitate which in control experiments turned out to be the drug on its own, and therefore does not represent a useful reagent for crystallisation experiments.
5. References


Coralyne has a preference for intercalation between TA-T triples in intramolecular DNA triple helices
Ana Ariana Moraru-Allen, Sarah Cassidy¹, Juan-Luis Asensio Alvarez², Keith R. Fox¹, Tom Brown³ and Andrew N. Lane²,*

Department of Chemistry, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JJ, UK, ¹Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Boldrewood, Bassett Crescent East, Southampton SO16 7PX, UK, ²Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and ³Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received February 13, 1997; Revised and Accepted March 27, 1997

ABSTRACT
Intercalating ligands may improve both the stability and sequence specificity of triple helices. Numerous intercalating drugs have been described, including coralyne, which preferentially binds triple helices, though its sequence specificity has been reported to be low [Lee, J.S., Latimer, L.J.P. and Hampel, K.J. (1993) Biochemistry, 32, 5591–5597]. In order to analyse the sequence preferences of coralyne we have used a combination of DNase I footprinting, UV melting, UV-visible spectrophotometry, circular dichroism and NMR spectroscopy to examine defined intermolecular triple helices and intramolecular triple helices linked either by hexaethylene glycol chains or by octandiol chains. DNase I footprinting demonstrated that coralyne has a moderate preference for triple helices over duplexes, but a substantial preference for TA-T triples compared with CG-C+ triplets. The drug was found to have essentially no effect on the melting temperatures of duplexes of the kind d(A)n d(T)n or d(GA)n d(TC)n. In contrast, it increased the \( T_m \) for triple helices of the kind d(T)n d(A)n d(T)n, but had little effect on the stability of d(TC)n d(GA)n d(TC)n at either low or high pH. On binding to DNA triple helices, there is a large change in the absorption spectrum of coralyne and also a substantial fluorescence quenching that can be attributed to intercalation. The changes in the optical spectra have been used for direct titration with DNA. For triple helices \( d(T)^n d(A)_6^d(T)_6 \), the \( K_d \) at 298 K was 0.5–0.8 µM. In contrast, the affinity for d(TC)n d(GA)n d(TC)n triple helices was 6- to 10-fold lower and was characterized by smaller changes in the absorption and CD spectra. This indicates a preference for intercalation between TA-T triples over CG-C+TA-T triples. NMR studies confirmed interaction by intercalation. However, a single, secondary binding was observed at high concentrations of ligand to the triple helix \( d(GAAGA- TCTTCT-L-TCTTCT) \), presumably owing to the relatively low difference in affinity between the TA-T site and the competing, neighbouring sites.

INTRODUCTION
External regulation of specific gene expression can in principle be achieved by several different means, including the use of substrate analogues to inhibit a gene product, DNA or RNA complementary to a particular mRNA sequence (antisense strategy and ribozymes), a small molecule that binds to particular DNA sequences or an oligonucleotide that binds to a specific DNA duplex sequence (antigene strategy). The last methods are to be preferred, as there is only one target (or a small number of similar genes) per cell, compared with the amplification achieved by both transcription and translation. However, small molecules, despite recent successes at achieving high affinity for DNA (1), generally do not have sufficient sequence selectivity to turn off a single gene in a large genetic background (2). Antigene oligonucleotides may offer much greater sequence selectivity, despite the requirement for runs of purines on one strand of the target in order to form the triple helix. One of the main difficulties with antigene methods is the relatively low thermodynamic stability of triple helices, compounded for some sequences by the need for protonation of cytosines (2). An alternative is to use a binary system of a triple-forming oligonucleotide in conjunction with a small ligand that preferentially binds the triple helix. Such a ligand will automatically improve triple helix stability and also provide additional sequence selectivity, either by the sequence preferences of the ligand itself or by permitting the formation of specific mismatches (decreasing the stringency of the triple helix).

Intercalating ligands that preferentially bind triple helices have been described and include the acridines, anthracene-9,10-diones, BePI, the naphthoquinolines and coralyne (3–11). Other aromatic systems that intercalate well into DNA duplexes may destabilize triple helices (e.g. ethidium; 2,7). One ligand that has been studied extensively is coralyne (7,12). This ligand is yellow and has convenient absorption bands at 330 and 450 nm, well removed from DNA absorption in the 260–280 nm range. It is also fluorescent and, as a planar ligand, has very little intrinsic circular dichroism. It can also be used to stabilize intermolecular triple helices, which then inhibit the action of DNase I (footprinting).

We have used DNase I footprinting, optical spectroscopy and NMR to characterize the binding of coralyne to defined triple helices and determine the dissociation constants. Previous work with

*To whom correspondence should be addressed. Tel: +44 181 959 3666; Fax: +44 181 906 4477; Email: a-lane@nimr.mrc.ac.uk
coralyne binding to polymeric DNA was reported to show that coralyne has no sequence specificity and that it also interacts with duplex DNA (7). We find significantly different behaviour, with both intra- and intermolecular triple helices.

**MATERIALS AND METHODS**

Strands of DNA were made using phosphoramidite chemistry and purified by HPLC as previously described (13). Covalently linked DNA oligomers were prepared using hexaethylene glycol as previously described (14). Such intramolecular triple helices are written as, for example, d(AAAAA-L-TTTTT-L-TTTTT), where L denotes the linker and the underlining shows the Watson–Crick duplex.

Plasmid pAG1, which contains the sequence d(AAAAGGGG-GGGG)-d(CCCCCCTTTTTTTT) cloned into the BamHI site of pUC19, was prepared as previously described (3,15). A fragment containing this region was prepared by cutting the plasmid first with HindIII, labelling at the 3'-end with reverse transcriptase and [α-32P]dATP and cutting again with EcoRI. This procedure labels the purine strand of the target triplex site. The fragment was resolved and separated from the remainder of the plasmid by electrophoresis on a native 8% polyacrylamide gel. The fragment was eluted from the gel slice, precipitated with ethanol and redissolved in 10 mM Tris–HCl, 0.1 mM EDTA, pH 7.5.

Coralyne was purchased from Sigma (Poole, UK). According to TLC, NMR and mass spectrometry, the compound was ≥98% pure and was used without further purification.

**DNase I footprinting**

DNase I footprinting was carried out as previously described (3,15). Aliquots of 1.5 μl labelled DNA (−10 nM) were mixed with 1.5 μl third strand oligonucleotide and, where present, 1.5 μl coralyne in a buffer containing 50 mM sodium acetate, 5 mM MgCl2, 2 mM MnCl2, pH 5.5, for the parallel Y-RY triple helices, or 10 mM Tris–HCl, 5 mM MgCl2, pH 7.5, for the antiparallel R-RY triple helices. The mixture was allowed to equilibrate at room temperature for at least 1 h before adding 2 μl DNase I solution (0.1 U/ml in 20 mM NaCl, 2 mM MgCl2). The reaction was stopped after 1 min by adding 4 μl formamide containing 10 mM EDTA and 0.1% (w/v) bromophenol blue. Samples were boiled for 3 min, rapidly cooled on ice and loaded onto 11% polyacrylamide gels containing 8 M urea. The 40 cm gels were run at 1500 V for 2 h, fixed in 10% (v/v) acetic acid, transferred to Whatman 3 MM paper and dried at 80°C. The dried gels were autoradiographed at −70°C using an intensifying screen. Bands were assigned by comparison with Maxam–Gilbert marker lanes specific for G+A.

**Optical spectroscopy**

UV melting studies were carried out as previously described (16). Optical spectra were recorded on a Hitachi U-3210 spectrophotometer with 1 ml cuvettes thermostatted at 25°C. Difference spectra were recorded using tandem quartz cuvettes (d = 4.4 mm). Titration and optical studies were carried out either in Tris–NaCl, pH 8.0, or in sodium acetate, 5 mM MgCl2, pH 5.0. Stock solutions of coralyne (~2 mM) were made up in water, as the ligand is less soluble in buffer, and kept in the dark. The stock solution was diluted with buffer for subsequent work. For titrations, small aliquots of a buffered solution of DNA were added serially to a solution of coralyne at 2–10 μM and the change in absorbance measured at 334 nm. There is no absorbance for DNA at this wavelength. Titration curves were analysed using the program Kaleidograph (Synergy Software, Reading, PA). Circular dichroism (CD) spectra were recorded on a JASCO spectropolarimeter as previously described (17).

NMR spectra were recorded at 14.1 T on a Varian Unity NMR spectrometer. For spectra in H2O, the Watergate method (18) was used for solvent suppression. The NMR spectra of the free triple helix (d(AGAAGA-L-TCTCT-L-TCTCT)), where L is hexaethylene glycol, were assigned by standard methods. The chemical shifts of the exchangeable protons were monitored as a function of the concentration of coralyne. The amino groups of the cytosines bases in the Hoogsteen strand have highly characteristic chemical shifts (19) and are good indicators of the presence of a triple helix.

**RESULTS**

**DNase I footprinting**

The interaction of coralyne with intermolecular DNA triple helixes was examined by its effect on the formation of a triple helix between d(TTTTTTCCCCC) and a DNA duplex fragment containing the sequence d(AAAAGGGG)-d(CCCCCCTTTTTT) (see Materials and Methods). This triplex is not stable, presumably because of its short length and the presence of five contiguous C+GC base triplets. However, the triple helix can be stabilized either by attaching an intercalating acridine molecule at one end (15) or by adding a triplex binding ligand such as a naphthylquinoline (3). Figure 1A shows a DNase I footprint of the duplex d(AAAAGGGGG)-d(CCCCCCTTTTTT) with increasing concentrations of coralyne in the absence of the third strand. As the concentration of coralyne was increased, the intensity of the bands in general decreased and the enzyme was strongly inhibited above 10 μM drug. It is not clear whether this was due to a direct inhibition of the enzyme or to intercalation of coralyne into the duplex. In the absence of coralyne, the third strand produced essentially no footprint even up to 50 μM (Fig. 1B). At 10 μM third strand, the DNase I pattern was altered in the presence of 10 μM coralyne and higher (Fig. 1B). Although there was a general attenuation of the band intensities in the presence of both ligand and oligonucleotide, the pattern was different from that produced by either molecule alone, suggesting that coralyne stabilizes the DNA triplex helix. In particular, there was a region of enhanced cleavage at the 3'-end of the target site at the positions AAAAAAGGGGG/G/GATC corresponding to the triplex junction. This enhanced cleavage has previously been shown to be characteristic of parallel triplex formation (15). The enhancement was not observed with ligand alone (Fig. 1A). Furthermore, the doublet at the upper, 5'-edge of the target site was more efficiently protected in the presence of oligonucleotide than with ligand alone. This indicates that coralyne significantly stabilizes this parallel DNA triplex helix. The changes in the intensity of other bands above and below the target site reflect the interaction of coralyne with the DNA duplex and are consistent with the generalized decrease in band intensity shown in Figure 1A.

Because it has been reported that coralyne, unlike other triplex-binding ligands, has no preference for TA-T or GC-C triplets (7), we have compared the ability of coralyne to stabilize triplexes formed from the same target site with either d(TTTTTTCCCCC) or d(TTCCTCCCCC), which consist primarily of TA-T and CG-C triplets respectively. Neither triplex was stable in the absence of...
added ligand (5), but 10 μM coralyne clearly stabilized the triplex with d(TTTTTTTCC) (Fig. 1C), as shown by the enhanced cleavage at the 3' triplex/duplex junction and loss of the doublet at the upper edge of the target site. The enhancements observed with d(TTTTTTTCC) were located at AAAAAGG/GGGGATC, three bases higher in the gel than with d(TTTTTTTCC) in the presence of coralyne. In contrast, no triplex was visible with d(TTTTTTTCC) in the presence of coralyne. This clearly indicates a significant binding preference for TAT triples over CG-C' triples, even at pH 5.5.

We have also studied the effect of coralyne on the antiparallel R'-R' triplex formed by d(GGGGGTTTT) at the same target site (Fig. 1D). In the absence of ligand this oligonucleotide did not produce a significant footprint (5). However, a footprint was observed with 50 μM oligonucleotide in the presence of 30 μM coralyne at pH 7.5, indicating that the antiparallel motif can also be stabilized by this ligand. The differences between the footprints produced by the TC and GT oligonucleotides in the presence of coralyne may reflect the different pH values in the experiments and the different conformations adopted by parallel and antiparallel triplexes.

UV melting

UV melting curves were recorded in the absence and presence of different concentrations of coralyne (not shown). In the absence of coralyne, d(AAAAAAAGGGGGG)-d(CCCCCCTTTTT) also melted cooperatively, with a Tm of 48.5°C (not shown). Only at lower concentrations of salt did the melting of d(AAAAAAAGGGGGG)-d(CCCCCCTTTTT) separate into two transitions (A.Moraru-Allen and T.Brown, unpublished data). This shows that the triplex is at least as stable as the duplex. Increasing the concentration of coralyne shifted the melting curve of the triplex to higher temperature, with a significant increase (7.8°C) in the Tm of the transition. In contrast, coralyne had a negligible effect on the Tm of the duplex (Table 1).

Hence, coralyne binds to triplexes made of d(A)8 sequences much more tightly than to the corresponding duplexes. A similar result was obtained for the antiparallel d(GGGGGTTTT) at pH 7.5. In this case, the melting curve was biphasic, as the triple helix was rather unstable. However, adding coralyne greatly stabilized the lower temperature transition, with no discernible effect on the higher temperature (duplex to strand) transition (Table 1). Stabilization of the triplex was remarkable; at 30 μM coralyne the Tm increased 24°C. Hence, coralyne can indeed greatly stabilize DNA triplexes consisting of TA-T triplets only and, furthermore, the stabilization is greater for antiparallel than for parallel triplexes.

In contrast, triplexes consisting of alternating CG-C' and TA-T triplets were not significantly stabilized by coralyne at either pH 5 (where the triplex is stable) or at pH 7 (where deprotonation of the cytosine results in an unstable triplex) (Table 1). Hence, coralyne does not bind strongly between CG-C' and TA-T triples at micromolar concentrations.
Table 1. Effect of coralyne on the melting temperatures of different triplexes.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>pH</th>
<th>Tm (°C) at 0 equivalents coralyne</th>
<th>Tm (°C) at 1.2 equivalents coralyne</th>
<th>∆Tm (1.2 - 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(AAAAAAXTTTTTTT)</td>
<td>7</td>
<td>49.5</td>
<td>55</td>
<td>7.5</td>
</tr>
<tr>
<td>d(AAAAAXTTTTTTT)</td>
<td>7</td>
<td>48.5</td>
<td>49</td>
<td>0.5</td>
</tr>
<tr>
<td>d(TTTTTTXXTTTTTXXAAA)</td>
<td>7</td>
<td>53</td>
<td>59</td>
<td>8</td>
</tr>
<tr>
<td>d(TTTTTTXXAAAAAXTTTTTT)</td>
<td>7</td>
<td>14</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>d(GAGAGAXTCTCCXCTCTCT)</td>
<td>5.5</td>
<td>58</td>
<td>59</td>
<td>2</td>
</tr>
<tr>
<td>d(GAGAGAXTCTCC)</td>
<td>5.5</td>
<td>60</td>
<td>61.6</td>
<td>2</td>
</tr>
<tr>
<td>d(GAGAGAXTCTCCXCTCTCT)</td>
<td>7</td>
<td>14.4</td>
<td>14.9</td>
<td>-2</td>
</tr>
<tr>
<td>d(GAGAGA(L-TCTCCT-TC)</td>
<td>5</td>
<td>11.3</td>
<td>11.9</td>
<td>1.5</td>
</tr>
<tr>
<td>d(AAAAXTTTTTTT)</td>
<td>7</td>
<td>24</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>d(GAGAGA(L-TCTTCCT-L-TCTTC)</td>
<td>5</td>
<td>16</td>
<td>5</td>
<td>1.5</td>
</tr>
<tr>
<td>d(TCTTCCTCXXCTCTCAGAGAG)</td>
<td>5</td>
<td>22</td>
<td>8.3</td>
<td>1.5</td>
</tr>
<tr>
<td>d(GAGAGA(L-TCTTCCT-L-TCTTC)</td>
<td>5</td>
<td>14</td>
<td>1.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>

*Tm* values were obtained from UV melting curves as described in Materials and Methods. X is the hexaethyleneglycol linker.

**Optical spectrophotometry**

It is difficult to determine dissociation constants for ligand binding by footprinting or UV melting. We have therefore utilized optical spectroscopy. Coralyne is a yellow compound that has absorption bands in the near UV that do not overlap with the absorption of DNA. Figure 2 shows a UV-Vis difference spectrum for coralyne binding to the intramolecular triplex d(AAAAAA-L-TTTTT-L-TTTTT). Under the conditions of this experiment the DNA and ligand are each ~80% in the bound form (see below). The largest differences were at ~330 nm, which corresponds to an absorption band of the free ligand (not shown). In addition, there were two positive bands at 405 and 434 nm, which correspond to weak bands of the free ligand. Hence, the ligand electronic transitions are perturbed on binding to the DNA triplex, indicating a substantial change in environment. These differences are particularly useful for titration, as DNA does not absorb above 300 nm. A typical titration curve is shown in Figure 2B and C. At high concentrations of ligand the stoichiometry appears to be 1:1. In principle, there are up to five essentially equivalent intercalation sites in this triplex helix. However, as the DNA concentration is increased, statistically only one site per triplex will be occupied, although not necessarily the same one on each molecule. The apparent dissociation constant and difference absorption coefficient were determined at two different concentrations of ligand, with essentially the same result. In this instance, *Kd* was determined as 0.5 μM, which is dominated largely by binding to the tightest site. A similar *Kd* value was obtained at pH 5 in the presence of Mg, showing that binding itself was sensitive to pH. In contrast, with d(GAGAGA(L-TCTTCCT-L-TCTCT) at pH 5 much weaker binding was observed (Kd ~ 5 μM), although the UV difference spectrum was qualitatively similar to that obtained with d(AAAAXTTTTTT-L-TTTTT). Thus, the TA-T/TA-T intercalation sites bind coralyne up to an order of magnitude more tightly than CG-C/TA-T or AT/C-GC sites (Table 2). This parallels the results obtained by UV melting and is consistent with the information obtained by Nase I footprinting (see above).

Because of the difference in affinity of coralyne for the different sequences, we made further oligonucleotides that contain a single sequential pair of AT-T/AT-T triplets, namely d(GAGAGA(L-CCTTCCLCTTC), d(GAGAGA-L-CCTTCCLCTTC) and d(GAGAGA(L-TCTTCCL-TCTTC), where L is hexaethyleneglycol. The neighbouring TA-T triplets should preferentially interact with the ligand. The UV-Vis difference spectra were qualitatively similar to those obtained with the other oligonucleotides, although the difference absorption coefficient was significantly smaller. For the d(GAGAGA(L-TCTTCCLCTTC) and d(GAGAGA(L-CCTTCCL-CCTTC) triplexes, the binding affinity was somewhat stronger than for d(AAAAXTTTTTT-L-TTTTT) and decreased significantly as the pH was raised to 7.9 (Table 2). The difference absorption coefficient was smaller at pH 8 than at pH 5 and it is possible that there is a different mode of binding to the duplex and (small amount of) triplex under these conditions. The d(GAGAGA(L-TCTTCCL-TCTTC) sequence showed intermediate affinity for coralyne at pH 5 and was only 2-fold lower at pH 8. For comparison, we have also determined the
dissociation constant for naphthoquinoline at pH 5 as 1.3 ± 0.2 μM using the same method. In comparison, this ligand seems to give a stronger DNase I footprint than coralyne (4,5).

To characterize the nature of the coralyne–DNA complexes further, we have used CD, which provides information on perturbation of the electronic environment of the ligand. Figure 3 shows CD spectra of different triplexes in the absence and presence of coralyne. There were substantial changes in the CD when coralyne binds to d(AAAAAA-L-TTTT TT-LTTTTTT), but much smaller ones when it was mixed with d(GAGAAG-

Figure 3. CD spectra. Spectra were recorded at 15°C as described in Materials and Methods. (A) 8.5 μM d(AAAAAA-L-TTTT TT-LTTTTTT) ± 10 μM coralyne in 20 mM NaPi, 100 mM KCl, pH 7. (B) 8.1 μM d(GAGAAG-L-TCTCTCTC-LCTCTCT) ± 10 μM coralyne in 10 mM NaPi, 100 mM KCl, 10 mM MgSO4, pH 5.5

L-CTTCTC-L-CTCTTC), even though the degrees of saturation should have been ~80 and 50% respectively under these conditions (cf. Table 2). Fluorescence quenching experiments were consistent with coralyne intercalating into TA.T triple helices (7) (see also below). It is possible that the drug does not intercalate into the GCC+ -containing triplexes as well as it does into the all-TA triplexes.

NMR spectroscopy

NMR spectra of d(AAAAAA-L-TTTT TT-LTTTTTT) and d(GAGAAG-L-TCTCTCTC-LCTCTCT) ± coralyne were recorded. Stoichiometric binding was observed for the former triplex (Fig. 4A). Under the conditions of these experiments the concentration of potential sites exceeds the concentration of the ligand, so that, on average, only one molecule of coralyne is bound per mole of triplex and excess free triplex is present. The large changes in chemical shifts are consistent with intercalation. However, the number of changes in chemical shifts suggests that more than one kind of complex is present, i.e. intercalation occurs at different sites within the DNA triplex, with somewhat different spectra. The relative intensity of the peaks indicates that these states are populated to a similar extent. Binding was also observed with d(GAGAAG-L-TCTCTCTC-LCTCTCT) (not shown), which was expected given a Kd value in the range 5–10 μM (Table 2). The NMR spectrum of the triplex d(AGAAGA-L-LCTCTCTC-LCTCTCT) was assigned by standard methods and the exchangeable protons are shown in Figure 4B. The amino protons of the
Hoogsteen cytosines resonated at a very characteristic shift (−10.2 and 9.8 p.p.m.) (19). We note that one of the two triplexes shows evidence of more than one conformation of the triplex state. The reasons for this structural heterogeneity are unclear, but seem to be a function of the sequence. On adding coralyne, initially one set of triple helix-specific resonances shifted up to a population of ~50%. After this, a second ligand began to bind to a different site (Fig. 4). This indicates that although the primary site (TA-T/TA-T) does have the highest affinity, neighbouring sites must have affinities that are not much lower. Under the conditions of the NMR experiments (~1 mM DNA) ligand binding should be essentially stoichiometric and the occupation of sites will occur in the ratio of their ligand dissociation constants.

DISCUSSION

The thermodynamic results in combination with the spectroscopic and footprinting data all indicate the same result, namely that coralyne preferentially interacts with TAT triples in triple-stranded DNA. Further, the apparent dissociation constant from d(AAAAA-L-TTTTTT-L-TTTTTT) determined by optical titration is ~0.5 μM. This is in close agreement with the value estimated from Tm measurements on polydisperse material (7). However, our results clearly indicate that this ligand binds nearly an order of magnitude less tightly to triplexes containing alternating CG-C+ and TAT triples. This result is in contrast to the previous conclusions (7) based on the effect of the drug on the disproportionation of polymeric d(GA)-d(TC). They found that in the absence of magnesium the affinity of coralyne for poly[d(TC)d(GA)-d(C+T)] was an order of magnitude greater than for poly[d(T)d(A)-d(T)] (7).

The affinity of coralyne for ‘single site’ triplexes of the kind d(AGAGA-L-TCTTCT-L-TCTTCT), while always higher than for the alternating GA sequence, depends on the position of CG-C+ triples (Table 2). This is may be a consequence of long range electrostatic interactions between the positively charged ligand and the positive charge on the protonated cytosine. We would expect the pK of a terminal GC-C+ to be much lower than that of an internal one (J.-L. Asensio Alvarez, J.Dhesai, S.Bergquist, T.Brown and A.N.Lane, unpublished data), such that at pH 5, terminal CG-C+ triples are only partially protonated and therefore give rise to less repulsion on average.

The binding of coralyne to these triplexes is relatively strong and is at least as tight as the best of the unfused aromatic cations, of which the best are the naphthoquinolines (3–5,11), and considerably tighter than the anthracene-9,10-diones (6).

The NMR data are useful for characterizing both the starting material and the end state. It is clear that it is possible to make a complex in which only one unique site per molecule is occupied. However, at NMR concentrations this requires an excess of DNA over ligand for the triplexes studied here, leading to a mixture of two complex, overlapping spectra. Nevertheless, it is clear from the data presented here that coralyne does indeed intercalate into TA-T triplexes, in agreement with fluorescence quenching experiments (7). It is possible that the weaker binding to 6CC+−containing triplexes (and duplexes) results in a different conformation, in which the planarophore is only partly intercalated. This is suggested by the optical spectra and would be consistent with charge repulsion between the protonated C in the third strand and the positive charge of the ligand. To analyse the conformation of the ligand–triplex complex further, it will be necessary to improve the sequence selectivity further, so that a single, unique complex accounts for >90% of the material in solution. The results presented here indicate that this goal should be achievable and will allow us to determine the details of an intercalated triplex–drug complex and provide essential information about the nature of the molecular interactions.
ACKNOWLEDGEMENTS

This work was supported by the MRC, the Cancer Research Campaign and the Wellcome Trust (Prize studentship to A.M.-A.). J.-L.A.A. gratefully acknowledges a fellowship from the Spanish Ministry of Education. We thank Dr J. Feeney for helpful comments on the manuscript.

REFERENCES