THE SYNTHESIS AND STUDY OF MODIFIED OLIGONUCLEOTIDES

by David M Lough

A thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

1996

Declaration.

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

To Mum, Dad, Murray and Jaq.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my Ph.D. supervisor, Professor Tom Brown whose guidance and encouragement kept me motivated for the past four-and-a-bit years.

I am also very grateful to Oswel DNA Service for the funding of my Ph.D., and to Dorcas Brown and the team at Oswel for their help and use of equipment.

Thanks are in order to the Brown Group troops, past and present, bearded or otherwise, for making the time fly a little too quickly. In particular I would like to thank Tom Barlow, Duncan Graham, John Grzybowski, and the young 'uns. I would also like to thank the Baxter Group, in particular Lisa McIver—everybody needs good neighbours.

^{1}H and ^{13}C NMR spectra were recorded by John Miller, Heather Grant and Wesley Kerr, ^{31}P NMR spectra were recorded Duncan Graham, and FAB MS were recorded by Alan Taylor—all of Edinburgh University Chemistry Department.

Thanks to Laura for some eventful summers.

Lastly, but by no slip of the imagination, least—special thanks to Rhona.

LIST OF ABBREVIATIONS

A	Adenosine
Ac	Acetyl
A_b	Absorbance
AMA	Ammonium hydroxide/aqueous methylamine solution (50:50)
AP	2-Aminopurine
Bz	Benzoyl
C	Cytidine
c^3A	3-Deaza-adenosine
c^7A	7-Deaza-adenosine
c^7G	7-Deaza-guanosine
CE	Capillary electrophoresis
CPG	Controlled pore glass
C_T	Oligonucleotide concentration (total)
c^7z^8G	7-Deaza-8-aza-guanosine
d	2'-Deoxy
D	2,6-Diaminopurine
dA	2'-Deoxyadenosine
Dbf	Di-N-butylformamide
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
dC	2'-Deoxycytidine
DCA	Dichloroacetic acid
DCC	N,N'-Dicyclohexylcarbodiimide
DCM	Dichloromethane
DCU	Dicyclohexylurea
dD	2'-Deoxy-2,6-diaminopurine (diaminopurine)
DEC	N'-(3-Dimethylaminopropyl)-N-ethylcarbodiimide
dG	2'-Deoxyguanosine
DIPEA	N,N-diisopropylethylamine

DMA	N,N-Dimethylacetamide
DMAP	4-Dimethylaminopyridine
Dmf	Dimethylformamide
DMF	N,N-Dimethylformamide (solvent)
DMSO	Dimethyl sulphoxide
DMTr	4,4'-Dimethoxytrityl
DMTrCl	4,4'-Dimethoxytrityl chloride
DNA	Deoxyribonucleic acid
dU	2'-Deoxyuridine
EA	Ethanolamine
ε	Extinction coefficient
εA	Etheno-2'-deoxyadenosine
EDTA	Ethylenediamine tetraacetic acid
Endproc.	End procedure of oligonucleotide synthesis
Et.U	5-Ethyl-uridine
F	Hyperchromicity factor
FAB	Fast atom bombardment
G	Guanosine
ΔG°	Stability (free energy) of duplex formation
HPLC	High performance liquid chromatography
ΔH°	Enthalpy of duplex formation
I	Inosine
IR	Infrared
Ib	Isobutyryl
LCAA	Long chain alkylamine
Mac	Methoxyacetyl
Me.C	5-Methyl-cytidine
mRNA	Messenger RNA
MSCl	Mesitylene sulphonyl chloride
N	Nebularine
NMR	Nuclear magnetic resonance

NPE	2-(2-nitrophenyl)ethyl
NTP	Nucleotide triphosphate
Pac	Phenoxyacetyl
PAO-TMG	Pyridine aldoximate-tetramethyl guanidine
Phac	Phenylacetyl
Ph_3P	Triphenyl phosphine
Pixyl	9-Phenylxanthenyl-9-yl
Pu	Purine
Py	Pyrimidine
r	Ribo
R	Gas constant
RNA	Ribonucleic acid
RNase H	Enzyme that hydrolyses RNA when part of a DNA-RNA hybrid
$\Delta S°$	Entropy of duplex formation
T	Thymidine
TBAZ	Tetrabutyl ammonium azide
TCA	Trichloroacetic acid
TDIC	Tolylene-2,6-diisocyanate
TEA	Triethylamine
Tfa	Trifluoroacetyl
TFAAn	Trifluoroacetic anhydride
THF	Tetrahydrofuran
TIPS	Tetraisopropyl-disiloxane
$TIPSCl_2$	1,3-Dichloro-1,1,3,3-tetraisopropyl-disiloxane
TLC	Thin-layer chromatography
T_m	Melting temperature
TPSCl	2,4,6-Tris-isopropyl sulphonyl chloride
tRNA	Transfer RNA
U	Uridine
X	Xanthosine
z^2I	2-Aza-inosine.

ABSTRACT

2-Aminoadenine or 2,6-diaminopurine (D), is an adenine analogue where three hydrogen bonds are formed in a base pair with thymine, instead of only two hydrogen bonds in the A.T base pair. Hence the thermodynamic stability of duplexes containing D.T base pairs should be increased relative to those containing A.T. More subtle interactions affect the structural transitions of duplexes containing diaminopurine.

Previous attempts at incorporating diaminopurine into DNA during routine oligonucleotide synthesis were hindered due to acid catalysed depurination (dependent on N^6-amino protection) and very poor lability of the N^2-amino protecting group during ammonia deprotection.

An improved N^2-amino and N^6-amino protection strategy was developed for the synthesis of the diaminopurine monomer. This monomer was used successfully during routine machine synthesis to make diaminopurine-containing oligonucleotides. Derivatised solid support was synthesised to allow 3'-incorporation of diaminopurine. A novel synthetic route to the free diaminopurine nucleoside was also developed.

The ability of nucleic acids to form very stable duplexes has many important applications in diagnostic and therapeutic molecular biology. The hybridisation of DNA probes or PCR primers to their targets may be increased, and hence the specificity; or allow shorter probes/primers to be employed. The same strategy may be applied to antisense therapy, or DNA sequencing primers.

In order to predict the physical properties of such duplexes, the thermodynamic stability (UV melting) of duplexes containing diaminopurine was determined, using oligonucleotides with all combinations of nearest neighbours. This data revealed that when diaminopurine replaced adenine, stability of the duplex was increased, with the exception of the TD/AT nearest neighbour interaction. This showed that the additional hydrogen bond does not always increase stability—more complex interactions are responsible for overall duplex stability.

The cleavage, after synthesis, of an oligonucleotide bound to its solid support, is a time-consuming process. Investigations were carried out on a phthaloyl linker, and preliminary results showed substantial reductions in the cleavage times, compared to the standard succinyl linker. This may give rise to a family of fast-cleaving linkers based on the phthaloyl group.

CONTENTS

PAGE

INTRODUCTION AND BACKGROUND 1

1.1 GENERAL INTRODUCTION 2

1.1.1 INTRODUCTION TO DNA 2

1.1.2 CHEMICAL SYNTHESIS OF OLIGONUCLEOTIDES 4

1.1.2.1 Protecting Groups 4

1.1.2.2 Phosphoramidite Solid Phase Synthesis 6

1.2 DUPLEX NUCLEIC ACID STRUCTURE 10

1.2.1 CONFORMATIONS OF DNA 10

1.2.1.1 Idealised Models of DNA 11

1.2.1.2 Other Double Helical Structures 13

1.2.2 THE INFLUENCE OF SEQUENCE ON DUPLEX STRUCTURE 13

1.2.2.1 The Effect of Base Composition 14

1.2.2.2 Environmental Effects 14

1.3 FACTORS AFFECTING DUPLEX STABILITY 15

1.3.1 THE SUGAR-PHOSPHATE CHAIN 15

1.3.1.1 Modifications to the Sugar-Phosphate Backbone 16

1.3.2 BASE PAIRING 16

1.3.2.1 Non-Watson-Crick Base Pairs 17

1.3.2.2 Modified Bases 19

1.3.3 BASE STACKING AND HYDROPHOBIC INTERACTIONS 19

1.3.4 ENVIRONMENTAL INTERACTIONS 20

1.3.4.1 The Role of the Solvent 20

1.3.4.2 Ionic Interactions 20

1.4 MODIFICATION OF THE HETEROCYCLIC BASES 21

1.4.1 PYRIMIDINE MODIFICATIONS 21

1.4.2 PURINE MODIFICATIONS 22

1.4.3 MODIFIED BASES FOUND IN NATURE 25

1.4.4 SYNTHETIC WATSON-CRICK BASE PAIRS 26

1.5 THE BASE ANALOGUE 2,6-DIAMINOPURINE 29

1.5.1 BASE PAIRING INVOLVING DIAMINOPURINE 30

1.5.2 THE EFFECT OF dD ON DUPLEX STABILITY 30

1.5.2.1 Initial Studies on Oligonucleotides Containing dD 30

1.5.2.2 The Spine of Hydration and (dA).(dT) Tracts 31

1.5.3 THE EFFECT OF dD ON DUPLEX STRUCTURE 32

1.5.3.1 Comparing (A).(T) and (D).(T) Tracts 33

1.5.3.2 Comparing (A.T) and (D.T) Tracts 33

1.5.3.3 Structural Transitions of Poly(dD.dT) 34

1.5.3.4 B-Z Structural Transitions 37

1.6 SYNTHESIS OF DIAMINOPURINE NUCLEOSIDES 38

1.6.1 A REVIEW OF DIAMINOPURINE SYNTHESIS 38

1.6.2 PROTECTION STRATEGY 41

1.6.2.2 N^6-Amidine Protecting Groups 43

1.6.2.3 The Effect of N^6 Protection on Depurination 44

1.6.2.4 N^6 Protection for Diaminopurine 46

1.6.2.5 N^2 Protection for Diaminopurine 48

1.7 DNA CONTAINING DIAMINOPURINE 49

1.7.1 DNA PROBES 49

1.7.1.1 Gene probing 49

1.7.1.2 Ambiguity in probing 50

1.7.2 PCR PRIMERS AND DNA SEQUENCING 50

1.7.3 ANTISENSE THERAPY 51
1.7.4 MOLECULAR RECOGNITION OF DIAMINOPURINE 52
1.7.4.1 Probing restriction enzymes 52
1.7.4.2 Replication of DNA containing diaminopurine 53
1.7.4.3 Transcription involving diaminopurine 54

1.8 BASE-LABILE LINKERS AND AMINO PROTECTION 55
1.8.1 THE SOLID SUPPORT AND COVALENT LINKER 55
1.8.1.1 The solid support 55
1.8.1.2 The covalent linker 57
1.8.2 BASE-RESISTANT LINKAGES 57
1.8.3 BASE-LABILE LINKERS 58
1.8.4 ULTRAFAST DNA SYNTHESIS 59
1.8.4.1 Base-labile protecting groups 60
1.8.4.2 More nucleophilic deprotection reagents 61

RESULTS AND DISCUSSION 63

2.0 SYNTHESIS OF dD PHOSPHORAMIDITES 64
2.1 N^2-PHENYLACETYL PROTECTION OF DIAMINOPURINE 64
2.1.1 Synthesis of the N^2-phenylacetyl-dD nucleoside **[5]** 64
2.1.2 Synthesis of the N^2-phenylacetyl-dD monomer **[7]** 66
2.1.3 Oligonucleotide synthesis using monomer **[7]** 66
2.2 N^2-ACETYL PROTECTION OF DIAMINOPURINE 70
2.2.1 Synthesis of the N^2-acetyl-dD monomer **[12]** 70
2.2.2 Oligonucleotide synthesis using **[12]** 70
2.2.3 Synthesis of N^6-dimethylformamide-dD monomer **[14]** 74
2.2.4 Synthesis of N^6-azido-dG monomer **[16]** 75
2.3 N^2-DMF-DIAMINOPURINE VIA 3',5'-BIS-ACETYL PROTECTION 77
2.3.1 Synthesis of N^2-dmf-N^6-azido-diaminopurine nucleoside **[21]** 77

2.4 SYNTHESIS OF 3',5'-TIPS-NUCLEOSIDES 80
2.4.1 Synthesis of 3',5'-TIPS-N^2-dimethylformamide-diaminopurine **[24]** 80
2.4.2 Synthesis of 3',5'-TIPS-N^2-trifluoroacetyl-diaminopurine **[27]** 82
2.5 SYNTHESIS OF THE N^2-DMF, N^6-DBF-dD MONOMER 85
2.5.1 Synthesis of the free-diaminopurine nucleoside **[28]** 85
2.5.2 Synthesis of diaminopurine monomer **[32]** 85
2.5.3 Stability of nucleosides to TCA 87
2.5.4 Oligonucleotide synthesis using monomer **[32]** 88
2.5.5 Synthesis of diaminopurine-succinyl CPG **[34]** 91
2.5.6 Oligonucleotide synthesis using succinyl CPG **[34]** 92
2.6 CONCLUSIONS AND FUTURE WORK 93

3.0 THE THERMODYNAMICS OF DNA CONTAINING dD 95
3.1 THERMODYNAMIC COMPARISON OF $d(TA)_4$ AND $d(TD)_4$ 95
3.1.1 UV Melting curves of $d(TA)_4$ and $d(TD)_4$ 95
3.2 THERMODYNAMIC PARAMETERS OF DNA CONTAINING dD 97
3.2.1 Calculating thermodynamic parameters from melting curves 98
3.2.2 Obtaining the data 98
3.2.3 Manipulation of the thermodynamic data 104
3.2.4 Interpretation of the thermodynamic data for diaminopurine 106
3.3 CONCLUSIONS AND FUTURE WORK 108

4.0 THE SYNTHESIS OF BASE-LABILE LINKERS- A PRELIMINARY STUDY 110
4.1 THE SYNTHESIS OF PHTHALOYL LINKERS 110
4.1.1 Synthesis of phthaloyl linker **[40]** 111
4.1.2 Synthesis of nitro-phthaloyl linkers **[41]** and **[42]** 113
4.1.3 Comparing the phthaloyl linkers with the succinyl linker 116
4.2 COMPARING THE PHTHALOYL AND SUCCINYL LINKERS 117
4.2.1 Comparing the quality of oligonucleotide synthesis 118

4.2.2 Cleavage study—a more detailed study 120
4.3 CONCLUSIONS AND FUTURE WORK 122

EXPERIMENTAL SECTION 124

5.1 SOLVENTS AND REAGENTS 125
5.2 ANALYTICAL TECHNIQUES 125
5.3 SYNTHESIS OF DIAMINOPURINE NUCLEOSIDES 128
5.4 OLIGONUCLEOTIDE SYNTHESIS 161
5.5 UV MELTING STUDIES 164
5.6 HYDROGENATION CONDITIONS 166
5.7 STABILITY OF NUCLEOSIDES TO TCA 167
5.8 SYNTHESIS OF PHTHALOYL LINKERS 168

REFERENCES 175

APPENDICES 184

APPENDIX 1: Predicted and determined stability of **Duplex 10**. 185
APPENDIX 2: Publication. 186

INTRODUCTION

1.1 GENERAL INTRODUCTION

1.1.1 INTRODUCTION TO DNA

Since the discovery of the nucleic acids in the latter part of the last century, several decades elapsed before the structures of the constituent purines and the corresponding nucleosides, and the nature of the internucleotide bonds in the polynucleotides were clarified. By 1952, the chemical structures of both classes of nucleic acids, RNA and DNA, had been established. Soon followed the Watson-Crick proposal for the DNA structure[1] and this heralded a new era in chemical, biochemical, and molecular biological studies.

Nucleic acids are biopolymers built from monomeric units known as nucleotides. These nucleotides are composed of a sugar-phosphate backbone, and a heterocyclic base which is one of four specific "letters" which makes up the genetic code. These bases- adenine (A), guanine (G), cytosine(C), and thymine(T)- are shown below.

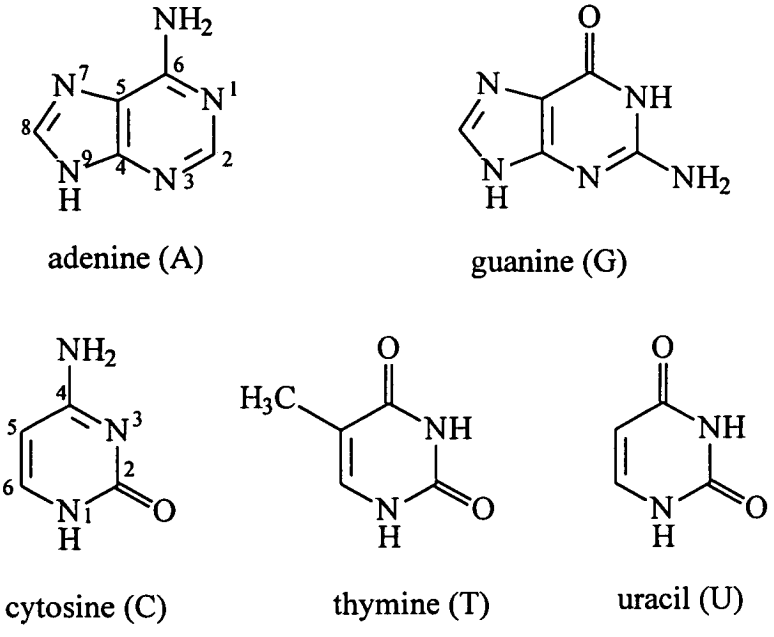
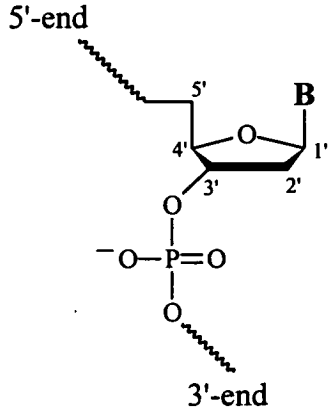


Figure 1.1. *The five bases found in nucleic acids.*
(Uracil replaces T in RNA).

The bases are joined to the D-2'-deoxyribose sugar via a β-glycosyl linkage, and the phosphodiester junction links the sugar groups at the 3' and 5' ends, to give the oligonucleotide a directional nature.



B= A, C, G or T.

Figure 1.2. *The Repeating Nucleotide Unit.*

The double-helical model for DNA proposed by Watson and Crick was instrumental in helping understand the role of DNA in biology.[1] The idea of two complementary strands was developed from the recognition that only two kinds of base pair can occur involving the major tautomeric forms of G with C, and A with T. This idea supported earlier work[2] that identified the ratios of guanine to cytosine, and adenine to thymine to always be very close to 1:1.

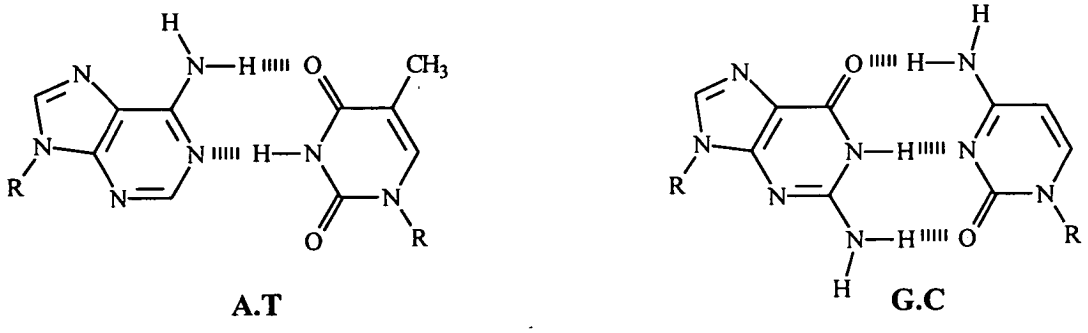


Figure 1.3. *Watson-Crick base pairing between A.T and G.C.*
(R denotes the D-2'-deoxyribose sugar).

The Watson-Crick model also required that the two strands should run in opposite directions (antiparallel), with the strands held together with hydrogen bonding.

DNA: The Central Dogma of Molecular Biology

DNA is the repository of genetic information for the vast majority of organisms on Earth. DNA is copied by the process of replication, using each complementary parent strand as a template for the daughter strands. DNA also serves as a template on which transcription of RNA occurs, this messenger RNA (mRNA) is then translated into a peptide chain in the ribosome, via transfer RNA (tRNA), which carries the individual amino acids to the ribosome.

1.1.2 CHEMICAL SYNTHESIS OF OLIGONUCLEOTIDES

As with other classes of biological macromolecules, interest in the synthesis of oligo- and polynucleotides increased very rapidly and systematic studies on the synthesis of polynucleotides were undertaken in the mid-1950's. Synthesis of oligonucleotides of defined sequences, in conjunction with enzymatic approaches enabled the synthesis of high molecular weight polymers that were successfully used in the elucidation of the genetic code and later in the total synthesis of genes fully functional *in vivo*.

Nucleic acids are highly sensitive to a wide range of chemical reactions. The heterocyclic bases are prone to alkylation, oxidation, and reduction. DNA is vulnerable to acidic hydrolysis, as opposed to RNA where alkaline hydrolysis is likely. Such considerations limit the synthesis of oligodeoxyribonucleotides to only the mildest of chemical reactions.

An oligonucleotide is synthesised by assembling monomeric blocks. Each block features at least a nucleophilic and an electrophilic function, ie. the 5'-OH and the 3'-function (phosphate, phosphoramidite, or phosphonate), for nucleotides. The nucleophilic and electrophilic sites are linked together in the coupling step. Protection is a necessity. It guarantees the chemoselectivity of coupling and the solubility of the synthons in organic solvents.

1.1.2.1 Protecting Groups

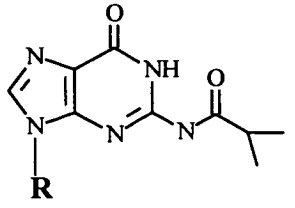
There are two classes of protecting groups: persistent and transient.[3]

The persistent protections remain on the oligonucleotide throughout the synthesis, and are cleaved at the very end. They cap the functional groups of the heterocyclic base and the phosphate oxygen. The transient protections block the functional groups to be coupled (5'-OH) and can be specifically cleaved before each coupling reaction.

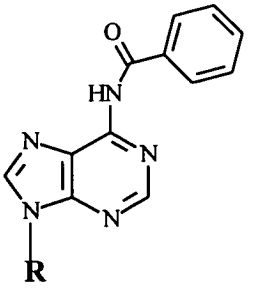
The most convenient way to assemble an oligonucleotide is to utilise preformed deoxynucleoside phosphates as basic building units and to couple these sequentially to a terminal nucleoside attached to a solid support. Since the primary (5')-hydroxyl is a more effective nucleophile than the secondary (3')-hydroxyl, the phosphate is best placed on the 3'-position. To achieve selectivity, it is necessary to protect both the heterocyclic amino groups and the 5'-hydroxyl group.

Heterocyclic bases

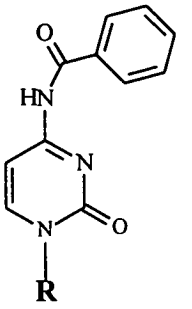
Since the initial work of Khorana,[4,5] most studies on the synthesis of oligonucleotides have used N-acyl protecting groups for adenine, cytosine, and guanine residues. The purine and pyrimidine amino groups are normally protected as N^6-benzoyl-2'-deoxyadenosine, N^2-isobutyryl-2'-deoxyguanosine, and N^4-benzoyl-2'-deoxycytosine.[6] These acyl groups are removed by alkaline hydrolysis during the ammonia treatment. Therefore there is orthogonality in the protection strategy: the permanent protection being base labile, while the sugar protection, which needs to be removed for each coupling cycle, is acid labile.



N^2 Isobutyryl-dG



N^6-benzoyl-dA

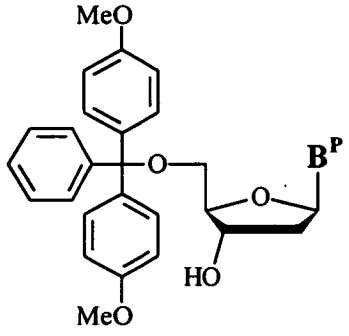


N^4-benzoyl-dC

Figure 1.4. *The acyl groups commonly used in persistent exocyclic amino protection.* (**R** denotes the D-2'-deoxyribose sugar).

The 5'-hydroxyl group

The 5'-hydroxyl functions are usually protected with the acid labile trityl ether (again introduced by Khorana[7]). Most often the 4,4'-dimethoxytrityl group (DMTr) is employed, due to its high acid lability which reduces the problem of acid-catalyzed depurination during deprotection of the 5'-hydroxyl at the start of each coupling cycle.



B^P denotes a protected base.

Figure 1.5. *Standard 4,4'-dimethoxytrityl protection.*

1.1.2.2 Phosphoramidite Solid Phase Synthesis

The phosphoramidite method is an improved version of the phosphite triester method pioneered by Letsinger in the 1970's.[8] Letsinger showed that 2-chlorophenyl phosphorodichloridate would phosphitylate a 5'-O-protected thymidine at low temperatures.[9] Such phosphorodichloridates were exceptionally reactive–too much so for common use.

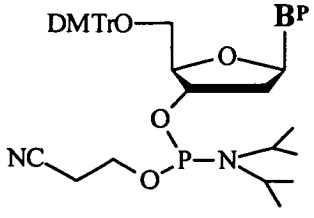


Figure 1.6. *A typical phosphoramidite monomer, showing the cyanoethyl and diisopropyl groups.*

However, very shortly afterward, Beaucage and Caruthers found that nucleoside-3'-O-phosphoramidites could be synthesised, stored, and activated for coupling by acid catalysis.[10] Oxidation then led to the phosphate diester. This technique has been developed into the most widely used method for automated oligonucleotide assembly,[6] with the phosphitylating agent: 2-cyanoethyl-N,N-diisopropylamino phosphorochloridate, first developed by McBride and Caruthers,[11] emerging as one of the most widely used. Figure 1.6 shows the persistent cyanoethyl function, and the N,N-diisopropyl function which is protonated by the addition of tetrazole during the coupling stage of phosphoramidite synthesis.

Solid Phase Synthesis

The pioneering work carried out by Merrifield[12] on the synthesis of peptides on a solid support, led to a similar breakthrough in oligonucleotide synthesis, following work by Letsinger.[13]

Solid phase synthesis lends itself to automation of the repetitive steps required for oligonucleotide assembly, and has the added advantages of speed, high yields, and ease of separation of products from impurities. All the required reagents are delivered to the growing oligonucleotide which is fixed to an inert support, usually controlled pore glass (CPG) beads,[14] via a succinyl linker[15]. Hence all the chemistry takes place within the reaction vessel (column), and work up between steps consists of flushing away excess reagents and by-products, then introducing reagents for the next step.

The first step in solid phase phosphoramidite synthesis is the detritylation of the solid support-bound nucleoside, using TCA. Step 2 involves activating the next monomer with tetrazole, and coupling this to the support-bound nucleoside, in step 3. In step 4, any unreacted 5'-OH are capped with acetic anhydride. In step 5, the labile phosphorus linkage is converted into a stable phosphate triester using I_2/H_2O. Steps 1-5 are repeated until the desired oligonucleotide chain is assembled, after which the oligonucleotide is cleaved from its support using conc. ammonia, and deprotected.

The synthesis cycle is shown below in Figure 1.7.

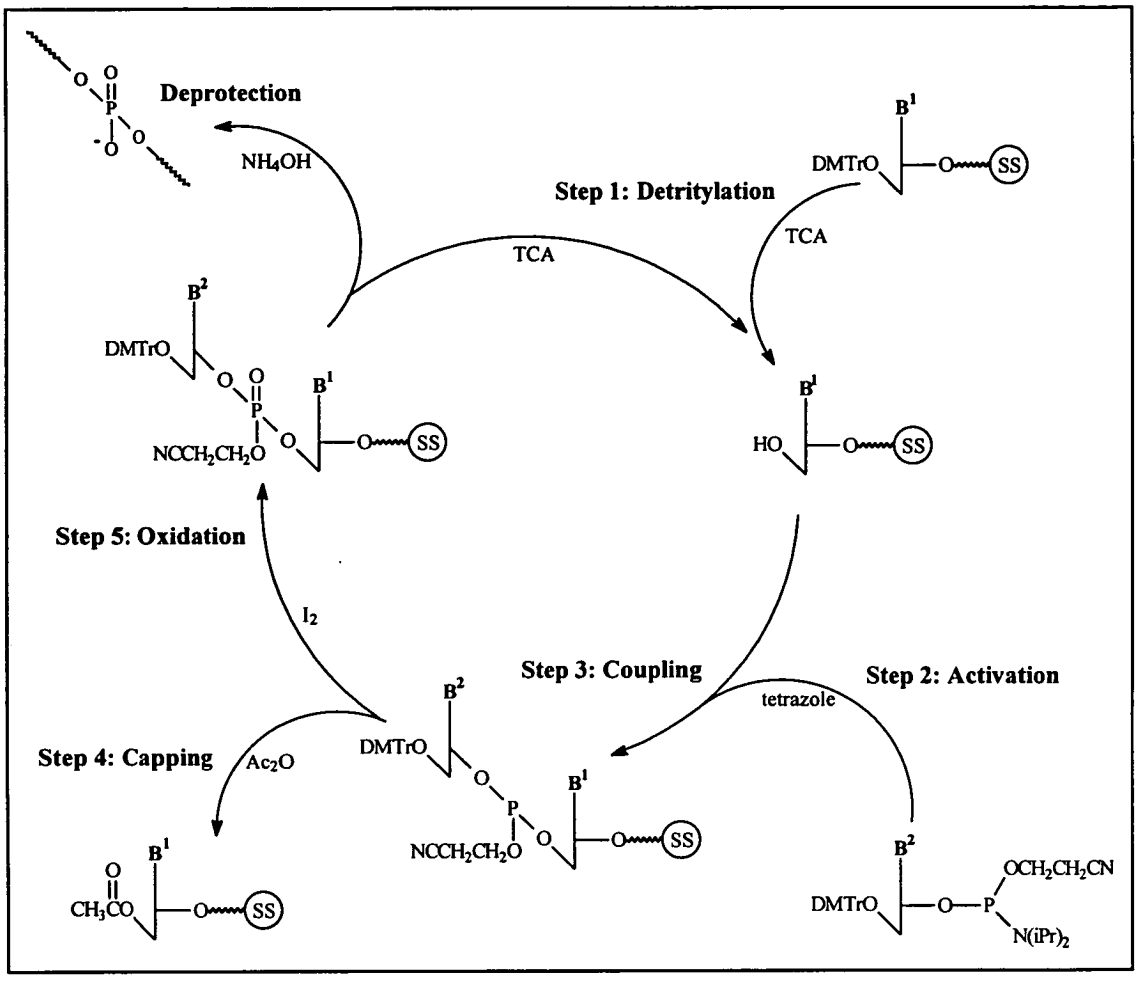


Figure 1.7. *Solid phase oligonucleotide synthesis, by the phosphoramidite approach.*

Future Developments

Synthetic oligonucleotides are used as primers in the polymerase chain reaction (PCR), in DNA sequencing, and in diagnostic probes (with or without labels), and these are applications which have now become routine.

Academic interest is currently in the development of modified DNA/RNA molecules, and in the improvement of RNA synthesis. As functions and types of heterocyclic residues used in oligonucleotide synthesis diversify, protection strategies have to be adapted.

Future developments include the application of large- and very large-scale synthesis to DNA therapeutics, (the antisense strategy). At the other end of the

spectrum, a much smaller scale synthesis is of great interest in developing arrays of oligonucleotides in hybridisation-based DNA "chips" for use in rapid sequencing techniques or diagnosis.[16]

1.2 DUPLEX NUCLEIC ACID STRUCTURE

DNA and RNA duplexes can assume several distinct structures that vary with such factors as the nature of the cations present, the humidity, and the base sequence.[17] Double-stranded DNA is conformationally variable; the major conformational states are discussed below.

1.2.1 CONFORMATIONS OF DNA

B-DNA, also known as the "Watson-Crick" structure, is regarded as the native form because its X-ray diffraction pattern resembles that found for the DNA in intact sperm heads. B-DNA consists of a right-handed double helix of antiparallel sugar-phosphate chains with ~10 base pairs per turn and with the bases perpendicular to the helix axis. The aromatic bases have van der Waals thicknesses of 3.4Å, and are partially stacked on each other, therefore the helix has a pitch (rise per turn) of 34Å. Bases on opposite strands hydrogen bond in a geometrically complementary manner to form exclusively A-T and G-C Watson-Crick base pairs. B-DNA has two deep exterior grooves, of unequal size, that wind between its sugar-phosphate chains. The *minor groove* is that in which the C1'-helix axis-C1' angle is <180°, whereas the *major groove* opens towards the opposite edge of each base pair.[17] (*See* Figure 1.8(b)).

A-DNA. When the relative humidity is reduced to 75%, B-DNA undergoes a reversible transformation to a wider, flatter right-handed double helix known as A-DNA. A-DNA has 11 base pairs per turn and a pitch of 28Å, leaving an axial hole. A-DNA's most striking feature, however, is that the planes of its base pairs are tilted into the helix axis. A-DNA has a deep major groove and a very shallow minor groove. Although B-DNA is the most common form found in nature, most self-complementary oligonucleotides of <10 base pairs crystallise in the A-DNA conformation.

Z-DNA. Complementary oligonucleotides, with alternating purine-pyrimidine base sequences, can take up the Z-DNA conformation at high salt concentrations. This helix has 12 Watson-Crick base pairs per turn, a pitch of 45Å, a deep minor groove and no discernible major groove. Z-DNA is unusual in that it adopts a left-handed double-helix, and the purines are in the *syn* conformation. The base pairs in Z-DNA are flipped 180° relative to those in B-DNA. As a consequence, the repeating unit of Z-DNA is a dinucleotide, and the phosphate groups follow a zigzag path around the helix. A high salt concentration stabilises Z-DNA by reducing the otherwise increased electrostatic repulsions between closest approaching phosphate groups on opposite strands (8Å in Z-DNA, 12Å in B-DNA).[18] The methylation of cytosine at the C5-residue also promotes Z-DNA formation since the hydrophobic methyl group in this position is less exposed to solvent in Z-DNA than it is in B-DNA.[19] It is thought that the reversible conversion of specific segments of B-DNA to Z-DNA may act as a switch in regulating genetic expression,[20] but the *in vivo* existence of Z-DNA has been difficult to prove.

1.2.1.1 Idealised Models of DNA

Table 1.1 below summarises the *average* structural features of ideal A-, B-, and Z-DNA. However, real DNA is not ideal. Individual residues depart significantly from the average conformation in a sequence-specific manner. For example, the self-complementary "Dickerson" dodecamer d(CGC GAA TTC GCG) crystallises in the B-conformation.[21] Overall, this molecule shows average helical parameters per residue nearly equal to that of ideal DNA. However, the helical twist per base pair in this dodecamer ranges from 28-42°, and other deviations include distortions due to propeller twisting. Indeed, X-ray and NMR studies of many other double helical DNA oligomers catalogue these sequence-specific irregularities.[22,23]

Figure 1.8. *Ideal A-, B-, and Z-DNA.* (Overleaf).
The striking difference between the three major types of DNA (apart from the left-handedness of Z-DNA), is that the A helix appears to be broader and more compressed along its axis relative to B and Z. In B-DNA, the base pairs sit almost directly on the axis so that the major and minor grooves are of equal depth. In A-DNA the pairs are displaced off-axis, making the major groove very deep, and the minor groove extremely shallow. In Z-DNA the base pairs are displaced in the opposite direction, yielding a major groove that is little more than a shallow depression, and a very deep minor groove.[24]

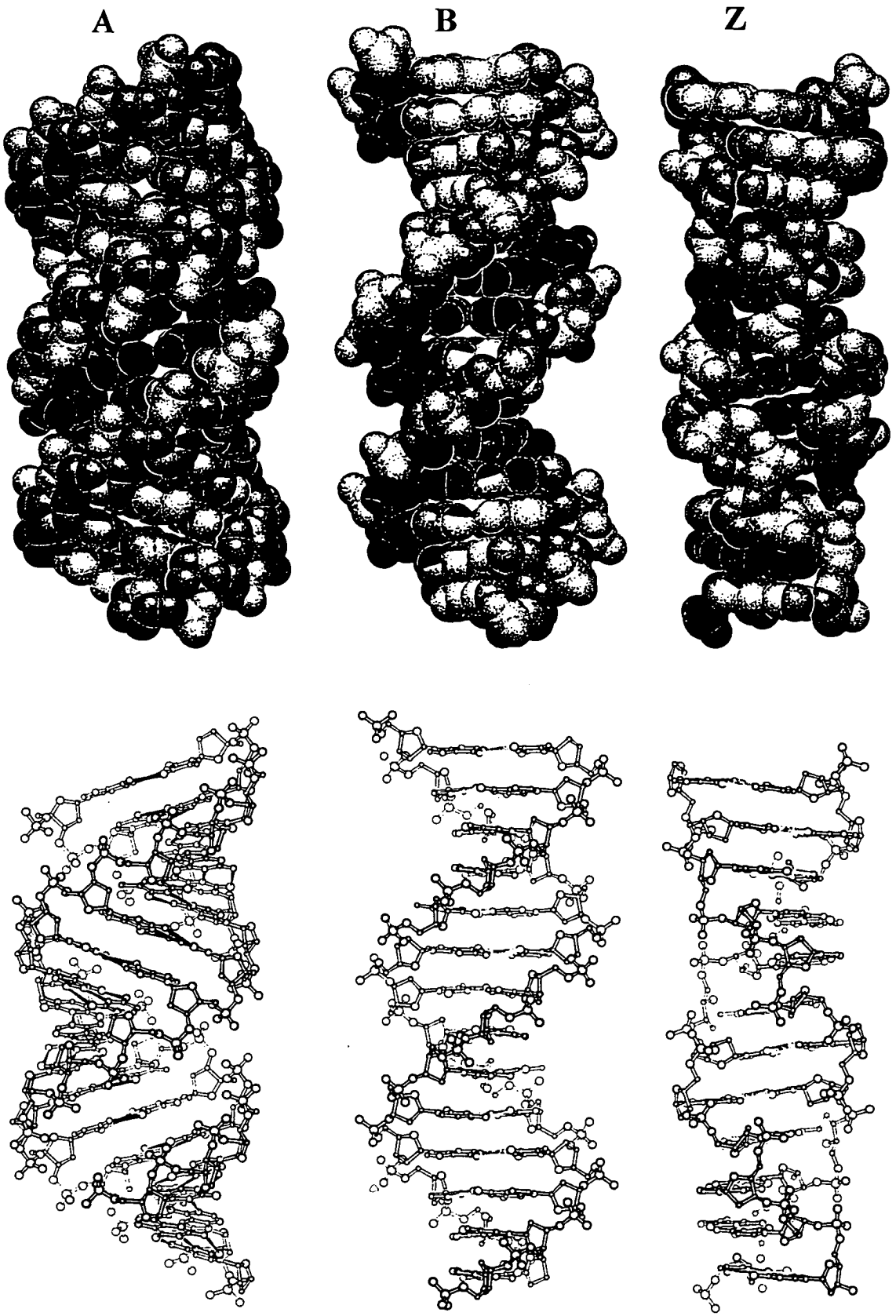


Figure 1.8. *Ideal **A**-, **B**-, and **Z**-DNA. The space-filled representations are shown along with ball and stick structures.*

Table 1.1. Comparing Ideal A-, B-, and Z-DNA.

	A	*B*	*Z*
Helical sense	Right handed	Right handed	Left handed
Base pairs per turn	11	10	12 (6 dimers)
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C3'-*endo*	C2'-*endo*	C2'-*endo* (py); C3'-*endo* (pu)
Glycosidic bond	*Anti*	*Anti*	*Anti* (py); *syn* (pu)

1.2.1.2 Other Double Helical Structures

A-RNA. Double helical RNA cannot assume a B-DNA-like conformation because its 2'-OH groups would cause steric clash. Instead, RNA adopts a conformation resembling A-DNA, known as A-RNA, which has 11 base pairs per turn, a pitch of 30Å, and a base pair inclination of ~14°.[17] Many RNA structures, for example, transfer and ribosomal RNAs, contain complementary sequences that form double helical stems.[25]

DNA-RNA Hybrids. These double helices, which consist of one strand each of DNA and RNA, possess both "A-like" and "B-like" characteristics.[26]

Segments of DNA-RNA hybrid helices must occur in the transcription of RNA on DNA templates;[27] in the initiation of DNA replication by short primers of RNA;[28] and in the use of heteroduplex molecules for antisense regulation of gene expression.[29] RNA may not only serve as the primer in DNA replication, but also, by inducing and stabilising an A conformation may play a role in longer-range interactions.[30]

1.2.2 THE INFLUENCE OF SEQUENCE ON DUPLEX STRUCTURE

Different structures may be obtained in duplexes of similar overall base composition, but different base sequence. For example poly(dA).(dT) possesses an unusual B-DNA conformation which does not undergo a B-A transition.[31]

Poly(dA.dT), on the other hand, is thought to exist as two conformers which more readily adopt a metastable A-form.[32] (*See* Section 1.5.3 for further discussion).

Another example of changing sequence, and consequently the structure, is seen in DNA-RNA hybrids. Melting studies of synthetic homopolynucleotides have shown that the hybrid: poly(dA-rU) is much less stable than the poly(rA-dT).[33] This work was followed up with studies on heptamer duplexes containing a central $(A)_5$-$(T/U)_5$ tract.[27] This work supports the idea that instability of the DNA-RNA hybrid plays a role in the termination of transcription. Other work with heteroduplexes[34,35] shows an order of stability which differs to that found for homoduplexes, where GC content influences the magnitude of relative stabilities. The structural conformations of double helical DNA-DNA and DNA-RNA, will be discussed in more detail, later, along with the transitions they may undergo. The examples given will be relevant to the structure and stability of duplexes containing diaminopurine. (*See* Sections 1.5.2, 1.5.3, 3.1 and 3.3).

1.2.2.1 The Effect of Base Composition

The most stable duplexes contain A.T/U and G.C Watson-Crick base pairs. Each possible duplex varies in intermolecular interactions, particularly hydrogen bonding. Along a helical chain, base-stacking interactions also depend on the exact sequence because of the asymmetry of the heterocyclic bases. The ordering of duplexes leads to the partition of thermodynamic stability between specific base pair hydrogen bonding and base-stacking, often called horizontal and vertical ordering, respectively. Sequence-dependent properties of double stranded nucleic acids are frequently discussed in terms of base stacking.[36]

1.2.2.2 Environmental Effects

There is a myriad of environmental parameters that can affect duplex stability. These parameters include: the pH of the solution, temperature, ionic strength, counter-ion type, and hydration patterns.[36]

The environmental and compositional effects are discussed along with other parameters, in Section 1.3 overleaf.

1.3 FACTORS AFFECTING DUPLEX STABILITY

Now that the structural families adopted by double helical nucleic acids have been discussed, I will describe the forces acting on these structures that contribute to the overall stability of the duplex. These forces, which may affect the structure, invariably affect the stability of the duplex.

1.3.1 THE SUGAR-PHOSPHATE CHAIN

The rotation of a base about its glycosidic bond is greatly hindered. Purines have two sterically allowed orientations relative to the sugar: *syn* and *anti.* Only pyrimidines readily form the *anti* conformation. In most double helical nucleic acids, all bases are in the *anti* conformation, with the exception of Z-DNA.

To relieve the crowding of the substituents in the eclipsed ribose ring, the ring puckers into a half-chair conformation, where the out-of-plane atom is endo-displaced. There are two main orientations of the ring, depending on which atom (C2' or C3') is out-of-plane. In DNA, C2'-*endo* pucker of the deoxyribose leads to the formation of mostly B-type helices. Steric hindrance, due to the O2' oxygen in RNA, alters the ribose conformation, resulting in a C3'-*endo* pucker. There are several possible interactions of the O2' oxygen which then stabilise the resulting A-type conformation of RNA. In Z-DNA, the purine nucleotides are all C3'-*endo* and the pyrimidine nucleotides are C2'-*endo*.

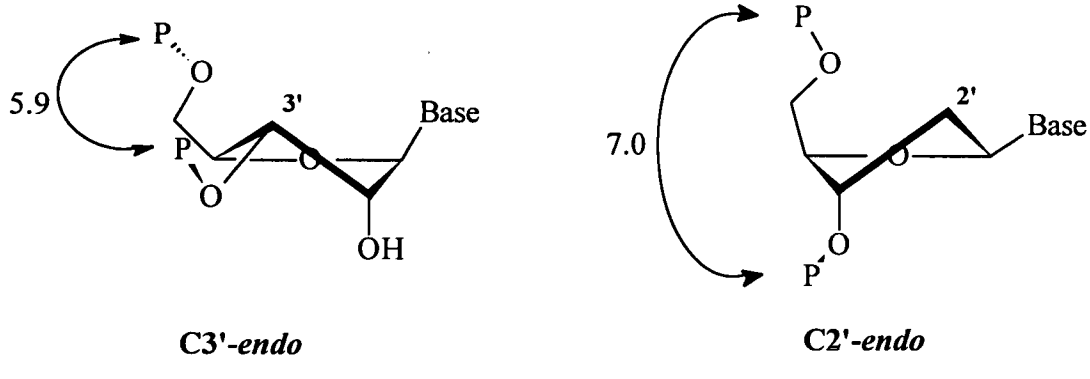


Figure 1.9. *Sugar puckers, showing the two main families of sugar conformations in A-RNA (left), and B-DNA (right). Phosphate-phosphate distances are indicated in Å.*

The various torsion angles in the sugar-phosphate chain are also sterically constrained. If the angles were completely free to rotate, there would be no stable nucleic acid structure. Nevertheless, the sugar-phosphate backbone is by no means rigid, so following denaturation, it assumes a random coil conformation. This relative flexibility also means that the backbone plays only a minor role in determining sequence-dependent structure.[37]

1.3.1.1 Modifications to the Sugar-Phosphate Backbone

There have been many studies carried out on the modification of the sugar and the phosphate groups, particularly for applications in antisense therapeutics.[38,39] Electronegative substituents on the C2' moiety stabilise the C3'-*endo* configuration. Therefore, DNA with certain C2'-substituents mimics RNA and can form more stable duplexes with target RNA,[39] since the favoured A-form duplex is adopted.

An example of backbone modification is the incorporation of amide groups, giving an achiral linkage which increases the propensity for duplex formation.[39] This is in contrast to the first generation of backbone-modified DNA where the phosphorothioate group destroyed the local centre of symmetry, and created an additional centre of chirality, leading to a reduction in binding affinity to the target RNA. Geometrical factors are predominant in influencing the affinity of a modified oligonucleotide for its target. Adjusting the distance between the sugar moieties and choosing the right degree of rigidity (pre-organisation) for the backbone are important parameters which influence the binding behaviour.[39]

1.3.2 BASE PAIRING

Base pairing is responsible for bringing together double-stranded nucleic acids. Hydrogen bonding between bases makes Watson-Crick geometry the preferred mode of base pairing in double helices—only the bases of Watson-Crick pairs (ie. G.C and A.T, shown in Figure 1.3) have a high mutual affinity. (Non-Watson-Crick base pairs are generally of low thermal stability). Hence, Watson-Crick base pairing is both geometrically and electronically favoured.

1.3.2.1 Non-Watson-Crick base pairs.

In biological systems, non-Watson-Crick base pairing may be split into two types: mispairing involving standard bases, and mispairing involving damaged or modified bases. Both types may lead to a mutagenic lesion; and both may have a profound effect on duplex stability.

Mispairing Involving Natural Bases

In addition to the standard G-C and A-T base pairs, there are eight possible mispairs of varying thermodynamic stability.[40] Examples of these wobble base pairs are shown below.

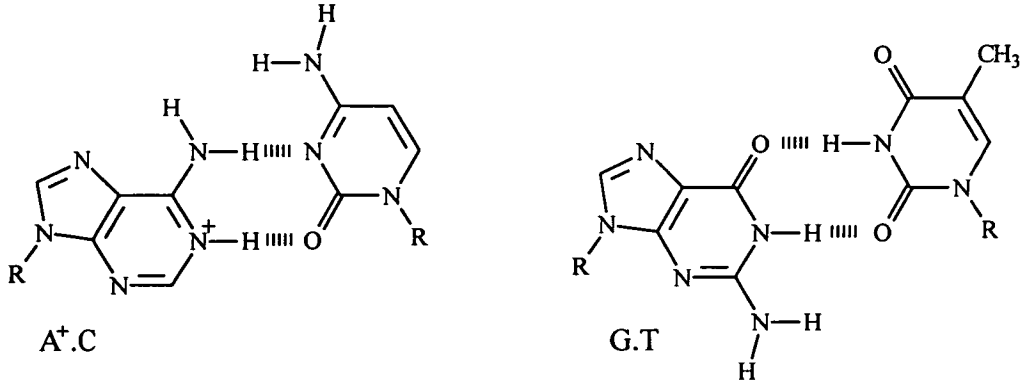
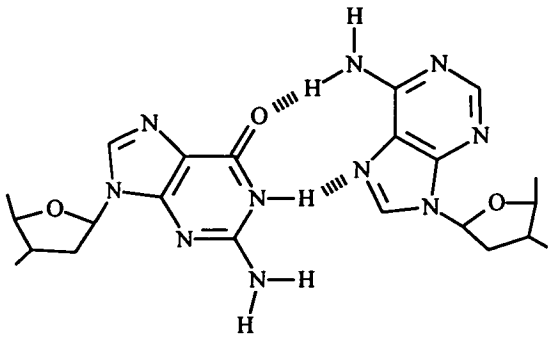


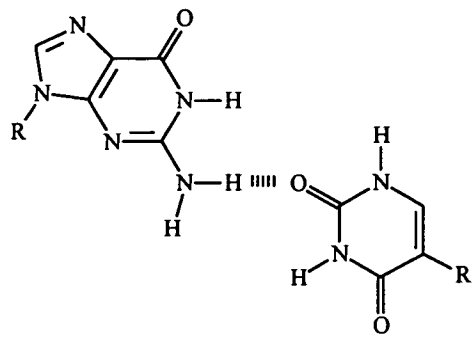
Figure 1.10. Examples of wobble base pairs.

In higher eukaryotes, there are highly efficient enzymatic defense mechanisms against such mismatches. The error rate is kept down to a level of 1 in 10^9, during replication in humans.[40]

Some of these non Watson-Crick base pairs may prevail and represent a mutagenic event which may be deleterious, beneficial or neutral to the host organism. Other non Watson-Crick base pairs may be present to fulfil a structural or other functional role within the host cell, e.g. the G-U and G-A "mismatches" play important roles in the secondary structure of RNA, such as internal and hairpin loops. Examples of other specialist base pairs include those involved in DNA triple-helices,[41] tRNA,[25] and guanine quartets in telomeres.[42]



G(*anti*).A(*syn*)

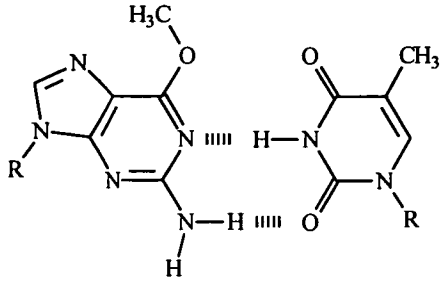


G.*pseudo*U

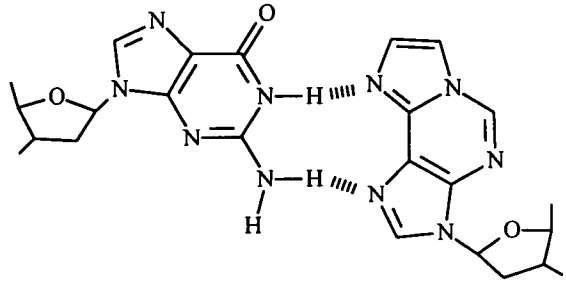
Figure 1.11. *Unusual base pairs involved in nucleic acid structure.*
Only one example for a G.A mismatch is given—there are four known conformations of this unusually stable base pair.[43]

Mutagenic Lesions

The proof-reading and repair enzymes mentioned above, along with other related enzymes have evolved to deal with another kind of non-Watson-Crick base pair. These may involve bases modified by chemical damage caused by endogenous mutagens, or by environmental mutagens such as UV light and ionising radiation.[44] These mutagenic lesions include: O^6-methyl-G, εA, and 8-oxo-purines such as 8-OH-G. O^6-Methyl-G and εA are shown below along with examples of non-Watson-Crick base pairs in which they may be involved.



O6-MeG.T



G(*anti*).etheno-A(*syn*)

Figure 1.12. *Examples of mispairing involving mutagenic lesions.*

1.3.2.2 Modified Bases

Modified bases may be present naturally or be the result of some chemical or physical insult (as mentioned above). In addition to the "natural" or "unnatural" bases, the development of nucleic acid chemistry and the advent of automated oligonucleotide synthesis has facilitated the incorporation of an increasing number of novel modified bases (and sugar-phosphate backbones) into oligonucleotides.

Changing the moieties involved in Watson-Crick base pairing directly affects the duplex stability; changing other moieties of the heterocyclic base may affect the structure and stability of the duplex in a more subtle manner.

The modifications made to the heterocyclic bases, and the consequent changes to the base pairing, is the subject of Section 1.4. The effects of such changes on the duplex structure and stability, are also discussed.

1.3.3 BASE STACKING AND HYDROPHOBIC INTERACTIONS

Purines and pyrimidines tend to form extended stacks of planar parallel molecules. The bases in these structures are usually partially overlapped, and can occur between individual bases, or intra-strand between bases in single- or double-stranded nucleic acids.

It is the stacking interactions between successive base pairs that provide the crucial link between sequence, structure and hence properties in double helical nucleic acids. Stacking interactions between aromatic molecules have been described as "π-π interactions". The total interaction involves dipole-dipole interactions, van der Waal's and atom-atom electrostatic interactions.[37]

The surfaces of the aromatic bases are poorly solvated by water and the resulting hydrophobic effect forces minimisation of surface area exposed. Environmental factors such as solvation and metal counterions also play an important role in nucleic acid structure and stability.

1.3.4 ENVIRONMENTAL INTERACTIONS

1.3.4.1 The Role of the Solvent

Sequence-specific hydration patterns and cross-strand water bridges which may be associated with the Watson-Crick moieties, have been proposed as important in determining the conformational preferences of base pair steps. For example, A-form DNA in fibres undergo a transition to B-form DNA by increasing the degree of hydration.[45]

The hydrophobic effect also becomes more important as water and salt content increases. This means that the base pair steps are forced into conformations which minimise surface area exposed to water. The fact that B-DNA is better solvated than A-DNA may also contribute to A-B transitions at high water content in fibres.[37]

1.3.4.2 Ionic Interactions

The melting temperature of nucleic acid duplexes increases with the cation concentration because these ions electrostatically shield the anionic phosphate groups from each other.[17] The presence of counterions, such as Na^+, changes the effects of hydration—for example, at high salt concentration, water activity is reduced, and B-A transitions occur. Salt concentration is also very important in other transitions, such as B-Z. Divalent cations, such as Mg^{2+}, specifically bind to phosphate groups—an Mg^{2+} ion has an influence on the DNA double helix comparable to that of 100 to 1000 Na^+ ions.[17]

Sequence-specific ionic interactions occur with the heterocyclic base, for example, metal ions can form cross-strand bridges by coordinating to carbonyl groups in the grooves.[46]

1.4 MODIFICATION OF THE HETEROCYCLIC BASES

Chemical modifications of heterocyclic bases are made in order to achieve one or more of the following objectives: to enhance their duplex stability (hybridisation efficiency), to maintain or increase their target specificity, to provide stability against nucleases, and to facilitate cellular uptake.[47] During the last decade, much of the work in oligonucleotide synthesis has been directed to investigating the suitability of various bases (such as mimics of natural bases, universal bases, and inert or ambiguous bases) for the incorporation into antisense oligonucleotides.[39,48]

The same general principles apply to modification of both pyrimidines and purines, and similar effects are observed when modified bases are found in nature (Section 1.4.3). Indeed many studies based on modified bases began after such bases had been originally discovered in natural nucleic acids. The sites of modification are shown below.

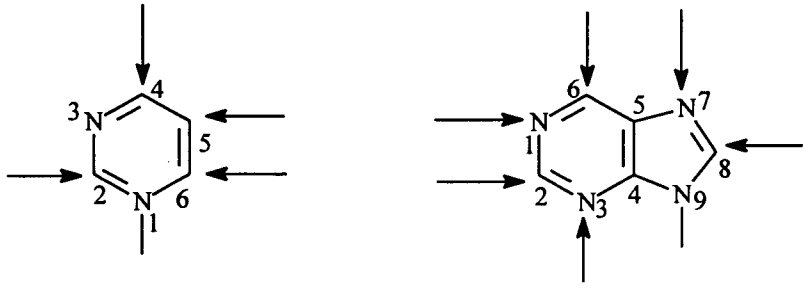


Figure 1.13. Sites of modification in pyrimidines and purines.

1.4.1 PYRIMIDINE MODIFICATIONS

There are four principal sites of modification in pyrimidines, as shown in Figure 1.13.

C2-Modification. There are few reported cases of modification at the C2, as this would interfere with hydrogen bonding, and could change steric requirements in the minor groove.

C4-Modification. The susceptibility of the 4-keto group to nucleophilic reagents, and the ease of its conversion into a leaving group, has led to several modifications at this position.[47]

C5-Modification. Modifications at this position lie in the major groove of the duplex, hence hybridisation will be relatively unaffected. Electrophilic reactions at this position are facile, and many 5-substituted pyrimidines have been incorporated into oligonucleotides, and their properties studied. The substitution of cytosines using 5-methyl- or 5-bromocytosines increased the stability of the duplex.[49] A similar behaviour was found for thymine replacing uracil, indicating beneficial hydrophobic interactions of the C5-methyl groups in duplex structures.[19] The introduction of triple bonds at C5 of pyrimidine bases dramatically influences the binding behaviour—oligonucleotides containing 5-(1-propynyl)-2'-dU and -dC significantly enhance the RNA binding affinity.[51] These substitutions are thought to increase π-π stacking in the duplex between the additional C5-substituent and the adjacent base on the 5'-side.

C6-Modification. Most substitutions at this position decrease the stability of duplexes, as the base, due to steric clashes, is forced to adopt the *syn* conformation, precluding Watson-Crick base pairing.

1.4.2 PURINE MODIFICATIONS

Certain substitutions in the heterocyclic base increase, or decrease the thermodynamic stability of duplexes, and usually provide a valuable insight into the properties affecting stability. Figure 1.14 shows examples of the purines described below.

N^1 Substituted Bases

Any modification of the N^1-position of purines would be expected to decrease the stability of resulting duplexes, due to disruption of a Watson-Crick site.

C-2 Substituted Bases

Two examples of C-2 substituted purines that increase duplex stability, are: 2-aza-inosine (z^2I), and 2,6-diaminopurine (D). The z^2I is expected to have an acceptor-donor-acceptor configuration, similar to xanthine (*see* Figure 1.16), but its pk_a is

higher, and so it could stabilise base pairing (xanthosine carries a negative charge at neutral pH). Diaminopurine would be expected to increase the stability of the duplex, as it possesses a donor-acceptor-donor configuration, which is a strong Watson-Crick base pair complement to thymine (*see* Section 1.4.4).

The base analogue diaminopurine, and its effect on structure and stability when incorporated into oligonucleotides, is studied in this thesis.

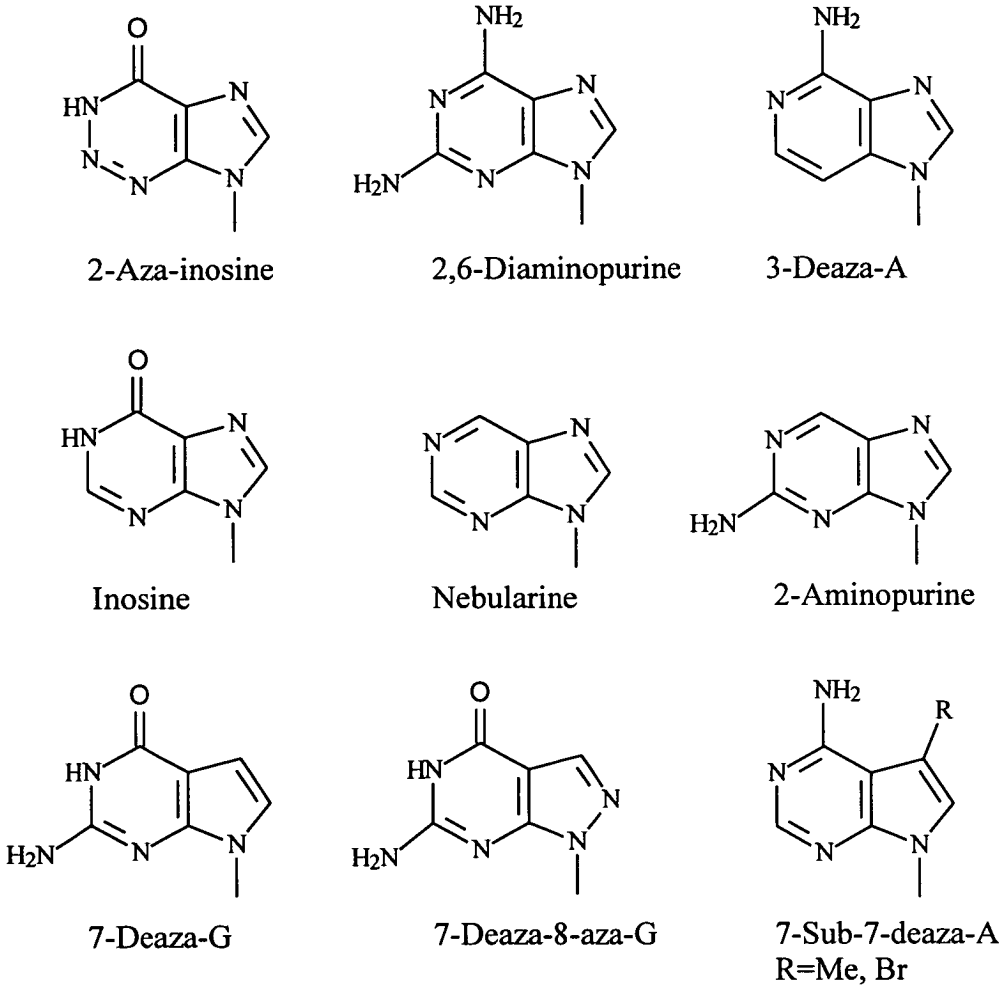


Figure 1.14. *Examples of purine modifications.*

N^3 Substituted Bases

3-Deaza-adenosine (c^3A) has been incorporated into DNA. It lacks an essential hydrogen bond acceptor (N^3 of A) from the minor groove, and hence was used to

probe protein-nucleic acid interactions. Study on duplexes containing a single c^3A residue in the middle of the sequence, showed no destabilising effect, compared with A.[50] The higher than expected T_m is explained in terms of the high basicity of c^3A (pk_a 6.80) versus A (pk_a 3.62).

C-6 Substituted Bases

The order of thermodynamic stability of base pairs containing inosine (I), is: G.C > A.T > C.G > I.C > I.A > T.A > G.A > I.G > G.G. This supports the idea that inosine might be an ambiguous base, because it pairs and stacks with any of the four normal bases, without substantially stabilising or destabilising the duplex.[52] DNA duplexes containing nebularine (N), where the 6-amino group of A is replaced by a hydrogen, show a large decrease in T_m's for both matched and mismatched duplexes.[53] The single hydrogen bond formed with thymine is thought to be crucial in fixing the base in a position optimal for base stacking. This "pairing" is less destabilising than expected, considering only one hydrogen bond is involed.

2-Aminopurine (AP), another candidate for base ambiguity, also prefers pairing to thymine, but destabilises any duplex formed.[54] Inosine still remains the best "multi-purpose" base ambiguity, although the development of improved bases continues.[47]

N^7 Substituted Bases

Seela and coworkers have focused on synthesising a whole series of 7-deaza- and 8-aza-7-deazapurines and studied the properties of the duplexes formed. Studies indicated that oligodeoxynucleotides containing 7-deaza-guanosine (c^7G) have a lower T_m value than those of the parent oligonucleotides. In contrast, T_m values were increased with oligonucleotides containing 7-deaza-8-aza-guanosine (c^7z^8G).[55]

Work was also carried out on oligonucleotides containing 7-substituted c^7A, (these are deoxy-derivatives of naturally-occurring tubercidin). These 7-substituted purines have a similar effect on the duplex as 5-substituted pyrimidines, and they similarly increase the thermodynamic stability, with Br- or Me-substitution. These moieties, which have steric freedom within the major groove, also increase its hydrophobicity.[56]

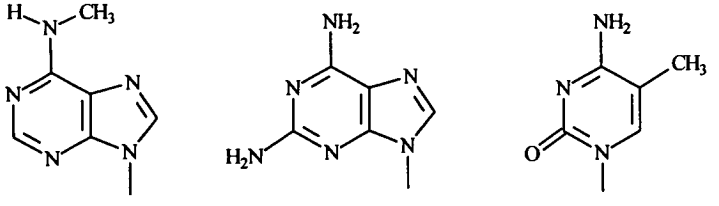
More recent work on 7-substituted-c^{7}A was carried out in order to investigate its effect on the homooligomer duplex: $d(A_{12}).d(T_{12})$, and the alternating oligomer duplex: $d(AT)_6.d(AT)_6$, as well as mixed-base duplexes. In the latter case the stabilising effect was less dramatic.[57] This work gives valuable insight into modifications carried out at non-base pairing moieties, and their effect on the structure and stability of duplex DNA.

C-8 Substituted Bases

8-Substituted purines decrease duplex stability because the substituents sterically disrupt the major groove. Analogous to 6-substituted pyrimidines, they probably force the base into an unfavourable *syn* conformation.

1.4.3 MODIFIED BASES FOUND IN NATURE

Modified DNA bases can be found naturally in eukaryotes, prokaryotes, and bacteriophages.[58] They may completely replace the standard base or replace only a small fraction. Their substituents vary from simple methyl or hydroxy groups to large moieties like amino acids. The majority of hypermodified bases identified thus far are present in bacteriophages where they can replace largely or completely one of the four common bases.



N6-Methyladenine 2,6-Diaminopurine 5-Methyl-C

Figure 1.15. *Examples of hypermodified bases found in natural DNA.*

The presence of these bases would seem to imply a purpose. It is thought that the changes in DNA-protein interactions play an important role in increasing phage resistance to host restriction endonucleases, or perhaps the modified bases act as transcription signals.

Bacteria themselves use modified bases: 5-methyl-cytidine is thought to be present due to the action of specific restriction-modification systems involved in combating phage infection.[58] 6-Methyl-adenosine may be used in the regulation of DNA repair, the timing of DNA replication, gene expression, and of DNA transposition.[58]

It is interesting to note the purpose of 5-methyl-C in mammalian genomic DNA: after enzymatic proof-reading, the parent strand is methylated at the 5-position of some of the cytosine residues. Repair enzymes recognise this methyl group "marker" and make changes only to the daughter strand after DNA replication.[40]

Other examples documented include the adoption of 5-methyl-C, 6-methyl-A and 5-OH-methyl-U in simple eukaryotes and trypanosomes.[58] It is also worth mentioning that in tRNA[25] and other RNA molecules, modified bases are present to fulfill specialist roles (*see* Figure 1.11).[59]

1.4.4 SYNTHETIC WATSON-CRICK BASE PAIRS

In order to find out why nucleic acids are composed from just four bases, and in preliminary attempts to extend the genetic alphabet, Piccirilli and coworkers synthesised a nucleoside analogue (K). This base hydrogen bonds with xanthosine (X), as shown below, to give a novel base pair whose structure is compatible with the double helix.[60] This X.K base pair is, along with the familiar G.C, among the possible distributions of strong base pairs (joined by three hydrogen bonds).

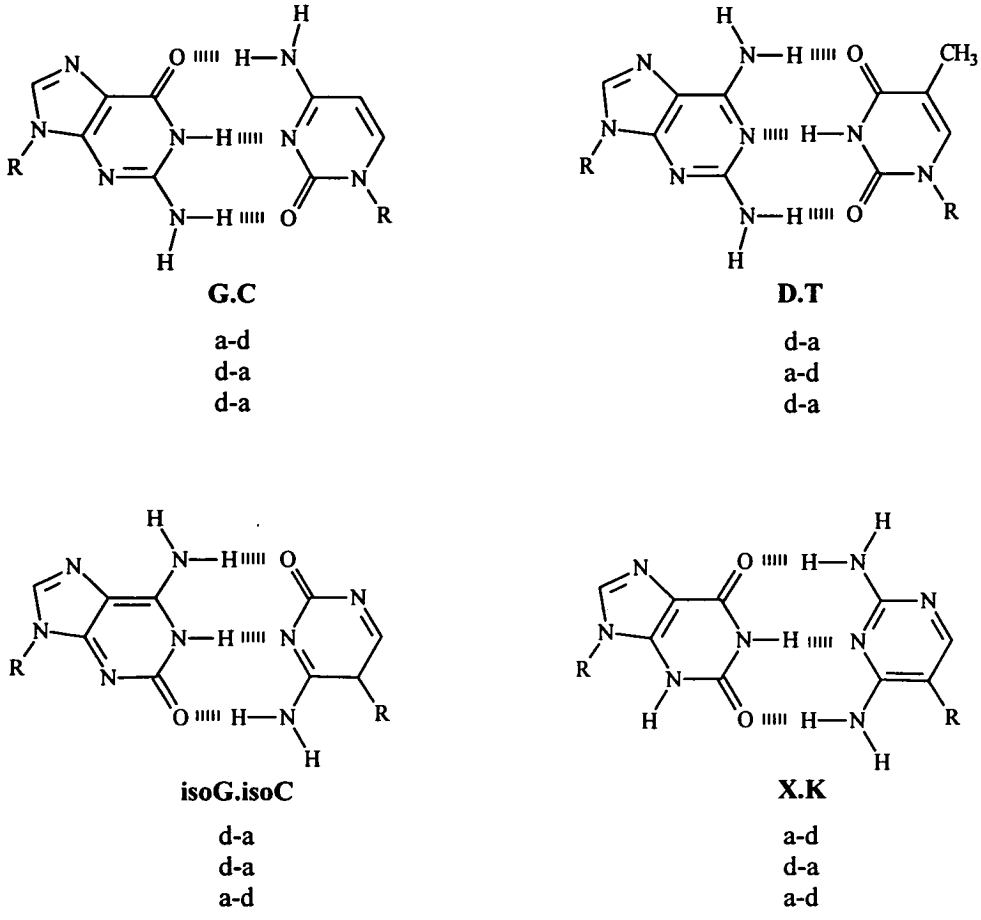


Figure 1.16. *Four mutually exclusive strong base pairs.*
Diagram shows hydrogen-bonding patterns: a=acceptor; d=donor.

The geometry of the Watson-Crick base pair can accommodate at least six, of eight, mutually exclusive hydrogen-bonding schemes. Four such examples are given in Figure 1.16.

The eight distributions of hydrogen donors and acceptors are defined as follows. The middle hydrogen always joins two heterocyclic nitrogen atoms, one of which carries a hydrogen atom. Each of the upper and lower hydrogen bonds joins an exocyclic donor amino group on one base to an exocyclic acceptor oxygen atom on the other.[61] This scheme is shown below.

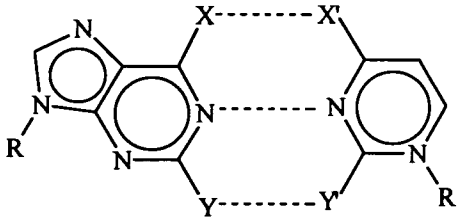


Figure 1.17. *An idealised strong base pair.*
If X is a donor NH_2 group, X' is an acceptor O atom, and vice versa. The same rule applies to Y and Y'.[61]

Piccirilli and coworkers also studied isoC.isoG but this did not form a satisfactory base pair (isoG can tautomerise to pair with U/T).[60]

To obtain the desired hydrogen bonding, several of the unnatural base pairs require a carbon-carbon bond joining the ribose ring with the heterocyclic base. Such structures are undoubtedly more difficult to prepare under prebiotic conditions than N-glycosides: perhaps explaining their absence from the repertoire of natural nucleotides. In the case of purines, this selection leaves only G, X, and D which are products of hydrocyanic reactions.[62]

Natural oligonucleotides use only two of the eight postulated strong base pairs—one of these incompletely, as adenine, (instead of diaminopurine, shown in Figure 1.16) as the complement to thymine.

Diaminopurine, and the strong base pair it forms with thymine, will be discussed in Section 1.5, along with the effects this change has on duplex structure and stability.

1.5 THE BASE ANALOGUE 2,6-DIAMINOPURINE

It was once thought that adenine (A), guanine (G), cytosine (C), and thymine (T)/ uracil (U) were the only bases found in nature. Studies have shown that although these bases are prevalent, they are not exclusive. (*See* Section 1.4).

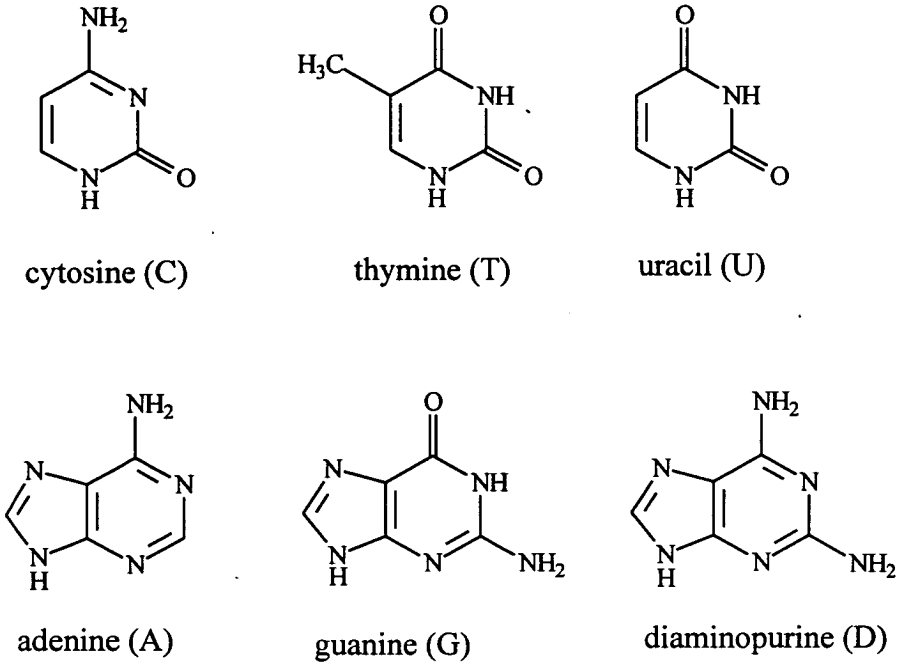


***Figure 1.18**. Comparing D with the common bases.*

2,6-Diaminopurine (D) or 2-aminoadenine was first synthesised and incorporated into RNA, in 1966 by Howard and coworkers.[63] They found that diaminopurine substantially stabilised the duplex compared to the native poly(rA)-(rU), and postulated the formation of three hydrogen bonds in a base pair involving rD and rU.

The initial finding of diaminopurine in naturally-occurring nucleic acids was detailed over a decade later. In 1977, Kirnos and coworkers reported the isolation and identification of diaminopurine, and its deoxyribonucleoside, from S-2L cyanophage DNA, where dA is fully replaced by dD.[64,65] The function of this total substitution was not investigated by the authors, although it may serve some structural or biochemical defense against the host bacteria.

1.5.1 BASE PAIRING INVOLVING DIAMINOPURINE

The introduction of the amino function at the purine 2-position inserts an NH_2 group into the minor groove of the base pair it forms. This allows the formation of three hydrogen bonds with thymine in DNA, or uracil in RNA, with normal Watson-Crick geometry. (The base pair formed between D and T, was mentioned in Section 1.4 as a member of the eight possible strong base pairs). The subsequent base pairs are A.T-like in the major groove, and G.C-like in the minor groove.

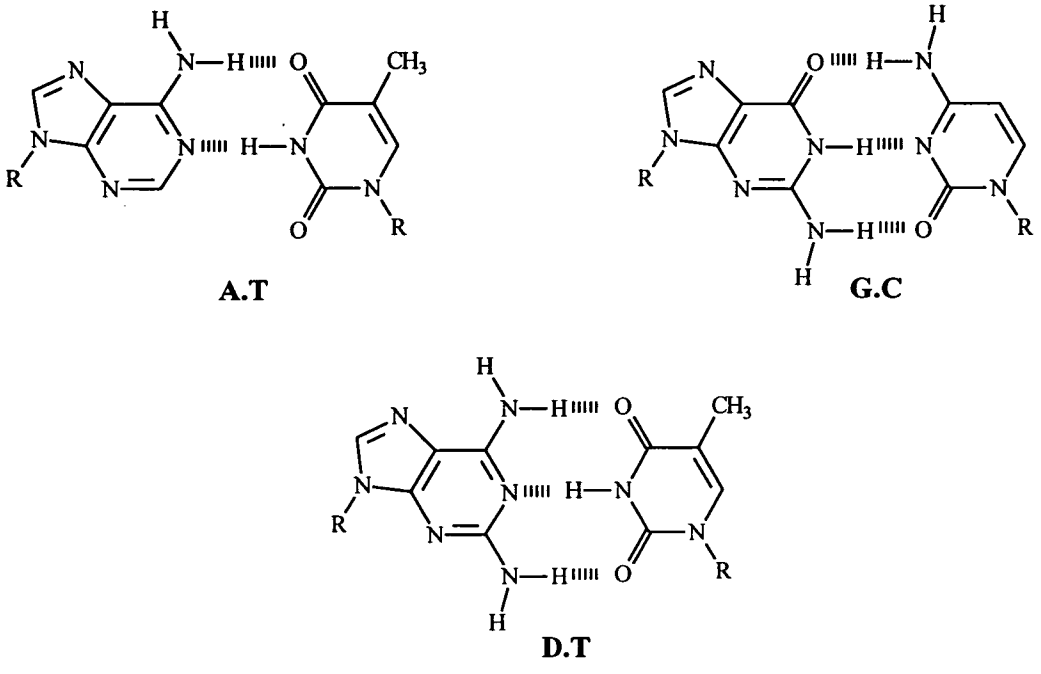


***Figure 1.19**. Comparing the standard A.T and G.C base pairs with D.T.*

1.5.2 THE EFFECT OF dD ON DUPLEX STABILITY

The ability of nucleic acids to form very stable duplexes has many important applications in biology: these applications will be discussed in Section 1.7.

1.5.2.1 Initial Studies on Oligonucleotides Containing dD

Three hydrogen bonds (as found in G.C and D.T) would be expected to be thermodynamically more stable than two (found in A.T)—*see* Figure 1.19. The first studies into the stability of duplexes containing diaminopurine were carried out by Howard and coworkers.[63] They found that the poly(rD).(rU) duplex to be

approximately 25°C more stable than the "native" poly(rA).(rU) duplex. They rationalised that the entropy change for dissociation of the two complexes should be nearly the same, suggesting that the higher stability of the rD-containing duplex to be attributed to a larger enthalpy of base-base interaction, conferred by the additional hydrogen bond.

Soon after this, Cerami et al[66] prepared an alternating DNA copolymer poly(dD.dT), and compared the melting temperature to the similarly-made poly(dA.dT). The native sequence was again found to have a much lower melting temperature (ΔT_m=20°C). This data was in agreement with that found for the RNA duplexes, and both studies showed duplex behaviour consistent with the "three hydrogen bond" proposal.

The identification of naturally-occurring dD in S2-L cyanophage DNA[64,65] prompted further study, and enzymatic and chemical methods were employed to allow the preparation of polynucleotides or oligonucleotides incorporating diaminopurine. For example, Scheit and coworkers[67] made poly dD and poly rD, and compared the duplexes formed with dA and rA respectively. In common with other workers, these enzymatically-made duplexes showed a large elevating effect where D replaced A.

The study of short synthetic oligonucleotides containing dD were first carried out in 1982 by Gaffney and coworkers on the self-complementary hexamer $d(TD)_3$.[68] The hexamer formed a stable duplex at temperatures where the native hexamer did not. This work led to the preparation of other hexamers, with regards to their duplex stability, classifying the dD.dT base pair as being more stable than the dA.dT base pair, but less stable than the dG.dC base pair.[69]

1.5.2.2 The Spine of Hydration and (dA).(dT) Tracts

Building on their earlier work, Howard and coworkers noted that the stabilising effect of D was much greater in RNA than in DNA.[70] They suggested that the lower increase in stability for DNA might be due to a change in the hydration pattern of DNA which was disrupted by the minor groove amino group.[70,71] For DNA, the

overall gain in stability was equal to the beneficial gain from forming three hydrogen bonds less the loss in stability due to changes in hydration patterns. These changes are particularly detrimental in the case of A.T tracts where Drew and Dickerson had previously reported finding a spine of hydration.[72]

Whilst examining the use of diaminopurine in synthetic probes, in particular the stability of duplexes formed, Chollet and coworkers reported that in most cases the dD-containing probes were more stable than the native probes.[73-75] However, they noted that the substitution of dD in A.T tracts, namely in d(CAA ADA AAG), the duplex formed was destabilised with respect to d(CAA AAA AAG).[74] In this case, the dD is thought to disrupt the network of hydrogen bonds (spine of hydration) formed between the N^6 of dA and the O4 of dT on opposite strands.[76]

While the use of dD adjacent to dA is unlikely, it is important to investigate the effect of substitution of dD for dA will have on the duplex structure and hence, causally, on the stability, where dD replaces dA.

This is one example of the complex relationship between the stability of duplex nucleic acids and the structure they adopt, which is in turn dependent on other factors such as counter ion and the degree of hydration.

1.5.3 THE EFFECT OF dD ON DUPLEX STRUCTURE

Diaminopurine has both adenine-like and guanine-like features, (in the major and minor grooves, respectively) and would therefore be expected to behave like either base, or a mixture of both. The presence, or absence, of the 2-amino group in particular, may have a profound effect on the structure adopted, and on subsequent structural transitions.

In order to build up a picture of the conformations and structural transitions that diaminopurine may influence, it will be compared with: adenine in homopolymer, and copolymer double-helices; and with guanine-containing double-helices. The structure adopted by mixed-base duplexes containing diaminopurine will also be discussed.

1.5.3.1 Comparing (A).(T), and (D).(T) Tracts

Tracts containing oligo(dA).(dT) have a B-form conformation of DNA that is significantly different from that of other deoxypolymers.[77] This B-form duplex bends in the helix axis, and has a certain stiffness in the A tract. It owes its particular stability to a propeller twist and buckle in the dA.dT base pair which helps maximise purine-purine stacking and creates a potential system of non-Watson-Crick hydrogen bonding. These bonds may form diagonally across the major groove inter-strand between N^6 of A and towards the O4 of T,[76] and form the spine of hydration mentioned in Section 1.5.2.2, above.

A striking consequence of this altered structure is its refusal to undergo a transition from B-form to the A-form of DNA, even at very low water content.[45] This can also be explained in terms of the AA/TT step, which has a strong calculated preference for the B-form.[37]

When diaminopurine replaces adenine to give dD.dT tracts, the conformational rigidity displayed in dA.dT is not observed. This is partly a consequence of the lack of spine of hydration. Howard and Miles[70] compared a variety of homopolymers containing (A).(T) and (D).(T) tracts. Of particular interest were: poly(dA).(dT), poly(dD).(dT), poly(dA).(rT), and poly(dD).(rT). DNA/RNA hybrids (duplexes of importance in antisense therapy—*see* Section 1.7) were also investigated more recently in oligonucleotides by Gryaznov and Schultz.[78] Both groups reached the same conclusion, that:

(i) dD-dT tracts show similar[78] or greater[70] stability to dA.dT, and

(ii) dD tracts bind strongly to rT, whereas dA tracts bind poorly to rT.

(In the latter case the dA's are reluctant to adopt the A-form helical structure, which is favoured by RNA-RNA duplexes and DNA-RNA hybrids).[27,30,35] In addition, Gryaznov and Schultz found that substitution with dD enhanced binding to RNA the most where the original relative destabilisation by dA residues was the greatest.

1.5.3.2 Comparing (A.T) and (D.T) Tracts

The alternating duplex, poly(dA.dT). is much more conformationally flexible than poly(dA).(dT), and has an alternating structure with a small helical twist of the AT steps and a larger twist at the TA steps.[57] This results in a polymorphic structure

which can adopt more than one conformation {compared to the "stiff" poly(dA).(dT)}, depending on the environment/ modifications present. Poly(dA.dT) may also undergo the transition to A-DNA more readily than poly(dA).(dT), due to the conformational flexibility of the former.

X-ray diffraction patterns on fibres of the copolymer poly(dA.dT) were first obtained by Davies and Baldwin.[79] This suggested a D-DNA structure. Anomalies involving poly(dA-dT) include the observation that the strands are cut into dinucleotides of pTpA instead of mononucleotides when exposed to DNase1.[80] Existence of regular alternation in the local helical parameters of d(A.T) tracts led to the proposal of an alternating B-DNA helix for poly(dA.dT).[81] In such a model the helical twist angle is smaller at the AT step than at the TA step. Yoon and coworkers[82] reported similar alternation in the central AT region of d(CGC ATA TAT GCG) to support the alternating B-DNA conformation.

These observations point to a structure maximising the dApdT overlap at the expense of the weaker dTpdA base stacking.

A 2D-nmr NOE study carried out on poly(dA.dT) showed both bases to adopt the *anti* conformation,[83] and the helix most likely to be right-handed. The authors proposed a right-handed B-DNA, and found evidence to exclude alternation. More recently, Kypr and coworkers have found poly(dA.dT) to coexist in two conformers.[32]

Less study has been carried out on (dD.dT) tracts, but like (dA.dT) tracts, the models developed for duplex structure have evolved over the years. This is especially true of the conformational transitions oligo and poly(dD.dT) may undergo, which differ to that for oligo and poly(dA.dT).

1.5.3.3 Structural Transitions of Poly(dD.dT)

Initial studies on CD spectra collected for oligo and poly(dD.dT) concluded that the high salt form was most likely Z-DNA.[68,69,84] This was deduced from the general inversion of the CD spectrum (going from low to high concentrations of NaCl), in a manner analogous to that seen for poly(dG.dC).

^{31}P nmr seemed to support a high-salt Z-DNA hypothesis. In low salt poly(dD.dT) has two clearly resolved signals, whose separation increases as the salt concentration increases. This is indicative of two different phosphorus environments, and was likened to that seen for poly(dA.dT), although it was felt that the downfield signal was closer to that found for poly(dG.dC). This was taken to indicate a Z-DNA structure for poly(dD.dT).[84,85]

These initial conclusions seemed to agree with the premise that alternating purine-pyrimidine, and the presence of the 2-amino group in dD (like dG), would give a salt induced B-Z transition. However, the picture is not quite as simple as this for Z-DNA promotion. This matter will be mentioned in more detail, and with reference to dD, in Section 1.5.3.4.

Borah and coworkers reached a different conclusion in later 2D nmr and ^{31}P nmr experiments.[86,87] From the 2D NOE spectroscopy structural information is obtained on the basis of spatial connectivities between nuclei coupled by dipolar interaction. In the archetypal B-Z transition, where poly(dG.dMeC) undergoes a transition in high salt, intense H^8-$H^{1'}$ intra-base cross peaks are observed from the resultant *syn*-dG. No such cross peaks are found in the high-salt form of poly(dD.dT) suggesting *anti*-dD and dT. A C2'-*endo* sugar pucker was proposed, possibly due to (H^6)T-$(H^{2''})$D interactions, which is typically shown in A-form DNA. This high-salt form of poly(dD-dT) was concluded to be a member of the A-family of DNA as it did not possess the B or Z conformation. Later studies showed similar results for poly(dD.dxU), where x=Br or I.[87]

Several years after the nmr studies, Vorlickova and coworkers recorded the CD spectra of poly(dD.dT) under the same conditions previously found to induce a B-A transition in poly(dA.dT).[88] They found that the CD spectra of the high salt and ethanol forms of poly(dD.dT) were the same. It was concluded this form was not A-DNA but X-DNA (previously reported for poly(dA.dT)). Poly(dA.dT) was not affected by, e.g. 70% MeOH, unless trace amounts of cesium cations were added. This produced the X-DNA profile. The conclusion reached was that the extra 2-

amino group performed the same role as the Cs^+, that is, it dehydrates the minor groove and relieves its conformational variability.

Poly(dD.dT) was then shown to form X-DNA under very mild conditions.[89] At low concentrations of the duplex, millimolar quantities of Mg^{2+} caused the B-X transition. This further exemplifies the fact that X-DNA and A-DNA are not the same—A-DNA is destabilised by divalent cations in aqueous solution.

The presence of alcohol, divalent cations, and the 2-amino group of dD are not the only factors which promote the transition to X-DNA. Substitution at the 5-position in pyrimidines enhances the transition to X-DNA. Indeed, recent work carried out by Vorlickova's group focused on the effect of the 5-methyl group in pyrimidines, namely in poly(dD.dT) duplexes.[50] The methyl group exerts a profound effect on the conformation of DNA, for example, poly(dA.dT) coexists in two conformers,[32] and poly(dA.dU) exists as a single conformation resembling RNA.[90] In addition, methylation of poly(rA.rU) to give poly(rA.rT) possesses some properties of DNA;[32] and methylation of poly(dG.dC) promotes the B-Z transition, but inhibits the A-Z transition in poly(rG.rC).[91]

The methyl group of dT, versus other small 5-position aliphatic groups, is the optimum for stabilisation of the B-type conformation of alternating purine-pyrimidine DNAs. Vorlickova et al[50] found that the putative A-form (previously X-form[89]) of poly(dD.dT) was strongly stabilised by the thymine methyl group. They realised that neither poly(dD.dEtU) nor poly(dD.dU) isomerised into the putative A-form in low salt aqueous solution with Mg^{2+}.

The conformational isomerism (B-putative A) is strongly promoted by the *combined* effects of the 2-amino group of purines with the 5-methyl group of pyrimidines. Only poly(dG.dMeC), not poly(dI.dMeC) or poly(dG.dC), adopts the Z-form relatively easy. The amino groups in the minor groove of the double helix cooperate with the methyl groups of the major groove, and with divalent cations, to stabilise the zigzag backbone of the form of poly(dD.dT).

In the poly(dD.dxU) family, the methyl group was found to *destabilise* the Z-form adopted by poly(dD.dT). The Z-form and putativative A-form are mutually exclusive[50]:

Putative A-form	⟵Reluctant to form⟶	Z-form
e.g. poly(dD-dT)		e.g. poly(dD-dEtU)

1.5.3.4 B-Z Transitions

As mentioned earlier, the transitions associated with poly(dD.dT) were thought to be B-Z, but the current body of evidence now points to a B-putative A-form. The ease of transition of poly(dD.dT) from its B-form to the putative A-form suggests that their base pair topologies are the same.[50] However, the putative A-form has a zigzag backbone which discriminates it from A-forms of poly(dA.dT), and other natural DNAs whose backbones are regular.[92] Poly(dD.dT) does adopt the Z-form, but only to a limited extent, as the putative A-form is the preferred conformation under conditions normally associated with B-Z transitions. What type of oligonucleotides undergo a B-Z transition?

Most study regarding non-polymeric, *mixed* sequences containing dD, have been concerned with their use as probes and other applications. Here, the thermodynamic and hybridisation properties of the duplexes formed have been of concern, (*see* Sections 1.2 and 1.3), and actual structural studies on such duplexes formed, are relatively few in comparison. However, a few mixed sequences have been studied which adopt, at least in part, the Z-form. These include the self-complementary duplex: d(CDC GTG)$_2$, this adopts 25% Z-DNA in the high salt NMR study, and clearly shows the Z-form at low humidity by IR, although CD showed no evidence.[93] The Z-form was found in two crystal structures by Coll and coworkers: d(CGT DCG)$_2$, and the duplex studied by the group above: d(CDC GTG)$_2$.[94]

1.6 SYNTHESIS OF DIAMINOPURINE NUCLEOSIDES

1.6.1 A REVIEW OF DIAMINOPURINE 2'-DEOXYNUCLEOSIDE SYNTHESIS

Via an N^1-Oxide

Initial synthesis of 2,6-diaminopurine involved the reaction of cyanogen bromide with adenosine N^1-oxide.[95] This yielded a cyclic N^1-N^6 product which with methanolic ammonia gave N^6-cyano-deoxyadenosine-N^1-oxide. Methylation followed by alkaline treatment resulted in a rearrangement leading to 2-amino-N^6-methoxyadenosine. Catalytic hydrogenation gave the free 2,6-diaminopurine nucleoside.

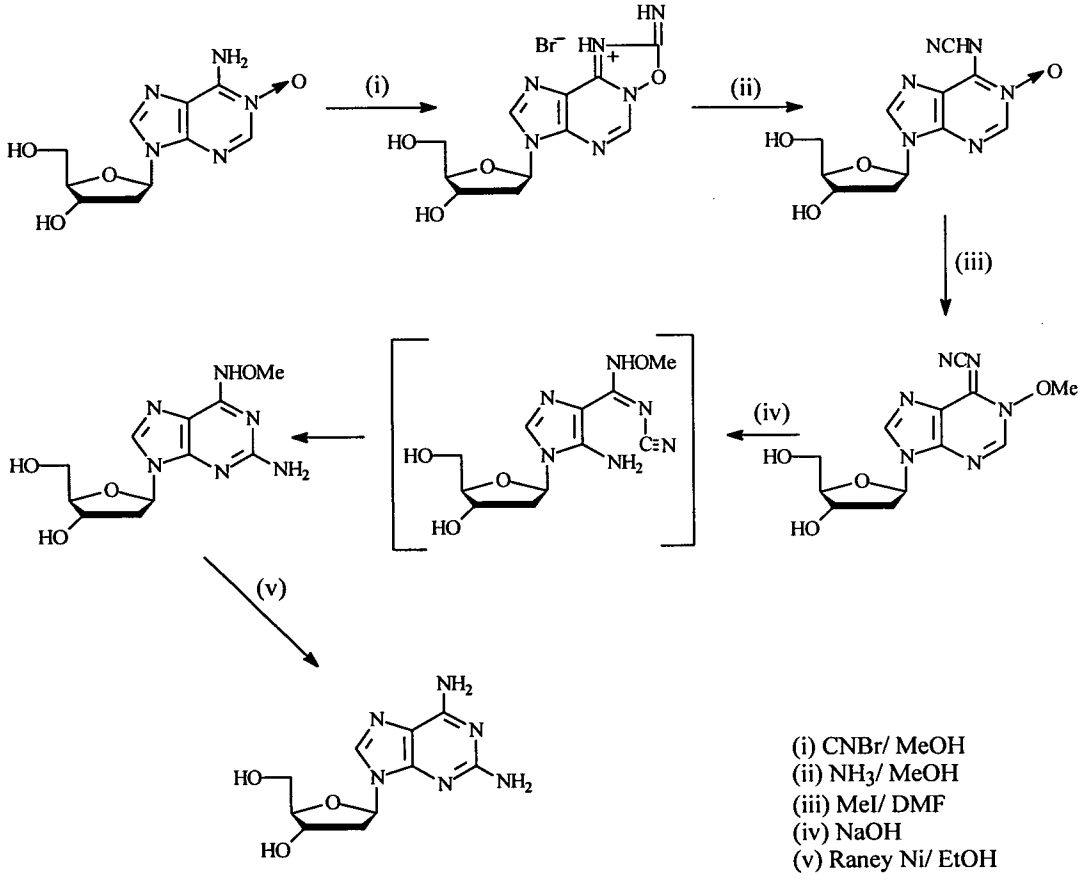


Figure 1.20. The synthesis of diaminopurine from dA-N^1-oxide.

Via N^6-Triazolo Derivatives

Another approach to synthesis of 2,6-diaminopurine[96] was provided by Sung who noticed that thymidine could be converted to 5-methyl-dC using relatively mild conditions.[97] The conversion of dT to 5-methyl-dC was achieved via an intermediate 4-triazolopyrimidone. This was prepared from the reaction of protected dT with p-chlorophenyl phosphorodichloridate and 1,2,4-triazole in pyridine; after work-up, treatment with ammonia gave 5-methyl-dC. (It is interesting to note that this reaction is similar to a known side reaction in phosphoramidite chemistry that converts dG to dD).[98]

This technique, based on post-synthesis treatment of the triazolo derivative of dG with ammonia, however, remained unsatisfactory, leaving a ratio of approximately 50:50 for dG:dD.[96] See below.

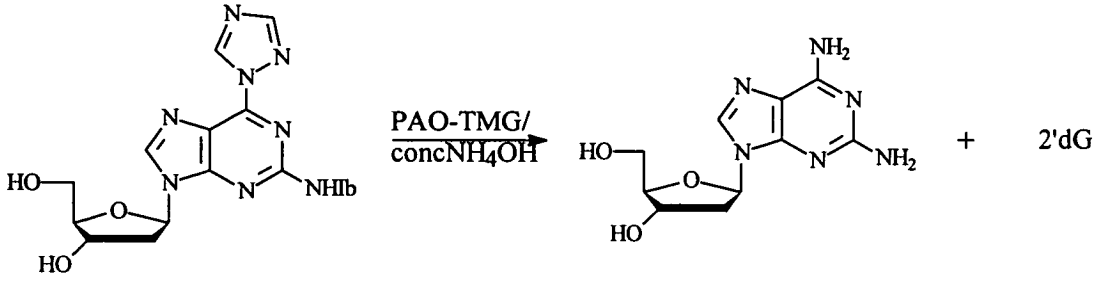


Figure 1.21. *Conversion of N^6-triazolo dG to dD.*
PAO-TMG= pyridine aldoximate-tetramethyl guanidine. Ib= isobutyryl.

Via O6 Sulphonation

Guanine nucleosides are known to be reactive at the O6 position with a variety of acylating, sulphonating, and phosphorylating agents.[99] This susceptibility causes problems in oligonucleotide synthesis,[99] but provides access to many guanine O6, N, and S derivatives.[99,100]

Gaffney and coworkers exploited the reactivity of the O6 position of dG towards sulphonating agents and alkylating agents as a general synthetic route in obtaining 6-substituted dG[101] or O6-protected dG[102] respectively. Sulphonation of the O6 position of dG had been reported to proceed easily[103,104] to give a sulphonated

nucleoside which readily underwent displacement reactions with nitrogen nucleophiles to introduce an N^6 function.

Gaffney and coworkers utilised the above findings to synthesise dD nucleosides.[69] Tris-isobutyryl deoxyguanosine was sulphonated with tris-isopropylbenzene-sulphonyl chloride (TPSCl) in triethylamine-containing dichloromethane solution, using dimethylaminopyridine as a catalyst. This derivative was then converted to the 2,6-diaminopurine derivative by liquid ammonia displacement of an intermediate trimethylamino group,[101,105] which is more readily displaced than the sulphonyl group. The N^6 protecting group of choice was the benzoyl group. However, the use of this group in dD nucleoside protection leads to depurination during oligonucleotide synthesis.

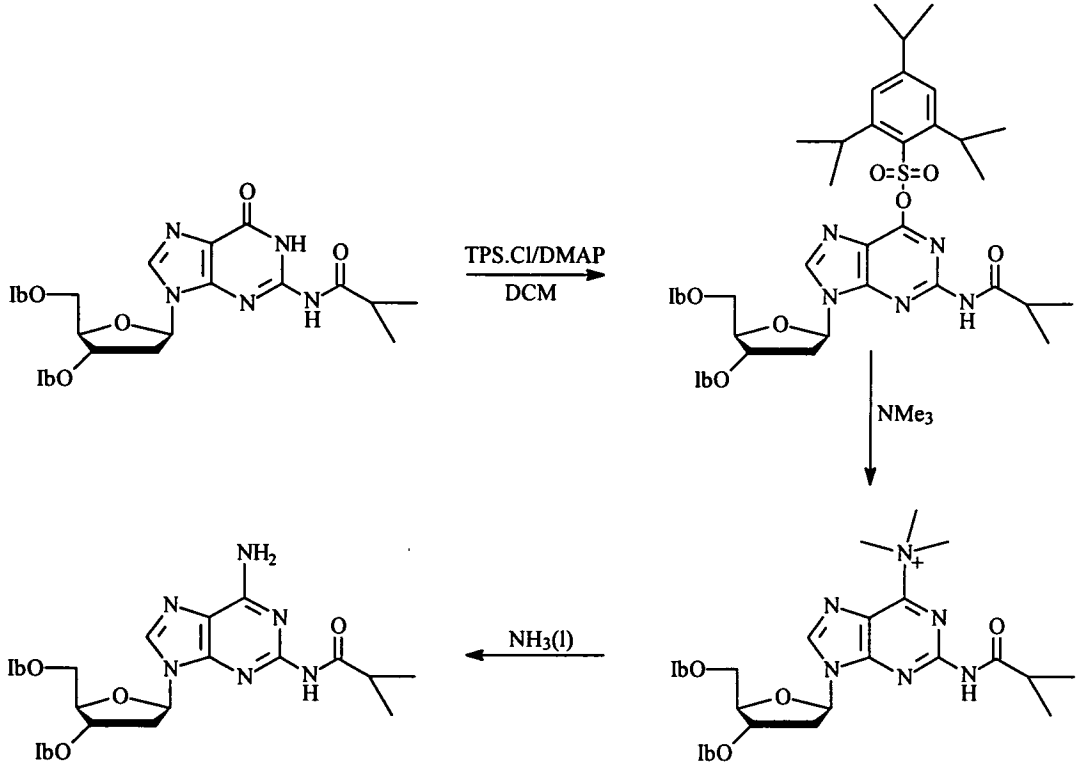


Figure 1.22. *Amination via Sulphonation.*

1.6.2 PROTECTION STRATEGY

2,6-Diaminopurine is unique among the heterocyclic bases in that two amino functions must be furnished with protecting groups.

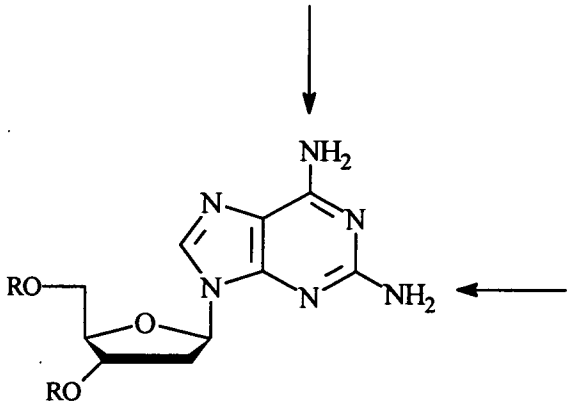


Figure 1.23. *The need for two amino-protecting groups.*

Different strategies lie behind the choice for the protection of the N^2 and N^6 moieties—firstly, the problem of acid-catalysed depurination, which is more serious in dD than dA, will be addressed.

1.6.2.1 Acid-Catalysed Depurination

Synthesis of sequences containing dD is subject to a major side reaction: acid-catalysed depurination of dD residues during TCA-induced detritylation of the 5'-terminus. The rate of depurination is highly dependent on the choice of N^6 protection, as noted for dA.[106] (The problem of depurination of dA during detritylation has been acknowledged for quite some time.[107]) Depurination is more severe when susceptible bases are situated towards the 3'-end of the sequence, and when the oligonucleotide is longer, since in both cases the purine residue is exposed to more acid detritylation steps.

Early attempts to synthesise dD oligonucleotides were handicapped by the effects of depurination,[68,69] even when zinc bromide was used for detritylation. This occurred because the N^6 protecting group employed was the benzoyl amide: the group of choice for dA. Following depurination, the oligonucleotide may be cleaved

at the apurinic site by the ammonia treatment at the end of synthesis. This cleavage of the 3'-OH side of the apurinic site is similar to that found in Maxam-Gilbert sequencing.[108] The product obtained from the synthesis is therefore a combination of the desired oligonucleotide, apurinic material, and oligonucleotide fragments.

There are three approaches that may be adopted to avoid the problem of depurination. These approaches involve changing either the: detritylation reagents; the 5'-OH protecting group; or the N^6 protection. Each of these alternatives will be considered in turn.

(a)Detritylation Reagents

The standard reagent used to remove the dimethoxytrityl group is a solution of TCA in dichloromethane. Other detritylating reagents which were used included benzenesulphonic acid in acetonitrile.[109] However, this also showed significant depurination. Attention then focused the use of aprotic reagents for detritylation. Saturated solutions of zinc bromide in nitromethane gave quantitative detritylation, and depurination did not occur as long as the nitromethane was scrupulously anhydrous.[110] Unfortunately, the 5'-OH group is unreactive to condensation reactions unless demetalated with mild alkaline hydrolysis. The deprotection stage then becomes a two step process: (1) a 30 minute treatment with sat. $ZnBr_2$/nitromethane, then (2) a 5 minute wash with n-butanol in THF/2,6-lutidine to regenerate the 5'-OH.[111] The complexity of the reagents required and the increase in cycle time make this an unattractive option compared to the use of TCA.

(b) 5'-OH Protection

Other groups used for 5'-OH protection include: tris(benzoyloxy)trityl, removed with NaOH;[112] trityloxyacetyl, removed with dilute NH_4OH;[113] t-butyldimethylsilyl, removed with F^- or AcOH;[114] 9-phenylxanthen-9-yl group (pixyl), removed by mild acid hydrolysis;[115] and o-dibromomethylbenzoyl, which is removed with silver perchlorate.[116] Whilst these are only a few examples from a long list of 5'-OH protecting groups, changing the group would incur a new set of problems. Cleavages involving basic conditions are not acceptable as orthogonality of base and sugar

protection would be lost if base-labile sugar protection is adopted. Also, if acid-labile protection is used, such as the pixyl group, depurination will still remain a problem. Other problems would be introduced, for example phosphate cleavage by the fluoride ion. The 4,4'-dimethoxytrityl has its own inherent advantages around which the standard synthesis of oligonucleotides is based, and it is important to adhere to these protocols.[117]

(c)N^6 Protecting Group

Changing the N^6 protection group is the most appropriate course of action, since there is no need to deviate from the well established protocols for phosphoramidite solid phase synthesis.

Many different groups have been developed for the protection of the exocyclic amines, with several aimed at reducing the amount of acid-catalysed depurination of dA on detritylation. A modified form of the dimethoxytrityl group has been proposed for the N^6 position.[118] The resulting N^6-4, 4' ,4"-tris(benzoyloxy)trityl-dA was found to offer up to four times the stability of N^6-benzoyl-dA in 2% DCA/dichloromethane and eight times as stable in 80% AcOH. The phthaloyl group offered better depurination resistance in DCA and AcOH.[119] At the same time the phthaloyl group was found to be more base labile than the benzoyl group. N^6-Succinyl-dA was also made, and showed improvements over phthaloyl protection.[120] In an effort to change the site of protonation of N^6-benzoyl-dA, a different approach was proposed by Morin.[121] This involved N^1-oxidation, which reduced the depurination during detritylation with benzene sulphonic acid.

The breakthrough in N^6 protecting group choice came with the application of amidine reagents to solid phase synthesis monomers.[122]

1.6.2.2 N^6-Amidine Protecting Groups

Amidine protection was originally used for ribonucleotides[123] before being reintroduced for deoxyribonucleotides.[124] The dimethylformamide (dmf) was easily placed on the N^6 of dA with dimethylformamide dimethylacetal in DMF to yield exclusively N^6-protected material in 90% yield. However, when exposed to

oligonucleotide detritylation reagents ($ZnBr_2$ in nitromethane/methanol, or 3% TCA in dichloromethane) the amidine was rapidly hydrolysed.[125] This led McBride and Caruthers to propose N^6-(N-methyl-2-pyrrolidine amidine)-dA for use in oligonucleotide synthesis.[125] The half life of this compound in 1M TCA was three times that of the benzoyl protected dA (60 minutes versus 20 minutes). In order to extend the range of usable amidines, and help increase their hydrolysis resistance, a series of N^6-dialkylformamides was prepared with increasing steric bulk in the dialkyl portion.[122]

N^6-Diisopropylformamide-dA was resistant to depurination in 2% DCA/ dichloromethane at 25°C with a half life of about 2 hours, but its deprotection with ammonia at 60°C took 12 hours. The group that emerged as most suitable was the N^6-di-n-butylformamide-dA which remained stable to all oligonucleotide synthesis conditions, readily deprotected with concentrated ammonia in 3 hours at 60°C, and offering an approximate 20-fold suppression of depurination.[122]

Formamide protection of the N^6 position of dA offers the following advantages. The protecting groups can be prepared from inexpensive starting materials. Unlike traditional (acyl) protection strategies, the nucleoside can be selectively N^6-protected without the inevitable transient protection of the 3'- and 5'-OH functions.[126] In addition, the formamide is easily removed under the standard deacylation conditions used, i.e. concentrated ammonia at 55°C. The most important property of the formamides is their ability to greatly increase the resistance of dA to acid-catalysed C1'-N^9 cleavage during the repetitive detritylation stage.

1.6.2.3 The Effect of N^6 Protection on Depurination

Observing the mechanism for acid-catalysed depurination does not make it obvious that the choice of N^6-protection will influence the rate of depurination. Depurination is known to proceed via an A1-type mechanism, involving the protonation of the purine in a pre-equilibrium step, followed by the rate determining step of C1'-N^9 glycosyl bond cleavage to form a cyclic glycosyl oxycarbenium ion.[127-129] This mechanism, shown below, is supported by a secondary deuterium isotope effect, and the influence of changes in the glycosyl moiety on the hydrolysis

rates as observed for the hydrolysis of ribosides versus deoxyribosides.[129] Deoxyadenosine has two reactive nitrogens as sites for protonation (N^1 and N^7), so the pre-equilibrium step may be the formation of either a mono- or di-cation. Both of these cases lead to C1'-N^9 bond cleavage. The 1-methyl-adenosine cation depurinates at about the same rate as adenosine, suggesting that it is the 1-protonated form of the latter that reacts with a second proton to result in a depurinating di-cation.[127]

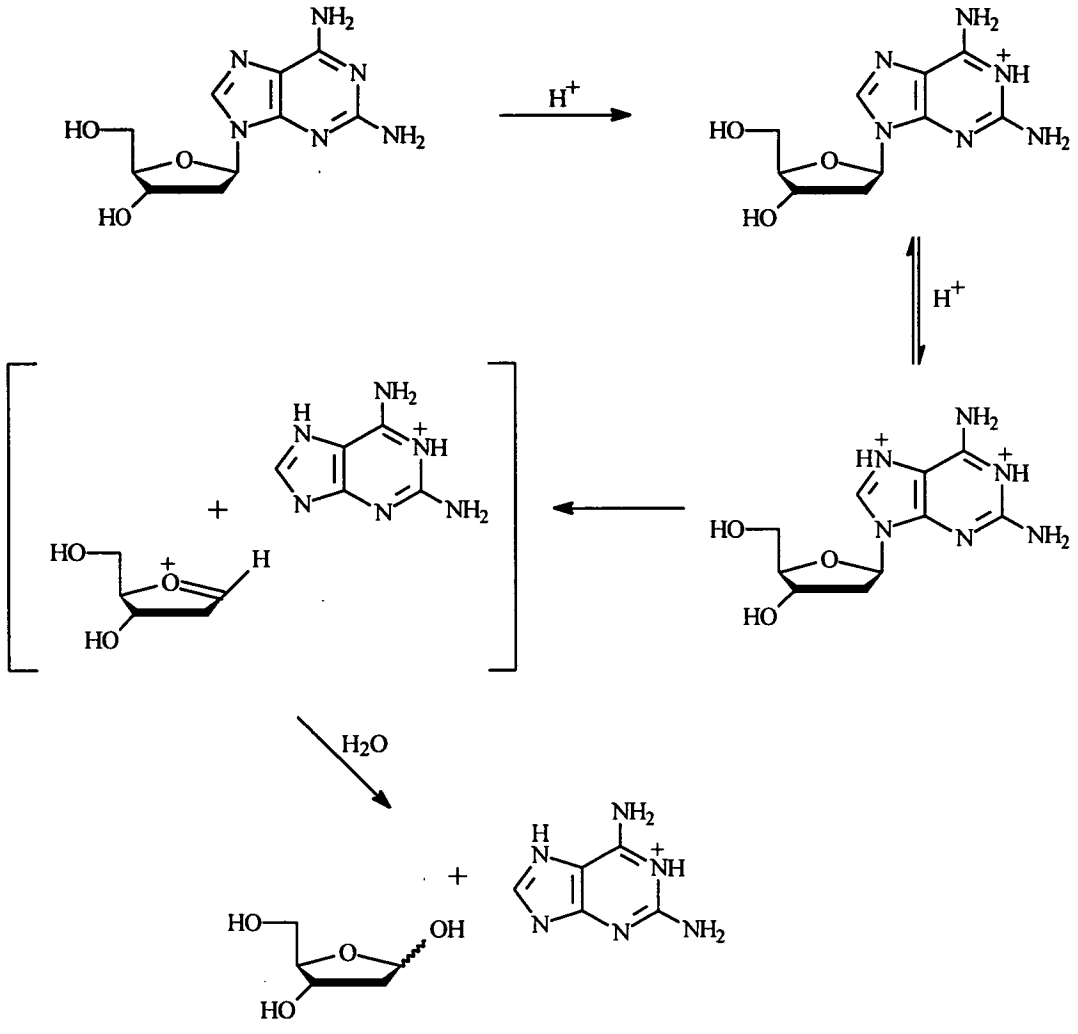


Figure 1.24. *Acid-catalysed depurination, shown for diaminopurine.*

What is the effect of the N^6-protection on the depurination rate? By observing the changes in the N^1 and N^7 basicities, ^{15}N nmr shifts in TFA the protonation sites can be determined.[130] 2'-Deoxyadenosine showed a change in only the N^1 shift, indicating N^1-protonation in agreement with the greater basicity of N^1, while N^6-

benzoyl-dA was protonated at N^1 and N^7. In other words the relative basicity of the nitrogens is changed in the latter case, probably due to a reduction in electron density of the pyrimidine ring caused by the electron-withdrawing N^6-substituent.

It has been proposed that N^7-protonation of dA leads to more rapid C1'-N^9 cleavage. The N^6-benzoyl-dA appears to depurinate more readily due to its preferential N^7-protonation. This may be due to N^7-protonated adenine being a more stable leaving group than the N^1 tautomer.[131] This change in protonation site had been observed previously in C8 hydrogen exchange experiments of N^6-substituted dA where protonation occurred at N^7 preferentially[132] with the degree of change depending on the acyl group present; benzoyl having a greater effect than the acetyl group.[132] Changes in nitrogen basicity influence depurination rates via the monocation route. The use of dialkylformamide N^6-protection for dA (and dD) such as the sterically hindered di-n-butylformamide group, would be expected to reduce the changes in protonation and N^1, N^7 relative basicity.

1.6.2.4 N^6 Protection for dD

The features outlined above make the di-n-butylformamide group an attractive option for the protection of the N^6-position of dD. Preliminary studies were carried out by Booth.[133]

There were two target monomers. In the first monomer, the N^2-protection employed was the isobutyryl group (by analogy to dG[126]); in the second monomer the N^2, N^6-bis-di-n-butylformamide-dD monomer was synthesised. The reaction scheme is shown below, starting from the 2,3',5'-protected nucleoside. The conversion of the O6 function to an N^6-amino function was carried out as outlined in Figure 1.25, (following Gaffney[69]).

The first monomer showed resistance to depurination in TCA, however the N^2-protection proved to be far too stable under standard deprotection conditions. The second monomer showed reduced coupling efficiency—this was probably caused by the large N^2-di-n-butylformamido group sterically hindering the 5'-OH reaction site, when the base is in the *syn* conformation.[133]

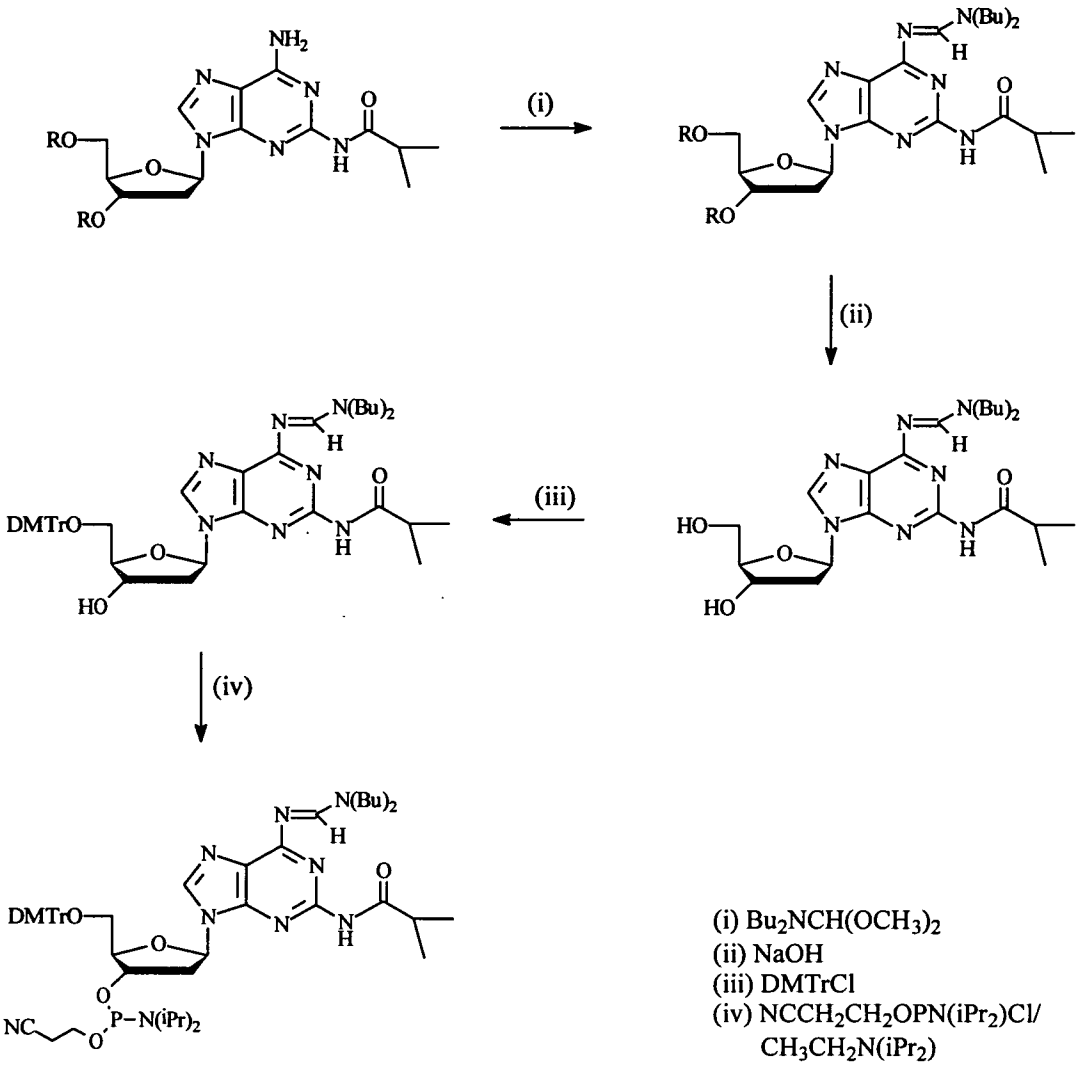


Figure 1.25. *Synthesis of the N^2-isobutyryl, N^6-di-n-butylformamidine monomer.* R= isobuyryl.

Stirring the 3',5'-OH, N^2-tris-isobutyryl-dD in dichloromethane with di-n-butylformamidine dimethylacetal afforded the N^6-formamidine protected nucleoside. This was saponified using 2MNaOH, to allow standard tritylation, then phosphitylation.

The free dD nucleoside was obtained by full deprotection of the isobutyryl protecting groups, by incubation in concentrated NH_4OH solution at 60°C for five days.[133] Suspending the free dD in dimethylacetamide (DMA) with di-n-butylformamide dimethylacetal for 18 hours gave the N^2, N^6-bis-protected

nucleoside; which was then tritylated and phosphitylated, to give the bis-protected monomer

In this project, the N^6-formamide strategy shall be adhered to, but the N^2-protection must be developed, with regards to improving its lability under standard deprotection conditions (concentrated NH_4OH).

1.6.2.5 N^2 Protection for dD

The standard choice of protecting the N^2-amino group (in dG) is the isobutyryl group.[134] As mentioned in Section 1.6.2.4, the isobutyryl group is unsuitable for N^2-protection of dD due to the extremely long deprotection times in concentrated NH_4OH; and N^2-di-n-butylformamide protection is unsuitable due to reduced coupling efficiency.[133]

Therefore a more base-labile protecting group must be found, which is not too sterically bulky. One aim of this project is to satisfy these criteria for N^2-protection of dD. The alternatives tested, and the results of these investigations will be discussed in Section 2.

1.7 DNA CONTAINING DIAMINOPURINE

The duplex stabilising properties of diaminopurine were described in the main in Section 1.5. The incorporation of diaminopurine in DNA and its use in hybridisation with complementary oligonucleotides are a useful consequence.

1.7.1 DNA PROBES

Synthetic oligonucleotides have numerous applications in molecular biology, including their use as hybridisation probes. Only perfectly matched probe-target DNA duplexes are formed under controlled hybridisation conditions.[135] The introduction of additional stabilisation to DNA duplexes would allow the use of more stringent hybridisation conditions and therefore increase the probe specificity for its target.

It has been shown that by incorporating base analogues, such as 2,6-diaminopurine, into oligonucleotides, the hybridisation properties can be modified to enhance probe selectivity and binding strength.[75]

1.7.1.1 Gene probing

Voss and coworkers,[136] found that replacement of a single adenine residue with 2-aminoadenine in a 20-mer oligonucleotide probe can be used to improve the identification and differentiation between the single base change associated with the sickle cell disease mutation. The DNA immobilised contained inserts of the homozygous normal, and the sickle cell diseased genotypes of the β-globin gene. The results indicated an increase in hybrid stability between the analogue-containing probe and the normal β-globin gene sequence.

Hence the use of this base analogue should not only facilitate the screening of other known base mutations for clinical genotyping, but should also enable the more stringent analysis of other gene sequences, such as those present in DNA libraries using oligonucleotide probes.

1.7.1.2 Ambiguity in probing

In determining the sequence of the synthetic DNA probe complementary to a region of a gene, from the amino acid sequence, an ambiguity problem arises due to the redundancy of the genetic code. This is not a trivial problem: a single mismatch can lower the melting temperature by as much as 10°C.[74] Many methods have been introduced to solve this problem.

An approach using probes carrying deoxyinosine (dI) at positions corresponding to ambiguous nucleotides was introduced to decrease the discriminative nature of the probes among ambiguous bases.[135] Other nucleotides which can make stable base pairs with ambiguous genomic nucleotides deserve study. 2,6-diaminopurine is a candidate as a base analogue to probe positions of sequence ambiguity. Cheong and coworkers[74] measured the stabilities of a set of DNA oligomer duplexes containing each of the four normal DNA bases paired with dD. The contributions of these dD base pairs to duplex stability was calculated, and compared with those for dI, dA, dG, dC, and dT.

The ideal analogue will not lower the melting temperature of a duplex (it is non-destabilising), and it will bind equally well to all four natural bases (it is non-discriminating). Cheong *et al* found that inosine was the best probe for the ambiguities of: A/C, C/G, C/T, A/C/G, A/C/T, C/G/T, and A/C/G/T; 2,6-diaminopurine was best for: A/T, G/T, and A/G/T; and thymine for A/G.[74]

1.7.2 dD IN PCR PRIMERS AND DNA SEQUENCING

Since its discovery in 1983 by Mullis,[137] the applications for the polymerase chain reaction (PCR) have spread throughout the biological sciences.

The oligonucleotide primers required for PCR are custom-synthesised and are usually 20 to 30 bases long. They hybridise to specific sequences on the target DNA segment. Diaminopurine-containing oligomers would be useful in primers which, by necessity, were too rich in adenine residues to hybridise with the template.

In sequencing technology, again the extra stability gained by replacing dA with dD, offers an attractive application for diaminopurine. This is particularly true in the recent developments in DNA arrays, where the maximisation of hybridisation, without sacrificing specificity, is crucial.[16]

1.7.3 ANTISENSE THERAPY

On transcription, every gene gives rise to a relatively large number of messenger RNA, which is then translated into a large number of proteins. Therefore it would seem that inhibition of gene expression[138] would be more efficient than inhibition of the resulting protein (this is the approach followed by conventional drug design). It is possible to achieve sequence-specific recognition of nucleic acids using synthetic oligonucleotides that bind specifically to complementary nucleic acids. These compounds are called antisense oligonucleotides based on their binding to the target sequence (sense strand). Since the sequence of a 17-mer oligonucleotide occurs statistically only once in the human genome, extreme selectivity should be possible with antisense oligonucleotides of at least this length.

Zamecnik and Stephenson were the first to propose the therapeutic use of synthetic antisense oligonucleotides.[139] They synthesised a 13-mer oligonucleotide complementary to the RNA of Rous sarcoma virus, and showed that this antisense strand inhibited the growth of the virus in a cell culture.

There are many requirements that an oligonucleotide must fulfil in order to function as an antisense strand. Therefore it is necessary to chemically modify the oligonucleotides. An example of this is that the phosphate backbone must be modified to avoid enzyme digestion when the strand enters a cell. Typically the approach by which phosphorothioate-backbone analogues of the oligonucleotides are made, avoids enzymatic degradation.[39]

Indeed, antisense technology could exploit the extra stability and selectivity in duplex formation afforded by diaminopurine replacing adenine residues. So far, the only way of using antisense DNA with diaminopurine has been by enzymatic incorporation via the triphosphate in a reaction catalysed by Klenow polymerase.[73,75]

Thus a facile and high-yielding method incorporating a diaminopurine phosphoramidite in solid phase oligonucleotide synthesis would prove useful in diagnostic and therapeutic molecular biology, as well as for structural studies of this novel base analogue.

1.7.4 MOLECULAR RECOGNITION OF dD

The substitution of bases by analogues in DNA is often a useful approach to investigate DNA biochemistry. DNA containing base analogues (such as dD) have been shown to be modified in thermal stability, base pairing pattern, or double-helical conformation. (*See* Sections 1.2-1.5). As a result, differences in hybridisation strength and selectivity, changes in protein recognition patterns, or mutational activity may occur.

Recognition of a specific DNA sequence by an enzyme is achieved by the formation of hydrogen bonds between specific amino acid side chains of the protein and determinants in the major and minor grooves of DNA. Alteration of bases in a recognition site sequence may therefore disrupt or modify the hydrogen bonding pattern by electronic or steric effects. This usually results in a loss or decrease in enzyme activity.

The ability of a protein to recognise a specific sequence of bases along a strand of double helical DNA lies at the heart of many fundamental biological processes, including the replication and expression of DNA. Other nucleic acid-protein interactions include: that of restriction endonucleases with double helical DNA, RNase H with DNA.RNA hybrids, and ligase enzymes with double- or single-stranded nucleic acids.

1.7.4.1 Probing restriction enzyme recognition sites

Restriction endonucleases are often used as a paradigm for protein-DNA interactions because of their high specificity for DNA sequences and the simplicity of the DNA site recognised.[140]

Chollet and Kawashima[75] used dD-substituted DNA to probe minor groove determinants, and demonstrated that substitution of adenine by dD in DNA has profound effects on the recognition and cleavage by some restriction endonucleases. All fragments obtained from cleavage of D-containing DNA were similar to the natural reference DNA. Cleavage of DNA by restriction endonucleases is accomplished in two steps: firstly the enzyme binds to the recognition site, and then the phosphodiester bond is cleaved. Whilst Chollet and Kawashima could not distinguish between these two steps, they were able to assess qualitatively the effect of the minor groove modification due to dD, and to demonstrate the complete inhibition of cleavage for several endonucleases

1.7.4.2 Replication of DNA

The faithful replication of DNA is dependent on the DNA polymerase enzyme recognising the primer-template region, and incorporating the correct substrate dNTP to complement the template.

Primer-templates with poly(dD.dT)

Sagi and coworkers[141] compared DNA polymerase reactions for poly(dA.dT) and poly(dD.dT) under conditions for the B-forms, and conditions known to induce the X-form of poly(dD.dT).* This was carried out in order to assess whether this unusual conformation has any effect on DNA replication. Kinetic data presented in this study show that Klenow DNA polymerase is sensitive to the formation of this unusual type of double helix. Contrary to Z-DNA[142], X-DNA is copied by the Klenow enzyme, although with a reduced efficacy as compared to its B-form. This DNA polymerase differentiation is *in vitro* enzymatic evidence of an unusual DNA conformation.

* The X-form (putative A-form) of poly(dD.dT), as discussed in Section 1.5.3.3, can be induced by millimolar concentrations of Mg^{2+}, when the monovalent cation concentration is low, and the pH is around 7. The DNA polymerase reactions were carried out in 60mM phosphate plus 6mM Mg^{2+} (B forms) and in 10mM phosphate plus 6mM Mg^{2+} {X-form of poly(dD.dT)}.

Diaminopurine and the fidelity of replication

Certain analogues of the dNTP's, modified in the sugar or base, are accepted by polymerases for pairing with the DNA template and are incorporated into DNA, but subsequently block further chain growth or interfere with nucleic acid function.[143]

Chollet and Kawashima[75] prepared dDTP and showed that dDTP is a substrate for *in vitro* DNA polymerase I mediated enzymatic synthesis. Also, it was shown that DNA containing dD are good substrates for the DNA modifying enzymes: T4 DNA ligase, polynucleotide kinase, and *E. coli* methylase. Enzymatic synthesis using dDTP as the substrate proceeded at 20% of the rate when the substrate was dATP; this could be improved up to 90% of the control rate by increasing the concentration of dDTP. The induction of mutations through the formation of D.C mispairs has been reported,[144] but the frequency of these mutations, while higher than spontaneous mutations, is insignificant.[145]

1.7.4.3 Transcription involving diaminopurine

In their studies of amino analogues of adenine, Rackwitz and Scheit[146] found that DTP, contrary to n^2c^6ATP, behaved as a true analogue of ATP in the transcription of poly(dA.dT), poly(dT), and calf thymus DNA. This, along with their other findings, led them to conclude that the 6-NH_2 group of adenine is an essential steric feature in the recognition of A.U base pairs during transcription.

The effect of the minor groove: 2-NH_2 of dD has also been studied, with regard to its substitution of dA in T7 RNA polymerase promoter sites—dD and hypoxanthine, which have minor groove alterations, prevent utilisation of the promoter. These analogues do not affect transcription which starts outside of the modified region.

1.8 BASE-LABILE LINKERS AND AMINO PROTECTION

1.8.1 THE SOLID SUPPORT AND COVALENT LINKER

The advantages of solid-phase synthesis (Section 1.1) are obvious. Despite this, the development of satisfactory procedures for oligonucleotide synthesis required almost 20 years from the first introduction of this technique, This was primarily because of two obstacles that had to be overcome. The first was the need for rapid and highly efficient coupling reactions, and the second was the need for a suitable solid-support and the important covalent linkage that binds the oligonucleotide to it.[147]

The insoluble material used to anchor oligonucleotides during solid phase synthesis cannot be considered as simply an inert carrier, since both its physical and chemical properties can have an important influence on the success of the synthesis. Such important features include: uniform surface structure, uniform particle size, and length and type of spacer. In addition, the type of covalent linkage joining the first nucleoside to the support is important in determining both the strategy required for the final deprotection and cleavage reactions; and the terminal composition of the final product.[147]

1.8.1.1 The solid-support

A large variety of materials have been tried as insoluble supports during the evolution of oligonucleotide synthesis. By far the most widely used support is controlled-pore glass (CPG) first used in 1972 by Köster,[148] and reintroduced in the early 1980's when phosphorochloridite coupling chemistry was first applied to the solid-phase synthesis of DNA.[111]

The solid-support to succeed the aminopropylated CPG was long-chain alkylamine CPG (LCAA-CPG). This material contains a primary amino group on the end of a 17-atom-long linker arm. This linker arm, along with the large (500Å) pore size of LCAA-CPG greatly improved coupling results up to oligonucleotides of sequence length 50-60 nucleotides.

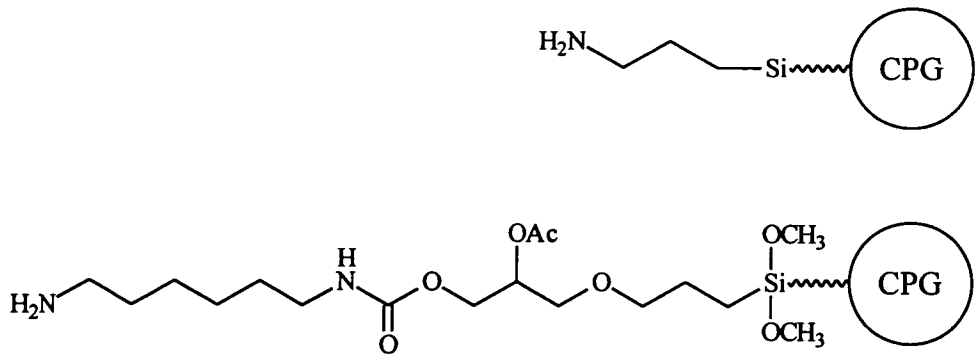


***Figure 1.20**. Comparing the aminopropylated CPG and LCAA-CPG.*

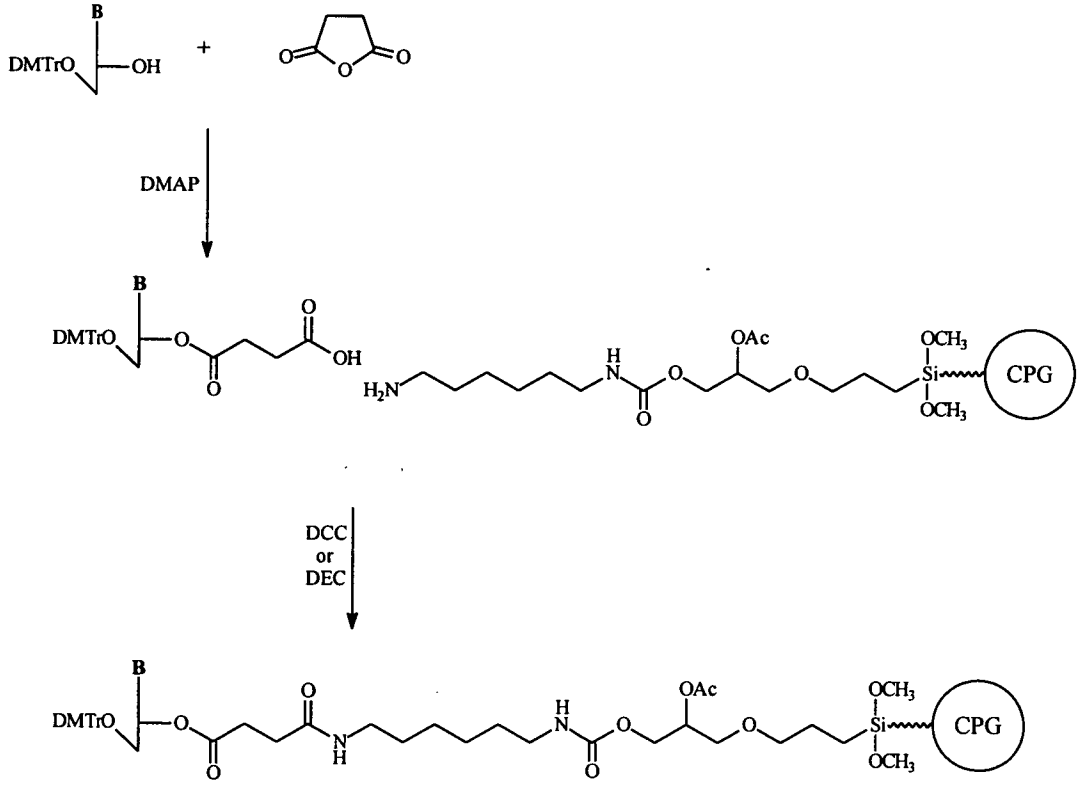


***Figure 1.21**. The synthesis of succinyl-derivatised LCAA-CPG.*
DCC=1,3-dicyclohexylcarbodiimide;DEC=1-(3-dimethylaminopropyl)-3-ethylcarbodiimide.

Much longer sequences can be prepared on similar 1000Å-2000Å CPG supports (wide-pore).[147]

1.8.1.2 The covalent linker

In almost all CPG supports, a nucleoside-3'-succinate is attached to the support via an amide bond. This succinate linkage is easily cleaved under basic conditions. This is usually desirable, because it allows cleavage from the support to be combined with the phosphate and base deprotections, which also require basic hydrolysis.

Alternatives to the succinyl linkage have been developed for three main applications: the development of base-stable linkages; the development of more base-labile linkages; and linkages which allow functionalisation of the 3'-terminus of oligonucleotides. The former two types of linker will be discussed below.

1.8.2 BASE-RESISTANT LINKAGES

The compatibility of the nucleoside linker with the persistent base-labile protecting groups precludes the synthesis of completely deprotected oligonucleotides on insoluble supports. This shortcoming has been overcome by several strategies that replace the nucleoside succinate attachment with more resistant linkages. These approaches allow easier purification of the final product, since fewer non-nucleotide impurities are present. They also allow the support-bound material to be used as a solid-phase affinity column.

The first base-resistant linkage on CPG was introduced by Sproat and Brown,[149] who used tolylene-2,6-diisocyanate (TDIC) to link nucleosides onto LCAA-CPG via two urethane linkages.

In another approach, by Brown et al,[150] LCAA-CPG was derivatised with a succinyl-sarcosyl linker which exhibited less than 5% cleavage after an overnight treatment with 10% DBU in dichloromethane. Under these conditions, the conventional succinate linkage was cleaved within 1 hour, presumably, as a result of the deprotonation of the amide function followed by intramolecular nucleophilic displacement at the ester carbonyl group.[150] Additional support for this mechanism

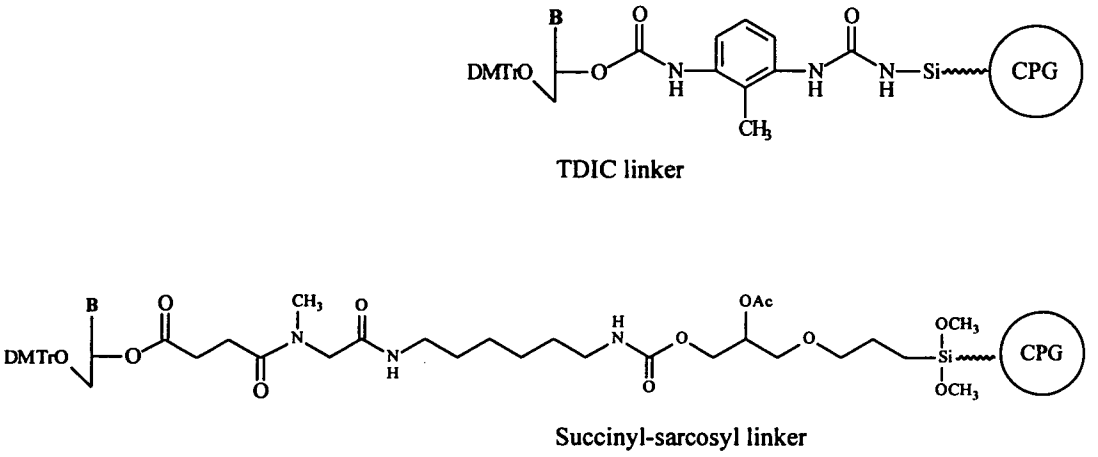


Figure 1.22. *Two base-resistant nucleoside linkages.*

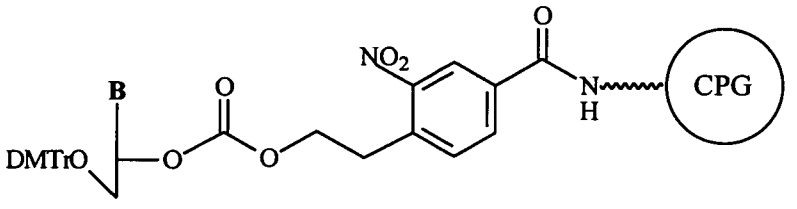
comes from studies on the phthaloyl linker. When the solid-support was treated with 10% DBU, liberation of the nucleoside was found to be extremely fast owing to the rigid conformation imposed by the aromatic ring, favouring the intramolecular reaction (50% cleavage in 3.5 minutes).[150] This observation led to the investigation of the phthaloyl group as a more base-labile linkage in this project. (*See* below, and Section 4).

1.8.3 BASE-LABILE LINKERS

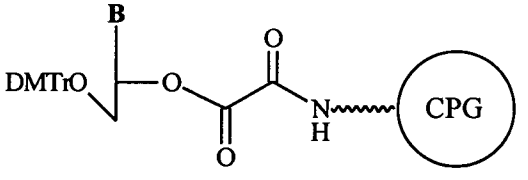
In contrast to the above applications, several types of nucleoside linker have been prepared that use milder conditions to release the oligonucleotide. These more base-labile supports are also supplemented by the introduction of either more labile protecting groups (see below), or coupling techniques that avoid base protection entirely,[151] so that overall more gentle conditions for deprotection can be employed.

One possibility for an easily cleavable linkage is to use a 2-(2-nitrophenyl)ethyl linkage (NPE).[152] Cleavage from the support was possible by treatment with DBU, therefore sequences made with the NPE-protected nucleoside 2-cyanoethyl phosphoramidites could be completely deprotected and cleaved from the support with a single DBU treatment.

An oxalyl linkage has also been used to attach nucleosides to LCAA-CPG.[153] In this method, oxalyl chloride replaces succinic anhydride as the linker that joins the nucleoside to the support. The oxalyl linker was stable to dry pyridine or triethylamine for at least 15 hours at ambient temperature. Conversely, 5'-O-DMTr-T was cleaved from the oxalyl linker within 5 minutes upon treatment with either wet triethylamine, 40% trimethylamine in methanol, or 5% ammonium hydroxide in methanol. Under these conditions, less than 5% of 5'-O-DMTr-T was cleaved from the standard succinyl LCAA-CPG support.[153]



Nitrophenylethyl linker



Oxalyl linker

Figure 1.23. *Easy-cleavable nucleoside linkages.*

1.8.4 ULTRAFAST DNA SYNTHESIS

Chemical synthesis of synthetic oligonucleotides has been completely automated during the last decade. However, deprotection, as well as purification, remains a manual and time-consuming process. At the same time the increasing demand for synthetic oligonucleotides in a variety of biological experiments has created much interest in their more rapid processing.

Commercial DNA synthesisers currently are capable of operating with fast cycles of 2 to 3 minutes. The synthesis of primers can, therefore, routinely be achieved in under an hour. However, a major bottleneck in processing the products is the cleavage and deprotection steps which can take anywhere from 1 to 7 hours. All this is before the isolation or purification strategies which take a minimum of 1 hour.

Using the standard base protection scheme (Figure 1.4), deprotection can be speeded up by elevating temperature,[154] or combining elevated temperature (80°C) with the addition of triethylamine to the ammonium hydroxide.[155] In addition to the potential hazard of such a high temperature, in both cases the time for both cleavage and deprotection is still at least 1 hour.

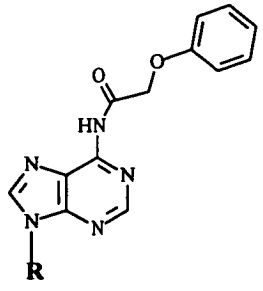
To speed up the rate of oligonucleotide deblocking, two approaches can be envisaged: firstly, utilise monomers which possess highly labile protecting groups[156-158] and linkers (see above), or, secondly, employ highly efficient deblocking reagents for the deprotection of classically-protected oligonucleotides.[159] The most effective technique, clearly, would involve both approaches, without sacrificing fidelity of deprotection.

1.8.4.1 Base-labile protecting groups

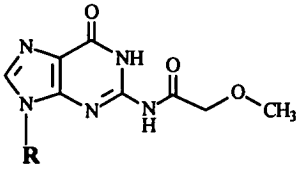
The main plausible candidates for fast deprotection developed in recent years include: the phenoxyacetyl (pac) group, methoxyacetyl (mac), t-butyl phenoxyacetyl (bpac), isobutyryl (ib), dimethylformamidine (dmf), and acetyl (ac). These groups were namely used as: pac-dA, mac-dG, and ib-dC by Schulof et al;[157] dmf-dA, dmf-dG, and ib-dC by Vu et al.[156] Sinha and co-workers developed bpac-dA, dG, and dC;[160] whilst Reddy and coworkers developed ac-dC.[158] These protecting groups are shown in Figure 1.24.

Each protection system introduced to date has exhibited flaws. Pac protection is a reasonable alternative to the standard base protecting groups, but a deprotection time of 30-60 minutes does not include the linker cleavage step (normally at least 45 minutes), and the pac-dG monomer is rather insoluble.[157] The use of dmf-protected nucleotides[156] allows deprotection comparable with pac, but the stability of dmf-dA during synthesis remains a problem. The bpac-protected monomers[160] (commercially

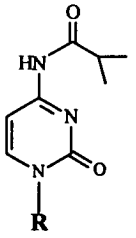
known as Expedite™) offer faster deprotection (15 minutes at 55°C) but again the cleavage from the support is still 45 minutes. Also, these monomers require a change in capping reagent (t-butylphenoxyacetic anhydride rather than the normal acetic anhydride). These deprotection times are compared in Table 1.2 below.



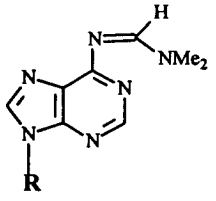
pac-dA



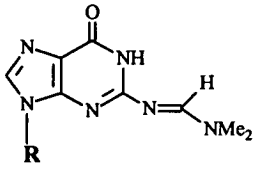
mac-dG



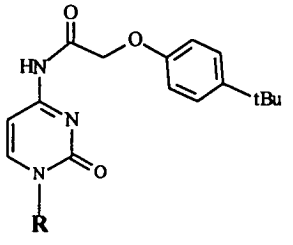
ib-dC



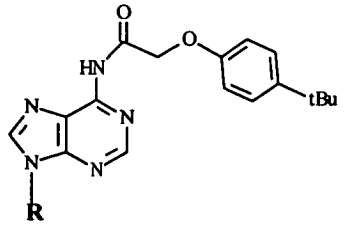
dmf-dA



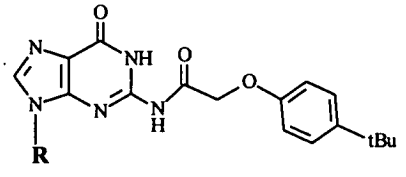
dmf-dG



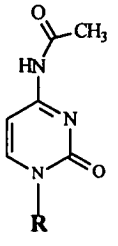
bpac-dC



bpac-dA



bpac-dG



ac-dC

Figure 1.24. *Easily-cleaved protecting groups.*

1.8.4.2 More nucleophilic deprotection reagents

The use of ethanolamine (EA) or particularly a mixture of ethanolamine/ hydrazine/methanol for deprotection of oligonucleotides was employed by Polushin and coworkers, in conjunction with oxalyl linkers and bpac-amino-protection.[161]

With EA, oligonucleotides were fully deprotected in 15-20 minutes at room temperature. Under the same conditions deprotection with the EA mixture required only 3-5 minutes. Potential drawback with this technique are the formation of a slight amount of modified C, as well as the volatility of EA.[161]

A commercial cleavage and deprotection system has been developed (UltraFAST),[158] requiring the novel ac-dC protection. All other monomer protecting groups remain the same. This seemingly minor change in protecting leads to an oligonucleotide which can be cleaved and deprotected in 10 minutes using the reagent mixture AMA: ammonium hydroxide/aqueous methylamine (50:50). Table 1.2, below compares this technique with other cleavage/deprotection strategies.[162]

Table 1.2. *Comparison of fast cleavage and deprotection techniques.*

Monomer	Reagent	Cleavage Time	Fastest Deprotection Temp.	Time	Total Time
Normal	NH_3	45min	80°	60min	105min
Normal	NH_3/TEA	45min	80°	30min	75min
Pac	NH_3	45min	70°	30min	75min
Dmf	NH_3	45min	55°	60min	105min
Bpac	NH_3	45min	55°	15min	60min
Ac-dC	AMA	5min	65°	5min	10min

RESULTS AND DISCUSSION

2.0 SYNTHESIS OF DIAMINOPURINE PHOSPHORAMIDITES

As detailed in Section 1.6, two separate protecting groups are required for the exocyclic amino moieties. A successful protecting group for the N^6-position is the di-N-butylformamidine group which helps suppress acid-catalysed depurination in adenine, and has been used for N^6-protection in dD.[133] This group will be used throughout this study—one exception being monomer **[12]** which was synthesised in order to assess N^6-dimethylformamide protection.

The main aim of this study was to develop a dD monomer stable to standard oligonucleotide synthesis conditions, with routine coupling efficiency (>98%), and with satisfactory deprotection rates. (The previous generation of diaminopurine monomers, employing N^2-isobutyryl protection, had a half-life of approximately 35 hours, for one diaminopurine residue, in concentrated ammonia at 55°C).[69,73,133] This is clearly unsatisfactory—N^2-isobutyryl-deoxyguanosine is *fully* deprotected in less than 12 hours under the same conditions. In addition, it was hoped a more convenient synthesis of the phosphoramidite could be developed.

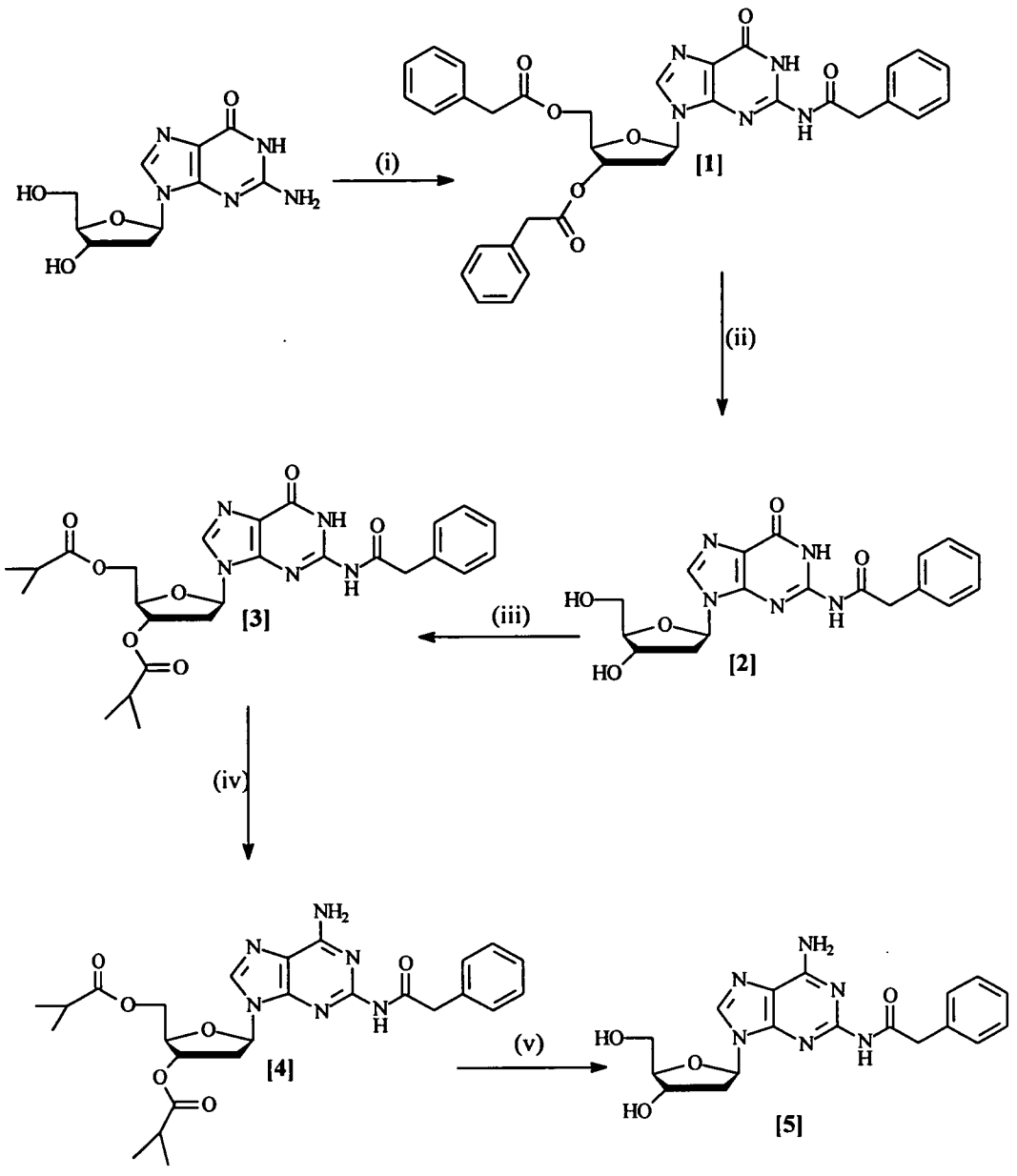
2.1 N^2-PHENYLACETYL PROTECTION OF DIAMINOPURINE

The N^2-phenylacetyl group was first employed for dG protection to help increase its solubility in organic solvents.[163] This group is more base-labile than the isobutyryl group, so it was hoped this group would make deprotection of oligonucleotides containing dD more facile.

2.1.1 Synthesis of the N^2-phenylacetyl-dD nucleoside [5] (Figure 2.1)

Reaction of 2'-deoxyguanosine with phenylacetic anhydride, with gentle reflux in pyridine, followed by work up and purification by silica gel chromatography gave compound **[1]** in 84% yield. This was dissolved in pyridine-methanol solution (3:1), cooled to 0°C, and treated with 2M NaOH solution. The reaction was quenched after 10min with H^+ form Dowex resin, filtered, and the solvent removed. This residue

was precipitated in dichloromethane-methanol (80:20), to give compound [2] in 88% yield.



(i) a:Phenylacetic anhydride in pyridine at 0°C. b:110°C.
(ii) 2M NaOH in pyridine/methanol at 0°C.
(iii) Isobutyryl chloride in pyridine at 0°C.
(iv) a:MSCl, TEA and DMAP in DCM. b:N-methyl pyrrolidine at 0°C. c:Liquid ammonia.
(v) 2M NaOH in pyridine/methanol.

Figure 2.1. *Synthesis of N^2-phenylacetyl diaminopurine nucleoside* ***[5].***

Reaction of **[2]** with isobutyryl chloride in pyridine gave compound **[3]** in 90% yield.This was dissolved in dichloromethane, with triethylamine (TEA) and DMAP, and reacted with: (i) mesitylene sulphonyl chloride, and (ii) N-methyl pyrrolidine. (iii) NH_3 gas was then condensed into the reaction vessel (fitted with a dry-ice/acetone cold finger). The resulting dark red-brown residue was worked up and purified to give **[4]** in 55% yield. This amination step is based on the method used by Gaffney,[69] but the more reactive mesitylenesulphonyl chloride replaces TPSCl, and the easier to handle N-methyl pyrrolidine (liquid) replaces trimethylamine (gas), (*see* Section 1.6). Compound **[4]** was saponified as for compound **[2]** to give dD nucleoside **[5]** in 83% yield.

2.1.2 Synthesis of N^2-phenylacetyl-dD monomer [7] (Figure 2.2)

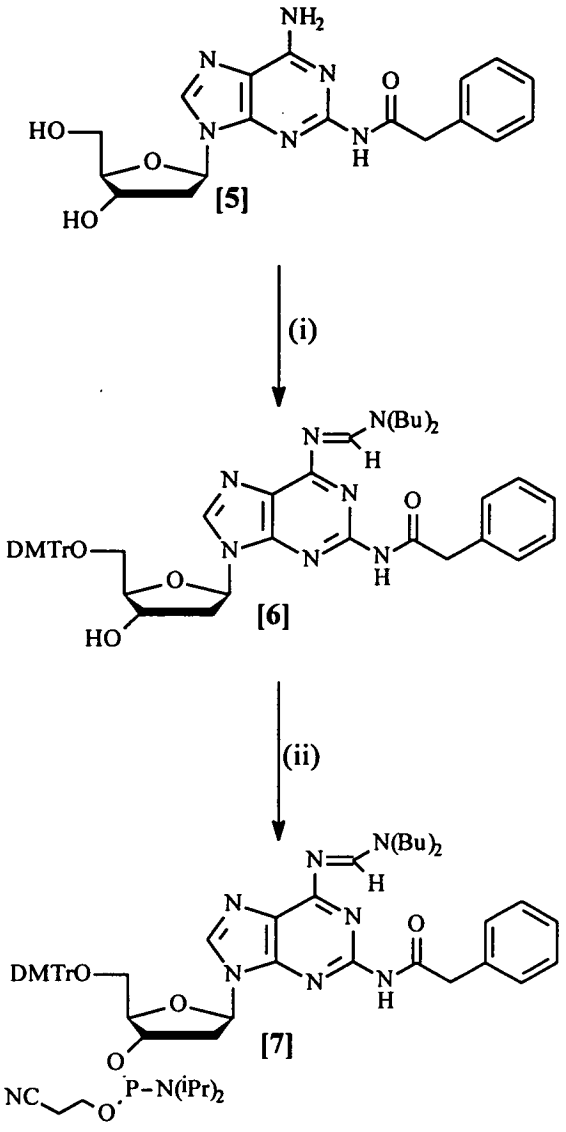
Diaminopurine nucleoside **[5]** was suspended in pyridine and reacted with di-N-butylformamide dimethylacetal. After concentration to a pale yellow oil, the residue was dissolved in dichloromethane with triethylamine and DMAP, and dimethoxytrityl chloride was added. The reaction mixture was then worked up and purified to leave compound **[6]** in 92% yield. This nucleoside was then dissolved in THF with N,N-diisopropylethylamine (DIPEA) and reacted with 2-cyanoethyl N,N-diisopropyl-chlorophosphoridate. After work up and purification, dD monomer **[7]** was obtained in 91% yield.

2.1.3 Oligonucleotide synthesis using monomer [7]

dD monomer **[7]** was typically dissolved in anhydrous acetonitrile, and used as a 0.12M solution during oligonucleotide synthesis

Coupling efficiency of monomer [7]

This monomer repeatedly gave coupling efficiencies of <95%, even after increasing coupling times, and increasing the monomer concentration to 0.15M. The exact reason for this poor coupling is not known. A possible explanation may lie with the bulky N^2-phenylacetyl group, hindering the approach of the incoming activated phosphoramidite.



(i) a:Di-N-butylformamide dimethylacetal in pyridine. b:DMTrCl and DMAP in pyridine.
(ii) 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite and DIPEA in THF.

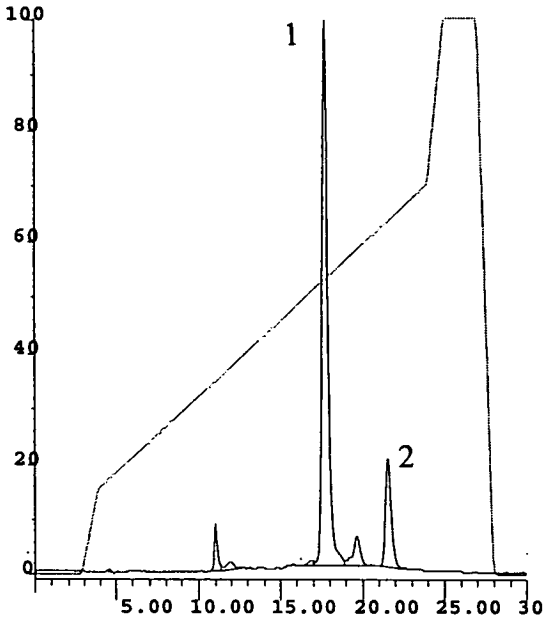
Figure 2.2. *Synthesis of the phenylacetyl diaminopurine monomer [7].*

Poor coupling was also observed in a previous study when the N^2-di-N-butylformamide group was employed.[133]

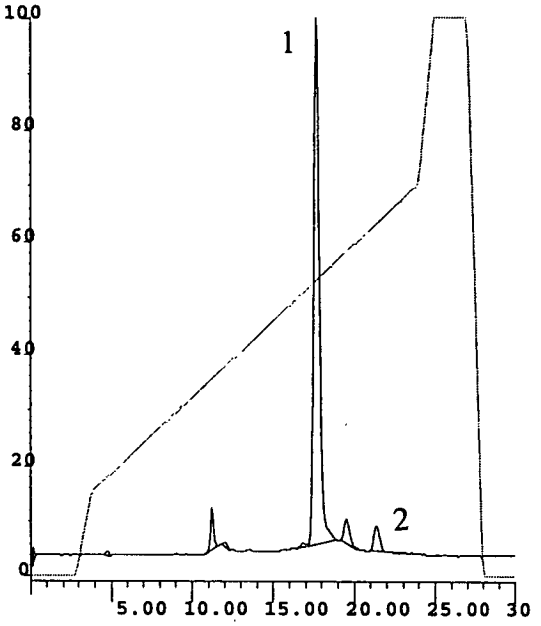
Deprotection of the N^2-phenylacetyl group

Although the coupling efficiency of monomer **[7]** was unsatisfactory, it was good enough to allow a study of the deprotection of the N^2-phac group. This group showed a vast improvement over the isobutyryl group previously used—the half-life of deprotection of an oligonucleotide containing 1 dD was 8 hours for phac, and 35 hours for ib. HPLC chromatograms showing deprotection of oligonucleotides containing one- and three-dD residues are shown in Figure 2.3.

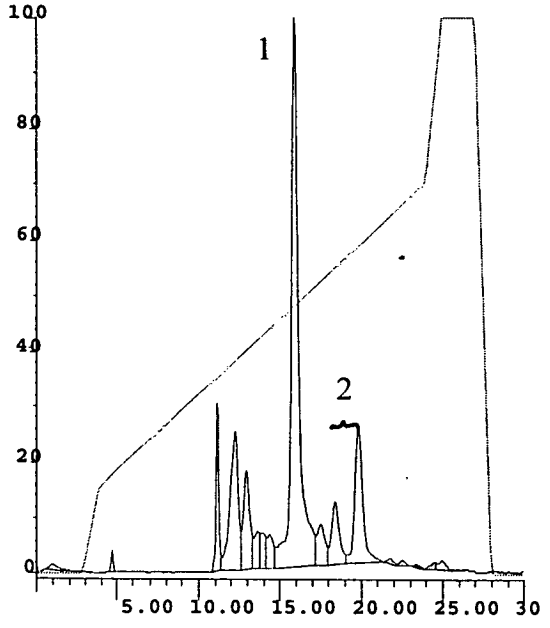
Although this group is not satisfactory, on the grounds of poor coupling, the deprotection times were much improved for single- or multiple-additions of dD residues. The search then had to focus on a less bulky group.



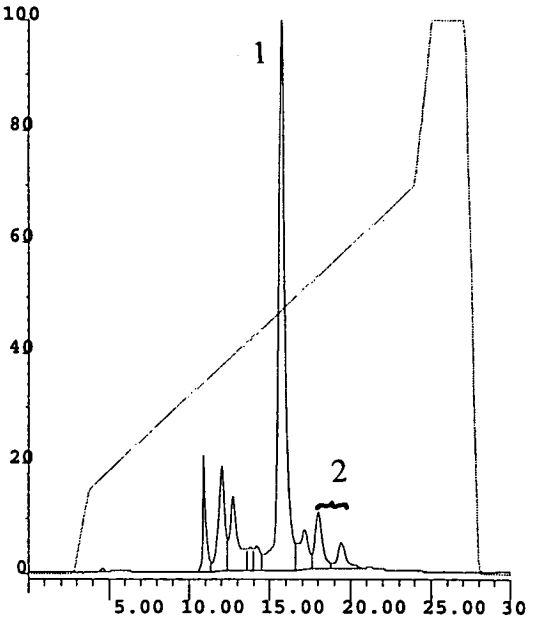
(a) TTT TTD TT at 55°C for 20 hours.



(b) TTT TTD TT at 55°C for 40 hours.



(c) TDD TDT T at 65°C for 20 hours.



(d) TDD TDT T at 65°C for 40 hours.

Figure 2.3. *The deprotection of N^2-phenylacetyl test oligonucleotides in conc. NH_3.* In each case, (a)-(d), the protected oligonucleotide(s) are denoted (2), and the fully deprotected oligonucleotides are denoted as (1). (Note the use of higher temperature for the oligonucleotide containing three diaminopurine residues).

2.2 N^2-ACETYL PROTECTION OF dD

Attention then turned to the acetyl group for N^2-amino protection. It was hoped that the acetyl group would deprotect rapidly, as well as couple with high efficiency.

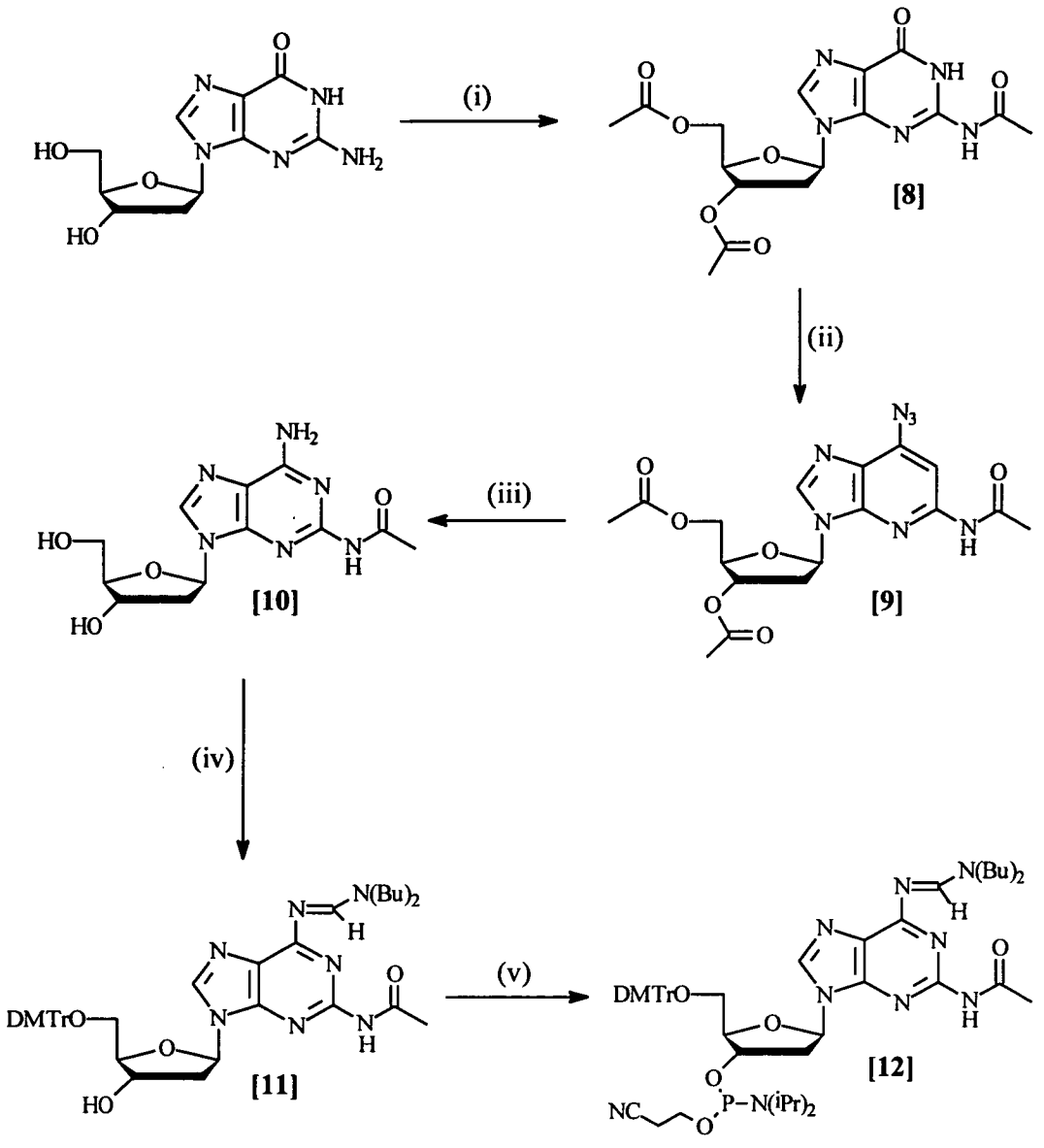
2.2.1 Synthesis of the N^2-acetyl monomer [12]

2'-Deoxyguanosine was refluxed with acetic anhydride and DMAP in pyridine to give a dark red-brown solution. This solution was treated with saturated aqueous $NaHCO_3$ solution, and concentrated *in vacuo*. The resulting solid residue was recrystallised from water and methanol to give compound **[8]** in 76% yield as pale pink crystals.

At this stage we decided to convert the lactam moiety to an amino group, via hydrogenation of an azide. It was hoped that this would increase the yield of the N^2-protected diaminopurine nucleoside, since conversion via the amination of the lactam function as previously discussed, rarely exceeded 55% (*see* Figure 2.1). Compound **[8]** was reacted with mesitylenesulphonyl chloride then N-methyl pyrrolidine in the normal manner, then tetrabutylammonium azide (TBAZ) was added. The dark brown reaction mixture was then worked up and purified by wet-flash chromatography to give compound **[9]** as pale brown solid in 80% yield. Saponification of compound **[9]**, followed by catalytic hydrogenation with palladium on activated charcoal at 4atm gave a light-brown solid on *in vacuo* evaporation of the solvent. This residue was washed with a small amount of dichloromethane to leave compound **[10]** as a white powder in 84% yield. Reaction with di-N-butylformamide dimethylacetal, followed by tritylation in pyridine, gave compound **[11]** as a white solid foam in 78% yield. This tritylated nucleoside was phosphitylated in THF to give compound **[12]** as a white foam in near quantitative yields.

2.2.2 Oligonucleotide synthesis using [12]

dD monomer **[12]** was dissolved in anhydrous acetonitrile to a concentration of 0.12M and used for oligonucleotide synthesis.



(i) a:Acetic anhydride and DMAP in pyridine at 0°C. b:Reflux at 110°C.
(ii) a:MSCl, TEA, and DMAP in dichloromethane. b:N-methyl pyrrolidine at 0°C. c:TBAZ.
(iii) a:2M NaOH in pyridine/methanol at 0°C. b.Pd/C in ethanol at 4atm hydrogen.
(iv) a:Di-N-butylformamide dimethylacetal in pyridine. b:DMTrCl and DMAP in pyridine.
(v) 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite and DIPEA in THF.

Figure 2.4. *Synthesis of the acetyl diaminopurine monomer [12].*

Coupling efficiency of monomer [12]

This monomer gave coupling efficiencies routinely >98% thus fulfilling the first expectation of the N^2-acetyl protection group.

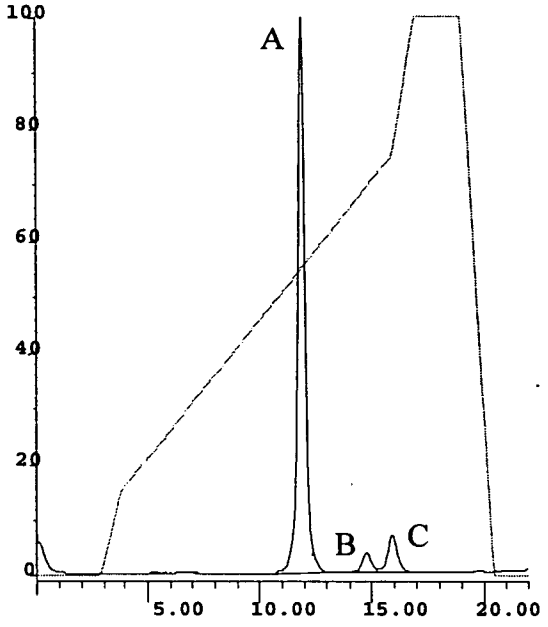
Deprotection of the N^2-acetyl group

The acetyl group offered a slight improvement over the isobutyryl group, but deprotected much more slowly than the phac group (half-life of 30 versus 8 hours). In addition, deprotection of the oligonucleotides synthesised could not be monitored by reverse phase HPLC—owing to the poor lipophilicity inherent with the acetyl group, the protected and deprotected oligonucleotides had exact same retention times, even when several ac-protected dD residues were present. This problem could not be resolved using capillary electrophoresis either, therefore the samples had to be desalted and subjected to enzyme digestion, *then* HPLC analysis in order to resolve the individual nucleosides. (Interestingly, the acetyl group was disregarded by Khorana for N^2-protection of dG, because it was not lipophilic enough during the block coupling stage of the diester method of oligonucleotide synthesis).[164]

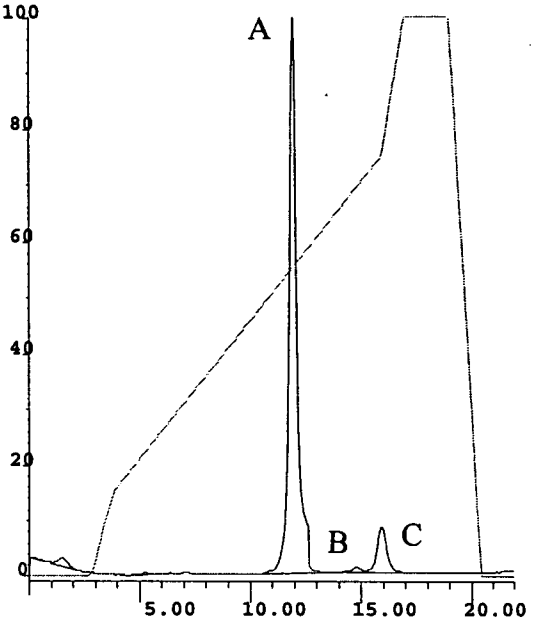
The HPLC chromatograms of the enzyme digests for the two test oligonucleotides are shown in Figure 2.5.

The slow deprotection, along with the difficulties involved in resolving the protected and deprotected oligonucleotides, deemed the acetyl-protected dD monomer unsuitable for use in oligonucleotide synthesis.

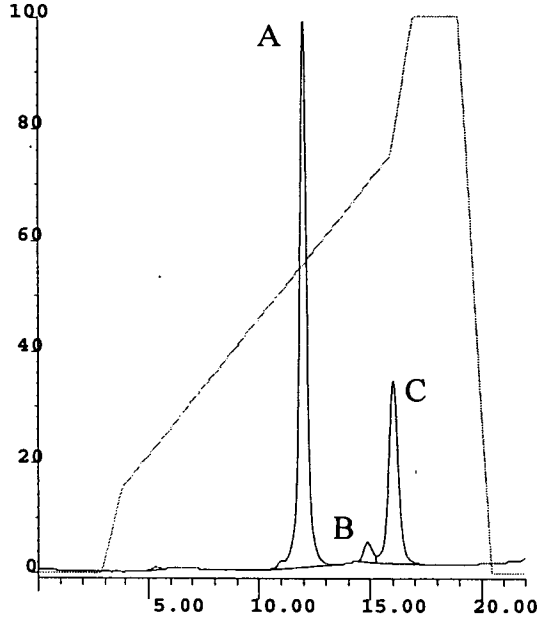
In the meantime, two other monomers were investigated in order to incorporate an N^6-azide or an N^6-dmf group, into oligonucleotides.



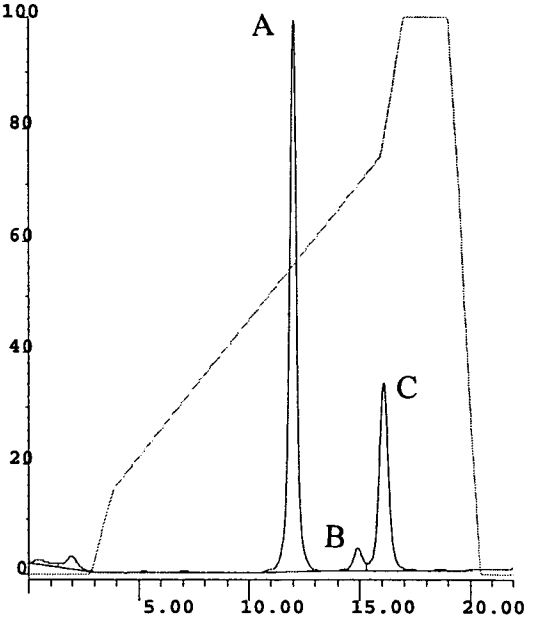
(a) TTT TTD TT at 55°C for 40 hours.



(b) TTT TTD TT at 55°C for 80 hours.



(c) TDD TDT T at 65°C for 60 hours.



(d) TDD TDT T at 65°C for 80 hours.

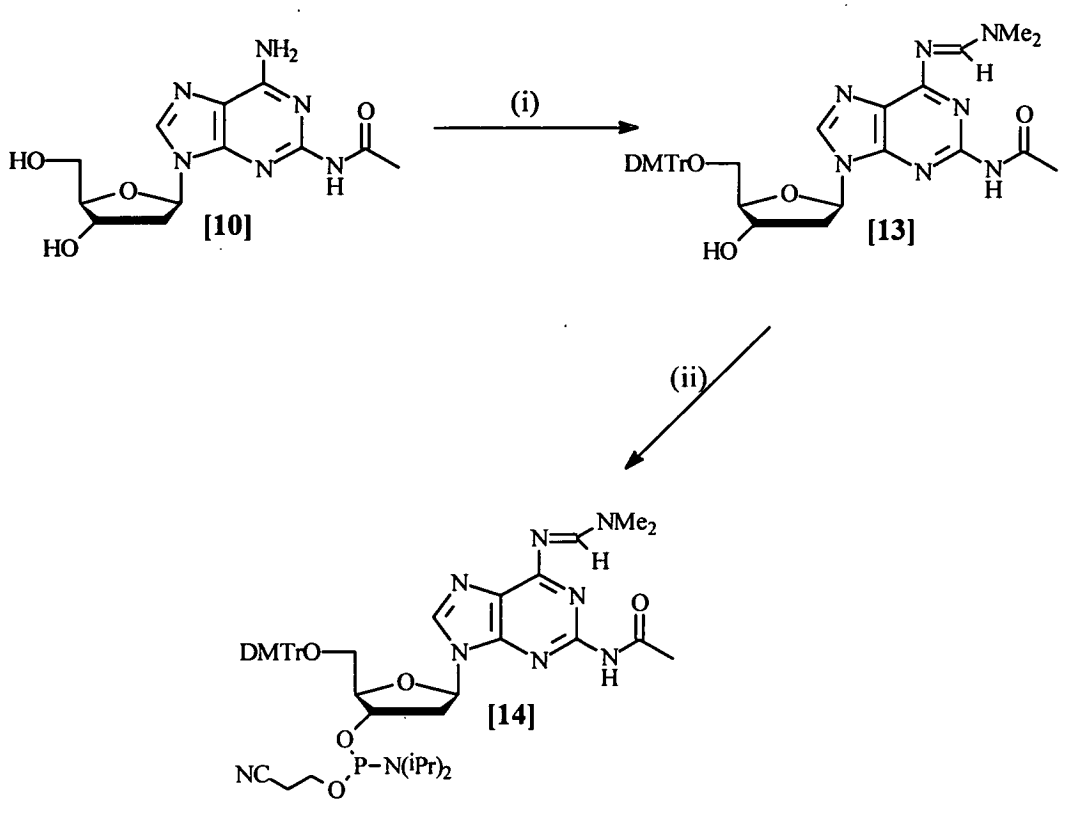
Figure 2.5. *Enzyme digests showing the deprotection of the acetyl test oligonucleotides.*

These traces show the emergence of the dD nucleoside (C) at 16 min, from N^2-acetyl-dD (B) at 15 min. The dT peak (A) is shown at 12 min.

2.2.3 Synthesis of N^6-dmf-dD monomer [14] (Figure 2.6)

Another novel nucleoside was synthesised—N^6-dimethylformamide (dmf) protected dD. This was carried out in order to evaluate the suitability of the dimethylformamide group, as a replacement for di-N-butylformamide protection.

Compound **[10]** was reacted with dimethylformamide dimethylacetal in pyridine to give the N^6-protected nucleoside. This was coevaporated with, and redissolved in pyridine, and reacted with dimethoxytrityl chloride to give compound **[13]** in 61% yield. Compound **[13]** was then phosphitylated to give compound **[14]** in 31% yield.



(i) a:Dimethylformamide dimethylacetal in pyridine. b:DMTrCl and DMAP in pyridine.
(ii) 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite and DIPEA in THF.

Figure 2.6. *Synthesis of the N^6-dimethylformamide monomer [14].*

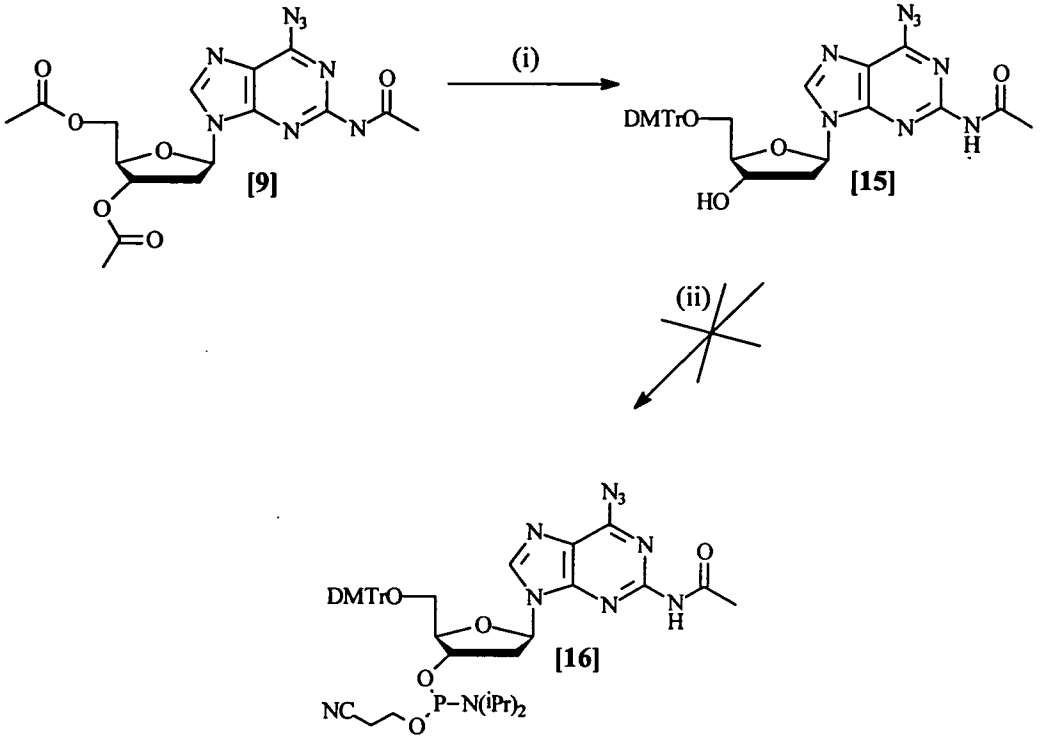
Oligonucleotide synthesis using monomer [14]

The low yield of monomer **[14]** may have been a contributing factor of the failure of oligonucleotide synthesis—less than 200mg was available for synthesis. At these low levels, oxidation by-products can become a serious problem. Alternatively, the low coupling efficiency obtained may have been due to the use of the N^6-dmf group, as a previous study had raised questions about the stability of N^6-dmf-dA during oligonucleotide synthesis.[165] The stability of the N^6-dmf-dD nucleoside to TCA was investigated (*see* Section 2.5.3) to help explain the failure of oligonucleotide synthesis.

2.2.4 Synthesis of N^6-azido-dG monomer [16] (Figure 2.8)

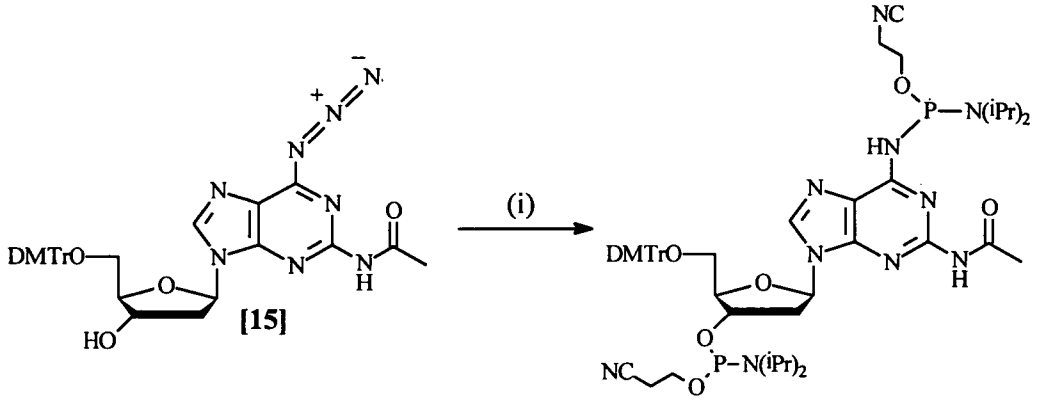
Many proteins that interact with DNA show a very high specificity for their particular target sequence. The use of reactive probes, where DNA containing modified bases which react with nucleophiles, is one method of studying the interaction of DNA with peptides and proteins.[166] This involves incorporating azido-, sulphur-, or halogen-modified nucleotides into the synthetic DNA of interest, and allowing the DNA-peptide interactions to take place.[167] Irradiation of the sample then crosslinks the two molecules together covalently. So far, the azido group has only been introduced into the sugar moiety of DNA, therefore the novel nucleoside **[16]** was the target of our attempts to synthesise a monomer containing an exocyclic azido group.

Compound **[9]** was saponified, then coevaporated with pyridine, and tritylated in the usual manner to give compound **[15]** as a solid white foam in 77% yield. This was then phosphitylated using the standard procedure, but the reaction failed as the N^6-azido group also reacted with the phosphitylating reagent. (This even occurred with the "milder" phosphitylating reagent: 2-cyanoethyl-bis-(N,N-diisopropyl) chlorophosphoramidite). In each case, on addition of the phosphitylating reagent a brown colour appeared, suggesting the degradation of the azido group. The reaction scheme is shown in Figure 2.7, and the likely product after phosphitylation is shown in Figure 2.8.



(i) a:2M NaOH in pyridine/methanol. b:DMTrCl and DMAP in pyridine.
(ii) 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite and DIPEA in THF.

Figure 2.7. *Attempted synthesis of the azido monomer* ***[16]***.



(i) 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite and DIPEA in THF.

Figure 2.8. *The putative azido phosphitylation product.*

A preliminary study into the introduction of the azido group after phosphitylation was attempted (results not shown) and seemed to offer a means to obtaining an N^6-azido-dG monomer.

2.3 N^2-DMF-PROTECTION OF dD VIA 3',5'-BIS-ACETYL-PROTECTION

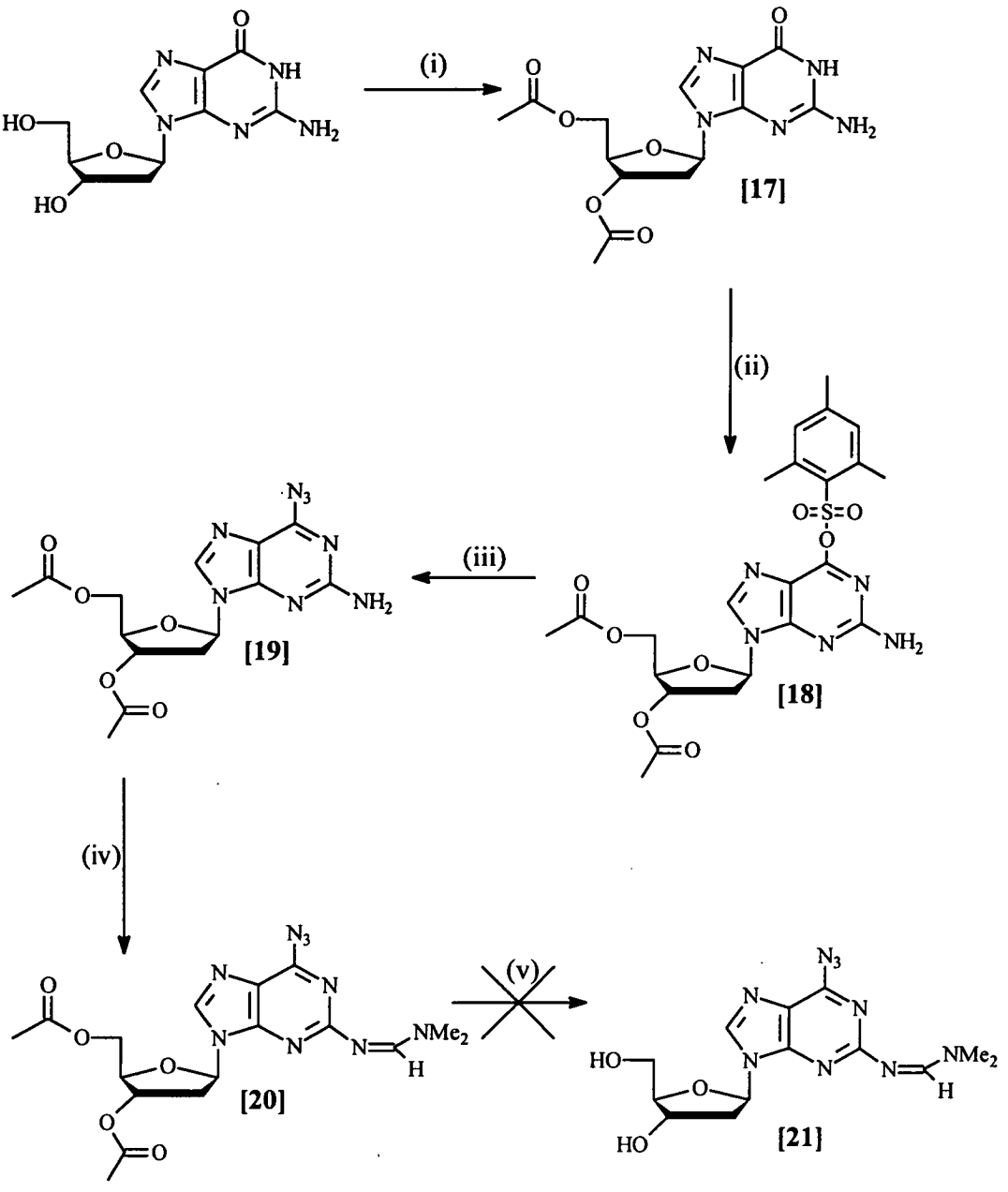
As mentioned in Section 1.8.4.1, the dmf group has been used to successfully replace the isobutyryl group in N^2-protection of deoxyguanosine in a fast-deprotecting monomer.[156] Therefore this group was the next choice for protection of the N^2-amino group of dD. This would be used in conjunction with the dbf group on the N^6-amino function, since the dmf group was found to be unsuitable as an N^6-amino protecting group.

2.3.1 Synthesis of N^2-dmf-N^6-azido-dD nucleoside [21] (Figure 2.9)

2'-Deoxyguanosine was suspended in pyridine and dimethylformamide and reacted with acetic anhydride. The resulting precipitate was filtered off , and ethanol was added to the supernatant, and this solution was concentrated *in vacuo* until precipitation commenced. Diethyl ether was added to encourage precipitation, and this white solid was recrystallised from ethanol and water to give compound **[17]** as a white crystalline solid in 92% yield.

Compound **[17]** was then reacted with MSCl in dichloromethane. However, the reaction mixture was seen to turn dark red. The dichloromethane was removed *in vacuo*, and the resulting dark residue was coevaporated with methanol. The solid obtained was washed with a solution of diethyl ether and methanol to leave compound **[18]** as a white powder in 58% yield.

Compound **[18]** was then reacted with N-methyl pyrrolidine and TBAZ in dichloromethane to give compound **[19]** as an off-white foam in 89% yield. This nucleoside was in turn reacted with dimethylformamide dimethylacetal to introduce the dmf group and give compound **[20]**, as an off-white foam in 68% yield.



(i) Acetic anhydride in pyridine/DMF.
(ii) Mesitylene sulphonyl chloride, TEA and DMAP in dichloromethane.
(iii) a:N-methyl pyrrolidine in DCM. b:TBAZ.
(iv) Dimethylformamide dimethylacetal in pyridine.
(v) 2M NaOH in pyridine/methanol.

Figure 2.9. *Attempted synthesis of the N^2-dmf, N^6-azido nucleoside **[21]**.*

However, in the next step, the dmf group was removed from **[20]**, along with the 3',5'-acetyl-groups. This cleavage is shown below in Figure 2.10. Whilst this was a set-back in this synthetic route, it did give a hint toward the higher base-lability of the dmf group compared to any previous N^2-protection. For the next approach, a base-stable sugar protecting group had to be employed in order to preserve the dmf group.

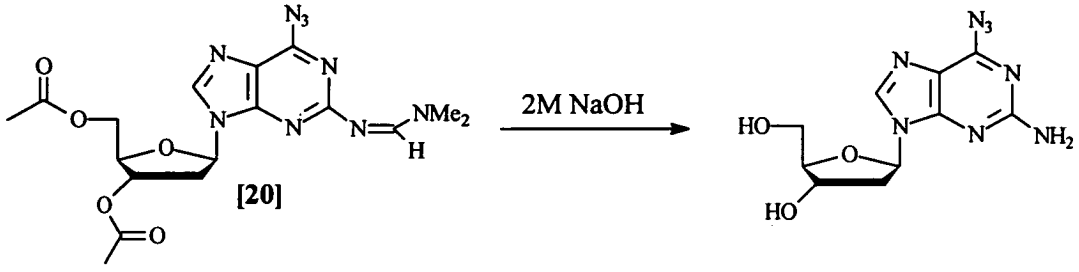


Figure 2.10. *Cleavage of the dimethylformamide group during ribose saponification.*

2.4 SYNTHESIS OF 3',5'-TIPS-NUCLEOSIDES

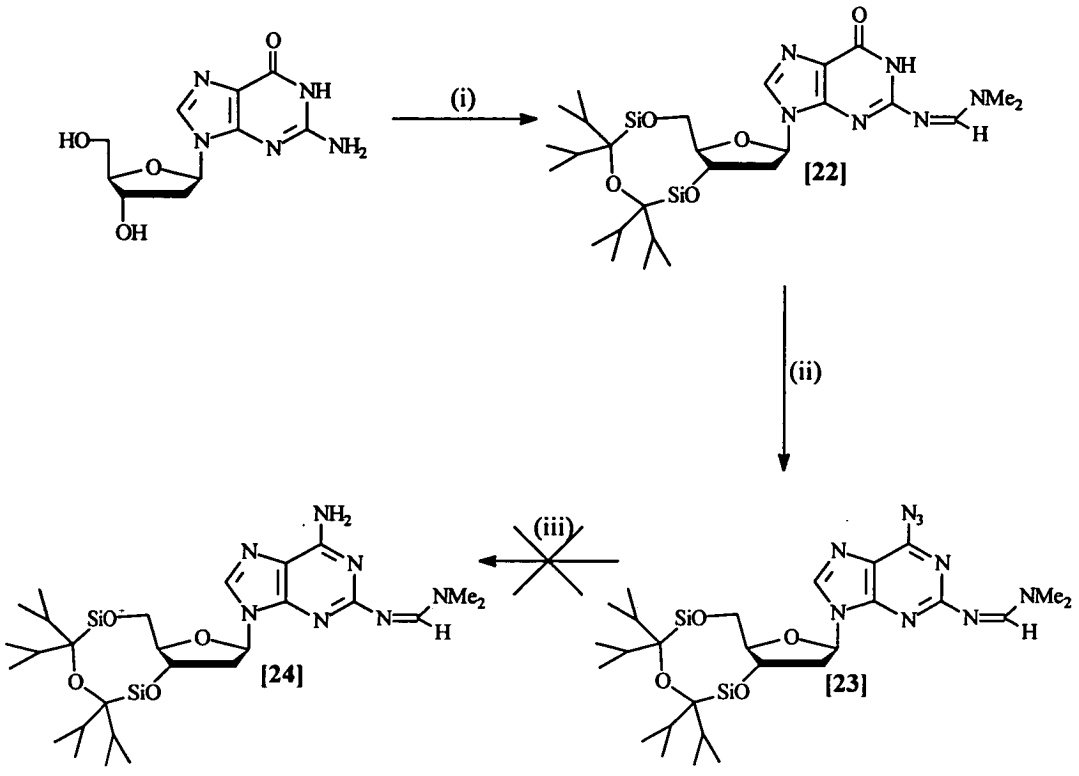
The 1,1,3,3-tetraisopropyl-1,3-disiloxanediyl group, (TIPS), was introduced by Markiewicz to protect the 3'- and 5'-positions in RNA nucleoside chemistry.[168] As it is cleaved by fluoride ions, it allows orthogonality of deprotection between the 3',5'-hydroxyl and the 2'-hydroxyl protection. TIPS should also be the ideal choice for selective deprotection of the 3',5'-O-protection in the presence of the N^2-dmf group in dD nucleosides, allowing non-basic deprotection of the sugar hydroxyl groups.

2.4.1 Synthesis of 3',5'-TIPS-N^2-dmf-dD [24] (Figure 2.11)

2'-dG was suspended in pyridine with DMAP and reacted with $TIPSCl_2$. After the reaction was complete (shown by tlc) dimethylformamide dimethylacetal was added to the cloudy-pale yellow reaction mixture. After completion, the solvents were removed *in vacuo*, the residue was then worked up and purified in the normal way, with the addition of 1% TEA added to the chromatography eluent to prevent degradation of the TIPS group. Compound **[22]** was obtained as an off-white solid in 85% yield.

The next stage involved the introduction of the azido function. Previous attempts at introducing the azido function in the presence of the basic N^2-dmf group gave very sluggish reactions. (Indeed, in Section 2.3.1, the N^6-azido group was introduced *before* the N^2-dmf group was introduced). The mesitylene sulphonyl chloride would only react with nucleoside **[22]** if *both* pyridine *and* TEA were present in the reaction mixture. Presumably the pyridine and TEA act in a synergic manner, and help more facile removal of the N^1-proton, allowing the MSCl to react at the O6-position more readily. It is also possible the basic mixture drives the keto-enol equilibrium more in favour of the enol tautomer.

Thus nucleoside **[22]** was suspended in pyridine along with DMAP and TEA. On addition of the MSCl, the reaction mixture was seen to turn purple briefly, then a dark red colour. After addition was complete, N-methylpyrrolidine and then TBAZ were added. The resulting dark-brown reaction mixture was filtered to remove precipitate, and the filtrate was removed *in vacuo*. The resulting residue was recrystallised from ethanol to leave compound **[23]** as white crystals in 60% yield.



(i) a:TIPSCl and DMAP in pyridine. b:Dimethylformamide dimethylacetal.
(ii) a:MSCl, TEA, and DMAP in pyridine. b:N-methyl pyrrolidine at 0°C. c:TBAZ.
(iii) Hydrogenation-*see* Table 2.1 for conditions.

Figure 2.11. *Synthesis of the TIPS-dimethylformamide nucleoside [24].*

The next stage of synthesis was the reduction of the azido group to leave the TIPS-N^2-dimethylformamide-dD nucleoside **[24]**. However, under the normal reducing conditions employed to convert the azide to the amine (catalytic hydrogenation using Pd on activated charcoal—*see* Section 2.2.1), the N^2-dimethylformamide group was removed almost quantitatively. This is shown in Figure 2.12 below.

As the dimethylformamide group was cleaved during catalytic hydrogenation, milder alternatives to the reduction of the azide were sought. Table 2.1 shows examples of these hydrogenating agents tested.

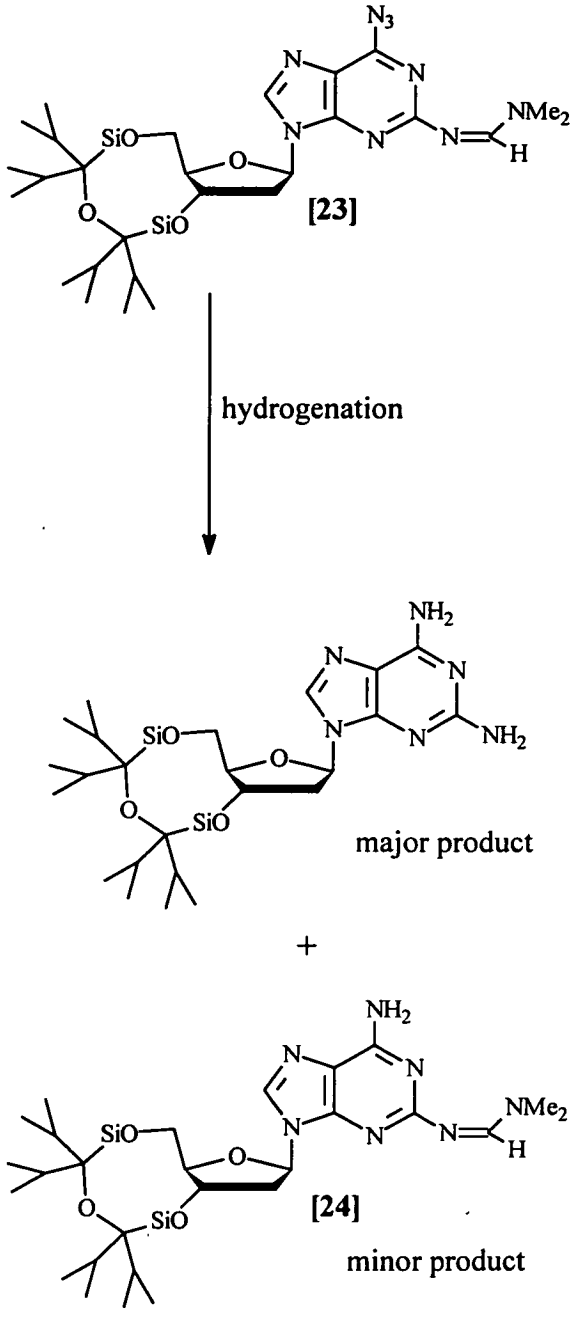


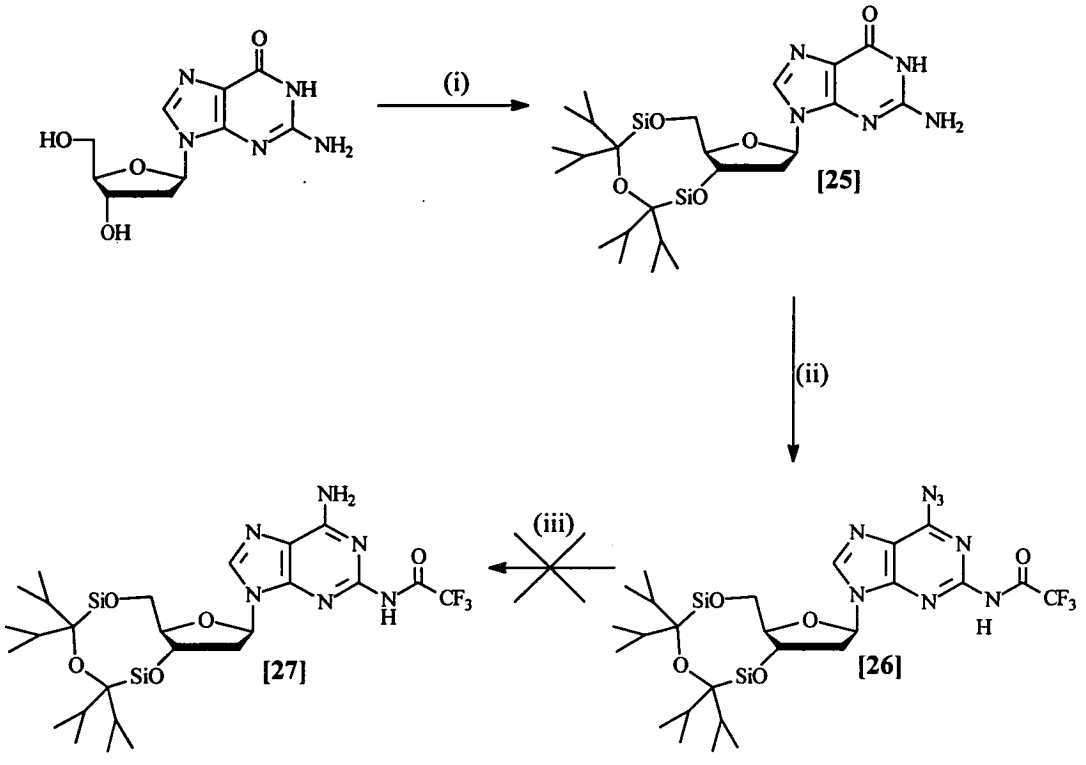
Figure 2.12. *Major and minor products formed during hydrogenation.*

2.4.2 Synthesis of 3',5'-TIPS-N^2-tfa-dD [27] (Figure 2.13)

At this point, also, dD nucleosides were developed where the N^2-protection consisted of the trifluoroacetyl (tfa) group. This group was employed at an earlier

stage, but was rejected along with the dmf group when both groups were found to be base labile during saponification of the acetyl esters (Section 2.3.1).

2'-Deoxyguanosine was suspended in pyridine along with DMAP, and reacted with $TIPSCl_2$. After concentration of the solvents, the residue was treated with water/methanol to leave a white powder on washing with methanol. This gave compound **[25]** in 89% yield.



(i) TIPSCl and DMAP in pyridine.
(ii) a:Trifluoroacetic anhydride in pyridine at 0°C. b:Sodium azide.
(iii) Hydrogenation *see* Table 2.1 for conditions.

Figure 2.13. *Synthesis of the TIPS-trifluoroacetyl nucleoside [27].*

The next stage involved the introduction of the tfa group for N^2-protection, and the azido group on the O6-position in the one pot. This was deemed possible as reaction of the lactam function of 2'dG with trifluoroacetic anhydride in pyridine results in the introduction of a pyridinium intermediate, which can then be displaced

by nucleophiles.[99,169] This novel method was applied, and a similar approach led to a new route to dD nucleosides (Section 2.5.1). Compound **[25]** was suspended in pyridine and trifluoroacetic anhydride was added to give a yellow solution. After the disappearance of the starting material to give a fluorescent spot on tlc under UV, sodium azide was added to give a purple solution. Excess sodium azide was filtered off and the reaction mixture was worked up and purified by flash-column chromatography. Compound **[26]** was obtained as a purple foam in 66% yield.

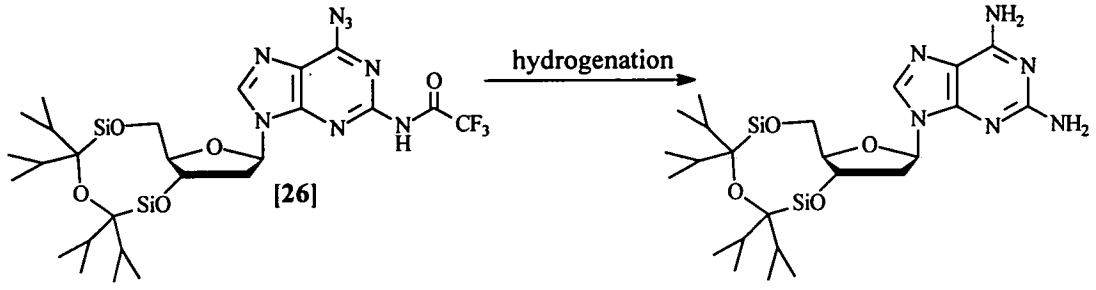


Figure 2.14. *Hydrogenation of the TIPS-trifluoroacetyl nucleoside.*

Compound **[26]** was subjected to reducing conditions as for the dmf analogue **[23]**, but again, the N^2-protecting group was cleaved, as shown in Figure 2.14. The extent of reduction was similar to that found for **[23]**—see Table 2.1 below.

Table 2.1. *The lability of dmf or tfa N^2-protection to hydrogenation conditions.*

	HYDROGENATION CONDITIONS			
N^2-PROTECTION	Pd/C, H_2	Raney Ni	$SnCl_2$	Ph_3P, H_2O
dmf or tfa	xx	xx	x	—

xx = almost quantitative cleavage; x = intermediate cleavage; — = no reduction.

An entirely different approach to synthesis of N^2-dmf-dD had to be made in order to avoid reduction of the N^2-protection. It was then decided that, initially the

synthesis of the free 2'-dD nucleoside would be carried out, followed by selective protection of the exocyclic amino groups.

2.5 SYNTHESIS OF THE N^2-DMF-N^6-DBF-dD MONOMER [32]

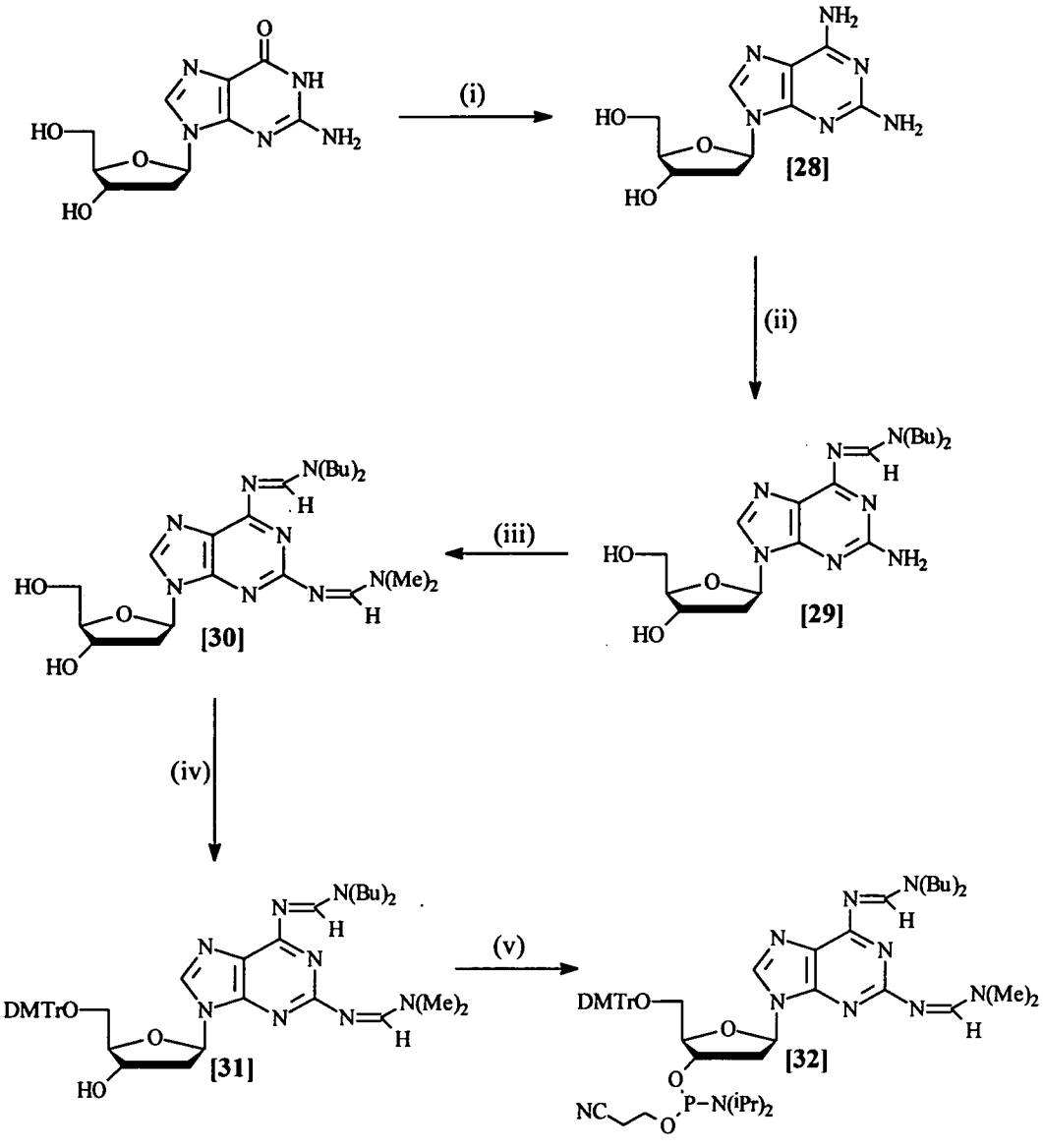
As the N^2-dmf group was labile under conditions that reduced the N^6-azido group, the next synthetic approach was to make the free-dD nucleoside, and enforce selectivity in protection of the N^2- and N^6-amino groups. This was necessary as bis-protection using N^2-di-N-butylformamide was unsuitable for oligonucleotide synthesis; and N^6-protection using the dimethylformamide gave ambiguous results. There was not enough time to further investigate the use of N^6-dmf protection, therefore the tried and tested di-N-butylformamide group was employed.

2.5.1 Synthesis of the free-diaminopurine nucleoside [28] (Figure 2.15)

The fact that Raney Ni in ethanol at 60°C cleaves the tfa group whilst reducing the azido group, (*see* Figure 2.14) was exploited in the development of a novel route to the synthesis of 2'dD. 2'-Deoxyguanosine was suspended in pyridine and reacted with trifluoroacetic anhydride to give a red solution. Sodium azide was then added and when reaction was complete, leaving a brown-coloured solution, the excess NaN_3 was filtered, and the filtrate was neutralised and worked up. The residue was then dissolved in ethanol and Raney Ni was added. This reaction was then gently refluxed at 60°C, then the catalyst was filtered off, and the filtrate was concentrated *in vacuo*, diethyl ether was added to the residue, and the mixture was left at 0°C to recrystallise. The resulting white crystals of compound **[28]** were obtained in 81% yield.

2.5.2 Synthesis of dD monomer [32] (Figure 2.15)

The next stage involved the selective protection of the N^6-amino group with dbf in the presence of the less nucleophilic N^2-amino group. This reaction proceeds to give N^6-dbf-dD, but care must be taken to avoid making the bis-dbf derivative.



(i) a:Trifluoroacetic anhydride in pyridine at 0°C. b:Sodium azide. c:Raney Ni in ethanol at 60°C.
(ii) Di-N-butylformamide dimethylacetal in pyridine.
(iii) Dimethylformamide dimethylacetal in pyridine at 50°C.
(iv) DMTrCl and DMAP in pyridine.
(v) 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite and DIPEA in THF.

Figure 2.15. *Synthesis of the N^2-dmf,N^6-dbf-diaminopurine monomer [32].*

Compound **[28]** was suspended in pyridine, and di-N-butylformamide dimethylacetal was then added to give compound **[29]** as a light-yellow foam in 54% yield after flash-column purification. Compound **[29]** was then suspended in pyridine and dimethylformamide dimethylacetal was added, and the reaction mixture gently

heated at 50°C. Column chromatography gave compound **[30]** in 62% yield as a light-yellow foam.

Compound **[30]** was tritylated then phosphitylated under normal conditions to give compounds: **[31]** in 66% yield, and **[32]** in 73% yield, respectively.

2.5.3 Stability of nucleosides to TCA

The ease at which protected dD nucleosides depurinate during the TCA-detritylation stage of oligonucleotide synthesis has hampered the development of dD monomers. As discussed in Section 1.6.2, depurination is particularly detrimental in oligonucleotides containing N^6-acyl protected dD. In order to assess the effectiveness of N^6-dbf protection, stability of this nucleoside was compared with several other related nucleosides in 3% TCA/DCM, and monitored with time. Decomposition of the nucleosides was followed by TLC, and the "native" nucleosides dA and dD were included as references. The results are shown below.

Table 2.2. *Stability of nucleosides to TCA.*

nucleoside	approximate half life
dA	45min
dD	30min
N^2-ib,N^6-bz-dD	<<5min[a]
N^2-ac,N^6-dmf-dD	<20min[b]
N^2-dmf,N^6-dbf-dD	>1.5hr

[a]Result taken from Booth.[n50]

[b]Non-depurination degradation observed, along with depurination.

As expected,[69] the N^6-benzoyl protection has a severe effect on the susceptibility of the dD nucleoside to depurination. Comparing this with studies carried out on the depurination of dA, the benzoyl protection may alter the N^1 and N^7 basicities relative to one another to favour the N^7-protonation. This species then acts as a good leaving

group according to the monocation depurination pathway. Such a low stability to TCA should preclude viable oligonucleotide synthesis.

On the other hand, dbf protection offers a much increased half life. This agrees with the studies carried out on N^6-dmf-dA,[122] where the formamide protection was found to be electron-donating, and did not to promote N^7-protonation to the same extent that acyl protection did. Here, the higher basicity of the N^1 position will favour protonation versus the less basic N^7.

The value of the half life for nucleoside **[32]** agrees well with a similar study carried out by Booth, where N^2-protection was isobutyryl or dbf, and N^6-protection was dbf.[133] He concluded that the choice of N^2-protecting group had no influence on depurination. From this position, there is no resonance stabilisation of the N^1 monocation, which is stabilised by delocalisation of the formamide moiety in the 6-position. Booth found similar half lives for the N^2-ib- and N^2-dbf-protected nucleosides.

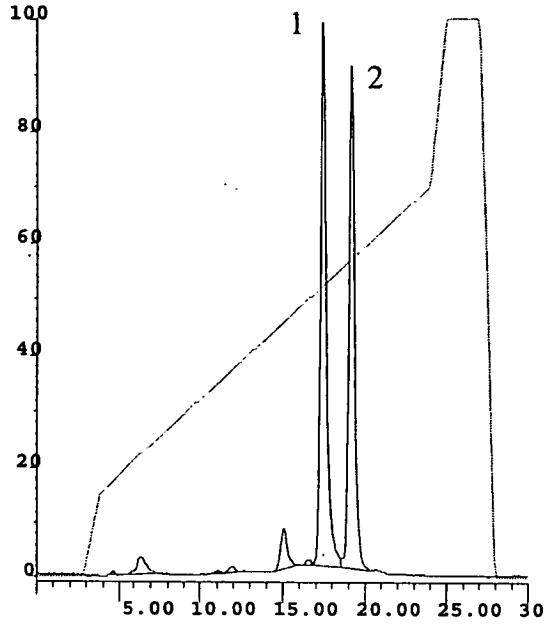
It is interesting to note that the N^6-dmf group was not as stable to the TCA conditions as di-N-butylformamide. Degradation of the nucleoside occurred, but not only of the depurination variety—the N^6-dmf group appeared to be slightly labile in TCA. This may help explain the failure of monomer **[14]** during machine synthesis—*see* Section 2.2.3. Again, similar behaviour to dA (N^6-dmf-dA gave ambiguous results during synthesis[165]) is observed for dD nucleosides. However, it was not clear if this was the cause of synthesis failure, and further study must be carried out to clarify this ambiguity.

2.5.4 Oligonucleotide synthesis using dD monomer [32]

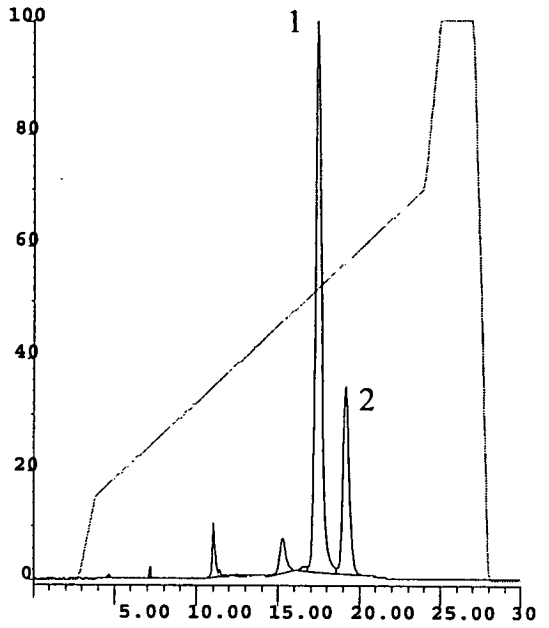
Monomer **[32]** was dissolved in anhydrous acetonitrile, and used as a 0.12M solution during oligonucleotide synthesis.

Coupling efficiency of monomer [32]

The coupling efficiency of oligonucleotide synthesis using monomer **[32]** was found to be routinely >98%. Therefore, this monomer fulfilled the initial criteron of use under standard synthesis conditions.

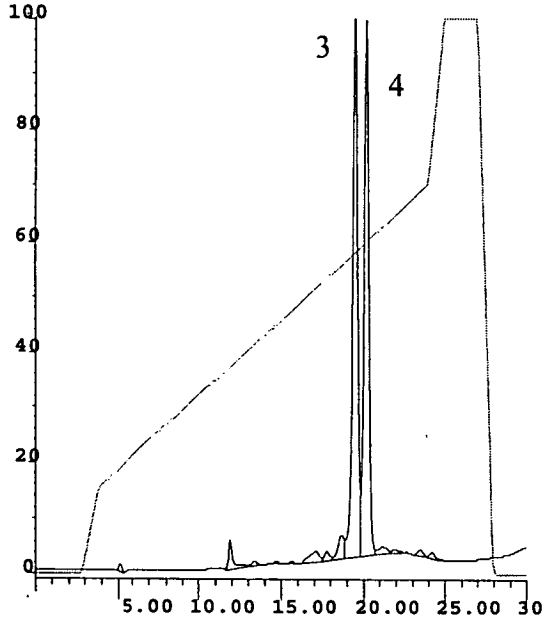


(a) **Isobutyryl** at 55°C for **40 hours**.

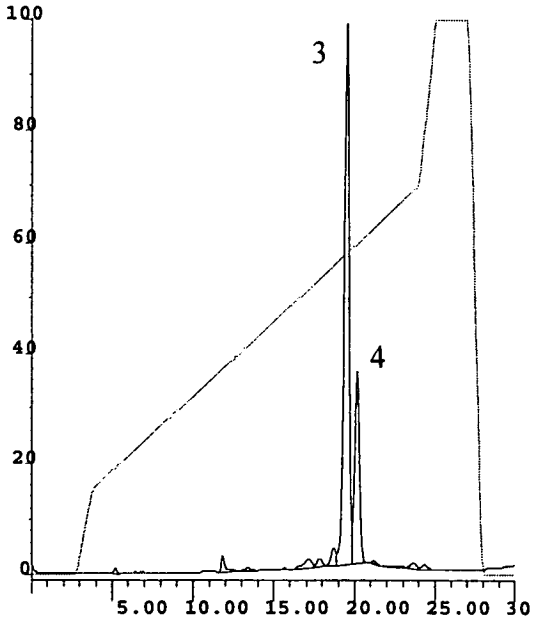


(b) **Isobutyryl** at 55°C for **90 hours**.

Chromatograms show (1) fully deprotected, and (2) isobutyryl protected oligonucleotides



(c) **Dmf** at 55°C for **2 hours**.



(d) **Dmf** at 55°C for **5 hours**.

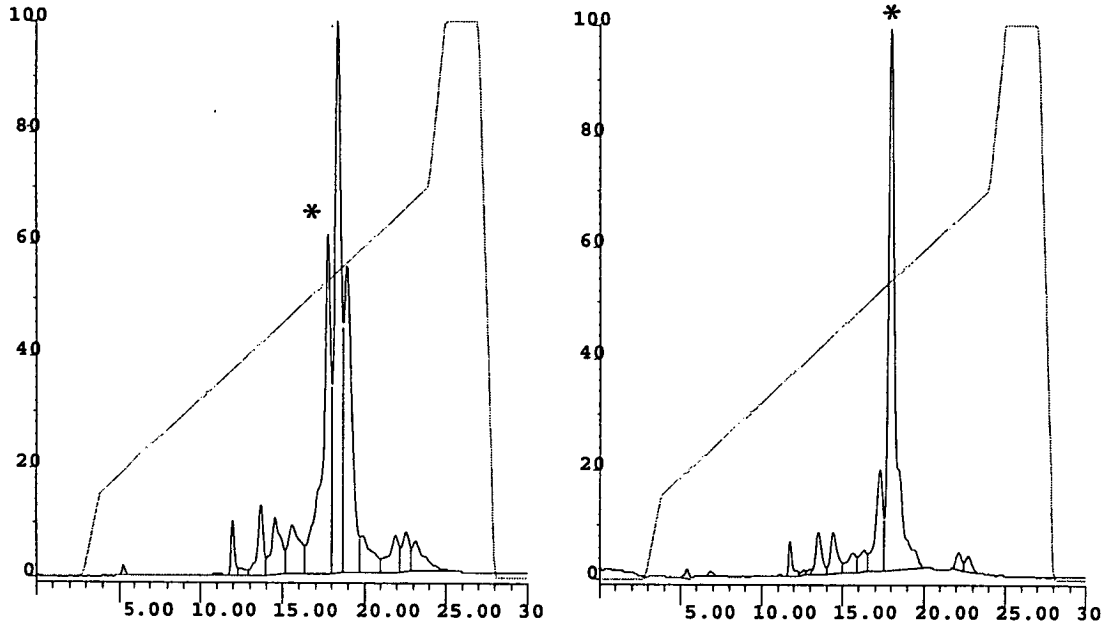
Chromatograms show (3) fully deprotected, and (4) dmf protected oligonucleotides.

Figure 2.16. *Contrasting the deprotection of the isobutyryl and dmf groups.*
These deprotections are for the oligonucleotide: TTT TTD TT.

Deprotection of the N^2-dimethylformamide group

In addition to an acceptible coupling efficiency, the N^2-dmf-N^6-dbf-dD monomer **[32]** deprotected faster than the previous N^2-protecting candidates, whilst allowing resolution of the deprotected and protected oligonucleotides. Figure 2.16 shows hplc traces for the deprotection of the test oligonucleotides containing 1-dD residue, contrasting the dimethylformamide group with the isobutyryl group.

Again, the test sequence TDD TTD T was synthesised, and for monomer **[32]** the deprotection is shown below in Figure 2.17.



(a) TDD TDT T at 55°C for 5 hours. (b) TDD TDT T at 55°C for 30 hours.

Figure 2.17. *Deprotection of the dmf test oligonucleotide in conc.* NH_3.
The asterisk marks the fully deprotected oligonucleotide in each case.

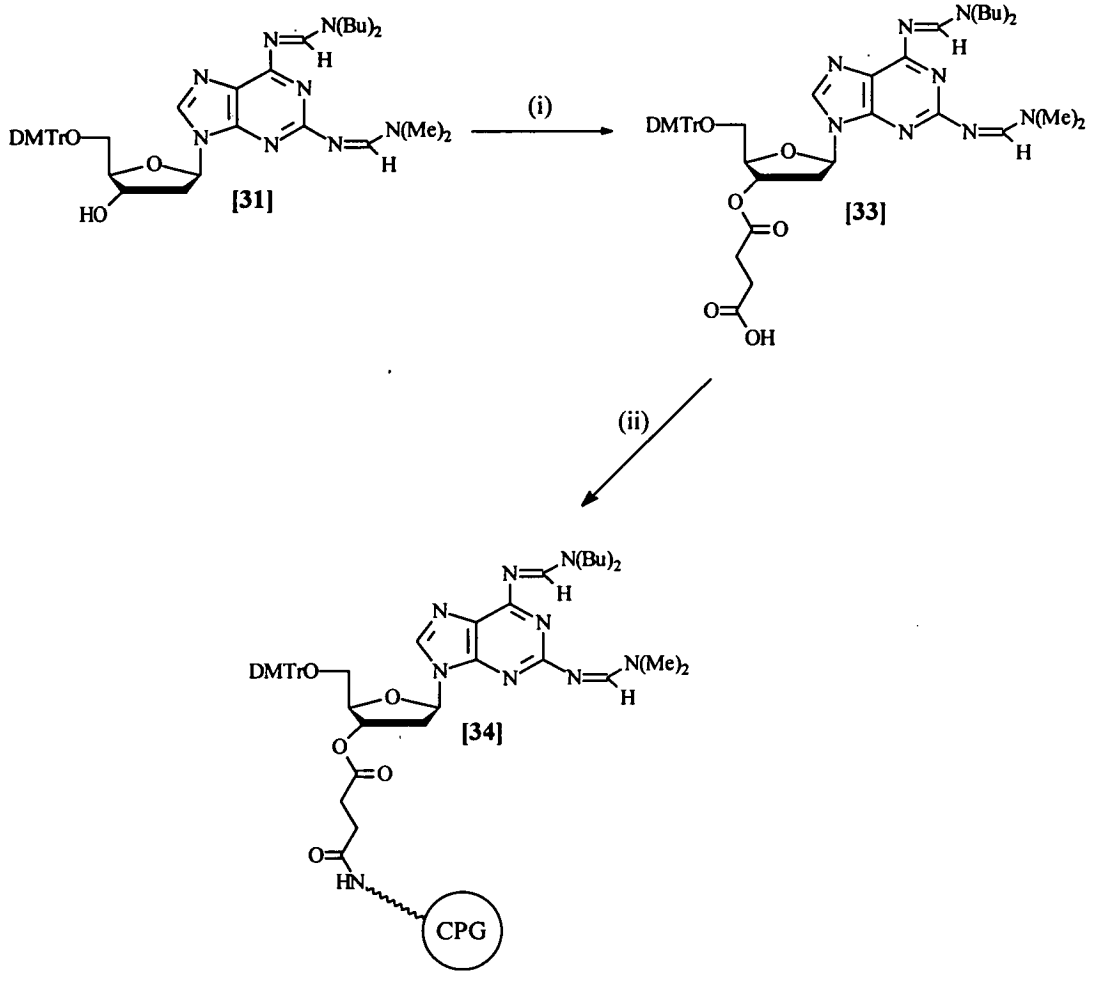
It is interesting to note that the "next best" labile protecting group was the phenylacetyl group, which showed comparable deprotection after the same time, but at the elevated temperature of 65°C, (*see* Figure 2.3(b)). The dmf oligonucleotide was fully deprotected at 65°C in ca. 18 hours, (results not shown).

Enzyme digestions of oligonucleotides containing monomer [32]

Enzyme digestions of oligonucleotides containing dD were carried out, as with previous monomers, in order to authenticate the incorporation of the modified base. Examples are given in Figure 2.19.

2.5.5 Synthesis of dD-succinyl CPG [34] (Figure 2.18)

As the N^2-dmf protection was deemed suitable for incorporation of dD into routine oligonucleotide synthesis, it was decided to make the dD-functionalised CPG.



(i) Succinic anhydride and DMAP in pyridine.
(ii) a:4-Nitrophenol and DCC in DCM. b:Activated CPG in DMF.

***Figure 2.18**. Synthesis of the diaminopurine-functionalised resin.*

This was carried out in order to incorporate dD residues at the 3'-terminus of oligonucleotides.

Compound **[31]** was dissolved in pyridine with DMAP, and reacted with succinic anhydride. The residue left after concentration was worked up with aqueous citric acid, and purified by column chromatography pre-equilibrated with TEA, and eluted with MeOH/DCM/TEA. Succinyl derivative **[33]** was obtained as a yellow-white foam in 77% yield.

The attachment of nucleside **[33]** to the solid support was carried out as described by Sproat and Gait,[170] to give CPG **[34]** with a loading after capping, of 35.0μmolg^{-1}.

2.5.6 Oligonucleotide synthesis using LCAA-CPG [34]

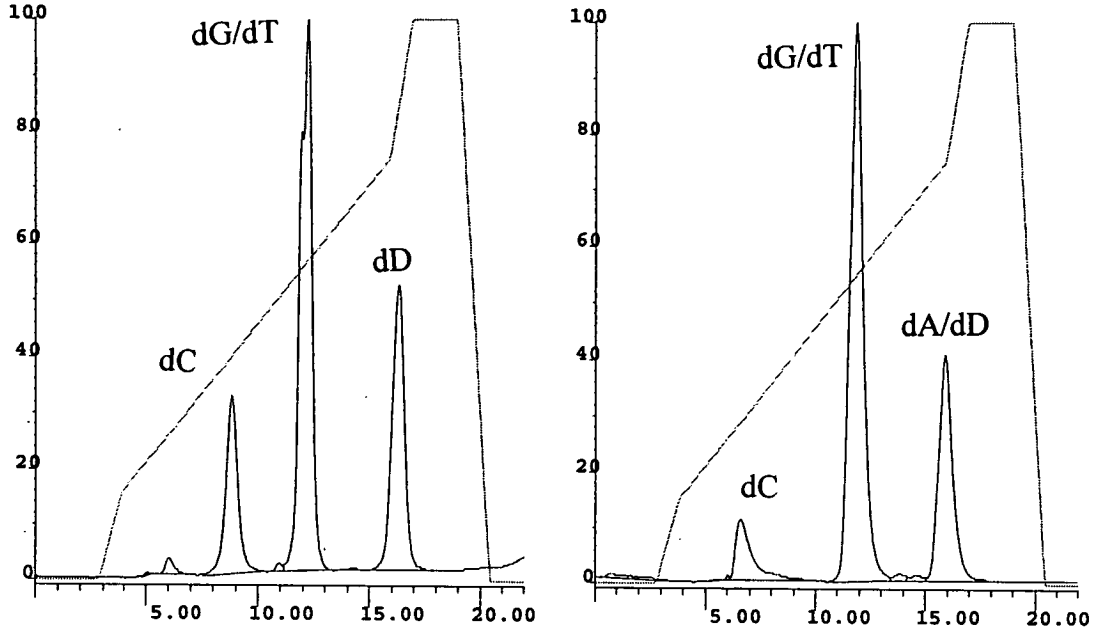
Appropriate amounts of the CPG were weighed out and packed in DNA synthesis columns, and used as for normal oligonucleotide synthesis.

Coupling efficiency and deprotection of CPG [34]

As with the dD monomer, oligonucleotides synthesised using dD-CPG **[34]** were found to exhibit the desirable coupling efficiency of >98%. The deprotection of the oligonucleotides was also similar to those synthesised in Section 2.5.3.

Enzyme digestions of oligonucleotides containing [34] (Figure 2.19)

As with monomer **[32]**, the oligonucleotides containing the 3'-terminal-dD residue were authenticated using enzyme digestion, and the HPLC chromatograms for the digests compared with that of the genuine 2'-dD. Figure 2.19 shows examples of enzyme digests of oligonucleotides made with dmf monomer **[32]** and dmf-CPG **[34]**.



(a) CGT DDC GT (b) GCA TTG AD

Figure 2.19. *Enzyme digests of diaminopurine-containing oligonucleotides.* The order of elution of the nucleosides is shown.

2.6 CONCLUSIONS AND FUTURE WORK

After synthesising several dD monomers containing different N^2-protecting groups, the best candidate to emerge was the dimethylformamide group. This fulfilled the requirement of coupling >98% during oligonucleotide synthesis, as well as being stable to all reagents during nucleoside and oligonucleotide synthesis. In addition, the deprotection time of this group in conc. NH_3 was a vast improvement over the isobutyryl group previously used. A comparison of the "old" and "new" protection groups is given in Table 2.2.

This monomer was ultimately synthesised via the free dD nucleoside, and a novel technique was devised for this approach, exploiting the simultaneous reduction of the N^6-azido intermediate, and the labile N^2-trifluoroacetyl group. In addition, the difference in nucleophilicity/basicity of the N^6- and N^2-amino moieties was taken advantage of during their selective protection with the dbf and dmf groups respectively.

Table 2.3. *A comparison of N^2-protecting groups used for dD.*
Deprotection is compared in conc. ammonia at 55°C.

MONOMER	Coupling Efficiency	approx. half-life
isobutyryl **[35]**	>98%	35 hours
phenylacetyl **[7]**	<95%	8 hours
acetyl **[12]**	>98%	30 hours
dmf **[32]**	>98%	2 hours

Recent work carried out by Gryaznov and Schultz[78] employed synthesis of the bis-protected nucleoside via the 5'-DMTr-dD. The protecting group of choice was the phenoxyacetyl group which, like the dmf group, is an established "fast-deprotecting" group (*see* Section 1.8.4.1). However, on the N^6-position this group will not have the same depurination suppression as the dbf group. Also, when this monomer was used in DNA synthesis, the diluent used was dichloromethane instead of the usual acetonitrile. This was compounded with the use of phenoxyacetic anhydride instead of acetic anhydride during the capping step.

Future work could involve using TIPS-protected dG **[25]**, instead of dG. This would enhance the solubility of the resulting dD nucleosides, and assist their subsequent purification. The N^6-dbf derivative of this would then provide an intermediate for the introduction of other candidates for the introduction of more base labile protection at the N^2-moiety. Such candidates would include the trifluoroacetyl group, and the phenoxyacetyl group. Owing to the inconclusive evidence, the bis-dimethylformamide-protected diaminopurine phosphoramidite should also be synthesised and evaluated.

Another approach to speeding up the deprotection of diaminopurine residues would be to use of deprotecting reagents more nucleophilic than conc. NH_3. Examples of these reagents were discussed in Section 1.8.

3.0 THE THERMODYNAMICS OF DNA CONTAINING dD

The affinity of an oligonucleotide for its complement can be assessed using spectrophotometric techniques. Typically, an oligonucleotide and its complement are mixed at 1:1 stoichiometry in an appropriate hybridisation buffer, and the absorbance at 260nm is measured as a function of temperature. As the temperature is raised, the duplex converts to single strands with a concomitant increase in absorbance (hyperchromicity).

The temperature at which half the molecules are duplex and half are single stranded is the T_m, or melting temperature. Analysis of the shape of the melting curve can also yield thermodynamic parameters for the coil to helix transition, provided the transition is two-state.[172] The manipulation of the data derived from melting curves will be discussed in Section 3.2 below.

3.1 THERMODYNAMIC COMPARISON OF $d(TA)_4$ AND $d(TD)_4$

Since duplexes containing the sequence: $d(TD)_n$ have been studied in less detail than their $d(TA)_n$ counterparts, melting curves were carried out comparing $d(TA)_4$ and $d(TD)_4$. These oligonucleotides were chosen as Booth had previously determined that the oligonucleotide $d(TD)_3$ readily formed a duplex, whereas the native hexamer $d(TA)_3$ failed to show any temperature dependence in its UV absorbance, indicating that it was not present in the duplex form.[133]

3.1.1 UV melting curves of $d(TA)_4$ and $d(TD)_4$

The UV absorbance, A_b, at 260nm was recorded at a heating rate of 0.9°C per minute for an appropriate length of time for both oligonucleotides. Melting curves for both oligonucleotides were determined at a concentration, C_T, of 40mM, using Eq. (1).

$$A_b = C_T.l.\varepsilon.F \qquad (1)$$

where l is the cell path length (1cm), ε is the extinction coefficient of the oligonucleotides, calculated using the values for dA and dT given by Sproat and Gait;[173] and the determined value of ε(dD) of $7.4x10^3 cm^2 \mu mol^{-1}$. The hyperchromicity factor, F, was taken to be 0.8, from measurements of strand and nucleoside absorbance at 260nm, before, and after enzyme digestion.

The first-order derivative of the UV absorbance measured as a function of time (temperature) gave the T_m at the maximum. The melting curves for $d(TA)_4$, **DAP(3)** and $d(TD)_4$, **DAP(4)** along with T_m's are given in Figure 3.1.

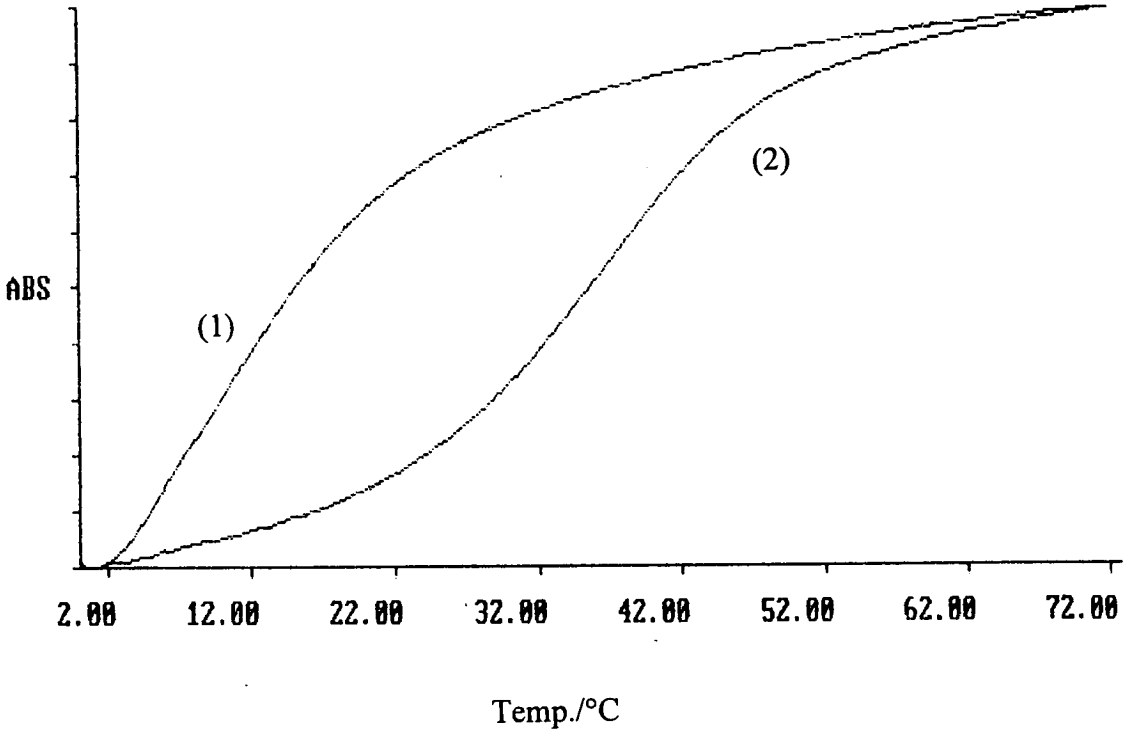


Figure 3.1. *Melting curves for (1) $d(TA)_4$ and (2) $d(TD)_4$ in 1M Na^+, 10mM phosphate buffer at pH 7.* T_m $d(TA)_4$ = 12°C and T_m $d(TD)_4$ = 37°C.

The dramatic difference in T_m (25°C) reflects the extra stability given by the D.T base pair, compared to the A.T base pair. Whilst this example serves to illustrate the difference in stability between A.T and D.T base pairs, this does not represent the full picture. As discussed in Section 1.5, poly d(DT) has a tendency to adopt the unusual putative A-form of DNA. This structure may offer extra stability for poly d(DT) over that of the native B-form, which may be too conformationally flexible. Only more detailed thermodynamic and comparison will help explain these differences.

Also, dD *destabilises* the DNA duplex when flanked by adenines, (disrupting the spine of hydration—*see* Section 1.5.2), and so the substitution of dD for dA clearly does not always increase the stability of the duplex. Again, obtaining full thermodynamic parameters for dD may allow prediction of the stability of duplexes containing dD.

3.2 THERMODYNAMIC PARAMETERS OF DUPLEXES CONTAINING dD

As outlined in Section 1.2.2, the stability of a DNA duplex depends on its base sequence. More specifically, the stability of a DNA duplex appears to depend primarily on the identity of the *nearest-neighbour* bases. The nearest-neighbours approach takes into account the "horizontal" and "vertical" interactions, in addition to other sequence-specific effects.

Ten different nearest-neighbour interactions are possible in any Watson-Crick DNA duplex structure. These pairwise interactions are: AA/TT; AT/TA; TA/AT; CA/GT; GT/CA; CT/GA; GA/CT; CG/GC; GC/CG; GG/CC. Results obtained, by Breslauer *et al*, on model systems, provided the thermodynamics of all ten nearest-neighbour interactions.[174] This data now provides an empirical basis for predicting the stability ($\Delta G°$) and the temperature-dependent behaviour ($\Delta H°$) of any DNA duplex region by inspection of its primary sequence.

The aim of this Section was to compile a similar library of nearest-neighbour interactions for dD. This data could then be used in conjunction with that obtained by Breslauer *et al*, in order to predict the thermodynamic parameters of duplexes containing dD.

3.2.1 Calculating thermodynamic parameters from melting curves

The formation or dissociation of complexes of molecularity greater than one will result in a concentration-dependent equilibrium. Such equilibria therefore can be characterised by determining the concentration dependence of the melting temperature.[175]

For a non-self complementary system at equilibrium

$$T_m = \Delta H°/\{\Delta S° + R\ln(C_T/4)\} \quad (2)$$

where $\Delta H°$ and $\Delta S°$ are, respectively, the enthalpic and entropy of duplex formation, C_T is the total oligonucleotide concentration, and R is the gas constant. Rearrangement of Eq. (2) gives

$$1/T_m = R/\Delta H° \ln(C_T/4) + \Delta S°/\Delta H° \quad (3)$$

$$y \qquad m \qquad x \qquad c$$

which, as indicated by the symbols in italics, corresponds to a straight line when the reciprocal of the melting temperature ($1/T_m$) is plotted against the natural logarithm of the strand concentration divided by four ($C_T/4$). The slope (m) of such a plot is equal to $R/\Delta H°$, and the intercept (c) is equal to $\Delta S°/\Delta H°$. Figure 3.3 shows a typical $1/T_m$ vs. $\ln(C_T/4)$ plot.

3.2.2 Obtaining the data

As seen in Section 3.1, absorbance vs. temperature profiles were measured for $d(TA)_4$, and $d(TD)_4$ and the T_m was obtained from the first derivative of the curve at a given concentration. In order to obtain concentration-dependent data, the melting curves for duplexes formed for oligonucleotides **DAP(5)** to **DAP(22)**, were obtained in triplicate for five concentrations covering a range of approximately 10-fold concentration. The other main difference in the following study was that none of the oligonucleotides were self-complementary—this avoided any possible intrastrand

D.T interactions as well as the added complication of having dD in both strands. (Incidentally, the latter event is unlikely in any application of dD in biology).

The oligonucleotides: **DAP(5)** to **DAP(22)** were carefully chosen so as not to contain any obvious self-complementary regions, and the 8-mer size was convenient for melting curve determination in the conditions outlined in Breslauer et al.[174] Adherence to these conditions was essential, since the values for the normal Watson-Crick nearest neighbour interactions were to be employed in this study. As a means of checking this compatibility, concentration-dependent melting curves, and the resulting thermodynamic parameters were obtained for the non-self complementary duplex: **Duplex 10**, which contains only A.T and G.C base pairs.

The oligonucleotides studied in the concentration-dependent melting curves were as follows.

DAP(5)	CGT ADC GT	**DAP(16)**	ACG TTA CG	**(Duplex 1)**
DAP(10)	CGT DAC GT	**DAP(16)**	ACG TTA CG	**(Duplex 2)**
DAP(7)	CGT DDC GT	**DAP(16)**	ACG TTA CG	**(Duplex 3)**
DAP(6)	CGT CDC GT	**DAP(17)**	ACG TGA CG	**(Duplex 4)**
DAP(9)	CGT TDC GT	**DAP(19)**	ACG TAA CG	**(Duplex 5)**
DAP(8)	CGT GDC GT	**DAP(18)**	ACG TCA CG	**(Duplex 6)**
DAP(11)	CGT DGC GT	**DAP(20)**	ACG CTA CG	**(Duplex 7)**
DAP(12)	CGT DTC GT	**DAP(21)**	ACG ATA CG	**(Duplex 8)**
DAP(14)	DCD CDC DC	**DAP(22)**	GTG TGT GT	**(Duplex 9)**
DAP(15)	CGT AAC GT	**DAP(16)**	ACG TTA CG	**(Duplex 10)**
DAP(13)	DCG TTA CG	**DAP(15)**	CGT AAC GT	**(Duplex 11)**

An example of treatment of this data is given for duplex formed between **DAP(5)** and **DAP(16)**, (**Duplex 1**).

DAP(5)	5'- CGT AD CGT	$\varepsilon = 78.5 \times 10^3\ cm^2 \mu mol^{-1}$
DAP(16)	3'- GCA TT GCA	$\varepsilon = 86.4 \times 10^3\ cm^2 \mu mol^{-1}$

Using Equation (1), and the average of the ε values, we obtain the concentration, at a given temperature, which is the average of three readings.

Figure 3.2 gives an example of one of the melting curves obtained for **Duplex 1**, and Table 3.1 shows the data obtained.

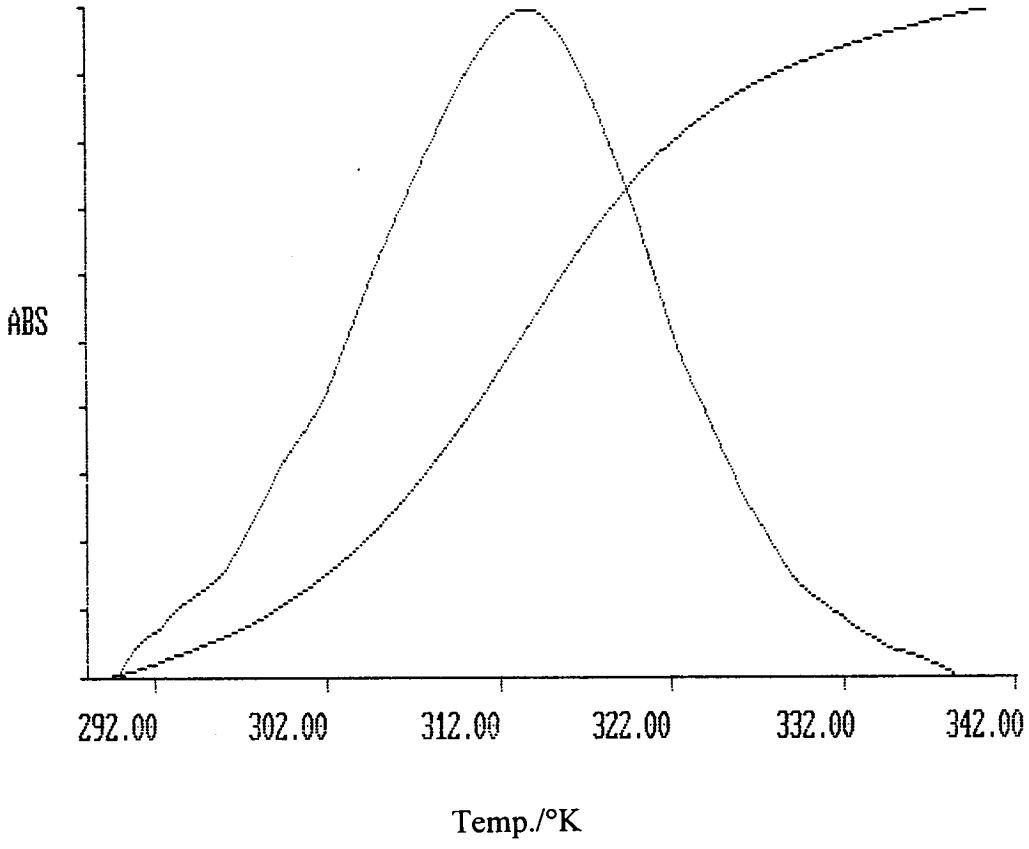


***Figure 3.2**. The UV melting curve for **Duplex 1** at 23.8μmol DNA concentration.*

Table 3.1. *Empirical data obtained from melting curves of* ***Duplex 1.***

$C_T/\mu mol^{-1}$	T_m(average)/°K	$\ln(C_T/4)$	$10^3/T_m$
62.6	318.75	-9.61	3.137
38.7	316.12	-10.10	3.163
23.8	313.73	-10.58	3.187
19.4	312.75	-10.79	3.197
7.8	308.40	-11.70	3.243

The concentration-dependence of the melting temperature is plotted as $\ln(C_T/4)$ vs. $1/T_m$, (Figure 3.3). The errors for the thermodynamic parameters are estimated from the standard error obtained for the slope and the intercept, based on the linear regression plot.

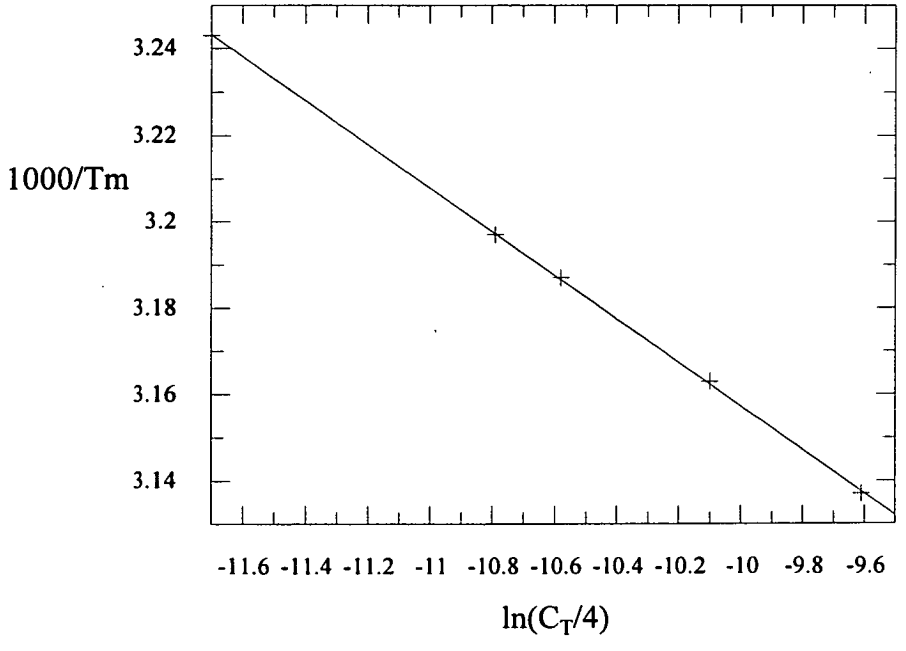


Figure 3.3. *The plot of* $\ln(C_T/4)$ *vs.* $1/T_m$ *for* ***Duplex 1.***

From a "least-squares" linear regression, the slope and intercept obtained for **Duplex 1** were

$$\text{Slope} = -0.0505 \pm 0.0005 \times 10^{-3}$$
$$\text{Intercept} = 2.5826 \pm 0.0058 \times 10^{-3}$$

and from Eq.(3),

$$\Delta H° = R/\text{slope} = -164.6 \pm 1.7 \text{ kJmol}^{-1}$$
$$\Delta S° = \Delta H°.\text{intercept} = -425.2 \pm 5.3 \text{ JK}^{-1}\text{mol}^{-1}$$

and from

$$\Delta G° = \Delta H° - T\Delta S° \qquad (4)$$

where T will be taken as 298.15K (25°C), we obtain

$$\Delta G° = -37.8 \pm 4.3 \text{ kJmol}^{-1}$$

All values will be given rounded up to three significant figures, hence:

$$\Delta H° = -165 \pm 2 \text{ kJmol}^{-1}, \Delta S° = -425 \pm 6 \text{ JK}^{-1}\text{mol}^{-1}, \text{ and } \Delta G° = -37.8 \pm 4 \text{ kJmol}^{-1}$$

The data obtained from the plots of $\ln(C_T/4)$ vs. $1/T_m$ for the rest of the duplexes were treated likewise, and are listed below in Table 3.2. (The errors obtained for the $\Delta G°$ values were, on average, ±10% for all duplexes studied.)

Table 3.2. Thermodynamic parameters for ***Duplexes 1-10.***
(ΔG° calculated at 25°C).

DUPLEX	ΔH°/ kJmol^{-1}	ΔS°/ JK^{-1}mol^{-1}	ΔG°/ kJmol^{-1}
1 CGT ADC GT GCA TTG CA	-165	-425	-37.8
2 CGT DAC GT GCA TTG CA	-193	-510	-40.6
3 CGT DDC GT GCA TTG CA	-190	-499	-41.0
4 CGT CDC GT GCA GTG CA	-205	-534	-46.4
5 CGT TDC GT GCA ATG CA	-201	-535	-41.0
6 CGT GDC GT GCA CTG CA	-208	-546	-46.0
7 CGT DGC GT GCA TCG CA	-242	-648	-48.5
8 CGT DTC GT GCA TAG CA	-205	-549	-41.0
9 DCD CDC DC TGT TGT TG	-237	-632	-48.6
10 CGT AAC GT GCA TTG CA	-172	-453	-37.2

Duplex 10 was included as a check on compatibility of the data obtained with that obtained by Breslauer *et al* for the 10 native nearest neighbour interactions.[174] The value for ΔG°(predicted) for **Duplex 10** agreed very closely with the experimentally obtained value. (*See* Appendix 1).

Duplex 11 (containing one 5'-terminal diaminopurine) was excluded from treatment of data since the correlation coefficient was below 0.99—all other duplexes had a value of greater than 0.99. It was decided that **Duplex 9** would be more "representative" of the DC/TG interaction, as the value obtained would be an average of four nearest neighbour values, not just relying on one 5'-terminal DC/TG

interaction, which may behave differently from internal DC/TG interactions. Ideally, thermodynamic parameters would have been determined for a 3'-terminal diaminopurine residue to help obtain an average of values. However, shortly after the determination of the 5'-terminal diaminopurine, the Peltier block malfunctioned, preventing further study of any duplexes.

3.2.3 Treatment of the data

The compositions of **Duplexes 1** to **10** for the nine possible nearest neighbour interactions for duplexes containing diaminopurine, as well as those for the ten native interactions, are given below in Table 3.3.

Table 3.3. *Nearest-neighbour frequencies.*

	Nearest neighbours present in duplex															
Duplex	*AD TT*	*DA TT*	*DD TT*	*CD GT*	*DC TG*	*GD CT*	*DG TC*	*TD AT*	*DT TA*	*AA TT*	*TA AT*	*CA GT*	*GT CA*	*GA CT*	*CG GC*	*GC CG*
1 CGT ADC GT GCA TTG CA	**1**	0	0	0	**1**	0	0	0	0	0	1	0	2	0	2	0
2 CGT DAC GT GCA TTG CA	0	**1**	0	0	0	0	0	**1**	0	0	0	0	3	0	2	0
3 CGT DDC GT GCA TTG CA	0	0	**1**	0	**1**	0	0	**1**	0	0	0	0	2	0	2	0
4 CGT CDC GT GCA GTG CA	0	0	0	**1**	**1**	0	0	0	0	0	0	0	2	1	2	0
5 CGT TDC GT GCA ATG CA	0	0	0	0	**1**	0	0	**1**	0	1	0	0	2	0	2	0
6 CGT GDC GT GCA CTG CA	0	0	0	0	**1**	**1**	0	0	0	0	0	1	2	0	2	0
7 CGT DGC GT GCA TCG CA	0	0	0	0	0	0	**1**	**1**	0	0	0	0	2	0	2	1
8 CGT DTC GT GCA TAG CA	0	0	0	0	0	0	0	**1**	**1**	0	0	0	2	1	2	0
9 DCD CDC DC TGT TGT TG	0	0	0	**3**	**4**	0	0	0	0	0	0	0	0	0	0	0
10 CGT AAC GT GCA TTG CA	0	0	0	0	0	0	0	0	0	1	1	0	3	0	2	0

From the tabulation of nearest neighbour interactions above, and from the known values for the normal Watson-Crick nearest neighbours,[174] the values of $\Delta G°(25°C)$ for diaminopurine-containing duplexes was reduced to doublets.

For example, from **Duplex 4**, we obtain

$$\Delta G°_{total} = DC/TG + CD/GT + 2(GT/CA) + 2(CG/GC) + GA/CT - (-20.9) = -46.4 kJmol^{-1}$$

where $-20.9 kJmol^{-1}$ represents the helix initiation free energy, and using the values calculated for GT/CA and CG/GC, and GA/CT (-5.4, -15.1, and $-6.7 kJmol^{-1}$)[174] this leaves the doublet:

$$\Delta G°(DC/TG + CD/GT) = -19.6 kJmol^{-1}$$

Values of the nearest neighbour doublets for all the diaminopurine containing duplexes were similarly obtained.

The next step was to "close" the data to give a value for each nearest neighbour interaction. This was done by creating an inequality—the duplex chosen was **Duplex 9**, which contains 4(DC/TG) and 3(CD/GT) nearest neighbours. The values were obtained as follows

$$\Delta G°_{total} = 4(DC/TG) + 3(CD/GT) - (-20.9) = -48.6 kJmol^{-1} \qquad (4)$$

Inserting this equation into Eq.(4) we obtain

$$\Delta G°(DC/TG) = -10.7 kJmol^{-1} \text{ and } \Delta G°(CD/GT) = -8.9 kJmol^{-1}$$

These values were then plugged into the appropriate doublet values, for example with **Duplex 1**

$$\Delta G°(AD/TT + DC/TG) = -13.9 kJmol^{-1}$$

therefore $\Delta G°(AD/TT) = -3.2$ kJmol^{-1}.

This process was repeated for all remaining nearest neighbour doublets containing diaminopurine. These $\Delta G°$ values are given in Table 3.4, along with the corresponding $\Delta G°$ values determined for the native nearest neighbour interactions.[174] The CG/GC, GC/CG and GG/CC nearest neighbours are also included.

Table 3.4. *$\Delta G°$ values for the 9 nearest neighbours containing diaminopurine, compared with the 10 native nearest neighbour interactions.*

Interaction	**$\Delta G°$/kJmol^{-1}**		Interaction	$\Delta G°$/kJmol^{-1}
AD/TT	**-3.1**	<	AA/TT	-7.9
DA/TT	**-12.8**	>	AA/TT	-7.9
DD/TT	**-7.9**	≡	AA/TT	-7.9
CD/GT	**-8.9**	>	CA/GT	-7.9
DC/TG	**-10.7**	>	AC/TG	-5.4
GD/CT	**-6.9**	≡	GA/CT	-6.7
DG/TC	**-13.2**	>	AG/TC	-6.7
TD/AT	**-2.3**	<	TA/AT	-3.8
DT/TA	**-11.9**	>	AT/TA	-6.3
			CG/GC	-15.1
			GC/CG	-13.0
			GG/CC	-13.0

3.2.4 Interpretation of thermodynamic data for diaminopurine incorporation

From Table 3.4, the most striking difference between nearest neighbour interactions with and without diaminopurine, are: the AD/TT interaction is much less stable than the corresponding AA/TT, whereas DA/TT is *more* stable than AA/TT. Also, the TD/AT interaction is less stable than TA/AT.

A previous study predicted that any duplex would be destabilised where diaminopurine is flanked by adenine, with respect to the duplex formed by adenines only.[74] It was reasoned that the substitution of dD for dA would disrupt the spine of hydration formed in d(A).d(T) tracts, (*see* Section 1.5.2). As can be seen from Table 3.4, the AD/TT nearest neighbour interaction is in agreement with this premise, however, the DA/TT interaction is actually more stable, by 4.9kJmol^{-1}, than the AA/TT interaction. Therefore there must be additional forces stabilising this particular interaction, which more than compensate for the disruption to the spine of hydration.

Also worthy of mention are the DT/TA and TD/AT interactions—the hierarchy of stability is analogous to that found for DA/TT and AD/TT, when compared to their respective native nearest neighbour interactions. Thus:

DT/TA	> AT/TA	> TA/AT	> TD/TA	
(-11.9)	(-6.3)	(-3.8)	(-2.3)	(kJmol^{-1})

A pattern emerges relating the actual position of substitution of dD for dA, and the inherent stability—DX/TX is more stable than XD/XT, where X= any base and its complement. In general, when comparing duplexes containing diaminopurine with duplexes containing adenosine:

$$DX/TX > XD/XT \geq AX/TX > XA/XT \qquad (5)$$

It is interesting to compare the DX/TX interaction with the AX/TX interaction—in the AX/TX case, steric clashes occur between the thymine methyl group and the 5'-neighbouring sugar.[37] This is the cause of the large negative propeller twist observed in A.T base pairs that blocks A-DNA conformations in AX/XT steps. It is possible that in DX/TX steps, since there already exists a tendency to adopt (putative)A-DNA conformation (*see* Section 1.5.3), there is enough flexibility to lessen the steric clash.

Since the DX/TX interactions offer substantial increases in stability compared to the AX/TX interactions, this usually compensates for the decrease in stability for the stability-decreasing TD/TA interaction. The exception being where the adenine substituted is 3' adjacent to a thymine, and is part of a terminal A-tract, thus 5'-N_nTA_n-3' is more stable than 5'-N_nTD_n-3' (by 1.5kJmol^{-1}). This "worst-case" example can be seen as negligible, as substitution of dD for dA in the N-region would, in most cases, compensate for the otherwise small decrease in stability when substituting dD for A in the 3'-A_n tail.

3.3 CONCLUSIONS AND FUTURE WORK

A thermodynamic library has been developed which may allow the prediction of stability of duplexes containing diaminopurine—this library is compatible with the data available for native nearest neighbour interactions.

There is only one condition where local destabilisation of the duplex will occur when diaminopurine replaces *all* adenines. This is where a T residue is followed by a terminal oligo-dA at the 3'-end. Development of a computer program would allow prediction of the thermodynamics of duplexes containing diaminopurine, and would "recommend" any instance where dD substitution would be inappropriate.

More oligonucleotides must be synthesised containing diaminopurine, and thermodynamics of the subsequent duplexes determined, in order to further test the data. Also, compiling a similar library of nearest neighbour interactions for DNA.RNA hybrids containing 2'-deoxydiaminopurine would prove useful. These hybrids, particularly relevant in antisense applications, have recently been shown to stabilise the resulting hybrid duplexes considerably more than the corresponding DNA.DNA duplex.[78] Normally DNA.RNA hybrids are *less* stable than the corresponding DNA.DNA duplex (*see* Section 1.5.3.1). The observation that diaminopurine-containing hybrids are very stable seems to agree with the observation that diaminopurine has a tendency to adopt the putative A-form—this form may well be compatible with the A-form which is adopted by DNA.RNA hybrids. This may also account for the increase in stability over the DNA.DNA duplex containing dD.

This study emphasises the importance of obtaining the thermodynamic parameters for oligonucleotides containing modified bases. This must be carried out before any predictions can be made about the effects this modification may have on duplex stability. Melting curves carried out for a selection of base pairs is insufficient—determination of the nearest neighbour interactions is essential. When such libraries of modified bases are compiled, then the design of "super-stable" duplexes can be envisaged, containing a pre-determined combination of base analogues, for use in applications such as antisense therapy or sequencing. Such analogues would include diaminopurine, and other duplex-stabilising bases such as 5-methyl-cytidine.

Study must also be carried out on the enzymatic recognition of diaminopurine where extension using a polymerase is carried out. This is of particular interest in primers containing 3'-diaminopurine residues for PCR and DNA sequencing reactions.

4.0 THE SYNTHESIS OF BASE-LABILE LINKERS— A PRELIMINARY STUDY

As discussed in Section 1.8, there is room for improvement with regards to speeding up the cleavage of an oligonucleotide from its solid support. This can be achieved by using a more nucleophilic base instead of ammonia in the final deprotection stage, or replacing the succinyl linker with a more base-labile linker.

The latter solution is the more attractive, as it means adhering to the well-established solid phase phosphoramidite synthesis protocol. This is important as employing cleavage/deprotection reagents other than ammonia may require the use of non-standard or extra equipment, such as non-corroding seals for use with the AMA reagent described in Section 1.8. These reagents may also be more toxic and difficult to handle than ammonia.

4.1 THE SYNTHESIS OF PHTHALOYL LINKERS

There is evidence to suggest that base cleavage of an oligonucleotide from its linker occurs via deprotonation of the amide nitrogen followed by intramolecular nucleophilic displacement at the ester carbonyl group. This was shown by Brown *et al*—when the amide proton is replaced by the methyl group, the ester function was essentially unchanged under the same conditions that cleaved the unsubstituted amide.[150]

The potential of the phthaloyl group as a base-labile group was realised by Brown *et al* whilst investigating the stability of a variety of linkers to 10% solutions of the base DBU in dichloromethane.[150] Under these conditions the phthaloyl linker was found to cleave much faster than the conventional succinyl linker, owing to the rigid conformation imposed by the aromatic ring. This conformation favours the intramolecular cleavage reaction, described above, and is illustrated by the phthaloyl group (*see* Figure 4.1).

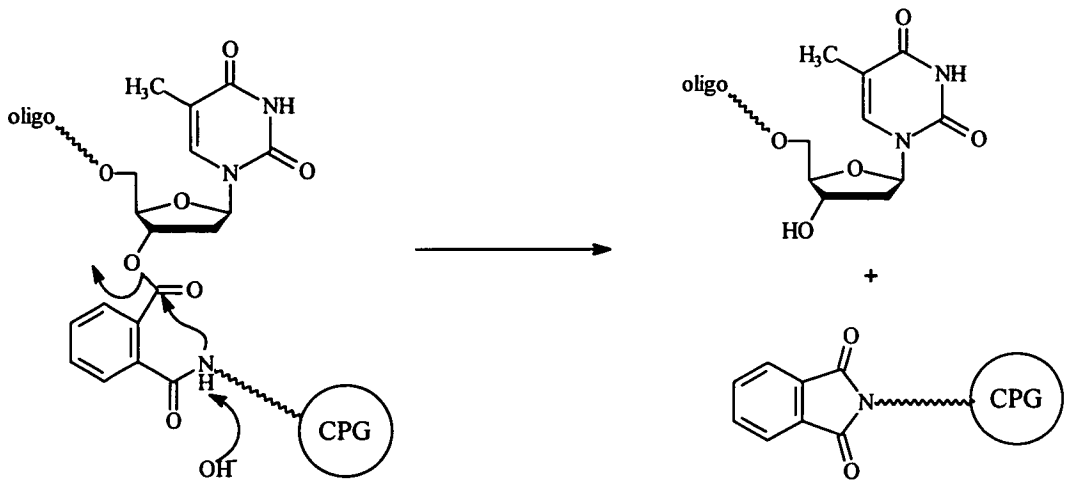


Figure 4.1. *The phthaloyl linker, showing the alignment for intermolecular cleavage by base.*

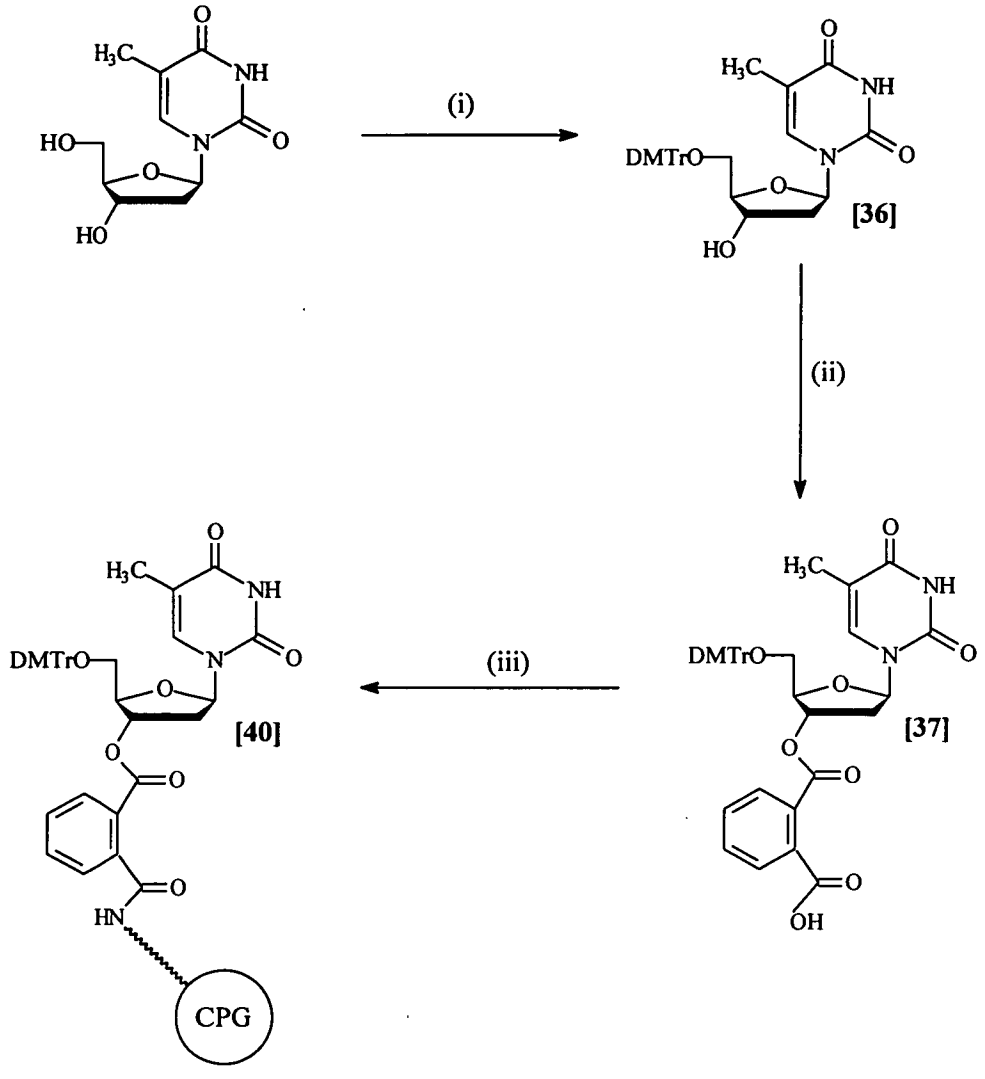
4.1.1 Synthesis of phthaloyl linker [40]

Following the discovery that the phthaloyl linker was cleaved much faster than the succinyl linker in DBU, it was important to compare the two in conc. ammonia solution, the reagent commonly used to cleave and deprotect synthetic oligonucleotides.

Firstly, thymidine was reacted with 4,4'-dimethoxytrityl chloride and DMAP in pyridine to give compound **[36]** as a solid white foam in 96% yield.

This was then reacted with phthalic anhydride and DMAP in pyridine to give compound **[37]** as a yellow-white foam in 71% yield. The attachment of the phthaloyl linker to the solid support was carried out as for the succinyl linker, described by Sproat and Gait.[170] This involves activating the 3'-O-phthaloyl nucleoside with 4-nitrophenol in pyridine, then reacting this with DCC in dichloromethane. Dicyclohexylurea is seen to precipitate, and after the reaction had gone to completion (tlc), the mixture was filtered, the filtrate concentrated, and

suspended in DMF. Activated CPG was added to the suspension to give phthaloyl CPG **[40]** with a loading, after capping, of 7.5μmolg^{-1}. This synthetic scheme is shown in Figure 4.2.



(i) DMTrCl and DMAP in pyridine.
(ii) Phthalic anhydride and DMAP in pyridine.
(iii) a:4-Nitrophenol and DCC in DCM. b:Activated CPG in DMF.

Figure 4.2. *The synthesis of phthaloyl CPG* ***[40]****.*

4.1.2 Synthesis of nitro-phthaloyl linkers [41] and [42]

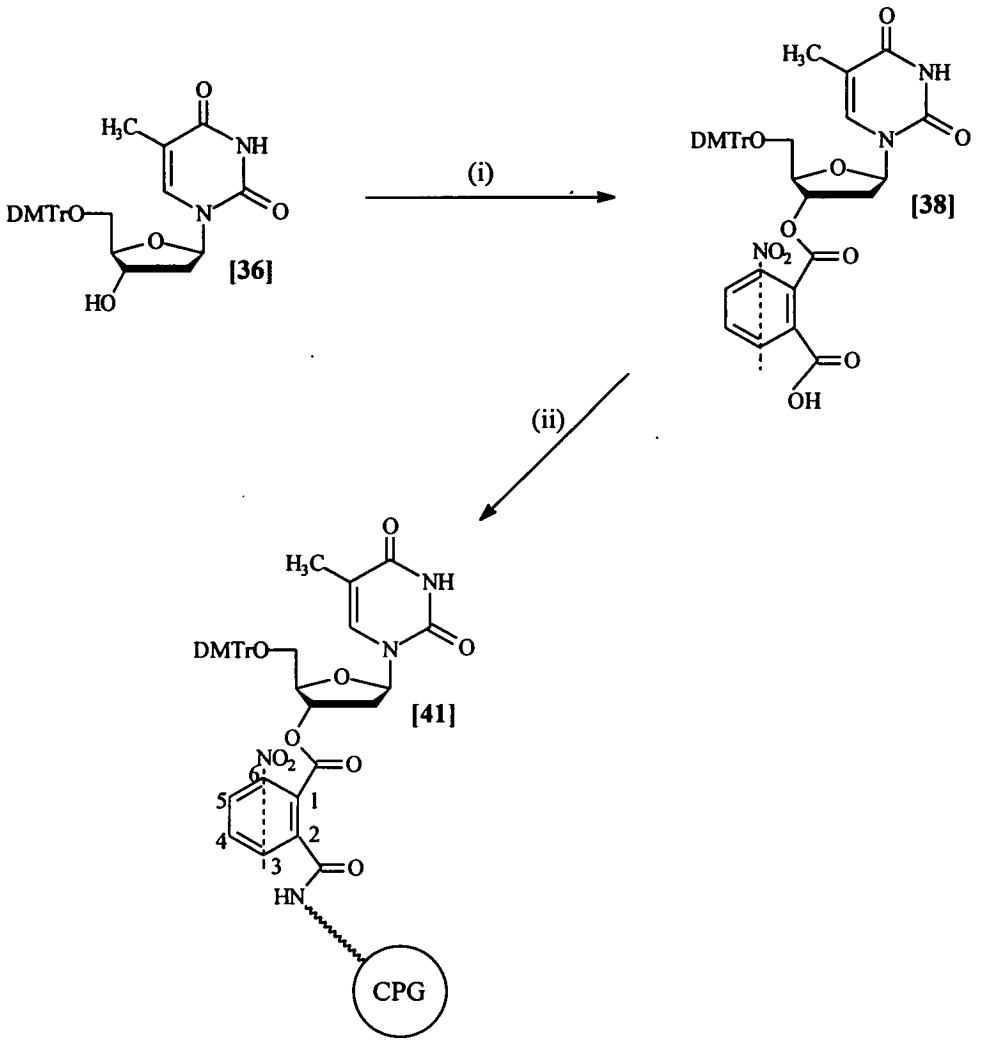
Nitro-phthaloyl CPG's **[41]** and **[42]** were also synthesised in order to determine the effect of the electron-withdrawing group on the cleavage rate of the linker.

Compound **[36]** was reacted with 3-nitro-phthalic anhydride and DMAP in pyridine, to give compound **[38]** as a yellow-white foam in 56% yield. This was then reacted as for the phthaloyl CPG to give nitro-phthaloyl CPG **[41]** with a loading, after capping, of 5.0μmolg^{-1}. This scheme is shown in Figure 4.3.

To synthesise the other nitro-phthaloyl linker, compound **[36]** was reacted with 4-nitro-phthalic anhydride and DMAP in pyridine to give two products (on tlc). The higher running product, **[39]** was isolated and obtained as a yellow-white foam, in 47% yield. (It was assumed that the lower running spot was due to the impurity, ca. 8%, of 3-nitrophthalic anhydride reacting with compound **[36]**. This impurity is present as a by-product in manufacture). Compound **[39]** was then reacted with activated CPG to give nitro-phthaloyl CPG **[42]** with a loading of 2.5μmolg^{-1}. The synthetic scheme is shown in Figure 4.4.

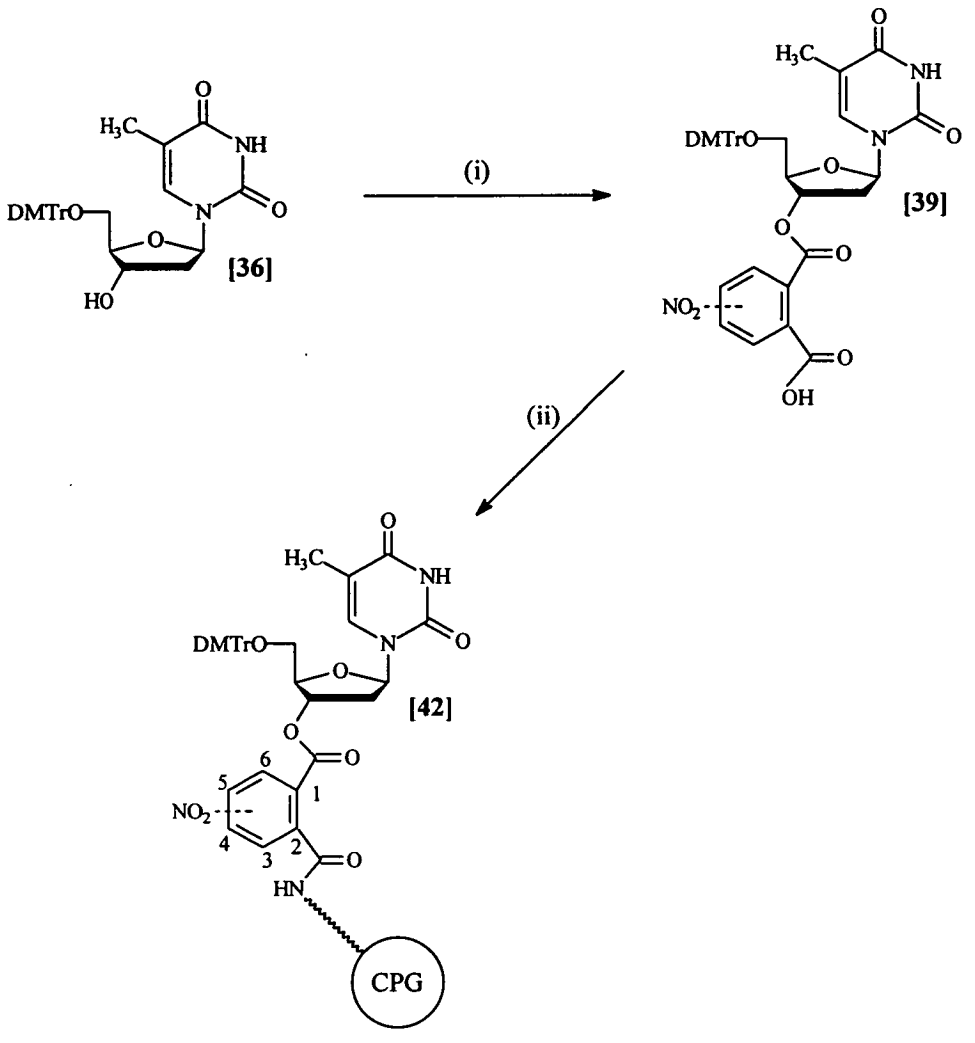
The loading of the nitro-phthaloyl resins **[41]** and **[42]** were both found to be considerably lower than phthaloyl resin **[40]**, which is already a "below-average" loading. Therefore there must be some intrinsic reason for such low loading in the phthaloyl series. This is likely due to steric bulk as compared to the succinyl linker where loadings of >30μmolg^{-1} may be obtained, if necessary. Another reason could be the low solubility of the phthaloyl nucleosides in DMF reducing reactivity towards the CPG.

Low loadings are not necessarily a bad thing, since there is less steric hindrance for the growing oligonucleotide. The main problem involving nucleosides **[38]** and **[39]**, was the inability to determine the actual orientation of the nitro groups by spectroscopic means.



(i) 3-Nitrophthalic anhydride and DMAP in pyridine.
(ii) a:4-Nitrophenol and DCC in DCM. b:Activated CPG in DMF.

Figure 4.3. *The synthesis of nitro-phthaloyl CPG* ***[41].***
The nitro group may be at the 3- *or* the 6-position of the phenyl ring.



(i) 4-Nitrophthalic anhydride and DMAP in pyridine.
(ii) a:4-Nitrophenol and DCC in DCM. b:Activated CPG in DMF.

Figure 4.4. *The synthesis of nitro-phthaloyl CPG* ***[42].***
The nitro group may be at the 4- *or* the 5-position of the phenyl ring.

4.1.3 Comparing the cleavage of the phthaloyl linkers with the succinyl linker

An initial study of the cleavage of oligonucleotides from succinyl CPG, and CPG's **[40]**, **[41]** and **[42]**, was carried out. These oligonucleotides were, respectively: **SUC-1**, **PTH-1**, **PTH-2** and **PTH-3**, and the sequence synthesised for each resin was:

5'-GAC TTG ACC TAT-3'

where the 3'-terminal T was attached to the linker.

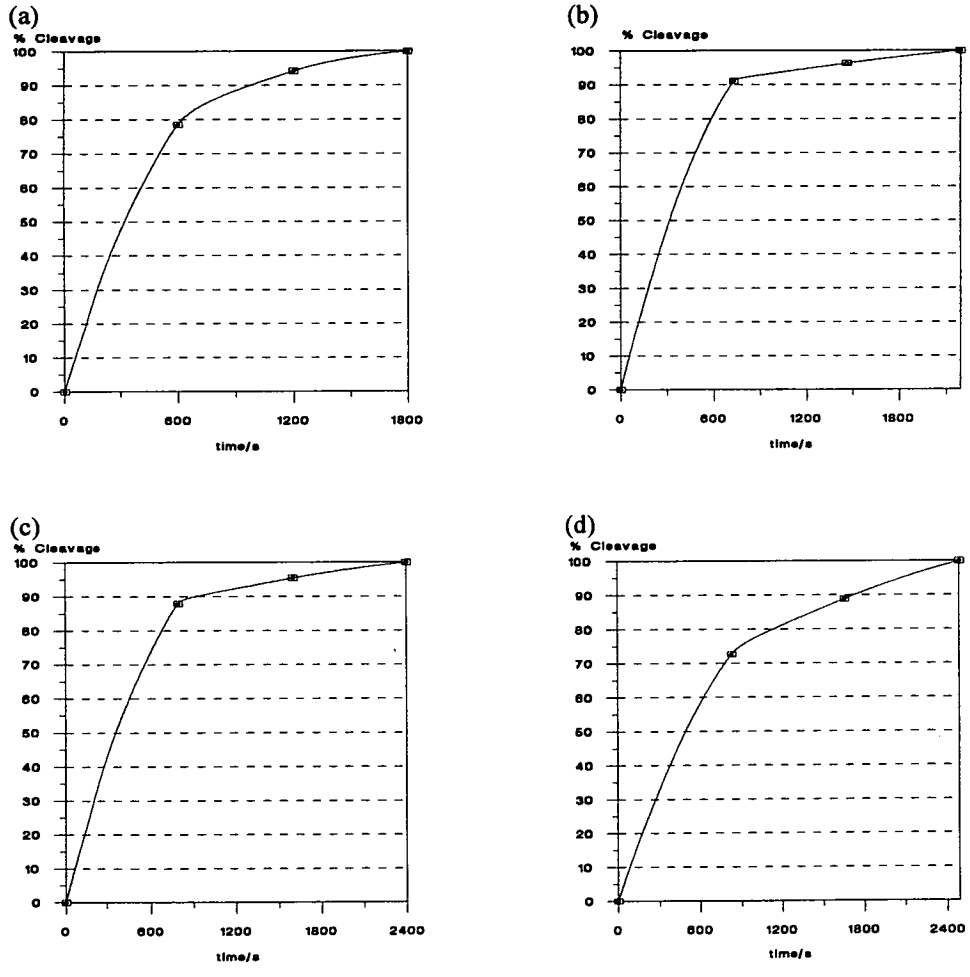


Figure 4.5. *The cleavage of linkers: (a) succinyl; (b) phthaloyl**[40]**; (c) nitro-phthaloyl**[41]**; and (d) nitro-phthaloyl**[42]**.*

This initial study was carried out by taking an aliquot from each stage of a 3x10 min Endproc., and measuring the absorbance. The cleavage of the four linkers is shown in Figure 4.5. This experiment was not very accurate, as the machine-delivered volumes of ammonia may vary, but this did give an idea of the cleavage times of the linkers.

The order of the time of cleavage*, with fastest first, was:

$$\text{phthaloyl}[\mathbf{40}] < \text{NO}_2\text{-phthaloyl}[\mathbf{41}] = \text{succinyl} < \text{NO}_2\text{-phthaloyl}[\mathbf{42}]$$

that is, both the nitro-phthaloyl linkers did not cleave any faster than the phthaloyl linker. In fact, the nitro-phthaloyl linkers appeared to cleave more *slowly* than the unsubstituted phthaloyl linker. The nitro-phthaloyl carboxylic esters were expected to cleave more rapidly than the unsubstituted phthaloyl linker since the electron withdrawing nitro group would induce a partial positive charge distribution, particularly at the *ortho* and *para* positions of the benzene ring. It is possible that the *reduction* in cleavage times are due to the steric bulk introduced by the nitro group, along with the increase in lipophilicity. Both of these effects would be expected to slow down the cleavage of the phthaloyl group in *aqueous* ammonia.

4.2 COMPARING THE PHTHALOYL AND SUCCINYL LINKERS

From the qualitative results discussed above, attention then focused on comparing the phthaloyl CPG **[40]**, with the succinyl CPG. (The nitro-phthaloyl linkers, bearing in mind the very poor loading of resins **[41]** and **[42]**, and their apparent slow cleavage, compared to the phthaloyl linker, eliminated them from further investigation).

In the following experiments, multiple tests were carried out, and more controlled conditions were employed—this gave a more accurate comparison of the two linkers.

---

* When the absorbance of the cleaved oligonucleotide had reached 90% of the final absorbance *measured*, this was taken as a working figure for comparing cleavage. This cut-off will be explained in Section 4.2.2.

4.2.1 The quality of oligonucleotide synthesis using phthaloyl or succinyl CPG

The quality of the oligonucleotides synthesised will be considered first. The oligonucleotides synthesised were compared, with respect to their average coupling efficiency during oligonucleotide synthesis, and, after cleavage and full deprotection, were analysed by HPLC and CE, including co-injections for authenticity checks.

Average coupling efficiency of oligonucleotide synthesis

In the 12- and 20-mer test sequences shown below, the average coupling efficiencies for the phthaloyl linker were found to be much the same as those for the succinyl linker. The average coupling efficiency for each of the oligonucleotides **SUC-2** and **-3**, and **PTH-4** and **-5** was found to be greater than 98.5%. Therefore the phthaloyl linker had no detrimental effect on oligonucleotide synthesis.

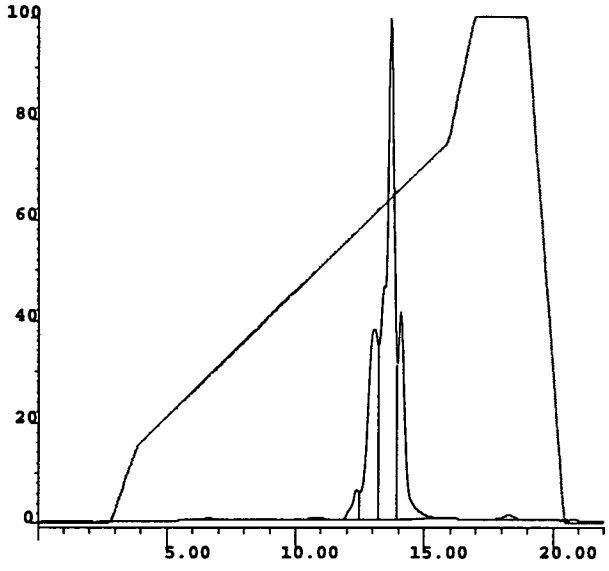
SUC-2	ATG CAA CCG AGT	using succinyl CPG
SUC-3	ATA GCT TGG CAA GTA TGA CT	using succinyl CPG
PTH-4	ATG CAA CCG AGT	using phthaloyl CPG **[40]**
PTH-5	ATA GCT TGG CAA GTA TGA CT	using phthaloyl CPG **[40]**

HPLC analysis

Next, the quality of the oligonucleotides synthesised was checked by HPLC. The chromatographs shown in Figure 4.6 are for the crude 20-mer test sequences.

As can be seen, far from affecting the synthesis in a detrimental manner, **PTH-5** seems to have been a better synthesis than **SUC-3**, with less failure sequences, and a more pure product. This may be a consequence of the low loading on the CPG—resulting in a less hindered, and better quality synthesis.

(a) **SUC-3.**



(b) **PTH-5.**

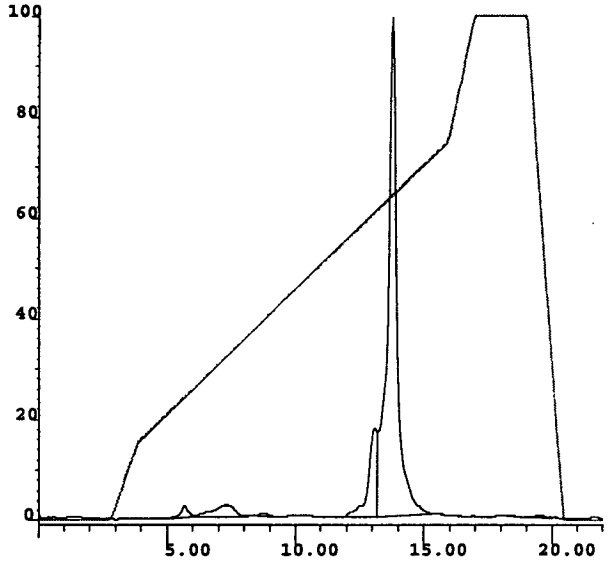


Figure 4.6. *HPLC chromatograms of the crude, deprotected 20-mer obtained after synthesis on: (a) the succinyl linker, and (b) the phthaloyl linker* ***[40]****.*

Capillary electrophoresis analysis

The 20-mer phthaloyl test sequence, after HPLC purification and desalting, was compared with the "authentic" succinyl oligonucleotide by capillary electrophoresis. As can be seen, the coinjection of the two purified oligonucleotides shows they are in fact identical. This was carried out as CE can differentiate between slight changes in

coinjections of oligonucleotides; for example a 119-mer can be resolved from a 120-mer.[176]

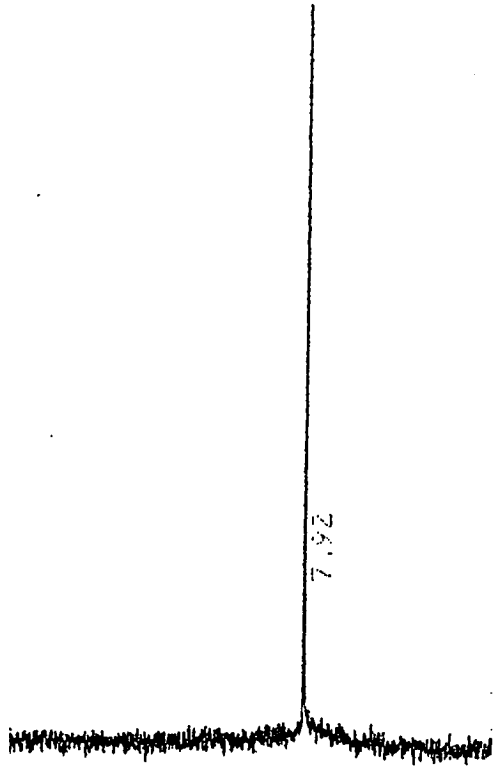


*Figure 4.7. The CE chromatogram of a coinjection of pure, desalted **SUC-3** and **PTH-5**.*

4.2.2 Comparing the cleavage of the phthaloyl and succinyl linkers

In order to study the cleavage of the succinyl linkers in a more controlled manner, CPG was taken from the column after a 0.2μmol, no Endproc. oligonucleotide synthesis. This CPG was then added to 2.5ml of concentrated ammonia solution in a 1cm UV cuvette. The UV absorbance at 260nm was measured every 5 minutes, straight after the UV cell had been inverted to allow mixing of the cleaved oligonucleotide with the ammonia. The readings were taken until the absorbance started levelling off.

This experiment was carried out three times for both of the linkers, to obtain an average value of cleavage time. An example of this absorbance at 260nm measured

with time, to give an indication of speed of cleavage, is given in Figure 4.8 for both linkers.

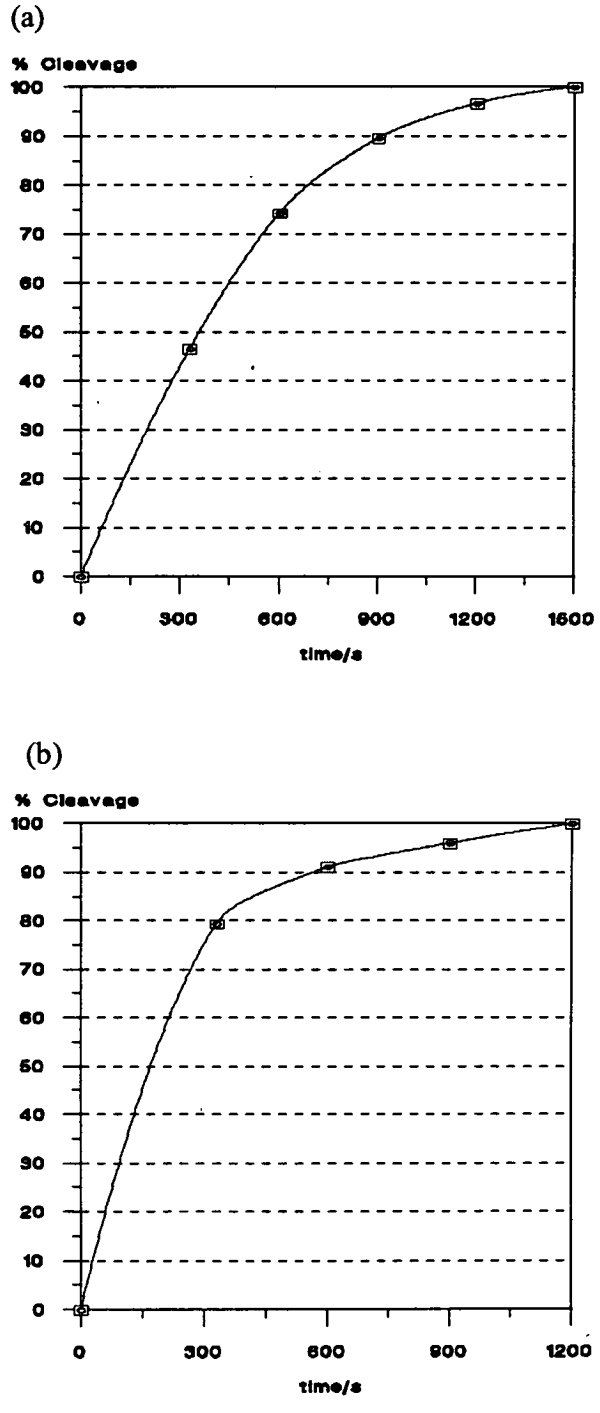


Figure 4.8. *Cleavage of the linkers as a function of time, for: (a)* ***SUC-2*** *and (b)* ***PTH-4****.*

This gives an average value of 9 minutes and 16 minutes for >90% cleavage from the solid support, for the phthaloyl and succinyl groups, respectively.

It must be noted that the point where 100% cleavage is reached is actually the point where the absorbance measured is no longer increasing geometrically. However, the absorbance does not level off, but continues to increase linearly after "true" levelling off has taken place. This is why the reasonable "cut off" point for cleavage of an oligonucleotide is taken to 90% of the maximum—to exclude the misleading increase in absorbance from adding to measured cleavage time.

It seems likely that after initial cleavage of the oligonucleotide from the linker, the resulting cyclic amide cleaving from the support adds to the absorbance measured, hence the linearity introduced into the graph. The size of this increase in absorbance was always proportionally greater for the phthaloyl group—this may reflect the higher absorbance at 260nm of this group over the cleavage of the succinimide group from the solid support.

In summary, the gradient for the phthaloyl plot is very steep in the first one or two measurements, and then switches to a linear rise. In contrast, the succinyl plot shows a gentler decrease in the steepness of the gradient over time. This is down to the relative slowness in cleavage compared to the phthaloyl group, combined with the higher absorbance of the phthaloyl group.

These observations would imply that scission of the phthaloyl linker was even faster than is shown in Figure 4.8(b).

4.3 CONCLUSIONS AND FUTURE WORK

From the finding that the succinyl linker is being given a cleavage time well in excess of that required, the Endproc. routinely used for the succinyl group is now 4x8 minutes in Oswel DNA Service. This is a saving of 28 minutes for every synthesis, over the previously used 4x15 minutes Endproc.

The results from the study of the phthaloyl linker suggest that an Endproc. of 4x2.5 minutes would be more than adequate for obtaining >90% of the

oligonucleotide bound to the solid support. However, further study on the phthaloyl linker should include the measurement of cleavage at 1 or 2 minute intervals to give a more accurate determination of the cleavage time, coupled with detailed analysis of the cleavage products. Hence the cause of the "linearity" in later absorbance measurements, could be identified. If the absorbance was indeed not due to cleavage of oligonucleotides, then the phthaloyl linker would need less time than presumed.

Again, this approach could be applied to the nitro-phthaloyl linkers in the event that the linearity problem was swamping the true measurement of cleavage. Further study on these groups would also determine the orientation of the nitro groups, in order that the actual charge distribution may be predicted. This would help in the design of other substituted phthaloyl linkers.

Phthaloyl linkers with electron withdrawing groups of less steric bulk, and lower lipophilicity should be synthesised, in order to test the hypothesis that the nitro groups are slowing down base cleavage in aqueous conditions.

EXPERIMENTAL SECTION

5.1 SOLVENTS AND REAGENTS

Solvents

Solvents used for the extraction and chromatography of phosphoramidites, and for the preparation of anhydrous solvents, were of HPLC grade. Reverse-osmosis purified water was used during oligonucleotide synthesis, analysis and purification. All other solvents used were of laboratory grade.

Anhydrous Solvents and Reagents

Dichloromethane was distilled over CaH_2; benzene was dried with sodium wire; diethyl ether was distilled from sodium/benzophenone; N,N-dimethylformamide was fractionally distilled from 4Å molecular sieves; pyridine was distilled from CaH_2; toluene was dried with sodium wire; triethylamine and N,N-diisopropylethylamine were dried over CaH_2. Anhydrous acetonitrile was purchased from Applied Biosystems Limited (ABI).

Other Reagents

4,4'-Dimethoxytrityl chloride was purchased from Courtaulds, all other chemicals were supplied by Aldrich, Fluka or Sigma. 2-Cyanoethyl-N,N-diisopropylchloro-phosphoramidite was synthesised by the procedure of Claesen et al.[177]

5.2 ANALYTICAL TECHNIQUES

Chromatography

Flash column chromatography was carried out using 60 mesh silica gel (Merck), under slight argon pressure.

Thin layer chromatography (TLC) was carried out on silica gel 60, F254, 0.2mm layer aluminium sheets (Merck) using the following solvent systems:

(A) Dichloromethane-methanol (9:1, v/v).

(B) Dichloromethane-methanol (95:5, v/v).

(C) Dichloromethane-methanol (8:2, v/v).

(D) Acetonitrile-ethyl acetate (50:50, v/v).

(E) Dichloromethane-methanol (98:2, v/v).

(F) Ethyl acetate-acetonitrile (98:2, v/v).

(G) Dichloromethane.

(H) Ethyl acetate-diethyl ether (50:50, v/v).

(I) Ethyl acetate-methanol-ammonia (5:1:1, v/v/v).

(J) Dichloromethane-methanol-acetic acid (10:2:1, v/v/v).

Products were visualised on TLC using the following techniques:

(i) UV absorption at 264nm for nucleosides.

(ii) Spraying with a solution of 4-methoxybenzaldehyde: glacial acetic acid: conc. sulphuric acid: ethanol (5:1:1:50). Compounds with a protected or unprotected 1,2-diol function gave a dark blue colour on strong heating.

(iii) Exposure to conc. hydrochloric acid vapour produced a strong orange colour for compounds containing a dimethoxytrityl group.

(iv) Purines containing an N6-azide gave a dark brown colour on strong heating.

(v) Purines containing an azide, nitrogen cation, or formamidine group fluoresced on absorption of UV at 264nm.

(vi) TLC plates were pre-treated with triethylamine for compounds containing the dimethoxytrityl or TIPS groups.

Combustion Analysis

CHN analysis was carried out on a Perkin Elmer 2400 Elemental Analyzer. All samples were dried at 60°C over P_2O_5 for at least 48 hours prior to analysis.

NMR Spectra

^{1}H-NMR spectra were recorded on a Brucker WP-200 spectrometer (200.13MHz), a Brucker WP-250 spectrometer (250.13MHz) and a Brucker WP-360 spectrometer (360.14MHz).

^{13}C-NMR spectra were recorded on a Brucker WP-200 spectrometer (50.32MHz), a Brucker WP-250 spectrometer (62.90MHz) and a Brucker WP-360 spectrometer (90.56MHz)

^{31}P-NMR spectra were recorded on a JEOL FX90Q spectrometer (90MHz).

Mass Spectra

Positive ion Fast Atom Bombardment (FAB) mass spectra were recorded on a Kratos MS50 TC spectrometer using a thioglycerol/acetonitrile or a 3-nitrobenzyl alcohol matrix.

UV Spectra

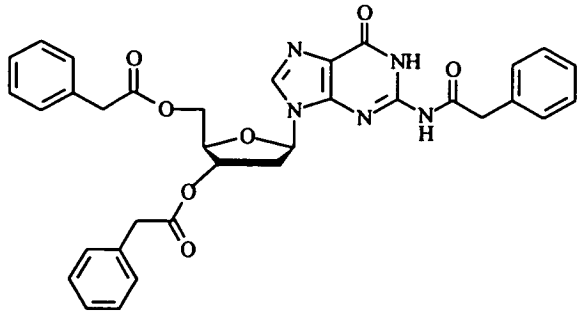
UV spectra were recorded on a Perkin Elmer Lamda 15 ultraviolet-visible spectrometer over the range 190-320nm, using PECSS2 software. The solvent used in each scan was acetonitrile.

IR Spectra

IR spectra were recorded on a Bio Rad FTS-7 fourier transform spectrometer controlled by a Bio Rad SPC 3200 microcomputer using KBr plates, using a nujol mull or as a film from dichloromethane.

5.3 SYNTHESIS OF dD NUCLEOSIDES

2-N, 3'-O, 5'-O-Tris-phenylacetyl-2'-deoxyguanosine [1].

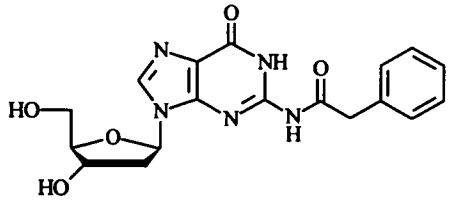


2'-Deoxyguanosine monohydrate (6.3g, 22mmol) was coevaporated three times with anhydrous pyridine and suspended in anhydrous pyridine (250ml). The reaction vessel was cooled to 0°C, and phenylacetic anhydride (5eq, 110mmol, 28.0g) was added dropwise and then the reaction mixture was heated to 110°C with stirring. After 1.5 hours the pyridine was removed *in vacuo*, and the residue was dissolved in dichloromethane, washed with saturated aqueous $NaHCO_3$ (250ml), saturated aqueous KCl (3x200ml) and dried with anhydrous Na_2SO_4. The organic layer was evaporated *in vacuo*, and the residue purified by wet flash column chromatography, eluting with dichloromethane-methanol (99:1, v/v) to give a pale cream foam. Yield: 11.5g, (84%).

Rf 0.69 (solvent A), Rf 0.24 (solvent B). Found: C,65.26; H,5.02; N,10.83%. Requires: C,65.73; H,5.03; N,11.27%. ^{1}H nmr (200.13MHz; $CDCl_3$): δ_H: 2.30-2.41 (1, ddd, J=2.0Hz, J=6.3Hz, J=14.4Hz, 2'-H), 2.62-2.74 (1, ddd, J=2.0Hz, J=6.3Hz, J=14.4Hz, 2"-CH), 3.58-3.60 (2, d, J=3.6Hz, 2-$PhCH_2$), 3.64 (4, s, 5'-$PhCH_2$), 3.65 (4, s, 3'-$PhCH_2$), 4.34-4.45 (3, m, 4'-H, 5'-CH_2), 5.27-5.31 (1, m, 3'-H), 6.05-6.12 (1, dd, J=6.2Hz, J=8.1Hz, 1'-H), 7.09-7.39 (15, m, aromatic-CH), 7.59 (1, s, 8-H), 9.41 (1, s, 2-NH), 11.88 (1, s, 1-NH). ^{13}C nmr (50.32MHz, $CDCl_3$) δ_C: 36.25 (CH_2), 40.80 (2xCH_2), 43.43 (CH_2), 63.88 (CH_2), 74.72 (CH), 82.11 (CH), 85.17 (CH), 121.92 (C), 127.16-129.13 (16xCH), 132.70 (C), 132.97 (C), 133.04 (C), 137.89 (C), 147.07

(C), 155.34 (C), 170.65 (C), 172.00 (C), 172.93 (C). m/z (FAB) 622.23013, $[(M+H)^+$ calc. for $C_{34}H_{32}N_5O_7$ 622.23015]. λmax 200, 253, 260, 286nm.

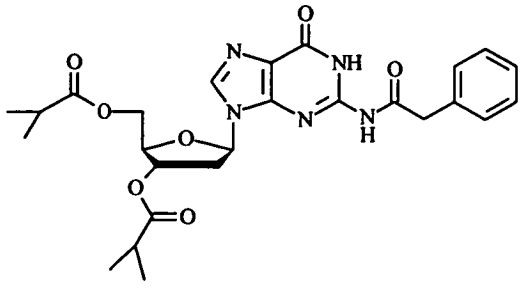
2-N-Phenylacetyl-2'-deoxyguanosine [2].



2-N, 3'-O, 5'-O-Tris-phenylacetyl-2'-deoxyguanosine **[1]** (9.4g, 15.2mmol) was dissolved in a pyridine-methanol solution (400ml, 3:1, v/v) and cooled to 0°C. To this mixture, 2M sodium hydroxide solution (3eq, 22.8ml) was added dropwise with stirring. After 10min, the reaction was quenched with Dowex resin (H^+ form). The Dowex resin was filtered off and washed with pyridine and methanol; the resulting filtrate was evaporated *in vacuo*. To the residue, dichloromethane-methanol (80:20 v/v) was added, and the resulting precipitate was filtered off to leave a white powder. Yield: 5.15g, (88%).

Rf 0.16 (solvent A), Rf 0.52 (solvent C). Found: C,56.31; H,5.30; N,18.13%. Requires: C,56.13; H,4.97; N,18.19%. ^{1}H nmr (200.13MHz; d6-DMSO) δ_H: 2.23-2.35 (1, m, 2'-H), 2.49-2.65 (1, m, 2"-H), 3.56-3.66 (2, m, 5"-CH_2), 3.81 (2, s, 2-PhCH_2), 3.83-3.89 (1, m, 4'-H), 3.39-4.40 (1, m, 3'-H), 4.99 (1, s, 5'-OH), 5.35 (1, s, 3'-OH), 6.19-6.26 (1, m, 1'-H), 7.22-7.36 (5, m, aromatic-CH), 8.25 (1, s, 8-H). ^{13}C nmr (50.32MHz; d6-DMSO) δ_C: 39.85 (CH_2), 42.70 (CH_2), 61.61 (CH_2), 70.62 (CH), 83.15 (CH), 87.89 (CH), 120.45 (C), 127.16 (CH), 128.60 (2xCH), 129.52 (2xCH), 134.36 (C), 137.66 (CH), 147.96 (C), 154.96 (C), 174.20 (C). m/z (FAB) 386.14641 $[(M+H)^+$ calc. for $C_{18}H_{20}N_5O_5$ 386.14641]. λmax 198, 254, 259, 286nm.

2-N-Phenylacetyl-3'-O,5'-O-bis-isobutyryl-2'-deoxyguanosine [3].

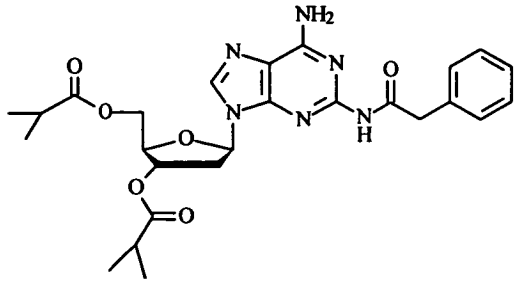


2-N-Phenylacetyl-2'-deoxyguanosine **[2]** (4.6g, 12mmol) was coevaporated three times with anhydrous pyridine and redissolved in anhydrous pyridine (250ml). The mixture was cooled to 0°C, and isobutyryl chloride (3.2eq, 38.2mmol, 4.07g, 4ml) was then added dropwise with stirring. After 20min the pyridine was evaporated *in vacuo*, the residue redissolved in dichloromethane and washed with saturated aqueous $NaHCO_3$ (150ml), water (3x150ml), and saturated aqueous KCl (100ml) and dried with anhydrous Na_2SO_4. The organic layer was then evaporated *in vacuo* to leave a stable foam. Yield: 5.8g, (90%).

R_f 0.53 (solvent A), Rf 0.85 (solvent C). ^1H nmr (200.13MHz; $CDCl_3$) δ_H: 1.06-1.25 (12, m, $(CH_3)_3C$), 2.40-3.04 (4, m, Me_2C-H, 2'-CH_2), 3.85 (2, s, Ph-CH_2), 4.23-4.36 (2, m, 5'-CH_2), 4.57-4.72 (1, m, 4'-H), 5.26-5.37 (1, m, 3'-H), 6.13-6.22 (1, m, 1'-H), 7.22-7.34 (5, m, aromatic-CH), 7.76 (1, s, 8-H), 9.98 (1, s, 2-NH), 12.00 (1, s, 1-NH). ^{13}C nmr (50.32MHz; $CDCl_3$) δ_C: 18.59 (2xCH_3), 18.73 (2xCH_3), 33.58 (CH), 33.73 (CH), 36.66 (CH_2), 43.66 (CH_2), 63.31 (CH), 74.02 (CH), 82.43 (CH), 85.26 (CH), 121.99 (CH), 127.49 (CH), 128.28 (CH), 128.73 (2xCH), 129.16 (2xCH), 132.71 (C), 137.82 (C), 147.22 (C), 155.30 (C), 172.93 (C), 176.11 (C), 177.39 (C).

m/z (FAB) 526.23015 [$(M+H)^+$ calc. for $C_{26}H_{31}N_5O_7$ 526.230216]. λmax 202, 253, 260, 286nm.

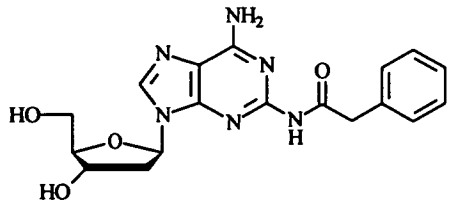
2-N-Phenylacetyl-3'-O,5'-O-bis-isobutyryl-2-amino-2'-deoxyadenosine [4].



2-N-Phenylacetyl-3'-O,5'-O-bis-isobutyryl-2'-deoxyguanosine **[3]** (5.5g, 10.5mmol) was coevaporated three times with, and redissolved in, anhydrous dichloromethane (100ml). To this, anhydrous triethylamine (4eq, 42mmol, 4.3g, 6.0ml), dimethylaminopyridine (0.05eq, 0.5mmol, 100mg) and mesitylene sulphonyl chloride (2.5eq, 26.3mmol, 5.8g) were added and this mixture was stirred at room temperature for 1 hour. The reaction vessel was cooled to 0°C and N-methyl pyrrolidine (30eq, 0.32mol, 26.8g, 33.5ml) was added dropwise. After 15min, anhydrous ammonia gas (ca. 150ml) was condensed into the mixture and the reaction was allowed to reflux under a dry-ice cold-finger for 5 hours (under argon flow) before allowing the ammonia to evaporate overnight. The reaction mixture was then washed with water (3x100ml), saturated KCl (100ml), dried with anhydrous Na_2SO_4, and the organic layer was concentrated *in vacuo* to leave a dark red-brown residue. Wet flash column chromatography, eluting with dichloromethane-methanol (99.2:0.8 v/v), gave the product as a light brown foam. Yield: 3.0g, (55%).

Rf 0.48 (solvent A), Rf 0.81 (solvent C). ^{1}H nmr (200.13MHz; $CDCl_3$) δ_H: 1.11-1.26 (12, m, Me_2C), 2.46-2.86 (4, m, 2'-CH_2, Me_2CH), 4.28-4.38 (5, m, 5'-CH_2, 4'-H, Ph-CH_2), 5.26-5.33 (1, m, 3'-H), 6.25-6.32 (1, dd, J=6.0Hz, J=7.9Hz, 1'-H), 6.81 (2, br s, 6-NH_2), 7.18-7.31 (5, m, aromatic-CH), 7.87 (1, s, 8-H), 10.18 (1, br s, 2-NH). ^{13}C nmr (50.32MHz; $CDCl_3$) δ_C: 18.57 (2xCH_3), 18.69 (2xCH_3), 33.54 (CH), 33.65 (CH), 37.58 (CH_2), 43.38 (CH_2), 63.39 (CH_2), 73.87 (CH), 82.30 (CH), 83.96 (CH), 116.70 (C), 126.63 (CH), 128.31 (2xCH), 129.30 (2xCH), 134.89 (C), 137.41 (CH), 149.48 (C), 152.90 (C), 156.54 (C), 173.55 (C), 176.17 (C), 176.42 (C). m/z (FAB) 525.24614 [$(M+H)^+$ calc. for $C_{26}H_{33}N_6O_6$ 525.24615]. λmax 199, 225, 272nm.

2-N-Phenylacetyl-2-amino-2'-deoxyadenosine [5].

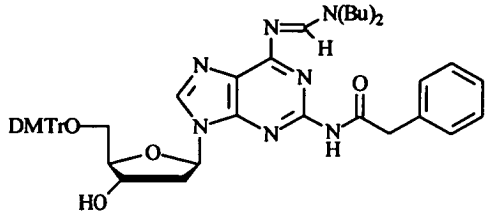


2-N-Phenylacetyl-3'-O,5'-O-bis-isobutyryl-2-amino-2'-deoxyadenosine **[4]** (1.9g, 3.6mmol) was dissolved in a pyridine-methanol solution (160ml, 3:1, v/v) and cooled to 0°C. To this mixture, 2M sodium hydroxide solution (3eq, 15.4ml) was added dropwise with stirring. After 5min, the reaction was quenched with Dowex resin (H^+ form). The Dowex resin was filtered off and washed with pyridine and methanol; the resulting filtrate was evaporated *in vacuo*, and dichloromethane-methanol was added (8:2, v/v) to give a precipitate which was filtered and washed to leave a white powder. Yield: 1.6g, (83%).

Rf 0.13 (solvent A), Rf 0.54 (solvent C). Found: C,53.24; H,5.39; N, 20.37%. $C_{18}H_{20}N_6O_4.(H_2O)$ requires: C,53.71; H,5.51; N,20.89%. ^1H nmr (200.13MHz; d6-DMSO) δ_H: 2.19-2.30 (1, m, 2'-H), 2.62-2.75 (1, m, 2"-H), 3.49-3.67 (2, m, 5'-CH_2), 3.82-3.86 (3, m, 4'-H, Ph-CH_2), 4.41-4.42 (1, m, 3'-H), 4.93-4.98 (1, dd, J=3.8Hz, 5'-OH), 5.31-5.33 (1, d, J=3.9Hz, 3'-OH), 6.24-6.31 (1, dd, J=7.3Hz, 1'-H), 7.19-7.33 (7, m, Ph-CH, 6-NH_2), 8.26 (1, s, 8-H), 10.14 (1, s, 2-NH). ^{13}C nmr (91.75MHz; d6-DMSO) δ_C: 40.05 (CH_2), 43.61 (CH_2), 61.86 (CH_2), 70.84 (CH), 83.10 (CH), 87.44 (CH), 115.54 (C), 126.07 (CH), 127.76 (2xCH), 128.67 (2xCH), 135.02 (C), 138.18 (CH), 149.06 (C), 151.61 (C), 154.87 (C), 168.54 (C).

m/z (FAB) 385.16242 $[(M+H)^+$ calc.for $C_{18}H_{21}N_6O_4$ 385.16243]. λmax 197, 225, 272nm.

2-N-Phenylacetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2-amino-2'-deoxyadenosine [6].

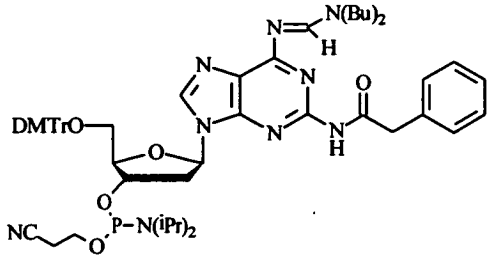


2-N-Phenylacetyl-2-amino-2'-deoxyadenosine **[5]** (0.95g, 2.5mmol) was suspended in pyridine and warmed to complete dissolution. Di-n-butylformamide dimethylacetal (1.3eq, 3.2mmol, 0.65g, 0.75ml) was added dropwise, and the solution was stirred for 2 hours at room temperature. The reaction mixture was then evaporated *in vacuo*, after coevaporation with toluene, to leave a pale yellow oil. This was dissolved in dry dichloromethane (25ml), with triethylamine (4eq, 10mmol, 1.4ml), and 4-dimethylaminopyridine (0.1eq, 0.25mmol, 31mg). 4,4'-Dimethoxytrityl chloride (1.2eq, 2.2mmol, 740mg) was added portionwise with stirring. After 2 hours, the reaction mixture was washed with water (2x50ml) and dried over Na_2SO_4, then the solvent was evaporated *in vacuo*. Wet flash column chromatography on silica gel, eluting with dichloromethane-triethylamine (99:1) gave the product which was evaporated *in vacuo* to leave a white solid foam. Yield: 1.9g, (92%).

Rf 0.58 (solvent D), Rf 0.49 (solvent E). Found: C,69.46; H,6.97; N,11.42%. Requires: C,69.84; H,6.72; N,11.88%. ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 0.90-0.98 (6, t, formamide CH_3), 1.15-1.64 (8, m, formamide $EtCH_3$, CH_3), 2.57-2.60 (2, m, 2'-CH_2), 3.28-3.43 (4, m, formamide $PrCH_2$), 3.58-3.68 (2, m, 5'-CH_2), 3.71 (6, s, CH_3O), (3, m, 3'-OH, $PhCH_3$), 4.21-4.23 (1, m, 4'-H), 4.67-4.68 (1, m, 3'-H), 6.62-6.69 (1, m, 1'-H), 6.71-6.77 (4, m, DMTr-H), 7.14-7.42 (H, m, DMTr-H), 7.97 (1, s, 8-H), 8.11 (1, s, 2-NH), 9.03 (1, s, formamide C-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 13.44 (CH_3), 13.69 (CH_3), 19.44 (CH_2), 19.95 (CH_2), 30.60 (CH_2), 40.85 (CH_2), 44.25 (CH_2), 44.73 (CH_2), 45.74 (CH_2), 51.63 (CH_2), 54.90 (2xCH_3), 64.00 (CH_2), 72.20 (CH), 83.68 (CH), 86.01 (C), 86.47 (CH), 112.86 (4xCH), 122.67 (C), 126.52 (CH), 127.01 (CH), 127.58 (2xCH), 127.94 (2xCH), 128.60 (2xCH), 129.30 (2xCH), 129.84 (4xCH), 134.08 (C), 135.65 (2xC), 139.40 (CH), 144.53 (C), 151.90 (C),

152.06 (C), 158.17 (3xC), 159.43 (CH), 160.21 (C). m/z (FAB) 826.42918 [$(M+H)^+$ calc. for $C_{48}H_{56}N_7O_6$ 826.42918]. λmax 201, 236, 260, 319nm.

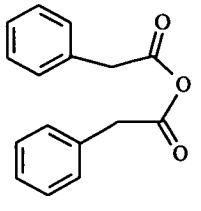
2-N-Phenylacetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2-amino-2'-deoxyadenosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite [7].



To 2-N-Phenylacetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2-amino-2'-deoxyadenosine **[6]** (460mg, 0.56mmol) dissolved in anhydrous tetrahydrofuran (8ml) was added dry diisopropylethylamine (4eq, 2.24mmol, 0.40ml) and, dropwise, 2-cyanoethyl N,N-diisopropyl-chlorophosphoridate[n54] (1.2eq, 0.67mmol, 0.16g, 0.15ml). After stirring at room temperature for 30min, the reaction was filtered under argon pressure, washed with 10% Na_2CO_3 (2x50ml), dried over Na_2SO_4, and taken to an oil. The oil was purified by flash column chromatography eluting with argon-purged ethyl acetate. (The column was pre-treated with triethylamine, and washed with ethyl acetate). The fractions were collected, and evaporated to give a white foam. Yield: 524mg, (91%).

Rf 0.69 (solvent F), Rf 0.89 (solvent G). ^{31}P nmr (90MHz; $CDCl_3$) δ_P: 149.98 (s). m/z (FAB) 1026.53945 $[(M+H)^+$ calc. for $C_{57}H_{73}N_9O_7P$ 1026.53706].

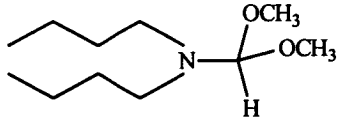
Phenylacetic anhydride.



Phenylacetic acid (60g, 0.44mol) was refluxed for 1 hour with acetic anhydride (2.5eq, 1.1mol, 112.5g, 104ml). Excess acetic anhydride was removed *in vacuo* to leave an oil. On scratching the oil crystallised to leave crude phenylacetic anhydride. The crude crystals were then recrystallised from sodium-dried diethyl ether. Yield: 47.8g, (86%).

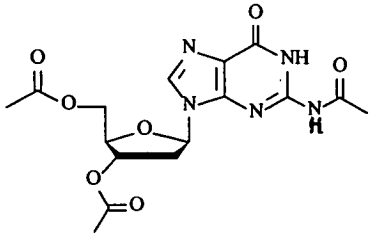
^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 3.74 (4, s, CH_2), 7.20-7.39 (10, m, Ph-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 41.85 (2xCH_2), 127.17 (2xCH), 128.60 (4xCH), 129.22 (4xCH), 131.80 (2xC), 166.78 (2xC). m/z (FAB) 255.10238 [$(M+H)^+$ calc. for $C_{16}H_{15}O_3$ 255.10211].

Di-N-butylformamide dimethylacetal.[122]



Di-N-butylamine (100ml, 0.59mol) and N,N-dimethylformamide dimethylacetal (1.2eq. 0.7mol, 85g, 94ml) were guarded from moisture, and heated together for five days at 100°C to give a straw-coloured solution. Distillation at reduced pressure gave the desired product which was stored in the freezer under argon. Yield: 35.0g, (29%). Bp. 92-94°C 10mmHg. ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 0.79-0.88 (6, t, J=7.2Hz, CH_3CH_2), 1.16-1.46 (8, m, $EtCH_2$, $MeCH_2$), 2.46-2.55 (4, m, $PrCH_2$), 3.23 (6, s, $CHOCH_3$), 4.43 (1, s, CH). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 13.66 (2xCH_3), 20.19 (2xCH_2), 31.73 (2xCH_2), 49.32 (2xCH_2), 77.10 (CH).

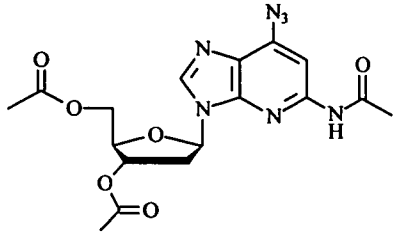
2-N, 3'-O, 5'-O-Tris-acetyl-2'-deoxyguanosine [8].



2'-Deoxyguanosine monohydrate (8.5g, 29.8mmol) was coevaporated three times with anhydrous pyridine (40ml), suspended in anhydrous pyridine (200ml), and cooled to 0°C. To this solution was added 4-dimethylaminopyridine (0.1eq, 3.0mmol, 370mg); and acetic anhydride (12eq, 358mmol, 33.7ml) was added dropwise with stirring. After the addition was complete, the reaction vessel was allowed to warm to room temperature. The reaction mixture was then refluxed at 110°C for 45min, with stirring, which resulted in a dark red-brown solution. This solution was concentrated *in vacuo*, coevaporated with toluene, and to the resulting residue was added a solution of saturated $NaHCO_3$ until effervescence subsided. The resulting solid was filtered, washed with water, and recrystallised from H_2O/MeOH (9:1) to give the product as pale pink-white crystals. Yield after 3 crops: 8.87g, (76%).

R_f 0.51 (solvent A), Rf 0.73 (solvent C). Found: C,65.26; H,5.02; N,10.83%. Requires: C,65.73; H,5.03; N,11.27%. ^1H nmr (200.13MHz; $CDCl_3$/d6-DMSO) δ_H: 1.87 (3, s, CH_3), 1.91 (3, s, CH_3), 2.07 (3, s, amide.CH_3), 2.32-2.42 (1, m, 2'-H), 2.67-2.81 (1, m, 2"-H), 4.07-4.24 (3, m, 5'-CH_2, 4'-H), 5.16-5.19 (1, m, 3'-H), 6.02-6.08 (1, t, 1'-H), 7.70 (1, br s, 8-H), 10.96 (1, br s, 2-NH, exch.), 11.99 (1, br s, 1-NH, exch.). 13C nmr (50.32MHz; $CDCl_3$/d6-DMSO) δ_C: 20.22 (CH_3), 20.34 (CH_3), 23.51 (CH_3), 36.41 (CH_2), 63.12 (CH_2), 73.85 (CH), 81.80 (CH), 83.86 (CH), 147.54 (CH), 155.0 (C), 169.56 (2xC), 170.02 (C), 172.70 (3xC). m/z (FAB) 394.13617 [$(M+H)^+$ calc. for $C_{16}H_{20}N_5O_7$ 394.13621]. λmax 206, 253, 259, 282nm.

2-N, 3'-O, 5'-O-Tris-acetyl-6-N-azido-2'-deoxyguanosine [9].

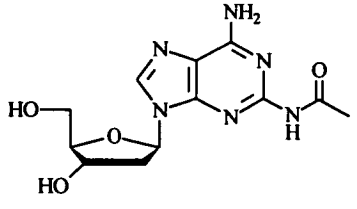


2-N, 3'-O, 5'-O-Tris-acetyl-2'-deoxyguanosine **[8]** (8g, 20.34mmol) was coevaporated three times with anhydrous dichloromethane (50ml), then dissolved in anhydrous dichloromethane (250ml). To this solution was added anhydrous

triethylamine (4eq, 81.36mmol, 11.3ml), 4-dimethylaminopyridine (0.1eq, 2.03mmol, 250mg), and mesitylene sulphonyl chloride (1.5eq, 30.51mmol, 6.67g) with stirring. After 2 hours the reaction vessel was cooled to 0°C and N-methyl pyrrolidine (15eq, 305mmol, 31.7ml) was added dropwise. After 15min, tetrabutylammonium azide (2eq, 40.68mmol, 11.6g) was added. After 5 hours stirring, (with the exclusion of light), the reaction mixture was then washed with saturated aqueous $NaHCO_3$, then saturated aqueous KCl, and the organic phase was dried over Na_2SO_4, filtered and concentrated *in vacuo* to leave a brown residue. This was purified by flash chromatography using methanol-dichloromethane (gradient 0-4% v/v) to give the product as a pale brown solid. Yield 6.8g, (80%).

Rf 0.68 (solvent A), Rf 0.17 (solvent H). Found: C,46.62; H,4.71; N,26.59%. Requires: C,45.93; H,4.34; N,26.79%. ^1H nmr (250.13MHz; $CDCl_3$) δ_H: 2.00 (3, s, CH_3), 2.07 (3, s, CH_3), 2.45 (3, s, amide.CH_3), 2.50-2.59 (1, m, 2'-H), 2.89-3.01 (1, m, 2''-H), 4.26-4.38 (3, m, 5'-CH_2, 4'-H), 5.36-5.41 (1, m, 3'-H), 6.28-6.33 (dd, 1, 1'-H), 8.01 (1, s, 8-H), 8.57 (1, s, 2-NH, exch.). ^{13}C nmr (62.90mhz; $CDCl_3$) δ_C: 20.51 (CH_3), 20.67 (CH_3), 25.02 (CH_3), 36.86 (CH_2), 63.41 (CH_2), 74.02 (CH), 82.32 (CH), 84.48 (CH), 120.81 (C), 141.34 (CH), 151.96 (C), 152.43 (C), 153.26 (C), 170.05 (C), 170.25 (2xC). m/z (FAB) 419.14326 [$(M+H)^+$ calc. for $C_{16}H_{19}N_8O_6$ 419.14276. λmax 199, 241, 298nm.

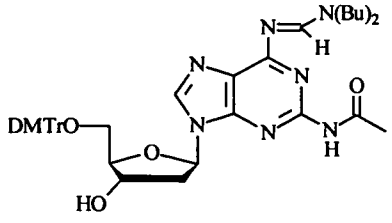
2-N-acetyl-2-amino-2'-deoxyadenosine [10].



2-N, 3'-O, 5'-O-Tris-acetyl-6-N-azido-2'-deoxyguanosine **[9]** (3.0g, 7.2mmol) was dissolved in pyridine-methanol (320ml, 3:1, v/v), and cooled to 0°C. To this solution was added 2M aqueous sodium hydroxide (3eq, 30.8ml) with stirring. The reaction vessel was allowed to warm to room temperature over 5min, after which time the reaction was quenched with Dowex resin (H^+ form). The Dowex resin was filtered

off and washed with pyridine and methanol and the solution was evaporated *in vacuo* to give a light brown foam. Palladium on activated charcoal (1g) was slurried with ethanol (10ml) in a 500ml hydrogenation reaction vessel, and the saponified product dissolved in methanol (200ml), was added. The vessel was attached to the hydrogenation apparatus, and slowly evacuated, opened to hydrogen to a pressure of 3Atm, and this process was repeated two times. At the end of the third time, the hydrogen pressure was allowed to reach 4Atm. The reaction vessel was agitated for 4 hours, after which time the vessel was evacuated slowly, opened to air, and the reaction mix was filtered through Celite and the catalyst was washed with hot methanol. (Caution: the catalyst should never be allowed to dry out). The resulting filtrate was evaporated in vacuo to give a light brown solid. This was washed with a small amount of dichloromethane to leave a white powder. Yield: 1.97g, (84%).
Rf 0.27 (solvent C), Rf 0.25 (solvent I). Found: C,46.26, H,5.38; N26.41%. Requires: C,46.75; H,5.23; N,27.26%. ^{1}H nmr (200.13MHz; d6-DMSO) δ_H: 2.20 (3, s, CH_3), 2.61-2.74 (1, m, 2'-H), 2.98-3.25 (1, m, 2''-H), 3.56-3.79 (2, m, 5'-CH_2), 3.85-3.86 (1, m, 4'-H), 4.43 (1, br s, 3'-H), 4.96 (1, br s, 5'-OH, exch.), 5.29-5.31 (1, m, 3'-OH, exch.), 6.24-6.31 (1, t, 1'-H), 7.21 (2, br s, 6-NH_2), 8.22 (1, br s, 8-H). ^{13}C nmr (90.56MHz; d6-DMSO) δ_C: 24.68 (CH_3), 39.82 (CH_2), 61.91 (CH_2), 70.97 (CH), 83.50 (CH), 87.89 (CH), 116.28 (C), 138.83 (CH), 150.11 (C), 152.99 (C), 156.21 (C), 169.78 (C). m/z (FAB) 309.13138 [$(M+H)^+$ calc. for $C_{12}H_{17}N_6O_4$ 309.13113]. λmax 197, 224, 274nm.

2-N-Acetyl-5'-O-4,4'-dimethoxytrityl-6-di-n-butylformamido-2-amino-2'-deoxyadenosine [11].

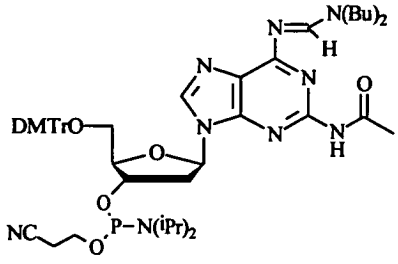


2-N-acetyl-2-amino-2'-deoxyadenosine **[10]** (620mg, 2.0mmol) was coevaporated from (3x10ml), and suspended in anhydrous pyridine (20ml). Di-n-butylformamide

dimethylacetal (4eq, 8.0mmol, 1.9ml) was added dropwise, and the reaction mixture was stirred for 4 hours at room temperature, then concentrated *in vacuo* to leave a foam. This was coevaporated three times with, and dissolved in anhydrous pyridine (20ml), with 4-dimethylaminopyridine (0.1eq, 0.2mmol, 24mg). 4,4'-Dimethoxytrityl chloride (1.2eq, 1.44mmol, 484mg) was added portionwise, and the reaction mixture was left stirring for 10 hours. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and washed three times with saturated aqueous KCl, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.5:99.5, v/v) until all the dimethoxytrityl alcohol had passed, then with a gradient of methanol in dichloromethane (0.5-1.5%, v/v). The product was obtained as a solid white foam. Yield: 0.70g, (78%).

Rf 0.86 (solvent A), Rf 0.53 (solvent D). Found: C,68.01; H,7.06; N,11.50%. Requires: C,67.27; H,6.86; N,13.08%. ^{1}H nmr (200.13MHz; $CDCl_3$) δ_H: 0.86-0.91 (6, t, form.CH_3), 1.25-1.40 (4, m, form.MeCH_2), 1.51-1.65 (4, m, form.EtCH_2), 2.36 (3, br s, amide.CH_3), 2.59-2.67 (2, m, 2'-CH_2), 3.31-3.38 (4, m, form.PrCH_2), 3.60-3.68 (2, t, 5'-CH_2), 3.72 (6, s, OCH_3), 4.18-4.19 (1, m, 4'-H), 4.63 (1, br s, 3'-H), 6.44-6.48 (1, m, 1'-H), 6.72-7.39 (13, m, DMTr-H), 7.94 (1, s, 8-H), 8.20 (1, br s, 2-NH, exch.), 8.99 (1, s, form.CH). ^{13}C nmr (50.32MHz; $CDCl_3$) δ_C: 11.09 (CH_3), 13.46 (CH_3), 13.70 (CH_3), 19.49 (CH_2), 19.96 (CH_2), 28.98 (CH_2), 30.65 (CH_2), 40.38 (CH_2), 44.98 (CH_2), 51.79 (CH_2), 54.96 (2xCH_3), 63.86 (CH_2), 71.86 (CH), 83.85 (CH), 86.12 (CH, C), 112.89 (4xCH), 122.61 (C), 126.60 (CH), 127.61 (2xCH), 127.88 (2xCH), 129.79 (4xCH), 135.53 (2xC), 139.24 (CH), 144.41 (C), 151.72 (C), 152.10 (C), 158.20 (3xC), 158.86 (CH), 160.11 (C). m/z (FAB) 750.39792 [$(M+H)^+$ calc. for $C_{42}H_{52}N_7O_6$ 750.39788]. λmax 200, 235, 259, 319nm.

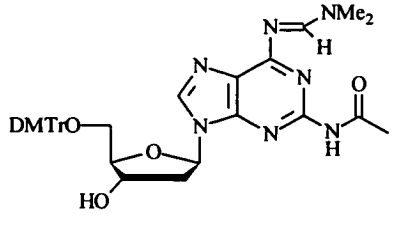
2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2-amino-2'-deoxyadenosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite [12].



2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2-amino-2'-deoxy-adenosine **[11]** (380mg, 0.51mmol), was coevaporated from anhydrous tetrahydrofuran three times, and dissolved in anhydrous tetrahydrofuran (4ml). Anhydrous diisopropylethylamine (4eq, 2.04mmol, 0.34ml) was added, and the reaction flask was kept under argon. 2-Cyanoethyl-N,N-diisopropylchloro-phosphoramidite[177] (1.2eq, 0.61mmol, 0.13ml) was added dropwise, and after a few minutes, a white precipitate (salt) was seen to appear. After 45min the reaction was quenched with by the addition of ethyl acetate (argon purged), and the organic layer was washed with saturated aqueous KCl before being quickly dried over anhydrous Na_2SO_4, filtered under argon, and evaporated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with ethyl acetate-triethylamine 99:1, v/v, then washed with ethyl acetate), eluting with ethyl acetate. The product was obtained as a white foam. Yield: 480mg, (99%).

Rf 0.70 (solvent A), Rf 0.92 (solvent B). ^{31}P nmr (90MHz; MeCN; $(CD_3)_2CO$ lock) δ_P: 147.67 (s). m/z (FAB) 951.51432 [$(M+H)^+$ calc. for $C_{51}H_{68}N_9O_7P$ 951.51359].

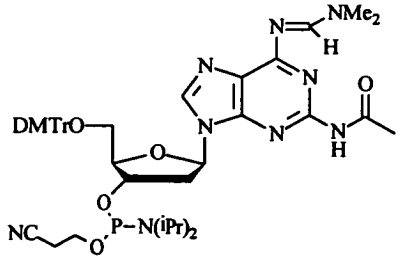
2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-dimethylformamido-2-amino-2'-deoxyadenosine [13].



2-N-acetyl-2-amino-2'-deoxyadenosine **[10]** (740mg, 2.4mmol) was coevaporated from (3x10ml), and suspended in anhydrous pyridine (25ml). Dimethylformamide dimethylacetal (4eq, 9.6mmol, 0.77ml) was added dropwise, and the reaction mixture was stirred for 2 hours at room temperature, then concentrated *in vacuo* to leave a foam. This was coevaporated three times with, and dissolved in anhydrous pyridine (25ml), with 4-dimethylaminopyridine (0.1eq, 0.24mmol, 30mg). 4,4'-Dimethoxytrityl chloride (1.2eq, 2.9mmol, 970mg) was added portionwise, and the reaction mixture was left stirring for 15 hours. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and washed three times with saturated aqueous KCl, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.5:99.5, v/v) until all the dimethoxytrityl alcohol had passed, then with a gradient of methanol-dichloromethane (0.5-2.0%, v/v). The product was obtained as a solid white foam. Yield: 978mg, (61%).

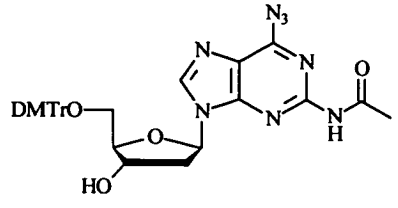
Rf 0.76 (solvent A), Rf 0.25 (solvent D). ^{1}H nmr (200.13MHz; $CDCl_3$) δ_H: 2.33 (3, s, amide.CH_3), 2.57-2.67 (2, m, 2'-CH_2), 3.10 (3, s, form.CH_3), 3.14 (3, s, form.CH_3), 3.26-3.35 (2, m, 5'-CH_2), 3.70 (6, s, OCH_3), 4.21-4.22 (1, m, 4'-H), 4.65 (1, br s, 3'-H), 6.47-6.50 (1, br t, 1'-H), 6.70-7.38 (13, m, DMTr-H), 7.94 (1, s, 8-H), 8.48 (1, br s, 2-NH, exch.), 8.92 (1, s, form.CH). ^{13}C nmr (50.32MHz; $CDCl_3$) δ_C: 24.70 (CH_3), 34.72 (CH_3), 40.42 (CH_2), 41.06 (CH_3), 54.84 (2xCH_3), 63.84 (CH_2), 71.92 (CH), 83.77 (CH), 85.97 (C), 86.22 (CH), 112.77 (4xCH), 122.48 (C), 126.48 (CH), 127.49 (2xCH), 127.77 (2xCH), 129.68 (4xCH), 135.44 (2xC), 139.31 (CH), 144.29 (C), 151.71 (C), 152.11 (C), 158.07 (3xC), 158.71 (CH), 159.55 (C). m/z (FAB) 666.29960 [$(M+H)^+$ calc. for $C_{36}H_{40}N_7O_6$ 666.30401].

2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-dimethylformamido-2-amino-2'-deoxyadenosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite [14].



2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-dimethylformamido-2-amino-2'-deoxy-adenosine **[14]** (500mg, 0.75mmol) was coevaporated from anhydrous tetrahydrofuran three times, and dissolved in anhydrous tetrahydrofuran (7ml). Anhydrous diisopropylethylamine (4eq, 3.0mmol, 0.50ml) was added, and the reaction flask was kept under argon. 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite (1.2eq, 0.90mmol, 0.19ml) was added dropwise, and after a few minutes, a white precipitate (salt) was seen to appear. After 45min the reaction was quenched with by the addition of ethyl acetate (argon purged), and the organic layer was washed with saturated aqueous KCl before being quickly dried over anhydrous Na_2SO_4, filtered under argon, and evaporated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with ethyl acetate-triethylamine (99:1, v/v), then washed with ethyl acetate), eluting with ethyl acetate. The product was obtained as a glassy oil. Yield: 200mg, (31%).
Rf 0.79 (solvent B), Rf 0.75 (solvent D). 865.40315 [$(M+H)^+$ calc. for $C_{45}H_{57}N_9O_7P$ 865.40404]. ^{31}P nmr - all sample used for oligo. synthesis.

2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-azido-2'-deoxyguanosine [15].

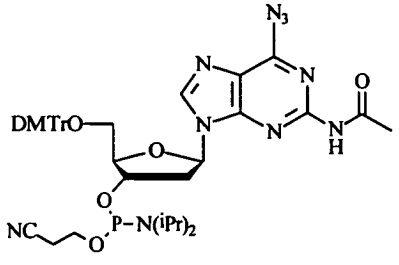


2-N, 3'-O, 5'-O-Tris-acetyl-6-N-azido-2'-deoxyguanosine **[9]** (380mg, 0.91mmol) was dissolved in pyridine-methanol (50ml, 3:1, v/v), and cooled to 0°C. To this solution was added 2M aqueous sodium hydroxide (3eq, 3.9ml) with stirring. The reaction vessel was allowed to warm to room temperature over 5min, after which

time the reaction was quenched with Dowex resin (H^+ form). The Dowex resin was filtered off and washed with pyridine and methanol and the solution was evaporated *in vacuo* to give a light brown foam. This was coevaporated three times with, and dissolved in anhydrous pyridine (10ml), with 4-dimethylaminopyridine (0.1eq, 0.1mmol, 12mg). 4,4'-Dimethoxytrityl chloride (1.2eq, 1.1mmol, 332mg) was added portionwise, and the reaction mixture was left stirring for 10 hours. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and washed three times with saturated aqueous KCl, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.5:99.5, v/v) until all the dimethoxytrityl alcohol had passed, then with a gradient of 0.5-4.0%, v/v methanol in dichloromethane. The product was obtained as a solid white foam. Yield: 446mg, (77%).

Rf 0.72 (solvent A). ^1H nmr (200.13MHz; $CDCl_3$) δ_H: 2.46 (3, s, CH_3), 2.70-2.75 (2, m, 2'-CH_2), 3.24-3.39 (2, m, 5'-CH_2), 3.73 (3, s, OCH_3), 3.80-4.05 (1, br s, 3'-OH, exch.), 4.17-4.18 (1, m, 4'-H), 4.60-4.63 (1, m, 3'-H), 6.39-6.46 (1, t, 1'-H), 6.71-7.37 (13, m, DMTr-H), 8.03 (1, s, 8-H), 8.20-8.60 (1, br s, 2-NH, exch.). m/z (FAB) 637.25241 [$(M+H)^+$ calc. for $C_{33}H_{33}N_8O_6$ 637.25229]. λmax 196, 226, 236, 300nm.

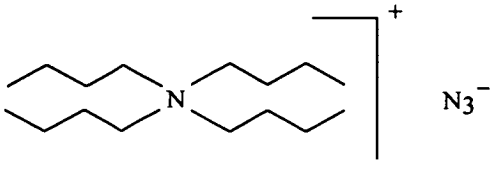
2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-azido-2'-deoxyguanosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite [16]. (Attempted phosphitylation of the 3'-OH).



2-N-Acetyl-5'-O-(4,4'-dimethoxytrityl)-6-N-azido-2'-deoxyguanosine **[15]** (320mg, 0.5mmol), was coevaporated from anhydrous tetrahydrofuran three times, and

dissolved in anhydrous tetrahydrofuran (4ml). Anhydrous diisopropylethylamine (4eq, 2mmol, 0.34ml) was added, and the reaction flask was kept under argon. 2-Cyanoethyl-N,N-diisopropylchlorophosphoramidite[n54] (1.2eq, 0.6mmol, 0.13ml) was added dropwise, and after a few minutes, a white precipitate (salt) was seen to appear. After 45min the reaction was quenched with by the addition of ethyl acetate (argon purged), and the organic layer was washed with saturated aqueous KCl before being quickly dried over anhydrous Na_2SO_4, filtered under argon, and evaporated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with ethyl acetate-triethylamine 99:1, v/v, then washed with ethyl acetate), eluting with ethyl acetate. The product obtained consisted of a long streak of two spots on tlc. (The N^6-azide was phosphitylated in addition to the 3'-OH). Furthermore, attempted oligonucleotide synthesis using this monomer failed.

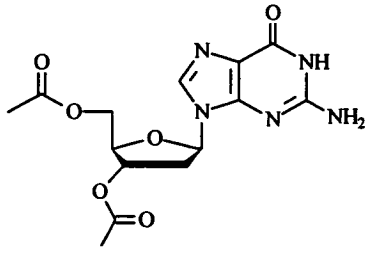
Tetrabutylammonium azide



A solution of tetrabutyl ammonium hydrogen sulphate (6.8g, 0.02 mol) in water (10.0ml) was stirred and treated at room temperature with aqueous NaOH solution (5.0ml, 10% w/v), then with a solution of sodium azide (2.6g, 0.04mol) in water (5.0ml). The resulting cloudy mixture was stirred at room temperature for 10min then extracted with dichloromethane (2x20ml) and the aqueous mother liquor discarded. The combined dichloromethane extracts were dried with anhydrous Na_2SO_4 and concentrated *in vacuo* to give the crude product as a pale yellow oil. This was azeotroped with anhydrous benzene to give the product as a cream-coloured waxy solid which was stored frozen, and in the dark. Yield: 5.6g, (98%).

^1H nmr (250.13MHz; $CDCl_3$) δ_H: 0.87-0.92 (12, t, CH_3), 1.27-1.42 (8, m, $MeCH_2$), 1.51-1.64 (8, m, $EtCH_2$), 3.21-3.28 (8, m, $PrCH_2$). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 13.27 (4xCH_3), 19.35 (4xCH_2), 23.46 (4xCH_2), 58.11 (4xCH_2). m/z (FAB) 242.28434 [M^+ calc. for $C_{16}H_{36}N$ fragment 242.28478].

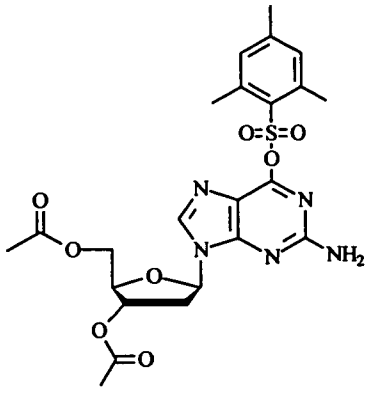
3'-O,5'-O-(Bis-acetyl)-2'-deoxyguanosine [17].



2'-Deoxyguanosine monohydrate (10g, 35mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (113ml) and anhydrous N,N-dimethylformamide (180ml). To this suspension was added acetic anhydride (12eq, 0.42mol, 39.7ml). The reaction mixture was stirred with the exclusion of light at room temperature for 48 hours. The precipitate (salt) was filtered off, and ethanol (20ml) was added to the supernatant, and stirred for 10min. This solution was concentrated *in vacuo* until crystallisation commenced. To this suspension was added diethyl ether (800ml) at 4°C to precipitate the product. The white solid was filtered off, and the product was recrystallised from ethanol-water (9:1, v/v) to give white crystals. Yield: 11.3g, (92%).

Rf 0.20 (solvent A), Rf 0.57 (solvent C), Rf 0.59 (solvent J). Found: C,45.72; H,5.16; N,18.61%. $C_{14}H_{17}N_5O_6.(H_2O)$ requires: C,45.51; H,5.19; N,18.97%. ^1H nmr (250.13MHz; d6-DMSO) δ_H: 2.03 (3, s, CH_3), 2.07 (3, s, CH_3), 2.40-2.51 (1, m, 2'-H), 2.85-2.97 (1, m, 2"-H), 4.14-4.28 (3, m, 4'-H, 5'-CH_2), 5.27-5.30 (1, "d", 3'-H), 6.10-6.16 (1, dd, 1'-H), 6.53 (2, s, 2-NH_2, exch.), 7.92 (1, s, 8-H), 10.74 (1, s, 1-NH, exch.). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 20.73 (CH_3), 20.97 (CH_3), 35.64 (CH_2), 63.82 (CH_2), 74.67 (CH), 81.67 (CH), 82.80 (CH), 116.94 (C), 135.42 (CH), 151.30 (C), 153.94 (C), 156.96 (C), 170.22 (C), 170.40 (C). m/z (FAB) 352.12480 [$(M+H)^+$ calc. for $C_{14}H_{18}N_5O_6$ 352.12571]. λmax 197, 255nm.

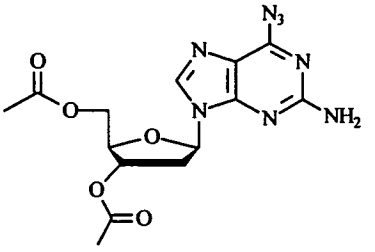
3'-O,5'-O-Bis-acetyl-6-O-mesitylene sulphonyl -2'-deoxyguanosine [18].



3'-O,5'-O-Bis-acetyl-2'-deoxyguanosine **[17]** (7.0g, 19.9mmol) was coevaporated three times with and then suspended in dichloromethane (160ml) with anhydrous triethylamine (4eq, 0.08mol, 11.1ml) and 4-dimethylaminopyridine. Mesitylene sulphonyl chloride (3eq, 59.7mmol, 13.025g) was added with stirring, and the solution was seen to turn dark red in colour. After 1 hour the solvents were evaporated *in vacuo,* coevaporated with methanol, and a solution of methanol-diethyl ether (8:2, v/v) was added to the dark red residue. The resulting solid was filtered and washed with diethyl ether to leave a white powder. Yield: 6.21g, (58%).

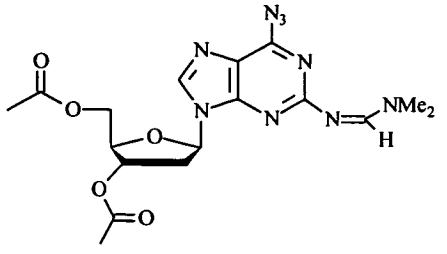
Rf 0.65 (solvent A), Rf 0.58 (solvent H). Found: C,52.20; H,5.48; N,13.16%. Requires: C, 51.58; H,5.46; N,13.08%. ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 2.23 (3, s, para-CH_3), 2.43-2.52 (1, m, 2'-H), 2.66 (6, s, ortho-CH_3), 2.84-2.96 (1, m, 2''-H), 4.26-4.40 (3, m, 5'-CH_2, 4'-H), 5.14 (2, s, 2-NH_2, exch.), 5.33-5.35 (1, m, 3'-H), 6.19-6.24 (1, t, 1'-H), 6.90 (2, s, Ar-H), 7.83 (1, s, 8-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 20.47 (CH_3), 20.62 (CH_3), 20.79 (CH_3), 22.42 (2xCH_3), 36.25 (CH_2), 63.44 (CH_2), 74.18 (CH), 82.06 (CH), 84.23 (CH), 116.34 (C), 131.40 (2xCH), 131.82 (C), 139.78 (CH), 139.97 (2xC), 143.75 (C), 154.73 (C), 155.12 (C), 158.32 (C), 170.06 (C), 170.34 (C). m/z (FAB) 536.17871 [$(M+3H)^+$ calc. for $C_{23}H_{30}N_5O_8S$ 536.18151]. λmax 204, 238, 291nm.

3'-O,5'-O-Bisacetyl-6-azido-2'-deoxyguanosine [19].



3'-O,5'-O-Bisacetyl-6-O-mesitylene sulphonyl-2'-deoxyguanosine **[18]** (6.2g, 11.6 mmol) was coevaporated three times with and then dissolved in dichloromethane (100ml). N-Methyl pyrrolidine (15eq, 174mmol, 18.4ml) was added with stirring, and after 20min, tetrabutylammonium azide (2eq, 23.2mmol, 6.60g) was added. After 5 hours, the reaction mixture was washed with saturated aqueous $NaHCO_3$ solution, then saturated aqueous KCl solution. The organic phase was then dried with anhydrous Na_2SO_4 and evaporated *in vacuo.* The resulting residue was purified by flash-column chromatography eluting with a methanol-dichloromethane gradient (0.5-2.0%, v/v). The product was obtained as an off-white foam. Yield: 3.9g, (89%). Rf 0.58 (solvent A), Rf 0.81 (solvent C), Rf 0.37 (solvent H). Found: C,44.26; H,4.45; N,28.52%. Requires: C,44.68; H,4.29; N,29.78%. ^1H nmr (250.13MHz; d6-DMSO) δ_H: 2.02 (3, s, CH_3), 2.09 (3, s, CH_3), 2.55-2.64 (1, m, 2'-H), 3.00-3.11 (1, m, 2"-H), 3.40 (1, s, 4'-H), 4.20-4.31 (2, m, 5'-CH_2), 5.35-5.37 (1, "d", 3'-H), 6.33-6.39 (1, dd, 1'-H), 8.37 (1, s, 8-H), 8.49 (2, s, 2-NH_2, exch.). ^{13}C nmr (62.90MHz; d6-DMSO) δ_C: 20.72 (CH_3), 20.98 (CH_3), 35.86 (CH_2), 63.78 (CH_2), 74.60 (CH), 81.97 (CH), 83.56 (CH), 112.28 (C), 138.14 (CH), 144.05 (C), 144.63 (C), 146.16 (C), 170.25 (C), 170.39 (C). m/z (FAB) 377.13426 [$(M+H)^+$ calc. for $C_{14}H_{17}N_8O_5$ 377.13219]. λmax 208, 270, 299nm.

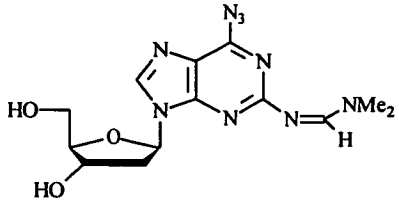
3'-O,5'-O-Bis-acetyl-2-N-dimethylformamido-6-N-azido-2'-deoxyguanosine [20].



3'-O,5'-O-Bis-acetyl-6-N-azido-2'-deoxyguanosine **[19]** (2.0g, 5.3mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (25ml). To this dimethylformamide dimethylacetal (5eq, 26.5mmol, 3.5ml) was added dropwise, and the reaction mixture was stirred for 1 hour at room temperature, then concentrated *in vacuo*, and coevaporated three times with toluene. The resulting residue was purified by flash-column chromatography eluting with a methanol-dichloromethane gradient (0.0-1.5%, v/v). The product was obtained as an off-white foam. Yield: 1.56g, (68%).

Rf 0.45 (solvent A), Rf 0.88 (solvent C), Rf 0.15 (solvent H). ^{1}H nmr (200.13MHz; $CDCl_3$) δ_H: 2.09 (3, s, acetyl.CH_3), 2.13 (3, s, acetyl.CH_3), 2.61-2.70 (1, m, 2'-H), 2.94-3.11 (1, m, 2''-H), 3.23 (3, s, form.CH_3), 3.30 (3, s, form.CH_3), 4.12-4.36 (3, m, 5'-CH_2, 4'-H), 5.51-5.54 (1, m, 3'-H), 6.28-6.34 (1, dd, 1'-H), 7.94 (1, s, 8-H), 8.69 (1, s, form.CH). ^{13}C nmr (50.32MHz; $CDCl_3$) δ_C: 20.37 (CH_3), 20.66 (CH_3), 35.67 (CH_3), 36.71 (CH_2), 41.47 (CH_3), 63.28 (CH_2), 73.62 (CH), 81.89 (CH), 84.49 (CH), 115.83 (C), 138.88 (CH), 142.46 (C), 146.09 (C), 147.35 (C), 158.21 (CH), 170.19 (C), 170.24 (C). m/z (FAB) 432.17482 [$(M+H)^+$ calc. for $C_{17}H_{22}N_9O_5$ 432.17437].

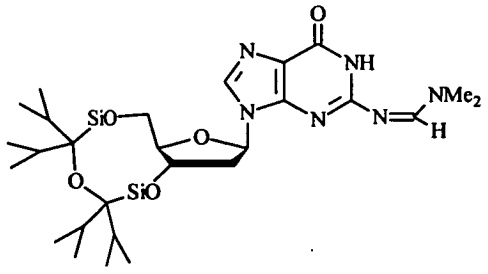
2-N-Dimethylformamido-6-N-azido-2'-deoxyguanosine [21].(Attempted deprotection of the 3',5'-O-groups).



3'-O,5'-O-Bis-acetyl-2-N-dimethylformamido-6-N-azido-2'-deoxyguanosine **[20]** (1.1g, 2.6mmol) was dissolved in pyridine-methanol (110ml, 3:1, v/v), and cooled to 0°C. To this solution was added 2M aqueous sodium hydroxide (3eq, 10.5ml) with stirring. The reaction vessel was allowed to warm to room temperature over 5min, after which time the reaction was quenched with Dowex resin (H^+ form). The Dowex resin was filtered off and washed with pyridine and methanol.

The N^2-dimethylformamidine protecting group was found to be cleaved after less than one minute in the conditions used to cleave the 3', 5'-O-protection.

3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-dimethylformamido-2'-deoxyguanosine [22].

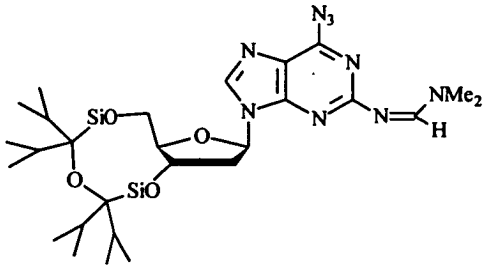


2'-Deoxyguanosine monohydrate (6g, 21.0mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (210ml), along with 4-dimethylaminopyridine (0.1eq, 2.1mmol, 250mg). 1,3-Dichloro-tetraisopropyl-disiloxane (1.2eq, 25.2mmol, 7.9ml) was added and the reaction mixture was stirred for 15 hours at room temperature, until the reaction reached completion (by tlc). Dimethylformamide dimethylacetal (5eq, 105.0mmol, 14.0ml) was added dropwise with stirring, and after 2 hours at room temperature, H_2O (20ml) was added to the cloudy-yellow reaction mixture. The solvents were removed *in vacuo*, and the residue was covaporated with toluene. The residue was then dissolved in dichloromethane and washed three times with saturated aqueous KCl, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v),eluting with a gradient of (0.0-2.0%, v/v) methanol in dichloromethane. The product was obtained as an off-white solid foam. Yield: 10.1g, (85%).

Rf 0.68 (solvent A), Rf 0.12 (solvent H). Found: C,51.71; H,8.61; N,13.52%. Requires: C,51.52; H,7.96; N,14.43%. ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 0.91-1.08 (28, m, TIPS-CH_3, TIPS-CH), 2.46-2.54 (2, m, 2'-CH_2), 3.02 (3, s, form.CH_3), 3.11 (3, s, form.CH_3), 3.76-3.82 (1, m, 4'-H), 3.94-3.98 (2, m, 5'-CH_2), 4.57-4.66 (1, dd, 3'-H), 6.15-6.19 (1, m, 1'-H), 7.75 (1, s, 8-H), 8.49 (1, s, form.CH). ^{13}C nmr

(62.90MHz; $CDCl_3$) δ_C: 12.19 (CH), 12.65 (CH), 12.81 (CH), 13.11 (CH), 16.60 (8xCH_3), 34.92 (CH_3), 41.08 (CH_2), 41.17 (CH_3), 61.49 (CH_2), 69.60 (CH), 81.58 (CH), 84.74 (CH), 119.98 (C), 135.28 (CH), 149.45 (C), 156.75 (C), 157.85 (CH), 158.32 (C). m/z (FAB) 565.30326 [$(M+H)^+$ calc. for $C_{25}H_{45}N_6O_5Si_2$ 565.29900]. λmax 198. 237, 305nm.

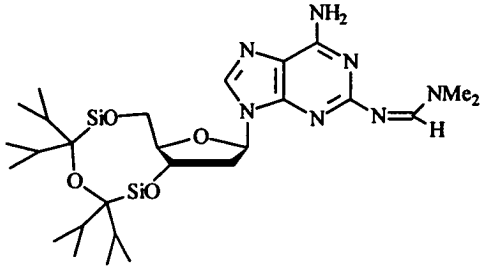
3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-dimethylformamido-6-N-azido-2'-deoxyguanosine [23].



3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-dimethylformamido-2'-deoxyguanosine **[22]** (5.0g, 8.85mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (90ml) along with 4-dimethylaminopyridine (0.1eq, 0.89mmol, 110mg) and anhydrous triethylamine (4eq, 35.4mmol, 4.9ml). To this was added mesitylene sulphonyl chloride (4eq, 35.4mmol, 7.7g) and left for 3 hours, stirring at room temperature. (The reaction mixture went briefly purple, then turned dark red in colour). N-Methyl pyrrolidine (30eq, 266mmol, 28.1ml) and tetrabutylammonium azide (4eq, 35.4mmol, 10.1g) were then added portionwise, and the reaction mixture was stirred for 20 hours at room temperature with the exclusion of light. The dark-brown reaction mixture was then filtered to remove any precipitate, and the filtrate was evaporated *in vacuo* to remove the solvents. The resulting residue was washed with saturated aqueous $NaHCO_3$ solution, then saturated aqueous KCl solution, and the organic phase was then dried with anhydrous Na_2SO_4 and evaporated *in vacuo.* The residue was coevaporated with toluene, then ethanol (three times), and treated with ethanol (50ml) and left for 15 hours at 0°C. The resulting solid was filtered, washed with ethanol, and recrystallised from ethanol, to leave white crystals. Yield: 3.1g, (60%).

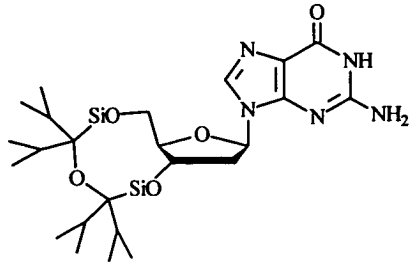
Rf 0.79 (solvent A). Rf 0.40 (solvent H). Found: C,50.48; H,7.69; N,21.21%. Requires: C,50.91; H,7.35; N,21.39%. ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 0.89-1.07 (28, m, TIPS-CH_3, TIPS-CH), 2.52-2.67 (2, m, 2'-CH_2), 3.20 (3, s, form.CH_3), 3.27 (3, s, form.CH_3), 3.79-3.84 (1, m, 4'-H), 3.97-3.98 (2, d, 5'-CH_2), 4.58-4.67 (1, m, 3'-H), 6.22-6.26 (1, m, 1'-H), 8.01 (1, s, 8-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 12.23 (CH), 12.66 (CH), 12.80 (CH), 13.13 (CH), 16.62-17.23 (8xCH_3), 35.58 (CH_3), 40.10 (CH_2), 41.63 (CH_3), 61.10 (CH_2), 69.05 (CH), 82.41 (CH), 84.88 (CH), 115.52 (C), 137.45 (CH), 142.21 (C), 146.19 (C), 147.21 (C), 158.00 (CH). m/z (FAB) 590.30338 [$(M+H)^+$ calc. for $C_{25}H_{44}N_9O4Si_2$ 590.30548]. λmax 198, 206, 287nm.

3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-dimethylformamido-6-amino-2'-deoxyguanosine [24]. (Attempted hydrogenation of N^6-azide).



3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-dimethylformamido-6-N-azido-2'-deoxyguanosine **[23]** was dissolved in an appropriate solvent and subjected to a variety of hydrogenation conditions and reagents in order to reduce the azide, without a reaction occurring at the N^2-dimethylformamido protecting group.
These conditions are outlined in Section 5.6.

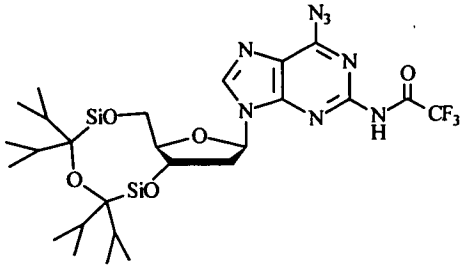
3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-deoxyguanosine [25].



2'-Deoxyguanosine monohydrate (4.5g, 15.8mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (160ml), along with 4-dimethylaminopyridine (0.1eq, 1.58mmol, 193mg). 1,3-Dichloro-tetraisopropyldi-siloxane (1.3eq, 20.5mmol, 6.5ml) was added and the reaction mixture was stirred for 12 hours at room temperature. Methanol (20ml) was added, and the solvents were evaporated *in vacuo*. The resulting residue was coevaporated with toluene, then methanol, and treated with H_2O-methanol (50:50, v/v) to give a white precipitate. This was filtered, and washed with methanol to leave a white powder. Yield: 7.16g, (89%).

Rf 0.28 (solvent A), Rf 0.57 (solvent C). Found: C,51.58; H,8.01; N,15.43%. Requires: C,51.84; H,7.72; N,15.70%. ^{1}H nmr (360.14MHz; d6-DMSO) δ_H: 0.74-0.90 (28, m, TIPS-CH_3, TIPS-CH), 2.45-2.50 (1, m, 2'-CH), 2.58-2.60 (1, m, 2"-CH), 3.91-3.92 (2, m, 5'-CH_2), 4.65-4.67 (1, dd, 4'-H), 5.61-5.62 (1, m, 3'-H), 6.04-6.07 (1, m, 1'-H), 6.43 (2, br s, 2-NH_2, exch.), 7.75 (1, s, 8-H), 10.68 (1, s, 1-NH, exch.). ^{13}C nmr (62.90MHz; d6-DMSO) δ_C: 12.19 (CH), 12.39 (CH), 12.72 (CH), 12.88 (CH), 13.07-17.41 (8xCH_3), 39.09 (CH_2), 62.47 (CH_2), 71.16 (CH), 81.13 (CH), 84.69 (CH), 116.90 (C), 134.70 (CH), 150.77 (C), 153.92 (C), 156.82 (C). m/z (FAB) 510.25718 [$(M+H)^+$ calc. for $C_{22}H_{40}N_5O_5Si_2$ 510.25680]. λmax 198, 256nm.

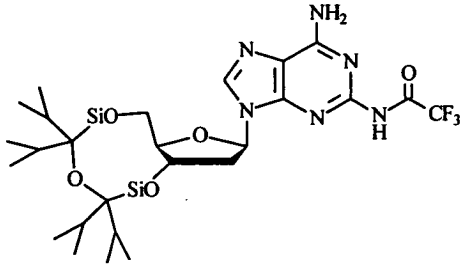
3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-trifluoroacetyl-6-N-azido- 2'-deoxyguanosine [26].



3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2'-deoxyguanosine **[25]** (3.5g, 6.87 mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (70ml). The reaction flask was then cooled to 0°C, and trifluoroacetic anhydride (5eq, 33.5mmol, 4.8ml) was added with stirring to give a

yellow solution. After 20min at room temperature, sodium azide (5eq, 33.5mmol, 2.23g) was added portionwise to give a purple solution, and the reaction mixture was stirred at room temperature, with the exclusion of light, for another 5 hours. The remaining sodium azide was filtered, and the volume of the filtrate was reduced to ca. 20ml. Saturated aqueous $NaHCO_3$ solution (150ml) was added, and the mixture was extracted with dichloromethane (3x250ml). The organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with a gradient of (0.0-1.5%, v/v) methanol in dichloromethane. The product was obtained as a purple foam. Yield: 2.87g, (66%).
Rf 0.37 (solvent A), Rf 0.72 (solvent C). ^1H nmr (250.13MHz; $CDCl_3$) δ_H: 0.89-0.99 (28, m, TIPS-CH_3, TIPS-CH), 2.05-2.06 (2, m, 2'-CH_2), 3.77-3.80 (1, m, 4'-H), 3.97 (2, m, 5'-CH_2), 4.60-4.67 (1, dd, 3'-H), 6.35-6.39 (1, m, 1'-H), 8.10 (2, m, 2-NH, exch.; 8-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 12.14 (CH), 12.52 (CH), 12.70 (CH), 13.01 (CH), 16.52-17.14 (8xCH_3), 40.19 (CH_2), 61.00 (CH_2), 68.84 (CH), 82.93 (CH), 84.90 (CH), 110.84-124.54 (quartet, J=287.6Hz, CF_3), 115.19 (C), 138.01 (CH), 142.87 (C), 145.81 (C), 146.76 (C), 159.66-161.29 (quartet, J=34.1Hz, $\underline{C}OCF_3$). m/z (FAB) 631.24823 [$(M+H)^+$ calc. for $C_{24}H_{38}N_8O_5Si_2F_3$ 631.24558]. λmax 201, 310nm.

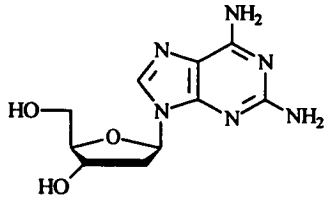
3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-trifluoroacetyl-6-amino-2'-deoxyguanosine [27]. (Attempted hydrogenation of N^6-azide).



3',5'-O-(1,1,3,3-Tetraisopropyl-1,3-disiloxanediyl)-2-N-trifluoroacetyl-6-N-azido-2'-deoxyguanosine **[26]** was dissolved in an appropriate solvent and subjected to a

variety of hydrogenation conditions and reagents in order to reduce the azide, without a reaction occurring at the N^2-trifluoroacetyl protecting group.
These conditions are outlined in Section 5.6.

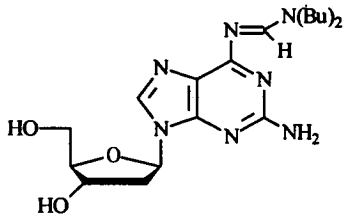
2-Amino-2'-deoxyadenosine [28].



2'-Deoxyguanosine monohydrate (3.5g, 12.3mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (120ml). The reaction flask was then cooled to 0°C, and trifluoroacetic anhydride (8eq, 98.4mmol, 13.8ml) was added dropwise with stirring; and left at room temperature after addition was complete (reaction mixture turned from dark yellow to red). After 30min, sodium azide (5eq, 61.5mmol, 2.8g) was added, and left for 1 hour at room temperature, with the exclusion of light, to leave a brown-coloured solution. The remaining sodium azide was filtered, and the volume of the filtrate was reduced to ca. 40ml. Saturated aqueous $NaHCO_3$ solution (250ml) was added, and the mixture was extracted with dichloromethane (3x400ml). The organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. After coevaporation with ethanol, the residue was dissolved in ethanol (120ml), and Raney nickel (7ml) was added portionwise with stirring (this reaction is exothermic), and then gently refluxed at ca. 60°C for 2 hours with stirring. The Raney nickel was filtered off using Celite filtering agent, washed with hot ethanol (200ml) (caution: Raney nickel should never be allowed to dry), and the ethanol was removed *in vacuo*. Diethyl ether (100ml) was added to the residue, and the mixture was left at 0°C for 10 hours (a little petrolium ether [40-60] was added to encourage crystallisation). The resulting white crystals/precipitate were filtered and washed with diethyl ether. Yield: 2.64g, (81%). Rf 0.10 (solvent A), Rf 0.34 (solvent C), Rf 0.67 (solvent I). Found: C,42.71; H,5.65; N,30.55%. $C_{10}H_{14}N_6O_3.(3/4H_2O)$ requires: C,42.92; H,5.59; N,30.05%. ^1H nmr

(250.13MHz; d6-DMSO) δ_H: 2.15-2.22 (1, m, 2'-H), 2.49-2.65 (1, m, 2"-H), 3.53-3.62 (2, m, 5'-CH_2), 3.85 (1, m, 4'-H), 4.36-4.37 (1, m, 3'-H), 5.36-5.37 (2, br "d", 5'-OH exch., 3'-OH exch.), 5.86 (2, br s, 2-NH_2, exch.), 6.15-6.20 (1, dd, 1'-H), 6.88 (2, br s, 6-NH_2, exch.), 7.93 (1, s, 8-H). ^{13}Cnmr (62.90MHz; d6-DMSO) δ_C: 39.6 (CH_2), 62.21 (CH_2), 71.27 (CH), 83.41 (CH), 87.91 (CH), 113.62 (C), 136.16 (CH), 151.39 (C), 156.40 (C), 160.27 (C). m/z (FAB) 267.12124 [$(M+H)^+$ calc.for $C_{10}H_{15}N_6O_3$ 267.12056]. λmax 217, 256, 280nm.

2-Amino-6-N-di-n-butylformamido-2'-deoxyadenosine [29].

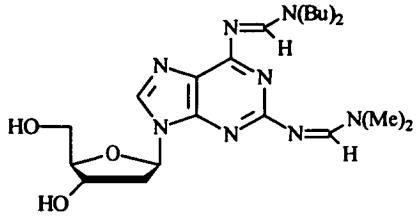


2-Amino-2'-deoxyadenosine **[28]** (2.46g, 8.65mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (90ml). Di-n-butylformamide dimethyl acetal (5eq, 43.25mmol, 10.5ml) was added dropwise with stirring over 30min. After 3 hours the reaction was quenched by the addition of water (20ml), and stirred for another 10min. The solvents were then evaporated *in vacuo*, coevaporated with toluene, and the product was purified by flash-column chromatography, eluting with a gradient of (0.0-3.0%, v/v) methanol in dichloromethane. The product was obtained as a light yellow foam. Yield: 1.89g, (54%).

Rf 0.67 (solvent C), Rf 0.37 (solvent I). ^{1}H nmr (250.13MHz; d6-DMSO) δ_H: 0.84-0.93 (6, m, CH_3), 1.20-1.37 (4, m, $MeCH_2$), 1.51-1.59 (4, m, $EtCH_2$), 2.16-2.23 (1, m, 2'-H), 2.54-2.65 (1, m, 2'-H), 3.31-3.63 (6, m, 5'-CH_2, $PrCH_2$), 3.85 (1, d, 4'-H), 4.37 (1, br s, 3'-H), 5.19 (1, br s, 5'-OH, exch.), 5.33 (1, br s, 3'-OH, exch.), 6.00 (2, br s, 2-NH_2, exch.), 6.20-6.26 (1, t, 1'-H), 8.04 (1, s, 8-H), 8.81 (1, s, form.CH). ^{13}C nmr (62.90MHz; d6-DMSO) δ_C: 13.69 (CH_3), 13.89 (CH_3), 19.31 (CH_2), 19.79 (CH_2), 28.81 (CH_2), 30.69 (CH_2), 39.58 (CH_2), 44.36 (CH_2), 51.00 (CH_2), 62.02 (CH_2), 71.07 (CH), 83.03 (CH), 87.75 (CH), 119.48 (C), 137.55 (CH), 153.23 (C),

157.84 (C), 160.02 (CH), 160.09 (C). m/z (FAB) 406.25880 [$(M+H)^+$ calc. for $C_{19}H_{32}N_7O_3$ 406.25666]. λmax 220, 252, 279nm.

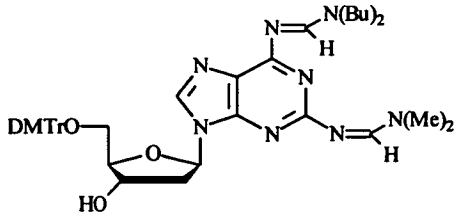
2-N-Dimethylformamido-6-N-di-n-butylformamido-2'-deoxyadenosine [30].



2-Amino-6-N-di-n-butylformamido-2'-deoxyadenosine **[29]** (1.23g, 3.03mmol) was coevaporated three times with anhydrous pyridine, then suspended in anhydrous pyridine (30ml). Dimethylformamide dimethyl acetal (10eq, 30.3mmol, 4.0ml) was added dropwise and the reaction flask was heated to 50°C with stirring. After 12 hours the reaction was quenched by the addition of water (10ml), and stirred for another 10min. The solvents were then evaporated *in vacuo*, coevaporated with toluene, and the product was purified by flash-column chromatography, eluting with a gradient of (0.0-5.0%, v/v) methanol in dichloromethane. The product was obtained as a light yellow foam. Yield: 0.86g, (62%).

Rf 0.30 (solvent A), Rf 0.56 (solvent C), Rf 0.57 (solvent I). Found: C,54.82; H,8.62; N,22.25%, $C_{22}H_{36}N_8O_3.(H_2O)$ requires: C,55.20; H,8.01; N,23.42%. ^1H nmr (250.13MHz; d6-DMSO) δ_H: 0.83-0.93 (6, m, butylform.CH_3), 1.20-1.38 (4, m, $MeCH_2$), 1.50-1.64 (4, m, $EtCH_2$), 2.20-2.29 (1, m, 2'-H), 2.63-2.74 (1, m, 2'-H), 3.00 (3, s, dmf.CH_3), 3.11 (3, s, dmf.CH_3), 3.38-3.66 (6, m, $PrCH_2$, 5'-CH_2), 3.87-3.91 (1, dd, 4'-H), 4.41-4.43 (1, m, 3'-H), 5.42 (2, br d, 5'-OH exch.,3'-OH exch.), 6.31-6.37 (1, dd, 1'-H), 8.24 (1, s, 8-H), 8.63 (1, s, dmf.CH), 8.93 (1, s, butylform.CH). ^{13}C nmr (62.90MHz; d6-DMSO) δ_C: 13.66 (CH_3), 13.88 (CH_3), 19.25 (CH_2), 19.78 (CH_2), 28.81 (CH_2), 30.58 (CH_2), 34.59 (CH_3), 39.60 (CH_2), 40.49 (CH_3), 44.38 (CH_2), 50.93 (CH_2), 62.06 (CH_2), 71.13 (CH), 83.56 (CH), 88.00 (CH), 122.22 (C), 139.74 (CH), 152.83 (C), 157.66 (CH), 158.10 (CH), 159.39 (C), 161.39 (C). m/z (FAB) 461.29834 [$(M+H)^+$ calc. for $C_{22}H_{37}N_8O_3$]. λmax 215, 291nm.

2-N-Dimethylformamido-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2'-deoxyadenosine [31].

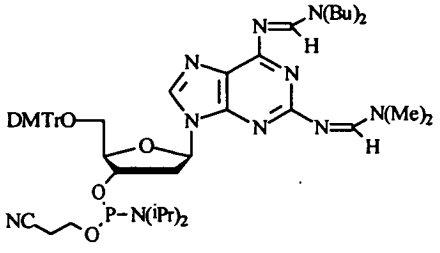


2-N-Dimethylformamido-6-N-di-n-butylformamido-2'-deoxyadenosine **[30]** (840mg, 1.8mmol) was coevaporated three times with, and dissolved in anhydrous pyridine (18ml), with 4-dimethylaminopyridine (0.1eq, 0.18mmol, 22.5mg). 4,4'-Dimethoxytrityl chloride (1.2eq, 2.16mmol, 740mg) was added portionwise, and the reaction mixture was left stirring for 5 hours. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and washed three times with saturated aqueous KCl, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.5:99.5, v/v) until all the dimethoxytrityl alcohol had passed, then with a gradient of methanol-dichloromethane (0.5-5.0%, v/v). The product was obtained as a solid white foam. Yield: 905mg, (66%).

Rf 0.85 (solvent A), Rf 0.28 (solvent H). Found: C,65.60; H,6.98; N,13.48%. $C_{43}H_{54}N_8O_5.(H_2O)$ requires: C,66.12; H,7.23; H,14.35%. ^1H nmr (250.13MHz; $CDCl_3$) δ_H: 0.87-0.95 (6, m, butylform.CH_3), 1.24-1.40 (4, m, $MeCH_2$), 1.55-1.64 (4, m, $EtCH_2$), 1.90-2.10 (2, m, 2'-CH_2), 3.06 (3, s, dmf.CH_3), 3.14 (3, s, dmf.CH_3), 3.30-3.35 (2, t, 5'-CH_2), 3.63-3.74 (10, m, $PrCH_2$, OCH_3), 4.03 (1, m, 4'-H), 4.37-4.39 (1, m, 3'-H), 6.61-6.68 (1, dd, 1'-H), 6.70-7.41 (13, m, DMTr-H), 7.81 (1, s, 8-H), 8.69 (1, s, dmf.CH), 9.00 (1, s, butylform.CH). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 13.49 (CH_3), 13.72 (CH_3), 19.54 (CH_2), 19.98 (CH_2), 28.96 (CH_2), 30.76 (CH_2), 34.80 (CH_3), 40.65 (CH_3), 44.57 (CH_2), 45.88 (CH_2), 51.34 (CH_2), 54.91 (2xCH_3), 63.75 (CH_2), 74.88 (CH), 82.96 (CH), 84.97 (CH), 86.01(CH), 112.85 (4xCH), 122.24 (C), 126.49 (CH), 127.56 (2xCH), 127.84 (2xCH), 129.70 (4xCH), 135.46

(C), 135.54 (C), 136.04 (CH), 144.40 (C), 153.09 (C), 157.93 (CH), 158.12 (2xC), 158.26 (CH), 160.03 (C), 162.48 (C). m/z (FAB) 763.42934 [$(M+H)^+$ calc. for $C_{43}H_{55}N_8O_5$ 763.42954]. λmax 215, 225, 274nm.

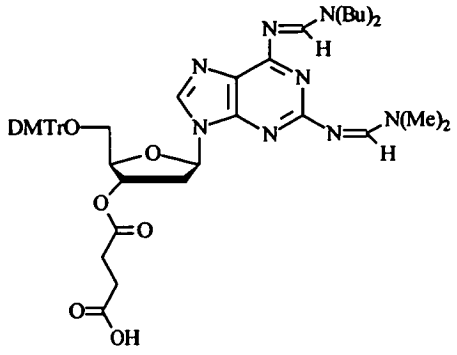
2-N-Dimethylformamido-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2'-deoxyadenosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite [32].



2-N-Dimethylformamido-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2'-deoxyadenosine **[31]** (510mg, 0.67mmol) was coevaporated from anhydrous tetrahydrofuran three times, and suspended in anhydrous tetrahydrofuran (7ml). Anhydrous diisopropylethylamine (4eq, 2.7mmol, 0.47ml) was added, and the reaction flask was kept under argon. 2-Cyanoethyl-N,N-diisopropylchloro-phosphoramidite (1.2eq, 0.80mmol, 0.18ml) was added dropwise, and after a few minutes, a white precipitate (salt) was seen to appear. After 2 hours the reaction was quenched with by the addition of ethyl acetate (argon purged), and the organic layer was washed with saturated aqueous KCl before being quickly dried over anhydrous Na_2SO_4, filtered under argon, and evaporated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with ethyl acetate-triethylamine (99:1, v/v), then washed with ethyl acetate), eluting with ethyl acetate. The product was obtained as a glassy oil. Yield: 470mg, (73%).

Rf 0.84 (solvent B), Rf 0.37/0.42 (solvent F), Rf 0.29 (solvent H). ^{31}P nmr (80MHz; MeCN; $(CD_3)_2CO$ lock) δ_P: 147.94 (s). m/z (FAB) 964.54302 [$(M+H)^+$ calc. for $C_{52}H_{72}N_{10}O_6P$ 964.54522].

2-N-Dimethylformamido-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-3'-O-succinyl-2'-deoxyadenosine [33].

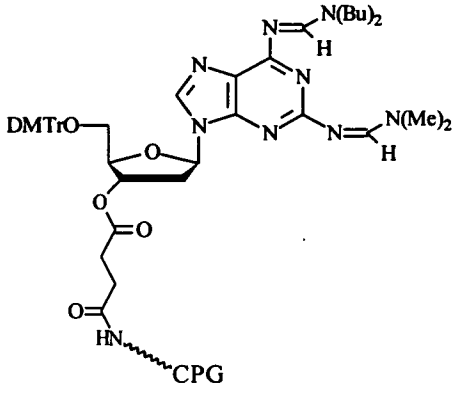


2-N-Dimethylformamido-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2'-deoxyadenosine **[31]** (240mg, 0.32mmol) was coevaporated three times with, and dissolved in anhydrous pyridine (3ml), with 4-dimethylaminopyridine (0.1eq, 0.03mmol, 39mg). Succinic anhydride (1.2eq, 0.38mmol, 38mg) was added portionwise with stirring, and the reaction mixture was left for 18 hours at room temperature. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and quickly washed three times with aqueous citric acid solution (10%, w/v), then with water, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.0-3.0%, v/v; with 1% triethylamine) The product was obtained as a yellow-white foam. Yield: 210mg, (77%).

Rf 0.28 (solvent A), Rf 0.59 (solvent C). ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 0.83-0.95 (6, m, butylform.CH_3), 1.31-1.38 (4, m, MeCH_2), 1.55-1.75 (4, m, EtCH_2), 2.54 (3, s, dmf.CH_3)?, 2.57 (3, s, dmf.CH_3)?, 2.80-2.90 (2, m, 2'-CH_2), 3.17-3.20 (2, m, 5'-CH_2), 3.30-3.40 (4, m, PrCH_2)?, 3.45-3.70 (4, m, succ.CH_2)?, 3.73 (6, s, OCH_3), 4.25 (1, m, 4'-H), 5.46 (1, m, 3'-H), 6.34 (1, dd, 1'-H), 6.72-7.36 (13, m, DMTr-H), 7.91 (1, s, 8-H), 8.25 (1, br s, acidic.OH), 8.98 (1, s, dmf.CH), 9.44 (1, s, butylform.CH). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 13.47 (CH_3), 13.72 (CH_3), 19.49 (CH_2), 19.97 (CH_2), 29.00 (CH_2), 30.65 (CH_2), 37.57 (CH_2), 44.85 (CH_2), 44.98 (CH_2), 51.76 (CH_2), 52.54 (CH_2), 54.98 (2xCH_3), 63.61 (CH_2), 74.90 (CH), 83.70 (CH), 83.84 (CH), 86.29 (C), 112.96 (4xCH), 123.22 (C), 126.69 (CH), 127.67 (2xCH), 127.86 (2xCH), 129.79

(4xCH), 135.31 (C), 135.33 (C), 138.81 (CH), 144.25 (C), 151.80 (C), 158.26 (2xC), 158.97 (CH), 160.34 (C), 163.23 (CH), 172.69 (C), 177.06 (C). λmax 202, 259nm.

Derivatisation of Long-Chain Alkyl Amino Controlled Pore Glass (CPG) solid support with Compound 33 [34].



Activation of the CPG.

CPG (0.5g, 150μmolg^{-1} loading) was dried *in vacuo* overnight over P_2O_5 and washed with a solution of anhydrous N,N-diisopropylethylamine in N,N-dimethylformamide (4x5ml; 10% v/v), followed by N,N-dimethylformamide (5x5ml).

Coupling the succinate onto the CPG.

2-N-Dimethylformamido-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-3'-succinyl-2'-deoxyadenosine **[33]** (1eq, 0.2mmol, 170mg) was coevaporated three times with, and dissolved in anhydrous dichloromethane (1ml). To this was added a solution of 4-nitrophenol (1.2eq, 2.4mmol, 33mg) in anhydrous pyridine (0.1ml) followed by a solution of N,N'-dicyclohexylcarbodiimide in dichloromethane (2.5eq, 0.5mmol; 103mg, 50% w/v). After a few minutes, dicyclohexylurea was seen to precipitate, and after the reaction had gone to completion (ca.4 hours) the dicyclohexylurea was removed by filtration. The filtrate was concentrated *in vacuo*, suspended in N,N-dimethylformamide (1ml), and activated CPG (ca.0.1eq, 150mg) was added to the suspension. The mixture was kept under argon and gently agitated for 9 hours, by which time the loading of the resin was 36.2μmol/g (see below for

testing of CPG loading). The CPG was filtered off, and washed with anhydrous: N,N-dimethylformamide (5x5ml), THF (5x5ml) and diethyl ether (5x5ml).

Capping the CPG.

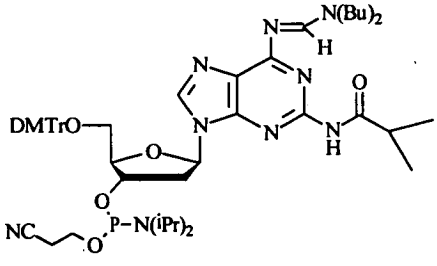
The functionalised resin was suspended in a solution of acetic anhydride-pyridine (4ml, 10% v/v) and gently agitated for 1 hour. The CPG was filtered off, washed successively with anhydrous: pyridine (5x5ml), THF (5x5ml), and diethyl ether (5x5ml); then dried *in vacuo* overnight over P_2O_5.

Calculation of the loading of the CPG.

A sample of the washed and dried resin (4-5mg) was weighed accurately, and placed in a 25ml volumetric flask. This was made up to the mark with a solution of trichloroacetic acid in anhydrous dichloromethane (3%, w/v). After 5min the absorbance of the solution at 504nm was measured to determine the concentration of dimethoxytrityl cations in solution and hence the loading of the resin. The final loading of CPG **[34]** was 35.0μmolg^{-1}. [The value of ε504=76mlcm^{-1}μmol^{-1} for the dimethoxytrityl cation, and general procedure for functionalisation, taken from Ref. 170].

2-N-Isobutyryl-5'-O-(4,4'-dimethoxytrityl)-6-N-di-n-butylformamido-2'-deoxy-adenosine-3'-O-(2-cyanoethyl)-diisopropylamido-phosphite [35]

This compound was synthesised according to the procedure outlined in Booth.[133]



5.4 OLIGONUCLEOTIDE SYNTHESIS

Oligonucleotide syntheses were performed on an Applied Biosystems 394 DNA synthesiser. All DNA synthesis reagents and cyanoethyl-phosphoramidite monomers

were purchased from ABI. Monomers were prepared as 0.1M solutions using anhydrous acetonitrile.

Coupling Efficiencies

Coupling efficiencies of phosphoramidites were measured either manually (by comparison of the absorbance at 498nm of the dimethoxytrityl cations produced in the detritylation steps of successive synthesis cycles after dilution to 25ml with 0.1M toluene-4-sulphonic acid in acetonitrile); or in the case of the ABI Model 394, trityl monitoring was carried out on-line using the built-in conductivity detector.[171]

Oligonucleotide Deprotection

After cleavage from the synthesis column, unmodified oligonucleotides were typically deprotected for 12 hours at 55°C and for 5 hours at 65°C. Oligonucleotides containing diaminopurine required longer deprotection times depending on the protecting groups employed, and on how many diaminopurine residues were present.

Purification of Oligonucleotides

Reversed-Phase HPLC analysis and purification were carried out on a Gilson model 306 HPLC system using ABI Brownlee Aquapore Octyl preparative-scale reverse phase columns. Following HPLC purification, oligonucleotides >12 bases long were passed through pre-equilibrated NAP-10 columns (Pharmacia, Sephadex G25) to desalt and remove small molecules. Shorter oligonucleotides were passed through a Sephadex G-10 column (16mm x 500mm). The pure, desalted oligonucleotides were then lyophilised to leave a white fluffy solid.

Buffer System

Buffer A: 0.1M NH_4OAc; Buffer B: 0.1M NH_4OAc/25% acetonitrile.

Gradient 1.Purification.

Time/min	Flow/mlmin^{-1}	%Buffer B
0	3	0

3	3	0
4	3	15
16	3	75
17	3	100
19	3	100
21	3	0
22	3	0

Gradient 2.Analysis.

Time/min	Flow/mlmin^{-1}	%Buffer B
0	3	0
3	3	0
4	3	15
24	3	70
25	3	100
27	3	100
28	3	0
30	3	0

Oligonucleotide Storage

Following desalting, oligonucleotides were stored at -20°C as frozen solutions or lyophilised solids.

Capillary Electrophoresis

The purity of HPLC-purified oligonucleotides was checked by Capillary Zone Electrophoresis (CE) on an Applied Biosystems Model 270A Capillary Electrophoresis system with a Microgel capillary, using 75mM tris phosphate buffer at pH 7.6 with 10% methanol.

Enzyme digests

The oligonucleotide to be digested (2 OD) was dissolved in 50 mM Tris buffer (0.3ml; 1M NaCl, pH=8.8). Phosphodiesterase I [type VII from Crotalus Atrox venom] (5µl=0.05units), and alkaline phosphatase [type VII from bovine intestinal mucosa] (1µl=0.15) were added. The digests were incubated for 10 hours at 37°C, and analysed by HPLC at 260nm using Gradient 1 with 20% MeCN in buffer B.

Extinction coefficient for dD at 260nm

2,6-Diaminopurine **[28]** (approx. 0.26mg, 0.01mmol) was accurately weighed out and dissolved in distilled water (25ml) and the UV spectrum recorded over the range 320-190nm (see Section 5.2). The absorbance at 260nm was used to calculate the extinction coefficient. The value was obtained in triplicate to give an average value of $\varepsilon_{dD}= 7.4x10^3 cm^2 \mu mol^{-1}$.

Oligonucleotides synthesised for deprotection and coupling efficiency tests

Oligonucleotides **DAP(1)** and **DAP(2)** were each synthesised on a 0.2µmol scale for the following monomers: **[7]**, **[12]**, **[14]**, **[16]**, **[32]**, and **[35]**.

DAP(1) TTT TTD TT

DAP(2) TDD TDT T

This was carried out to determine coupling efficiency of the given monomer, check the suitability of the monomer in oligonucleotide synthesis, and to monitor deprotection times of the particular N^2-protection.

5.5 ULTRAVIOLET MELTING STUDIES

For the determination of concentration-dependent thermodynamic parameters, the melting temperatures of the duplexes were measured in triplicate at 260nm for 5 points over a 10-fold concentration range on a Perkin Elmer Lambda 15 ultraviolet spectrometer equipped with a Peltier Block and controlled by an IBM PS2

microcomputer. A heating rate of 0.9°C per minute was used throughout, with data collection at 10 second intervals and the crude data was collected and processed using a PECSS-2 software package. The absorbance versus time curves were stored and converted to absorbance versus temperature curves (degrees K). The T_m value for each curve was obtained from the first derivative of the absorbance versus temperature curve

The oligonucleotides were dissolved in aqueous buffer {sodium chloride (1.0M), sodium dihydrogen orthophosphate (10mM), EDTA (1mM), and sodium cacodylate (10mM) which was adjusted to pH 7.0 by the addition of sodium hydroxide}. Oligonucleotides were heated to 80°C until the absorbance had stabilised to melt the duplexes, then allowed to anneal slowly to approx. 10°C.

Extinction coefficients of oligonucleotides

The extinction coefficient for each oligonucleotide was calculated using the following values at 260nm:

T= 8.8; C= 7.3; G= 11.7; A= 15.4;[173] and D= $7.4 \times 10^3 cm^2 \mu mol^{-1}$. To allow for suppression of absorbance due to single strand base stacking etc., the sum of ε values for each oligonucleotide were multiplied by 0.8 (hypochromicity factor, F). This value was determined from an average of F values, taken from measurements of the absorbance before, and after, an enzyme digest of the sample oligonucleotide was carried out.

($F=A_{[single\ strands]}/A_{[nucleosides]}$).

Oligonucleotides synthesised for UV melting

The following oligonucleotides were synthesised at 1.0μmol scale:

DAP(3) TDT DTD TD

DAP(4) TAT ATA TA

DAP(5)	CGT ADC GT	and complement:	**DAP(16)**	ACG TTA CG
DAP(6)	CGT CDC GT		**DAP(17)**	ACG TGA CG

DAP(7) CGT DDC GT
DAP(8) CGT GDC GT
DAP(9) CGT TDC GT
DAP(10) CGT DAC GT
DAP(11) CGT DGC GT
DAP(12) CGT DTC GT
DAP(13) DCG TTA CG
DAP(14) DCD CDC DC
DAP(15) CGT AAC GT
DAP(16) ACG TTA CG
DAP(18) ACG TCA CG
DAP(19) ACG TAA CG
DAP(16) ACG TTA CG
DAP(20) ACG CTA CG
DAP(21) ACG ATA CG
DAP(15) CGT AAC GT
DAP(22) GTG TGT GT
DAP(16) ACG TTA CG

Testing dD CPG

The following oligonucleotides were synthesised using CPG **[34]**, at 0.2μmol scale:

DAP(23) GCA TTG AD
DAP(24) ATG CAA CCG AGD

along with the native sequences:

DAP(25) GCA TTG AA
DAP(26) ATG CAA CCG AGA

The sequences were then compared on reverse phase HPLC (Gradient 2), and CE, (*see* Section 5.4).

5.6 HYDROGENATION CONDITIONS

Hydrogenation of nucleosides **[23]** and **[26]** to give **[24]** and **[27]** respectively, was unsuccessful employing catalytic hydrogenation with Pd on activated charcoal at 4atm H_2. The problem was due to reaction at the N^2-protection—dmf and tfa, for **[23]** and **[26]**, respectively.[178]

In order to try to prevent reaction at the N^2-moiety, various alternative reagents were tried to reduce the N^6-azide under milder conditions.

Ph_3P. Nucleosides **[23]** and **[26]** were stirred in anhydrous pyridine, and triphenylphosphine was added.[179] The azide reacted to give the iminophosphorane, but subsequent hydrolysis with H_2O did not occur.

$SnCl_2$. Nucleosides **[23]** and **[26]** were suspended in EtOH/H_2O and gently heated (60°C) until dissolution was complete. $SnCl_2$ was then added sparingly until azide reduction was complete (by tlc). These conditions also cleaved the N^2-protection to an extent.[180]

Raney Ni. Nucleosides **[23]** and **[26]** were suspended in a mixture of EtOH/H_2O (9:1) and gently heated (60°C) until dissolution was complete. Raney nickel was then added sparingly until azide reduction was complete (by tlc).[181] These conditions also cleaved the N^2-protection to a large extent.

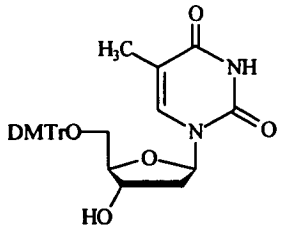
None of the conditions above selectively reduced the azide moiety without cleaving the N^2-dmf (or -tfa) group present.

5.7 STABILITY OF dD-NUCLEOSIDES TO TCA

A small amount of each nucleoside: **dA**, **dD [28]**, 2-Ac-6-dmf-dD **[13a]**, and 2-dmf-6-dbf-dD **[30]**, were dissolved in a 3% solution of trichloroacetic acid in dichloromethane (100µl). Samples were taken at time zero and at appropriate intervals thereafter, and examined by thin layer chromatography (solvent A for **[13a]**, solvent C for **dA** and **[30]**, and solvent I for **dD**) for signs of degradation against the starting material. Examination of the plates revealed loss of starting material, and build up of other products. From this, approximate half-lives for the stability of the nucleosides in TCA/DCM were obtained: **dA**: 45min; **dD**: 35min; **[13a]**: <20min; and **[30]**: >1.5hrs.

5.8 SYNTHESIS OF PHTHALOYL LINKERS

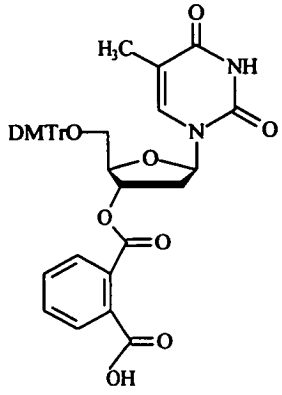
5'-O-(4,4'-Dimethoxytrityl)-thymidine [36]



Thymidine (3.17g, 13.1mmol) was coevaporated from (3x20ml), and suspended in anhydrous pyridine (30ml) with 4-dimethylaminopyridine (0.1eq, 0.13mmol, 160mg). 4,4'-Dimethoxytrityl chloride (1.1eq, 14.4mmol, 4.7g) was added portionwise, and the reaction mixture was left stirring for 10 hours. When complete, the reaction was quenched with methanol (20ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and washed three times with saturated aqueous KCl, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.5:99.5, v/v) until all the dimethoxytrityl alcohol had passed, then with a gradient of methanol in dichloromethane (0.5-1.5%, v/v). The product was obtained as a solid white foam. Yield: 6.7g, (96%).

Rf 0.71 (solvent A), Rf 0.30 (solvent H). Found: C,68.44; H,6.25; N,5.16%. Requires: C,68.37; H,5.92; N,5.15%. ^{1}H nmr (200.13MHz; $CDCl_3$) δ_H: 1.42 (3, s, 5-CH_3), 2.26-2.47 (2, m, 2'-CH_2), 3.32-3.47 (2, dd, 5'-CH_2), 3.75 (6, s, OCH_3), 4.09-4.10 (1, m, 4'-H), 4.57 (1, m, 3'-H), 6.22 (2, br s, 3'-OH, exch., 3-NH, exch.), 6.41-6.48 (1, dd, 1'-H), 6.80-7.41 (13, m, DMTr-H), 7.63 (1, s, 6-H). ^{13}C nmr (50.32MHz; $CDCl_3$) δ_C: 11.0 (CH_3), 40.76 (CH_2), 45.47 (CH_2), 54.97 (2xCH_3), 71.97 (CH), 84.61 (CH), 86.18 (C), 86.58 (CH), 111.01 (C), 112.99 (4xCH), 126.84 (CH), 127.72 (2xCH), 127.88 (2xCH), 129.84 (4xCH), 135.12 (C), 135.22 (CH), 135.61 (C), 144.13 (C), 150.63 (C), 158.38 (2xC), 164.16 (C). m/z (FAB) 545.23081 [$(M+H)^+$ calc. for $C_{31}H_{32}N_2O_7$ 545.23081]. λmax 213, 236, 266nm.

3'-O-Phthaloyl-5'-O-(4,4'-dimethoxytrityl)-thymidine [37]

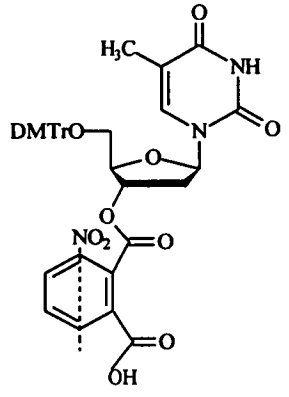


5'-O-(4,4'-Dimethoxytrityl)-thymidine **[36]** (1.0g, 1.8mmol) was coevaporated three times with, and dissolved in anhydrous pyridine (10ml), with 4-dimethylaminopyridine (0.1eq, 0.18mmol, 22mg). Phthalic anhydride (2eq, 3.60mmol, 533mg) was added portionwise with stirring, and the reaction mixture was left for 15 hours at room temperature. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and quickly washed three times with aqueous citric acid solution (10%, w/v), then with water, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.0-5.0%, v/v; with 1% triethylamine) The product was obtained as a yellow-white foam. Yield: 880mg, (71%).

Rf 0.40 (solvent A), Rf 0.59 (solvent C). Found: C,65.59; H,6.92; N,4.11%. $C_{39}H_{36}N_2O_{10}.(H_2O)$ requires: C,65.89; H,5.39; N,3.94%. ^1H nmr (250.13MHz; $CDCl_3$) δ_H: 1.39 (3, s, 5-CH_3), 1.78-1.94 (2, m, 2'-CH), 2.29-2.45 (2, m, 2''-CH), 3.40-3.50 (2, m, 5'-CH_2), 3.69-3.74 (6, s, OCH_3), 4.52-4.58 (1, m, 4'-H), 5.83-5.92 (1, m, 3'-H), 6.39-6.49 (1, m, 1'-H), 6.67-6.82 (4, m, phthal.CH), 7.08-7.40 (13, m, DMTr-H), 7.50-7.60 (1, m, 3-NH), 7.78-7.88 (1, s, 6-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 11.44 (CH_3), 37.09 (CH_2), 54.99 (2xCH_3), 63.78 (CH_2), 77.84 (CH), 82.68 (CH), 86.52 (C), 87.65 (CH), 110.60 (C), 113.04 (4xCH), 127.70-130.00

(9xDMTr-CH; 4xAr-CH), 133.60 (C), 133.89 (C), 134.72 (C), 134.93 (C), 135.14 (CH), 144.12 (C), 150.60 (C), 158.40 (2xC), 162.73 (C), 171.44 (C), 172.43 (C).
m/z (FAB) 692.23401 [$(M+H)^+$ calc. for $C_{39}H_{36}N_2O_{10}$ 692.23700]. λmax 216, 263nm.

3'-O-(Nitro-phthaloyl)-5'-O-(4,4'-dimethoxytrityl)-thymidine [38]

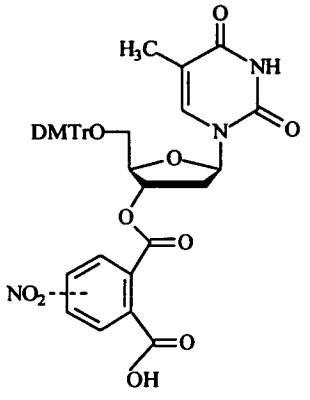


5'-O-(4,4'-Dimethoxytrityl)-thymidine **[36]** (400mg, 0.75mmol) was coevaporated three times with, and dissolved in anhydrous pyridine (5ml), with 4-dimethylaminopyridine (0.1eq, 0.08mmol, 10mg). 3-Nitro-phthalic anhydride (4eq, 3.0mmol, 290mg) was added portionwise with stirring, and the reaction mixture was left for 20 hours at room temperature. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and quickly washed three times with aqueous citric acid solution (10%, w/v), then with water, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.0-4.0%, v/v; with 1% triethylamine) The product was obtained as a yellow-white foam. Yield: 310mg, (56%).

Rf 0.41 (solvent A), Rf 0.67 (solvent C). ^{1}H nmr (250.13MHz; $CDCl_3$) δ_H: 1.30 (3, s, 5-CH3), 2.40-2.55 (1, m, 2'-H), 2.70-2.85 (1, m, 2"-H), 3.45-3.65 (2, m, 5'-CH2), 3.74 (6, s, OCH3), 4.60 (1, m, 4'-H), 5.75-5.77 (1, "d", 3'-H), 6.35-6.41 (1, m, 1'-H), 6.80 (2, s, DMTr-H), 6.83 (2, s, DMTr-H), 7.10-7.42 (11, m, 9xDMTr-H; Ar-H; 3-

NH), 7.63 (1, s, 6-H), 8.11-8.39 (2, m, Ar-H). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 11.29 (CH_3), 37.12 (CH_2), 54.97 (2xCH_3), 63.64 (CH_2), 76.40 (CH), 82.75 (CH), 84.19 (CH), 86.82 (C), 111.18 (C), 113.02 (4xCH), 125.12 (CH), 126.89 (CH), 127.73 (2xCH), 127.92 (2xCH), 128.73 (CH), 129.16 (CH), 129.86 (2xCH), 129.90 (2xCH), 135.03 (C), 135.09 (C), 136.02 (CH), 138.72 (C), 144.12 (C), 145.55 (C), 147.46 (C), 150.37 (C), 158.44 (2xC), 163.85 (C), 166.58 (C), 168.61 (C).
m/z (FAB) 738.22942 [$(M+H)^+$ calc. for $C_{39}H_{36}N_3O_{12}$ 738.22989].

3'-O-(Nitro-phthaloyl)-5'-O-(4,4'-dimethoxytrityl)-thymidine [39]

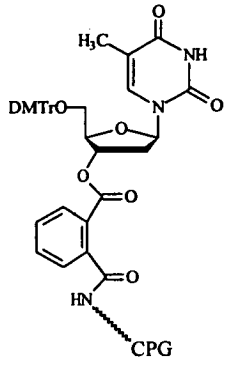


5'-O-(4,4'-Dimethoxytrityl)-thymidine **[36]** (400mg, 0.75mmol) was coevaporated three times with, and dissolved in anhydrous pyridine (5ml), with 4-dimethylaminopyridine (0.1eq, 0.08mmol, 10mg). 4-Nitro-phthalic anhydride (4eq, 3.0mmol, 290mg) was added portionwise with stirring, and the reaction mixture was left for 20 hours at room temperature. When complete, the reaction was quenched with methanol (10ml) and concentrated *in vacuo*. The residue was dissolved in dichloromethane and quickly washed three times with aqueous citric acid solution (10%, w/v), then with water, the organic phase was dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The product was purified by flash-column chromatography (the column was pre-equilibrated with dichloromethane-triethylamine 99:1, v/v), eluting with methanol-dichloromethane (0.0-9.0%, v/v; with 1% triethylamine) The product was obtained as a yellow-white foam. Yield: 260mg, (47%) [top spot].

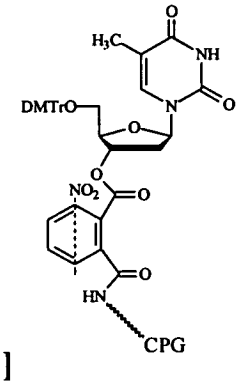
Rf 0.52 (solvent A), Rf 0.73 (solvent C). ^{13}C nmr (62.90MHz; $CDCl_3$) δ_C: 11.28 (CH_3), 37.12 (CH_2), 54.95 (2xCH_3), 63.64 (CH_2), 76.39 (CH), 76.90 (CH), 82.95 (CH), 84.08 (CH), 86.81 (C), 111.37 (C), 112.99 (4xCH), 123.19 (CH), 124.27 (CH), 126.91 (CH), 127.72 (2xCH), 127.88 (2xCH), 127.98 (CH), 129.85 (4xCH), 134.91 (C), 134.99 (C), 135.22 (CH), 137.72 (C), 140.04 (C), 143.96 (C), 148.26 (C), 150.42 (C), 158.44 (2xC), 163.81 (C), 168.09 (C), 169.26 (C).

m/z (FAB) 738.22957 [$(M+H)^+$ calc. for $C_{39}H_{36}N_3O_{12}$ 738.22989].

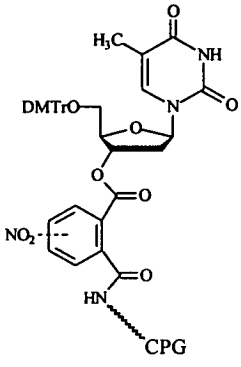
Derivatisation of Long-Chain Alkyl Amino Controlled Pore Glass (CPG) solid support with Compounds: 37, 38, and 39, to give: [40], [41], and [42]



[40]



[41]



[42]

Activation of the CPG.

CPG (1.5g, 150μmolg^{-1} loading) was dried in vacuo overnight over P_2O_5 and washed with a solution of anhydrous N,N-diisopropylethylamine in N,N-dimethylformamide (4x5ml; 10% v/v), followed by N,N-dimethylformamide (5x10ml).

Coupling the 3'-O-Phthaloyl Nucleoside onto the CPG.

The 3'-O-phthaloyl nucleoside (1eq, 0.3mmol) was coevaporated three times with, and dissolved in anhydrous dichloromethane (1ml). To this was added a solution of 4-nitrophenol (1.2eq, 3.6mmol, 50mg) in anhydrous pyridine (0.1ml) followed by a solution of N,N'-dicyclohexylcarbodiimide in dichloromethane (2.5eq, 0.2ml; 20% w/v). After a few minutes, dicyclohexylurea was seen to precipitate, and after the reaction had gone to completion (ca. 5 hours) the dicyclohexylurea was removed by filtration. The filtrate was concentrated *in vacuo*, suspended in N,N-

dimethylformamide (1ml), and activated CPG (0.1eq, 300mg) was added to the suspension. The mixture was kept under argon and gently agitated for ca. 20 hours. The CPG was filtered off, and washed with anhydrous: N,N-dimethylformamide (5x5ml), THF (5x5ml) and diethyl ether (5x5ml).

Capping the CPG.

The functionalised resin was suspended in a solution of acetic anhydride-pyridine (4ml, 10% v/v) and gently agitated for 1 hour. The CPG was filtered off, washed successively with anhydrous: pyridine (5x5ml), THF (5x5ml), and diethyl ether (5x5ml); then dried *in vacuo* overnight over P_2O_5.

Calculation of the loading of the CPG.

A sample of the washed and dried resin (4-5mg) was weighed accurately, and placed in a 25ml volumetric flask. This was made up to the mark with a solution of trichloroacetic acid in anhydrous dichloromethane (3%, w/v). After 5min the absorbance of the solution at 504nm was measured to determine the concentration of dimethoxytrityl cations in solution and hence the loading of the resin.

The final loading of the CPG was found to be: 7.5μmolg^{-1} for **[40]**, 5.0μmolg^{-1} for **[41]**, and 2.5μmolg^{-1} for **[42]**.

[Using the value of ε_{504}=76mlcm^{-1}μmol^{-1} for the dimethoxytrityl cation].[170]

Linker cleavage studies

The following oligonucleotides were synthesised, on a 0.2μmol scale to test cleavage times from CPG, following machine synthesis:

SUC-1	GAC TTG ACC TAT	using succinyl CPG
PTH-1	GAC TTG ACC TAT	using phthaloyl CPG **[40]**
PTH-2	GAC TTG ACC TAT	using NO_2-phthaloyl CPG **[41]**
PTH-3	GAC TTG ACC TAT	using NO_2-phthaloyl CPG **[42]**

The end proc. consisted of 3x10 minute cycles. After each cycle of cleavage, the absorbance of the resulting aliquot was measured at 260nm, with the time noted.

The following oligonucleotides were synthesised:

SUC-2	ATG CAA CCG AGT	using succinyl CPG
SUC3	ATA GCT TGG CAA GTA TGA CT	using succinyl CPG
PTH-4	ATG CAA CCG AGT	using phthaloyl CPG **[40]**
PTH-5	ATA GCT TGG CAA GTA TGA CT	using phthaloyl CPG **[40]**

The oligonucleotides were synthesised (**SUC-2**, **SUC-3**, **PTH-4**, and **PTH-5**) on a 0.2μmol scale, without an End proc., to allow more detailed study of the cleavage times. Approximately 1/10 of the post-synthesis CPG was weighed out, and added to a "blanked" 1cm cuvette containing 2.5ml of conc.NH_4OH. The cuvette was stoppered, and shaken once every 5 minutes—the absorbance reading (260nm) was taken once the CPG had settled. A plot of the time versus absorbance gave the $t_{1/2}$ etc.

Machine synthesis-compatibility of phthaloyl linkers

Oligonucleotides **SUC-2**, **SUC-3**, **PTH-4**, and **PTH-5** were also used to check the coupling efficiency, and the quality of synthesis of oligonucleotides containing a 3'-phthaloyl linker. The resulting oligonucleotides from **SUC-2** and **SUC-3** were compared to **PTH-4**, and **PTH-5** via reverse-phase HPLC and CE (*see* Section 5.4).

REFERENCES

(1) Watson,J.D. and Crick,F.H.C. (1953) *Nature* **171**, 737-738.

(2) Zamenhof,S., Brawerman,G., and Chargaff,E. (1952) *Biochim. Biophys. Acta* **9**, 402-405.

(3) Sonveaux,E. (1994). In *Methods in Molecular Biology, Vol 26.* (Ed. Agrawal,S.). Humana Press, Totowa, NJ.

(4) Gilham,P.T. and Khorana,H.G. (1958) *J. Am. Chem. Soc.* **80**, 6212-6216.

(5) Rammler,D.H. and Khorana,H.G. (1962) *J. Am. Chem. Soc.* **84**, 3112-3122.

(6) Beaucage,S.L. and Iyer,R.P. (1992) *Tetrahedron* **48**, 2223-2311.

(7) Smith,M., Rammler,D.H., Goldberg,I.H. and Khorana,H.G. (1962) *J. Am. Chem. Soc.* **84**, 430-440.

(8) Reese,C.B. (1978) *Tetrahedron* **34**, 3143-3179.

(9) Finnan,J.L., Varshney,A. and Letsinger,R.L. (1974) *Nucleic Acids Res. Symp.* No.7,133-145.

(10) Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Lett.* **22**, 1859-1862.

(11) McBride,L.J. and Caruthers,M.H. (1983) *Tetrahedron Lett.* **24**, 245-248.

(12) Merrifield,B. (1963) *J. Am. Chem. Soc.* **85**, 2149-2154.

(13) Letsinger,R.L. and Mahadevan,V. (1965) *J. Am. Chem. Soc.* **87**, 3526-3527.

(14) Köster,H. (1972) *Tetrahedron Lett.* 1527-1530.

(15) Atkinson,T. and Smith,M. (1984). In *Oligonucleotide Synthesis, a Practical Approach*, (ed. Gait,M.J.), pp69. IRL Press, Oxford.

(16) Bains,W. (1995) *Chem. in Britain* **F**, 122-125.

(17) Voet,D. and Voet,J.G. (1995). *Biochemistry*, 2nd Ed. Chapter 28. Wiley, NY.

(18) Johnston,B.H. (1992) *Methods in Enz.* **211**, 127-157.

(19) Kypr,J., Sagi,J., Szakonyi,E., Ebinger,K., Penazova,H., Chladkova,J. and Vorlickova, M. (1994) *Biochemistry* **33**, 3801-3806.

(20) Rich,A., Nordheim,A. and Wang,A.H-J. (1984) *Annu. Rev. Biochem.* **53**, 791-846.

(21) Wing,R., Drew,H., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1980) *Nature* **287**, 755-758.

(22) Kennard,O. and Hunter,W.N. (1989) *Q. Rev. Biophys.* **22**, 327-379.

(23) Feigon,J., Sklenar,V., Wang,E., Gilbert,D.E., Macaya,R.F. and Schultze,P. (1992) *Methods in Enz.* **211**, 235-254.

(24) Dickerson,R.E. (1992) *Methods in Enz.* **211**, 67-111.

(25) Saenger,W. (1984). *Principles of Nucleic Acid Structure* Chapter 15.

Springer-Verlag: NY.

(26) Chou,S-H., Flynn,P. and Reid,B. (1989) *Biochemistry* **28**, 2435-2443.

(27) Martin,F.H. and Tinoco,I. Jr. (1980) *Nucleic Acids Res.* **8**, 2295-2299. And references cited therein.

(28) Kornberg,A. (1992) *DNA Replication*, pp446-447. W.H Freeman, San Francisco.

(29) Cohen,J.S. (1989). In *Oligonucleotides, Antisense Inhibitors of Gene Expression.* CRC Press, Boca Raton, FL.

(30) Egli,M., Usman,N., Zhang,S. and Rich,A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 534-538.

(31) Zimmerman,S.B. (1982) *Ann. Rev. Biochem.* **51**, 395-427.

(32) Kypr,J., Sagi,J., Szabolcs,A., Ebinger,K., Otvos,L. and Vorlickova,M. (1990) *Gen. Physiol. Biophys.* **9**, 415-418.

(33) Riley,M., Maling,B. and Chamberlin,M.J. (1966) *J. Mol. Biol.* **20**, 359-389.

(34) Chamberlin,M.J. (1969) *Proc. Nucleic Acids Res.* 513-519.

(35) Hall,K.B. and McLaughlan,L.W. (1991) *Biochemistry* **30**, 10606-10613. And references cited therein.

(36) Cheng,Y-K. and Pettitt,B.M. (1992) *Prog. Biophys. Molec. Biol.* **58**, 225-257.

(37) Hunter,C.A. (1993) *J. Mol. Biol.* **230**, 1025-1054.

(38) Beaucage,S.L. and Iyer,R.P. (1992) *Tetrahedron* **48**, 2223-2311.

(39) De Mesmaeker,A., Haner,R., Martin,P. and Moser,H.E. (1995) *Acc. Chem. Res.* **28**, 366-374. And references cited therein.

(40) Brown,T. (1995) *Aldrichimica Acta* **28**, 15-20. And references cited therein.

(41) Saenger,W. (1984). *Principles of Nucleic Acid Structure*, pp271-272. Springer-Verlag: NY.

(42) Kang,C.-H., Zhang,X., Ratliff,R., Moyzis,R. and Rich,A. (1992) *Nature* **356**, 126-131.

(43) Leonard,G.A., McAuley-Hecht,K.E., Ebel,S., Lough,D.M., Brown,T. and Hunter,W.N. (1994) *Structure* **2**, 483-494.

(44) Lindahl,T. (1993) *Nature* **362**, 709-715.

(45) Leslie,A.G.W., Arnott,S., Chandrasekaran,R. and Ratliff,R.L. (1980) *J. Mol. Biol.* **143**, 49-72.

(46) Hunter,W.N. (1992) *Methods in Enz.* **211**, 221-233.

(47) Sanghvi,Y.S. (1993) In *Antisense Research Applications* (Crooke,S.T. and Lebleu,B. eds.), Chapter 15. CRC Press Inc., Boca Raton, FL.

(48) Uhlmann,E. and Peyman,A. (1990) *Chem. Rev.* **90**, 543-584.

(49) Sanghvi,Y.S., Hoke,C.T.D., Freier,S.M., Zounes,M.C., Gonzales,C., Cummins,L., Sasmor,H. and Cook,P.D. (1993) *Nucleic Acids Res.* **21**, 3197-3203.

(50) Cosstick,R., Li,X., Tuli,D.K., Williams,D.M., Connolly,B.A. and Newman,P.C. (1990) *Nucleic Acids Res.* **18**, 4771-75. And refs. cited therein.

(51) Froehler,B.C., Wadwani,S., Terhorst,T.J. and Gerrard,S.R. (1992) *Tetrahedron Lett.* **33**, 5307-10.

(52) Martin,F.H., Castro,M.M., Aboul-ela,F. and Tinoco,I. (1985) *Nucleic Acids Res.* **13**, 8927-8932.

(53) Eritja,R., Horowitz,D.M., Walker,P.A., Ziehler-Martin,J.P., Boosalis,M.S., Goodman,M.F., Itakura,K. and Kaplan,B.E. (1986) *Nucleic Acids Res.* **14**, 8135-8139.

(54) Eritja,R., Kaplan,B.E., Mhaskar,D., Sowers,L.C., Petruska,J. and Goodman,M.F. (1986) *Nucleic Acids Res.* **14**, 5869-5873.

(55) Seela,F. and Driller,H. (1989) *Nucleic Acids Res.* **17**, 901-910.

(56) Seela,F. and Thomas,H. (1994) *Helv. Chim. Acta* **77**, 897-903.

(57) Seela,F. and Thomas,H. (1995) *Helv. Chim. Acta* **78**, 94-108. And references cited therein.

(58) Gommers-Ampt,J.H. and Borst,P. (1995) *FASEB J.* **9**, 1034-1042. And references cited therein.

(59) Golankiewicz,B., Ostrowski,T., Boryski,J. and DeClerq,E. (1991) *J. Chem. Soc. Perkin Trans.I* 589-593.

(60) Piccirrilli,J.A., Krauch,T., Moroney,S.E. and Benner,S.A. (1990) *Nature* **343**, 33-37.

(61) Orgel,L.E. (1990) *Nature* **343**, 18-20.

(62) Eschenmoser,A. and Lowenthal,E. (1992) *Chem. Soc. Reviews* **21**, 1-16.

(63) Howard,F.B., Frazier,J. and Miles,H.T. (1966) *J. Biol. Chem.* **241**, 4293-4295.

(64) Kirnos,M.D., Khudyakov,I.Y.A., Alexandrushkina,N.I. and Vanyushin,B.F. (1977) *Nature* **270**, 369-370.

(65) Khudyakov,I.Y.A., Kirnos,M.D., Alexandrushkina,N.I. and Vanyushin,B.F. (1978) *Virology* **88**, 8-18.

(66) Cerami,A., Reich,E., Ward,D.C. and Goldberg,I.H. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 1036-1042.

(67) Scheit,K.H. and Rackwitz,H-R. (1982) *Nucleic Acids Res.* **10**, 4059-4069.

(68) Gaffney,B.L., Marky,L.A. and Jones,A.J. (1982) *Nucleic Acids Res.* **10**, 4351-

4362.

(69) Gaffney,B.L., Marky,L.A. and Jones,A.J. (1984) Tetrahedron (1984) 40, 3-13.

(70) Howard,F.B. and Miles,H.T. (1984) *Biochemistry* **23**, 6723-6732.

(71) Howard,F.B. and Miles,H.T. (1983) *Biopolymers* **22**, 597-600.

(72) Drew,H.R. and Dickerson,R.E. (1981) *J. Mol. Biol.* **151**, 535-556.

(73) Chollet,A., Chollet-Damerius,A. and Kawashima,E.H. (1986) *Chemica Scripta* **26**, 37-40.

(74) Cheong,C., Tinoco,I.Jr. and Chollet,A. (1988) *Nucleic Acids Res.* **16**, 5115-5122.

(75) Chollet,A. and Kawashima,E. (1988) *Nucleic Acids Res.* **16**, 305-317.

(76) Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) *Nature* **330**, 221-226. And references cited therein.

(77) Zimmerman,S.B. (1982) *Ann. Rev. Biochem.* **51**, 395-427.

(78) Gryaznov,S. and Schultz,R.G. (1994) *Tetrahedron Lett.* **35**, 2489-2492.

(79) Baldwin,R.L. and Davies,D.R. (1963) *J. Mol. Biol.* **6**, 251-255.

(80) Scheffler,I.E., Elson,E.L., and Baldwin,R.L. (1968) *J. Mol. Biol.* **36**, 291-304.

(81) Viswamitra,M.A., Shaked,Z., Jones,P.G., Sheldrick,G.M., Salisbury,S.A. and Kennard,O. (1982) *Biopolymers* **21**, 513-533.

(82) Yoon,C., Prive,G.G., Goodsell,D.S. and Dickerson,R.E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6332-6336.

(83) Assa-Munt,N. and Kearns,D.R. (1984) *Biochemistry* **23**, 791-796.

(84) Howard,F.B., Chen,C.W., Cohen,J.S. and Miles,H.T. (1984) *Biochem. Biophys. Res. Commun.* **118**, 848-853.

(85) Jovin,T.M., McIntosh,L.P., Arndt-Jovin,D.J., Zarling, D.A., Robert Nicoud,M., van de Sande, Jorge,K.F. and Eckstein,F. (1983) *J. Biomol. Str. Dyn.* **1**, 21-55.

(86) Borah,B., Cohen,J.S., Howard,F.B. and Miles,H.T. (1985) *Biochemistry* **24**, 7456-7462.

(87) Borah,B., Howard,F.B., Miles,H.T. and Cohen,J.S., (1986) *Biochemistry* **25**, 7464-7470.

(88) Vorlickova,M., Sagi,J., Szabolics,A., Szemzo,A., Otvos,L. and Kypr,J. (1988) *Nucleic Acids Res.* **16**, 279-289.

(89) Vorlickova,M., Sagi,J., Szabolics,A., Szemzo,A., Otvos,L. and Kypr,J. (1988) *J. Biomolec. Str. Dyn.* **6**, 503-510.

(90) Kypr,J. and Vorlickova,M. (1985) *Biochem. Biophys. Res. Commun.* **132**, 95-

99.

(91) Jovin,T.M., Soumpasis,D.M. and McIntosh,L.P. (1987) *Annu. Rev. Phys. Chem.* **38**, 521-560.

(92) Kypr,J., Sklenar,V. and Vorlickova,M. (1986) *Biopolymers* **25**, 1803-1812.

(93) Taboury,J.A., Adam,S., Tallandier,E., Neumann,J-M., Tran-Dinh,S., Huynh-Dinh,T., Langlois D'Estaintot,B., Conti,M. and Igolen,J. (1984) *Nucleic Acids Res.* **12**, 6291-6305.

(94) Coll,M., Wang,A.H-J., Van der Marel,G.A., Van Boom,J.H. and Rich,A. (1986) *J. Biomolec. Str. Dyn.* **4**, 157-172.

(95) Ueda,T., Miura,K. and Kasai,T. (1978) *Chem. Pharm. Bull.* **26**, 2122-2127.

(96) Huynh-Dinh,T., Langlois d'Estaintot,B., Allard,P. and Igolen,J. (1985) *Tetrahedron Lett.* **26**, 431-434.

(97) Sung,W.L. (1981) *J.C.S. Chem. Comm.* 1089.

(98) Eadie,J.S. and Davidson,D.S. (1987) *Nucleic Acids Res.* **15**, 8333-8349.

(99) Fathi,R., Goswami,B., Kung,R.P., Gaffney,B.L. and Jones,R.A. (1990) *Tetrahedron Lett.* **31**, 319-322. And references cited therein.

(100) Kung,P-P. and Jones,R.A. (1991) *Tetrahedron Lett.* **32**, 3919-3922.

(101) Gaffney,B.L. and Jones,R.A. (1982) *Tetrahedron Lett.* **23**, 2253-2256.

(102) Gaffney,B.L. and Jones,R.A. (1982) *Tetrahedron Lett.* **23**, 2257-2260.

(103) Bridson,P.K., Markiewicz,W.T. and Reese,C.B. (1977) *J. C. S. Chem. Comm.* 791-792.

(104) Daskalov,H.P., Sekine,M. and Hata,T. (1980) *Tetrahedron Lett.* **21**, 3899-3902.

(105) Kiburis,J. and Lister,J.H. (1980) *Tetrahedron Lett.* **21**, 2265-2268.

(106) Remaud,G., Zhou,X. and Chattopadhyaya,J. (1987) *Tetrahedron* **43**, 4453-4461.

(107) Schaller,H. and Khorana,H.G. (1963) *J. Am. Chem. Soc.* **85**, 3828-3835.

(108) Maxam,A.M. and Gilbert,W. (1980) *Methods in Enz.* **65**, 499-504.

(109) Myoshi,K., Huang,T. and Itakura,K. (1980) *Nucleic Acids Res.* **8**, 5591-5505.

(110) Matteucci,M.D. and Caruthers,M.H. (1980) *Tetrahedron Lett.* **21**, 3243-3246.

(111) Matteucci,M.D. and Caruthers,M.H. (1981) *J. Am. Chem. Soc.* **103**, 3185-91.

(112) Sekine,M. and Hata,T. (1983) *J. Org. Chem.* **48**, 3011-3014.

(113) Neilson,T. and Werstuik,E. (1973) *J. Am. Chem. Soc.* **96**, 2295-2297.

(114) Corey,E.J. and Venkateswarlu,A. (1972) *J. Am. Chem. Soc.* **94**, 6190-6191.

(115) Chattopadhyaya,J.B. and Reese,C.B. (1978) *J. C. S. Chem. Comm.* 639-640.

(116) Chattopadhyaya,J.B., Reese,C.B. and Todd,A.H. (1979) *J. C. S. Chem. Comm.* 987-988.

(117) ABI (1992) *User Bulletin for Synthesiser Models 392 and 394.*

(118) Sekine,M., Masuda,N. and Hata,T. (1985) *Tetrahedron* **41**, 5445-5453.

(119) Kume,A., Sekine,M. and Hata,T. (1983) *Chem. Letters* 1597-1600.

(120) Kume,A., Iwase,R., Sekine,M. and Hata,T. (1984) *Nucleic Acids Res.* **12**, 8525-8538.

(121) Morin,C. (1983) *Tetrahedron Lett.* **24**, 53-56.

(122) Froehler,B.C. and Matteucci,M.D. (1983) *Nucleic Acids Res.* **11**, 8031-8036.

(123) Zemlicka,J., Chládek,S., Holy,A. and Smrt,J. (1966) *Coll. Czech. Chem. Comm.* **31**, 3198-3211.

(124) Zemlicka,J. and Holy,A. (1967) *Coll. Czech. Chem. Comm.* **32**, 3159-3168.

(125) McBride,L.J. and Caruthers,M.H. (1983) *Tetrahedron Lett.* **24**, 2953-2956.

(126) Ti,G.S., Gaffney,B.L. and Jones,R.A. (1982) *J. Am. Chem. Soc.* **104**, 1316-1319.

(127) Garrett,E.R. and Mehta,P.J. (1972) *J. Am. Chem. Soc.* **94**, 8532-8541.

(128) Remaud,G., Zhou,X. and Chattopadhyaya,J. (1987) *Tetrahedron* **43**, 4453-61.

(129) York,J.L. J. (1981) *Org. Chem.* **46**, 2171-2171.

(130) Gonella,N.C. and Roberts,J.D. (1982) *J. Am. Chem. Soc.* **104**, 3162-3164.

(131) Sygula,A. and Buda,A. (1983) *J. Mol. Str.* **92**, 267-277.

(132) Maki,Y., Suzuki,M., Kameyama,K. and Sako,M. (1981) *J. Chem. Soc. Chem. Comm.* 658-659.

(133) Booth,E.D. (1991) *Ph.D. Thesis,* Edinburgh University.

(134) Buchi,H. and Khorana,H.G. (1972) *J. Mol. Biol.* **72**, 251-288.

(135) Wallace,R.B., Shaffer,J., Murphy,R.F., Bonner,J., Hirose,T. and Itakura,K. (1979) *Nucleic Acids Res.* **6**, 3543-3557.

(136) Voss,U.B., Chollet,A. and Malcolm,A.D.B. (1989) *Biochem. Soc. Trans.* **17**, 913.

(137) Mullis,K.B. and Faloona,F.A. (1987) *Methods in Enz.* **155**, 335-350.

(138) Uhlmann,E. and Peyman,A. (1990) *Chem. Rev.* **90**, 543-584.

(139) Zamecnik,P.C. and Stephenson,M.L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 280-284.

(140) Bodnar,J.W., Zempsky,W., Warder,D., Bergson,C. and Ward,D.C. (1983) *J.*

Biol. Chem. **258**, 15206-15213. And refs cited therein.

(141) Sagi,J., Vorlickova,M., Kypr,J. and Otvos,L. (1989) *Biochem. Biophys. Res. Commun.* **161**, 1204-1212.

(142) Ramesh,N., Shouche,Y.S. and Brahmachari,S.K. (1986) *J. Mol. Biol.* **190**, 635-638.

(143) Krayevsky,A.A. (1992) *Mol. Biology* **26**, 497-509.

(144) Poltev,V.I., Shulyupina,N.V. and Bruskov,V.I. (1977) *Molekulyana Biol.* 822-828.

(145) Alderson,T. (1964) *Mutation Res.* **1**, 205-208.

(146) Rackwitz,H.R. and Scheit,K.H. (1977) *Eur. J. Biochem.* **72**, 191-200.

(147) Pon,R.T. (1993). In *Methods in Molecular Biology*, Vol 20 (Ed. Agrawal,S.) Humana Press, Totowa, NJ. And references cited therein.

(148) Köster,H. (1972) *Tetrahedron Lett.* 1527-1530.

(149) Sproat,B.S. and Brown,D.M. (1985) *Nucleic Acids Res.* **13**, 2979-2987.

(150) Brown,T., Pritchard,C.E., Turner,G. and Salisbury,S.A. (1989) *J. C. S. Chem. Comm.* 891-893.

(151) Gryaznov,S.M. and Letsinger,R.L. (1991) *J. Am. Chem. Soc.* **113**, 5876-5877.

(152) Eritja,E., Robles,J., Fernandez-Forner,D., Albericio,F., Giralt,E. and Pedroso,E. (1991) *Tetrahedron Lett.* **32**, 1511-1514.

(153) Alul,R.H., Singman,N., Zhang,G. and Letsinger,R.L. (1991) *Nucleic Acids Res.* **19**, 1527-1532.

(154) Reynolds,T.R. and Buck,G.A. (1992) *Biotechniques* **12**, 518.

(155) Cruachem Inc. (1991) *Technical Bulletin* No.041R.

(156) Vu,H., McCollum,C., Jacobson,K., Thiesen,P., Vinayek,R., Spiess,E. and Andrus,A (1990) *Tetrahedron Lett.* **31**, 7269-7272.

(157) Schulof,J.C., Molko,D. and Teoule,R. (1987) *Nucleic Acids Res.* **15**, 397-416.

(158) Reddy,M.P., Hanna,N.B. and Farooqui,F. (1994) *Tetrahedron Lett.* **35**, 4311-4314.

(159) Polushin,N.N., Pashkova,I.N. and Efimov,V.A. (1991) *Nucleic Acids Res Symp. Ser.* **24**, 49-50.

(160) Sinha,N.D., Davies,P., Usman,N., Perez,J., Hodge,R., Kremsky,J. and Casale,R. (1993) *Biochimie* **75**, 13-23.

(161) Polushin,N.N., Morocho,A.M., Chen,B. and Cohen,J.S. (1994) *Nucleic Acids Res.* **22**, 639-645.

(162) Glen Research, (1993) *The Glen Report* Vol.6, No.2.

(163) Benseler,F. and McLaughlin,L.W. (1986) *Synthesis-Stuttgart* **1**, 45-46.

(164) Weber,H. and Khorana,H.G. (1972) *J. Mol. Biol.* **72**, 219-249.

(165) McBride,L.J., Kierzek,R., Beaucage,S.L. and Caruthers,M.H. (1986) *J. Am. Chem. Soc.* **108**, 2040-48.

(166) Connolly,B.A. (1984). In Oligonucleotides and Analogues (ed. Eckstein,F.), Chapter 7. IRL Press, Oxford.

(167) Chen,A. (1995) Ph.D. Thesis. Edinburgh University.

(168) Markiewicz,W.T. (1979) *J. Chem. Research* (S), 24-25.

(169) Adamiak,R.W., Biala,E. and Skalski,B. (1985) *Angew. Chem. Int. Ed. Engl.* **24**, 1054-1055.

(170) Sproat,B.S. and Gait,M.J. (1984). In *Oligonucleotide Synthesis, a Practical Approach*, (ed. Gait,M.J.), pp91. IRL Press, Oxford.

(171) Kaufman,J., Le,M., Ross,G., Hing,P., Budiansky,M., Yu,E., Campbell,E., Yoshimura,V., Fitzpatrick,V., Nadimi,K. and Andrus,A.

(172) Freier,S.M. (1993) In *Antisense Research and Applications*, (Eds Crooke,S.T. and Lebleu,B.), Chapter 5. CRC Press, Boca Raton, FL.

(173) Sproat,B.S. and Gait,M.J. (1984). In *Oligonucleotide Synthesis, a Practical Approach*, (ed. Gait,M.J.), pp109. IRL Press, Oxford.

(174) Breslauer,K.J., Frank,F., Blocker,H. and Marky,L.A. (1986) *Proc. Natl. Acad.Sci. USA* **83**, 3746-3750.

(175) Breslauer,K.J. (1994). *In Methods in Molecular Biology, Vol.26*, (Ed Agrawal,S.), Chapter 14, Humana Press Inc., Totowa, NJ.

(176) Andrus,A. (1994). In *Methods in Molecular Biology*, Vol.26, (Ed Agrawal,S.), Chapter 11, Humana Press Inc., Totowa, NJ.

(177) Claesen,C.A.A., Segers,R.P.A.M. and Tesser,G.I. (1985) *Recl. Trav. Chim. Pays- Bas* **104**, 119-122.

(178) Ciszewski,K., Celewicz,L. and Golankiewicz,K. (1995) *Synthesis-Stuttgart* **7**, 777-779.

(179) Mungall,W.S., Greene,G.L., Heavner,G.A. and Letsinger,R.L. (1975) *J. Org. Chem.* **10**, 1659.

(180) Maiti,S.N., Singh,M.P. and Micetich,R.G. (1986) *Tetrahedron Lett.* **27**, 1423-1424.

(181) Schubert,G., Schneider,G., Schade,W. and Dombi,G. (1982) *Acta. Chim. Acad. Sci. Hungary* **111**, 173.

APPENDICES

APPENDIX 1.

Duplex 10, a native oligonucleotide of the type studied, was included in the thermodynamic study in order to check the compatibility of the data obtained with that obtained by Breslauer *et al.*[174]

CGT AAC GT
GCA TTG CA **Duplex 10**

This duplex contains 1(AA/TT), 1(TA/AT), 3(GT/CA), and 2(CG/GC) nearest neighbours, and using the values from Table 3.4 (p106), with T=25°C:

$$\Delta G^\circ_{predicted} = (-7.9) + (-3.8) + 3(-5.4) + 2(-15.1) + \underline{(20.9)} = -37.2 kJmol^{-1}$$

The calculation assumes that the free energy of a duplex results from the sum of its nearest neighbour interactions. The underlined quantity corresponds to the helix initiation free energy.[174]

From a plot of $\ln(C_T/4)$ versus $1/T_m$, as described in Section 3, the following results were obtained:

$$\Delta H^\circ_{observed} = -172.5 kJmol^{-1}$$
$$\Delta S^\circ_{observed} = -452.7 JK^{-1}mol^{-1}$$

and from Eq.(4), and using T=25°C,

$$\Delta G^\circ_{observed} = -37.2 kJmol^{-1}$$

Therefore the results are in very close agreement with those predicted.

APPENDIX 2.

Publication:

"Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(*anti*).A(*anti*) base pairs."

Structure 15 June 1994, **2**: 483-494.

Leonard,G.A., McAuley-Hecht,K.E., Ebel,S., Lough,D.M., Brown,T. and Hunter, W.N.

Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(*anti*)·A(*anti*) base pairs

Gordon A Leonard[1], Katherine E McAuley-Hecht[2], Susanne Ebel[2], David M Lough[2], Tom Brown[2] and William N Hunter[1]*

[1]Department of Chemistry, University of Manchester, Oxford Road, Manchester, M13 9PL, UK and [2]Edinburgh Centre for Molecular Recognition, Chemistry Department, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK

Background: Non-Watson–Crick base pair associations contribute significantly to the stabilization of RNA tertiary structure. The conformation adopted by such pairs appears to be a function of both the sequence and the secondary structure of the RNA molecule. G·A mispairs adopt G(*anti*)·A(*anti*) configurations in some circumstances, such as the ends of helical regions of rRNAs, but in other circumstances probably adopt an unusual configuration in which the inter-base hydrogen bonds involve functional groups from other bases. We investigated the structure of G·A pairs in a synthetic RNA dodecamer, r(CGCGAAUUAGCG), which forms a duplex containing two such mismatches.

Results: The structure of the RNA duplex was determined by single crystal X-ray diffraction techniques to a resolution in the range 7.0–1.8 Å, and found to be an A-type helical structure with 10 Watson–Crick pairs and two G·A mispairs. The mispairs adopt the G(*anti*)·A(*anti*) conformation, held together by two obvious hydrogen bonds. Unlike analogous base pairs seen in a DNA duplex, they do not exhibit a high propeller twist and may therefore be further stabilized by weak, reverse, three-center hydrogen bonds.

Conclusions: G(*anti*)·A(*anti*) mispairs are held together by two hydrogen bonds between the O6 and N1 of guanine and the N6 and N1 of adenine. If the mispairs do not exhibit high propeller twist they may be further stabilized by inter-base reverse three-centre hydrogen bonds. These interactions, and other hydrogen bonds seen in our study, may be important in modelling the structure of RNA molecules and their interactions with other molecules.

Structure 15 June 1994, **2**:483–494

Key words: G·A mismatch, hydration, RNA structure, three-centre hydrogen bonding, X-ray crystallography

Introduction

Since the development of techniques to routinely synthesize large quantities of DNA, the crystallography of short oligodeoxyribonucleotides has become well established [1,2]. In addition, the detailed crystallographic analysis of tRNA represents a major success in structural biology [1]. The crystallography of synthetic RNA oligonucleotides has lagged behind because of the greater difficulties associated with synthesis and purification. To date only two RNA duplex structures have been fully determined [3–5] and the structures of four chimaeric RNA–DNA hybrid structures have also been reported [6–8]. Recently, a preliminary X-ray analysis of the synthetic RNA oligonucleotide corresponding to domain A of *Thermus flavus* 5S rRNA has been made [9] and structural detail should be available in the near future. NMR-derived structures are more common and this method has been applied in several cases [10].

In both the structure of DNA and that of RNA [2,11–15] the G·A mispair has been shown to be conformationally variable (Fig. 1). In DNA this conformational variability is both sequence and pH-dependent [11,13,16]. In RNA the structure of G·A mismatches appears to depend on both the sequence and the secondary structure of the RNA molecule. NMR studies suggest that G·A mispairs occurring at the ends of helical regions of rRNAs tend to adopt G(*anti*)·A(*anti*) configurations as do those in the sequence CGCAGGCG [15]. On the basis of NMR spectroscopic studies of G·A base pairs in the sequence GGCGAGCC, at the loop end of the stem in extra-stable tetraloops and in ribozyme models it has been concluded that these base pairs adopt an unusual configuration in which the interbase hydrogen bonds involve functional groups on the major and minor groove sides of the purine bases (Fig. 1d) [14,17–20].

As part of an ongoing project into the structure of oligonucleotides, in particular those containing non-Watson–Crick base pairs, we have synthesized for crystallographic study the dodecamer r(CGCGAAUUAGCG) which contains two G·A mispairs. We describe the synthesis, purification, crystallization and the single-crystal X-ray analysis of r(CGCGAAUUAGCG). We present a detailed structural analysis of the duplex, with an emphasis on the conformation of the mismatches, and make some comparisons with the only other RNA duplex crystal structure for which full details have been given [3].

Results and discussion

Structure solution and refinement

The RNA dodecamer was synthesized, purified and crystallized as described in Materials and methods.

*Corresponding author.

 ISSN 0969-2126

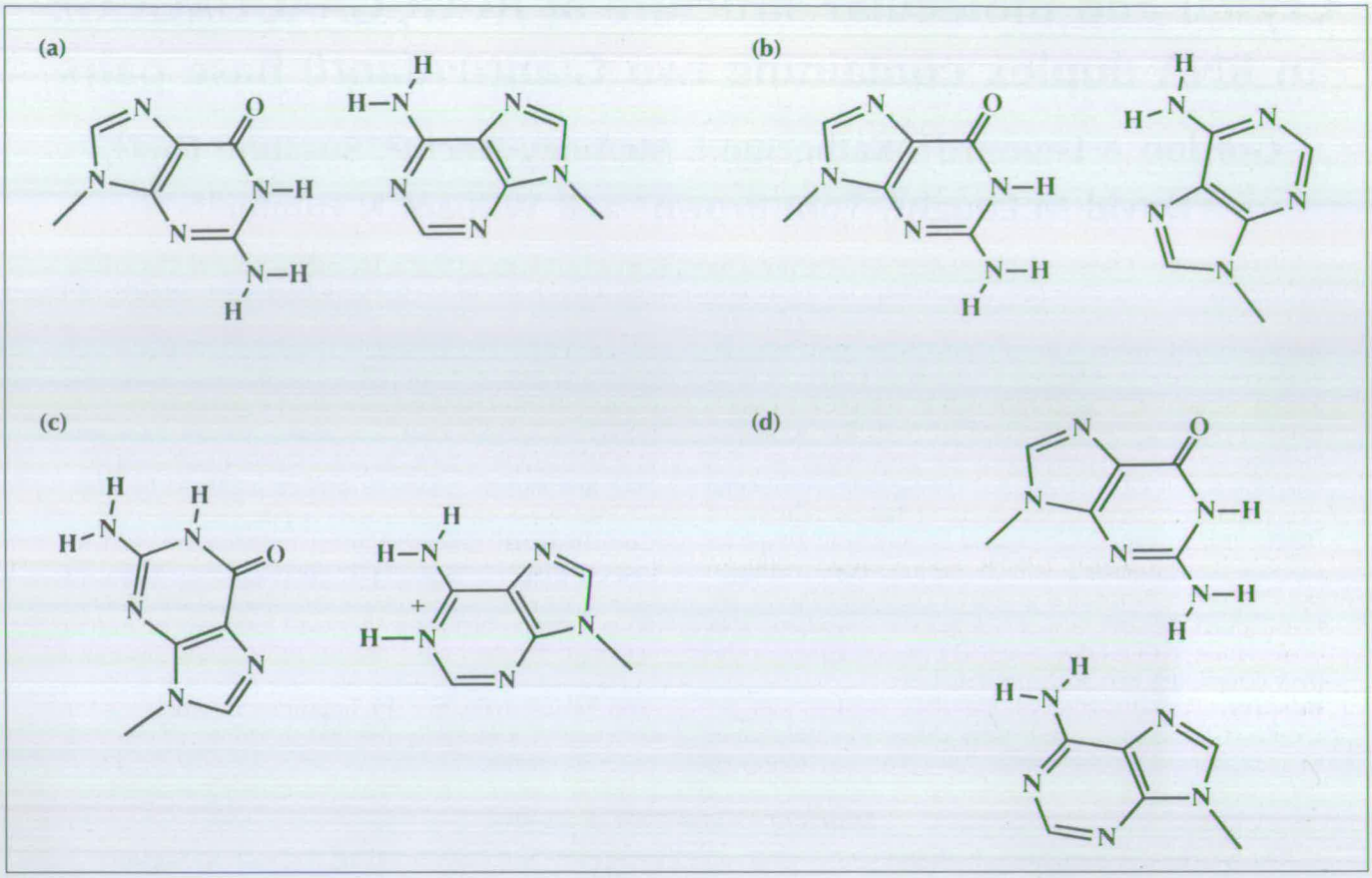


Fig. 1. Four base-pairing arrangements for GA base pairs. **(a)** The G(*anti*)·A(*anti*) conformation. **(b)** G(*anti*)·A(*syn*). **(c)** The protonated G(*syn*)·AH$^+$(*anti*). **(d)** The G·A N7-amino, amino-N3 base pair. Possible hydrogen bonding interactions have been omitted.

The space group is monoclinic, $P2_1$, with unit cell dimensions a = 41.69(1) Å, b = 34.62(1) Å, c = 32.13(3) Å, β = 127.6(1)° after post-refinement (numbers in brackets indicate standard errors). The asymmetric unit consists of one RNA dodecamer duplex. The structure was solved using the multi-dimensional search program ULTIMA [21] and refined using a combination of rigid-body techniques [22], slow-cooling simulated annealing [23,24] and restrained least-squares methods [25]. The final model consists of the RNA duplex (510 atoms) and 97 water molecules (each treated as an oxygen atom). Refinement converged at an R-factor of 19% for 3147 reflections with $F \geq 3\sigma(F)$ in the resolution range 7.0–1.8 Å (R-factor $\Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors respectively). The residues on strand 1 are labelled C1 to G12 (5′ to 3′ direction) and those on strand 2 are labelled C13 to G24 (5′ to 3′). Water molecules are numbered from 1 to 97 inclusive with the label HOH. The geometry of the final model is good with root mean square (rms) deviations from ideal bond lengths of 0.014 Å for sugars/bases and 0.030 Å for phosphate groups. For angle-associated distances the rms deviations from ideality are 0.016 Å and 0.023 Å respectively. The rms deviation from planarity for the nucleotide bases is 0.013 Å. For chiral volumes, which were not restrained, the rms deviation from ideality is 0.279 Å^3. Weak restraints were placed on the Watson–Crick hydrogen bonds in the structure but no such restraints were applied to the mispairs. An example of the fit of the model to the final calculated electron density is presented in Fig. 2.

Fig. 2. The electron density map ($2F_o - F_c$, α_{calc}; green chicken wire) for the G4·A21 base pair in the r(CGCGAAUUAGCG) duplex with the final refined model. The contour level is set at approximately one rms deviation from the mean value observed in the unit cell. The bases are depicted as sticks coloured as follows; carbon yellow, oxygen red and nitrogen blue, with key functional groups labelled.

The global structure of the RNA double helix and crystal packing

The antiparallel RNA double helix consists of 10 Watson–Crick base pairs and two G·A mispairs (G4A21, G16A9) both of which adopt a G(*anti*)·A(*anti*) conformation. The duplex has a global helical rise of 2.62 Å and global helical twist of 34.2°(10.5 residues per full helical turn) which place it in the A-form family [1,2] as do the values observed for the sugar–phosphate backbone torsion angles (Table 1; see below) and the geometrical parameters of the individual base pairs and base steps (Tables 2 and 3). The average value for the torsion angle δ is 78.5° and the standard deviation is 8° indicating that the conformation of all the ribose moieties in the structure is close to C3′-*endo*. The mean distance between adjacent phosphorus atoms on the same strand is 5.8 Å and this is also fairly typical of the A-form nucleic acid duplex. The double helix has a very deep, narrow major groove and a rather shallow, though wide, minor groove (Fig. 3). Where they can be measured, the average widths of the major groove and minor grooves are 3.4 Å and 10.8 Å respectively. These are comparable to the values of 4.1 Å and 11.3 Å derived from a fibre model of poly(A)·poly(U) [26] and also to those values found in the structures of the acceptor stem and anticodon stems of $tRNA^{asp}$ [27] and $tRNA^{phe}$ [28,29]. The global values quoted above, whilst allowing the overall cataloging of the duplex, do tend to mask structural variation at a local level. Such variation will be mentioned when the base-stacking interactions are described.

Table 1. Sugar–phosphate backbone and glycosyl torsion angles (°) for r(CGCGAAUUAGCG).

Residue	χ	α	β	γ	δ	ξ	ζ
C1	−162			30	83	−172	−62
G2	−156	−62	189	45	83	−168	−57
C3	−162	−95	191	76	76	−116	−106
G4	−165	−60	135	64	69	−155	−73
A5	−173	−74	177	76	71	−136	−90
A6	−151	−43	157	45	79	−139	−83
U7	−169	−71	158	72	83	−135	−95
U8	−155	−29	163	31	74	−138	−76
A9	−165	−91	157	73	72	−149	−71
G10	−168	−71	169	71	81	−142	−84
C11	−156	−52	165	47	80	−154	−78
G12	−157	−49	172	40	77		
C13	−180			170	91	−114	−103
G14	−163	−45	155	53	70	−129	−87
C15	−150	−9	182	−2	91	−154	−80
G16	−166	−106	171	75	82	−156	−76
A17	−171	−54	171	54	82	−135	−98
A18	−146	−12	154	16	83	−169	−55
U19	−158	−100	187	84	63	−137	−74
U20	−159	−47	179	27	96	−136	−94
A21	−161	−40	147	41	66	−156	−68
G22	−157	−49	168	45	77	−161	−65
C23	−166	−92	176	78	80	−142	−81
G24	−159	−45	164	46	76		

Main chain torsion angles are defined by O3′-P-α-O5′-β-C5′-γ-C4′-δ-C3′-ξ-O3′-η-P-O5′. The glycosyl torison angle χ is defined by 04′-C1′-N1-C2 for pyrimidines and 04′-C1′-N9-C4 for purines.

As we had predicted when considering the solution of the structure (see Materials and methods) the duplexes are arranged in the crystal so that they pack parallel to the c-axis of the unit cell and form a pseudo C-centred arrangement (Fig. 4). Each duplex interacts with six symmetry-related entities. Duplexes pack one on top of the other in a similar fashion to the structure of r(GGACUUCGGUCC) [4] to form a series of helical

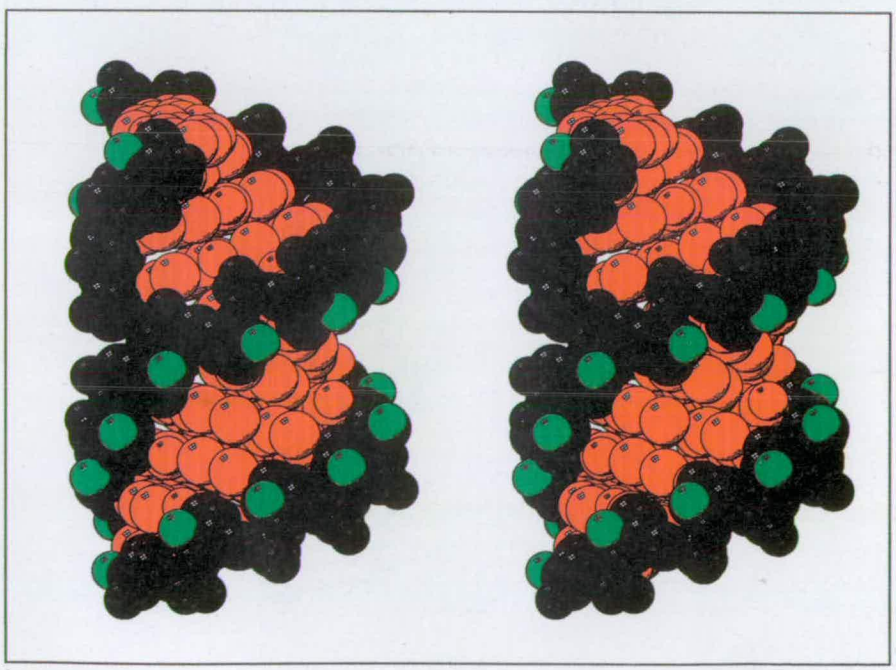


Fig. 3. A CPK representation in stereo of the RNA duplex. Base atoms are red, the sugar–phosphate atoms are black with the exception of the O2′ hydroxyl which is coloured green. The selected view has the minor groove at the bottom of the diagram, the major groove at the top. (Figs 3, 5, 7, 8 and 9 were obtained using MOLSCRIPT [53]).

columns parallel to the c-axis. At the junctions of adjacent helices in these RNA columns there is a psuedo GpC step. The helical rotation for this psuedo step is only 8.9° while the helical rise is 2.86 Å. In the duplex proper there are two real GpC steps which have helical rotations of 33.5°and 35.6°and helical rises of 2.45 Å and 2.55 Å (Table 2, discussed below). The rise and rotation observed for the pseudo GpC step at the junction of two duplexes are therefore a consequence of the lack of chain connectivity rather than a true reflection of the rise and twist for GpC steps. The van der Waals interactions between the aromatic bases of the symmetry-related helices are supplemented by van der Waals interactions involving atoms of the attached ribofuranose rings.

There are four helices positioned around the side of each duplex. The side-by-side array of helices is held together by numerous van der Waals contacts between symmetry-related duplexes all of which involve atoms on the sugar–phosphate backbone. Additionally, there are five inter-duplex hydrogen bonds associated with each RNA strand. These hydrogen bonds are all to functional groups on a symmetry-related partner strand, so strand 1 only interacts with strand 2 of symmetry-related helices and *vice versa*. Hence, in total there are 10 inter-duplex hydrogen bonds to stabilize the packing arrangement. These hydrogen bonds involve the O2′ and O3′ of C(3) interacting with O2′A(21), O2′A(6) with O2′G(24), O2′A(9) with O2′C(15) and O2′G(12) with O2′A(18). An example of the interduplex hydro-

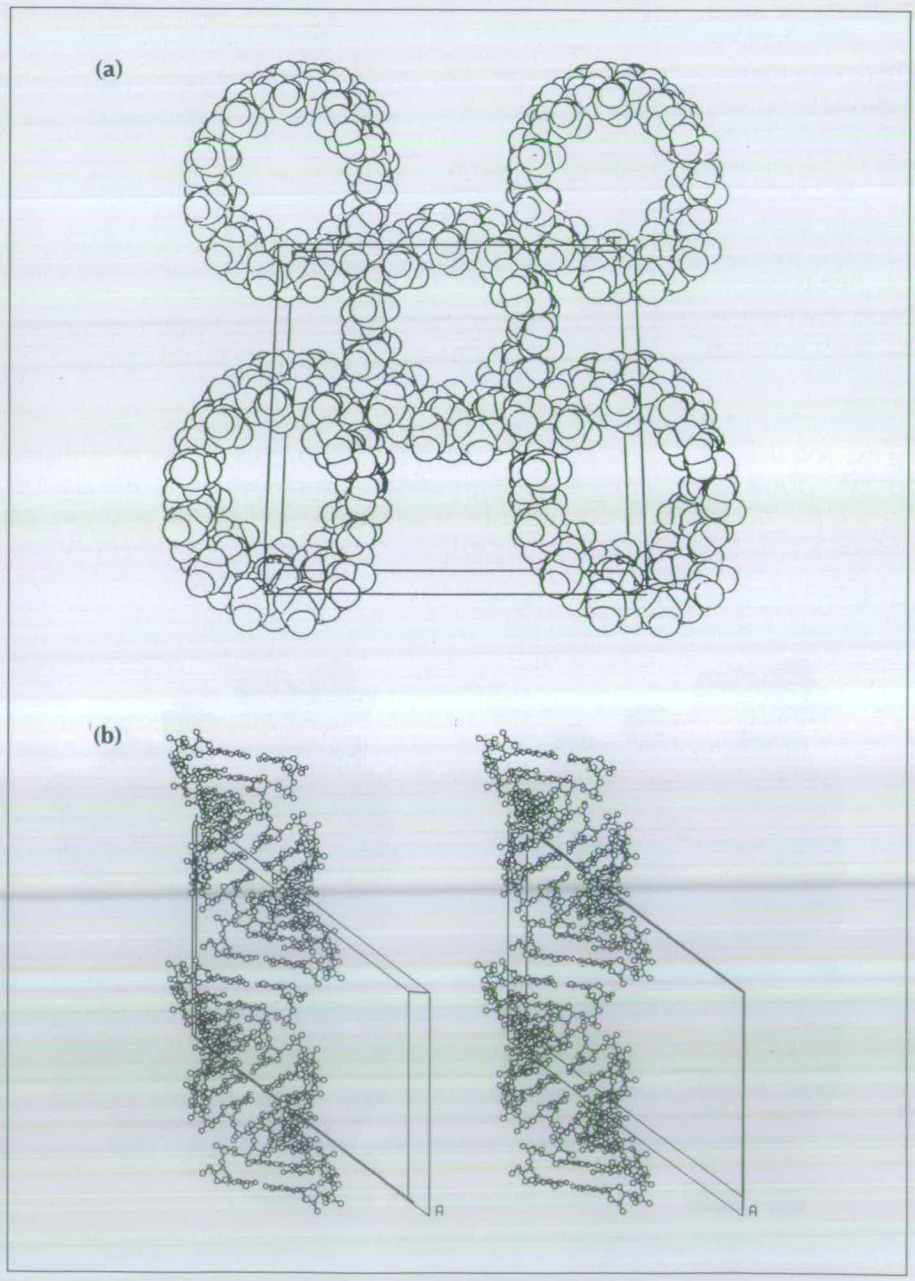


Fig. 4. The arrangement of r(CGCGAAUUAGCG) duplexes in the unit cell. **(a)** The duplexes viewed approximately parallel to their helix axes to show the cross-section of the cell perpendicular to the c-axis. For clarity, only the sugar–phosphate backbone atoms are shown. Note that the arrangement closely mimics a C-centred motif and that the central helix interacts with four others in this plane. **(b)** A stereoview illustrating how the helices are aligned along the c-axis to form a continuous tube of RNA.

Table 2. Geometric parameters for the base steps in the r(CGCGAAUUAGCG) duplex.[a]

Base step	Twist(°)	Rise (Å)	Slide (°)	Roll (°)
1	34.4	2.4	−1.9	1.6
2	33.5	2.5	−1.2	−4.7
3	27.2	3.1	−1.1	−0.1
4	30.4	2.7	−1.4	3.4
5	38.9	2.6	−0.6	−4.0
6	32.8	2.4	−1.0	−3.3
7	40.3	2.5	−1.1	−2.9
8	28.3	2.7	−1.4	3.8
9	28.0	3.0	−1.0	−1.4
10	35.6	2.6	−1.2	−2.0
11	39.6	2.5	−1.5	1.3

[a]Parameters were calculated using the NEWHEL92 program distributed by RE Dickerson and available from the Brookhaven Protein Data Bank.

Table 3. Geometric parameters for the base pairs in the r(CGCGAAUUAGCG) duplex.[a]

Base pair	Propeller twist (°)	C1′–C1′ (Å)	Inclination (°)
C1·G24	−12.5	10.4	17.9
G2·C23	−20.1	10.6	17.7
C3·G22	−9.1	10.6	19.4
G4·A21	−9.6	12.4	18.9
A5·U20	−12.4	10.6	19.0
A6·U19	−27.0	10.1	21.1
U7·A18	−23.1	10.4	20.3
U8·A17	−17.8	10.9	18.4
A9·G16	−10.4	12.9	18.7
G10·C15	−6.8	10.7	18.3
C11·G14	−22.6	10.4	18.0
G12·C13	−13.3	10.7	17.8

[a]Parameters were calculated using the NEWHEL92 program distributed by RE Dickerson and available from the Brookhaven Protein Data Bank.

gen bonding is presented in Fig. 5. There are also several solvent mediated contacts which principally link sugar–phosphate backbones.

The crystal structure of $U(UA)_6A$ [3] also uses the O2′ hydroxyl groups in the formation of the crystal lattice but in this case hydrogen bonds are formed to both the sugar–phosphate backbone and to the minor groove of the duplex. The importance of the O2′ hydroxyl group in the ribozyme recognition of RNA has been brought to prominence by Pyle and Cech [30]. These crystal structures may mimic aspects of this recognition process.

The conformation of the G·A base pairs

The G·A base pairs in the structure of r(CGCGAAUUAGCG) adopt a G(*anti*)·A(*anti*) conformation (Figs 1a and 2). This configuration is different from the G(*anti*)·A(*syn*) G·A mispairs in the analogous DNA structure d(CGCGAATTAGCG) [31], but similar to the configuration of the G·A base pairs observed in the structure of the synthetic B-form DNA decamer d(CCAAGATTGG) [12,32]. As in the d(CCAAGATTGG) duplex, the C1′–C1′distances for the G(*anti*)·A(*anti*) base pairs in the r(CGCGAAUUAGCG) double helix are increased by about 2 Å (Table 3) compared with the value of 10.5 Å associated with Watson–Crick base pairs. From an analysis of the sugar–phosphate torsion angles observed for the r(CGCGAAUUAGCG) double helix it would seem that the larger G(*anti*)·A(*anti*) base pairs are incorporated into the duplex with very little change to the structure of the phosphodiester backbone (see below). An examination of phosphorus atom separations across the helix suggests that, as in the structure of d(CCAAGATTGG) [12], there is slight bulging of the duplex as a result of the G(*anti*)·A(*anti*) conformation adopted by the mispairs. This bulging of the backbone occurs only at points immediately adjacent to the mispairs and the rest of the duplex is apparently unaffected.

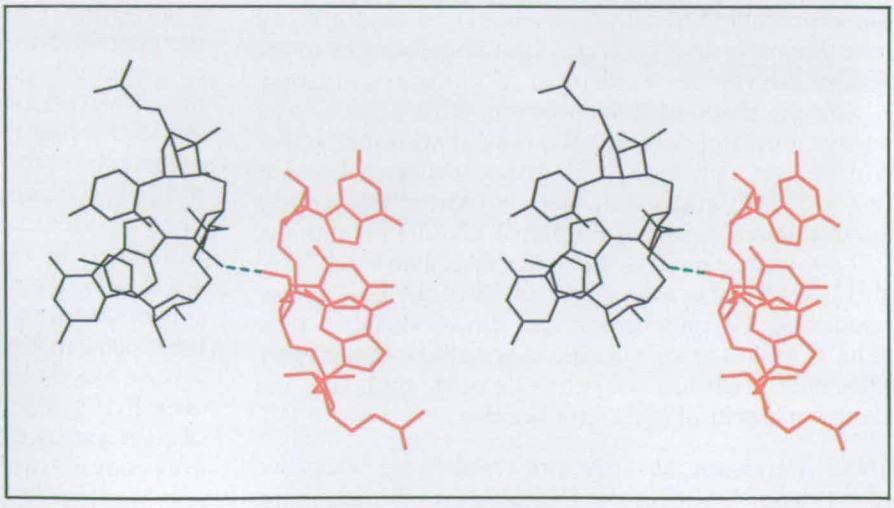


Fig. 5. Stereoview to show one of the O2′ to O2′ hydrogen bonds that help to stabilize the crystal lattice. In black is the trinucleotide r(UAG), residues 8, 9 and 10 of the asymmetric unit. In red, r(GCG) residues 14, 15 and 16 of a symmetry related duplex (1−x, 1/2+y, −z). The dashed green line indicates the hydrogen bond formed between O2′A(9) and O2′C(15).

Although the global conformations of the G(*anti*)·A(*anti*) mispairs in r(CGCGAAUUAGCG) are similar to those observed in d(CCAAGATTGG) there are a number of significant differences in the fine structure of the base pairs. The most obvious is in the degree of propeller twist that the base pairs exhibit. In the structure of r(CGCGAAUUAGCG) the G4·A21 and G16·A9 base pairs have fairly low propeller twists of 9.1° and 10.4°, respectively (Table 3). In the structure of d(CCAAGATTGG) the G·A mispairs have high propeller twists of about 25° [12]. In the latter structure, the mispair guanines are involved in unusual minor groove cross-strand hydrogen bonding. The guanine N2 amino group hydrogen that is directed towards the adenine in the G·A base pair forms a cross-strand hydrogen bond to the adjacent thymine O2. This feature is quite definately sequence dependent. Since the other hydrogen atom on the amino group is free to form hydrogen bonds to solvent, this means that the hydrogen-bonding potential of the guanine N2 amino group can be completely fulfilled. In our RNA duplex, the sequence in which the G·A mispairs are embedded makes it impossible for similar hydrogen bonds to form. A consequence of this is that, at first sight, one of the guanine N2 amino group hydrogen atoms in the G·A base pairs in the structure of r(CGCGAAUUAGCG) (circled in Fig. 6) cannot form a hydrogen bond. Note that there is one amino group hydrogen that can interact with solvent and this is indeed observed for G16, where HOH42 is 3.4 Å from N2. We have not, however, identified a well-ordered water hydrogen bonding to N2(G4).

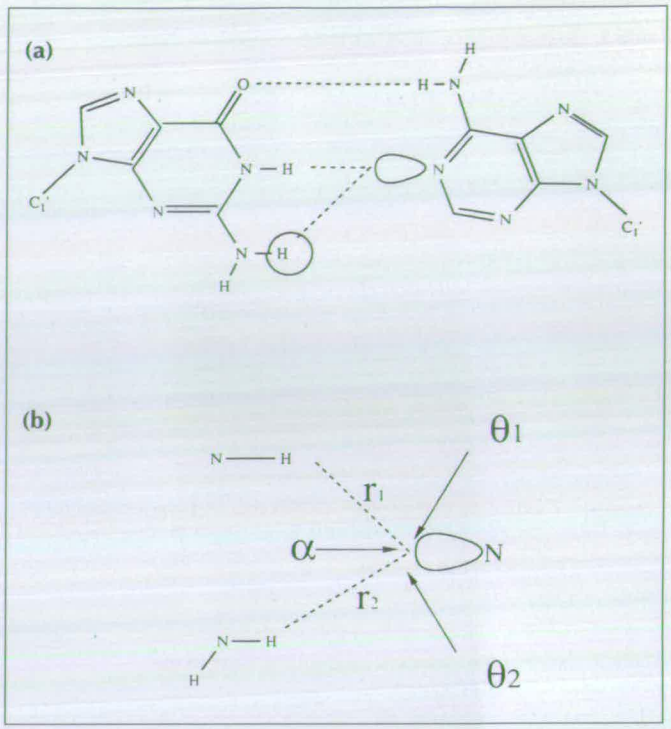


Fig. 6. A representation of the reverse three-centre hydrogen bonding scheme that may help to stabilize the G·A base pairs in the structure of r(CGCGAAUUAGCG). **(a)** The G(*anti*)·A(*anti*) bases. The lone pair on N1 of the adenine is depicted as a lobe, the guanine proton that could interact with this lone pair is circled. **(b)** The definitions of geometrical parameters tabulated in Table 2 which are useful to analyze three-centre hydrogen bonding possibilities. Parameters r1 and r2 are distances, α, θ1 and θ2 are angles.

It was Crick, in 1966, who recognized that base pairs containing unfulfilled hydrogen bond donors or acceptors would be destabilized and that this effect might even prevent base pairing [33]. With this in mind we have studied the possible hydrogen bonding patterns in this mispair, paying particular attention to the guanine N2 amino group. We remind readers that the electrostatic nature of the hydrogen bond means that the attractive force changes in very small decrements as the interatomic separation increases [34] and it is not necessary to restrict hydrogen-bonding distances to the sum of the van der Waals radii of the atoms involved. A difficulty arises in nomenclature; what actually constitutes a hydrogen bond? We concur with the Steiner and Saenger definition [35] that a hydrogen bond is "any cohesive interaction X–H···Y, where H carries a positive and Y a negative (partial of full) charge and that the charge on X is more negative than on H". We also believe that in a detailed analysis of such a cohesive interaction the dimensions and directionality of lone pairs as well as atomic positions should be considered. This latter point has not generally been applied in the characterization of hydrogen bonds.

There are clearly two distances which are indicative of hydrogen bonds in the G(*anti*)·A(*anti*) base pairs. These involve N6(A) to O6(G) (2.96 Å for the 4·21 pair and 2.66 Å for the 9·16 pair) and N1(G) to N1(A) (2.62 Å and 2.96 Å respectively). These are the primary determinants of the base alignments. We postulate that the inter-base hydrogen bonding scheme in the G·A base pairs may also involve an unconventional reverse three-centre hydrogen bond we previously proposed for other base pairs. In this scheme both the guanine imino hydrogen and the N2 amino group hydrogen atom form hydrogen bonds to the adenine N1 nitrogen atom (Fig. 6) and the hydrogen-bonding potential of the guanine N2 amino group can be fulfilled. To ascertain whether such a hydrogen-bonding scheme could occur we calculated ideal positions for both hydrogen atoms and the adenine N1 nitrogen lone pair [36] and analyzed the geometry of the potential hydrogen-bonding scheme. A full description of our protocol has been published elsewhere [37]. Distances and angles, as defined in Fig. 6, are presented in Table 4. They are close to the values expected for three-centre hydrogen bond formation [37–40] and suggest that the G·A base pairs in this RNA duplex could be stabilized by reverse three-centre hydrogen bonds. We caution however, that in energy and structural terms the presence of such a weak electrostatic interaction will be of minor importance compared with the two principal hydrogen bonds described above. An NMR study of the G·A mispair in d(GCCACAAGCTC)·d(GAGCTGGTGGC) by

Carbonnaux *et al.*, [41] revealed the bases were *anti* and an unusual chemical shift of an imino proton of the guanine suggested the formation of inter-base bifurcated hydrogen bonds. Our results are in accord with their study.

Table 4. The geometrical parameters associated with the three-centre hydrogen bonds found in the structure of r(CGCGAAUUAGCG).[a]

	r_1(Å)	r_2(Å)	θ_1(°)	θ_2(°)	α (°)	Σ (°)
Reverse three-centre hydrogen bonds						
G4·A21	1.7	2.9	151	110	54	315
G16·A9	2.0	2.6	150	123	76	349
Conventional three-centre hydrogen bonds						
A5·U20 : A6·U19	1.6	2.7	155	107	97	359
A17·U8 : A18·U7	1.6	3.0	152	111	85	348

[a]Both conventional three-centre hydrogen bonds [40] and reverse three-centre hydrogen bonds [36] are considered to be formed if $r_1 < r_2 \leq 3$ Å and $\Sigma \approx 360°$ (i.e. $\theta_1 + \theta_2 + \alpha \approx 360°$).

Conformation of the phosphate–furanose backbone

The torsion angles of the phosphate–furanose backbone and glycosyl bonds are presented in Table 1. In general the values observed correlate well with the expected range of values for an A-form duplex [1,2,26]. So α and γ assume *gauche*$^-$/*gauche*$^+$ configurations with the exception of γ for C13 which is *trans* and subject to end effects. Values of β and ξ are *trans*, while ζ is *gauche*$^-$. The glycosyl torsion angle, χ, ranges from $-146°$ to $-180°$. There is no apparent difference in χ values for purines or pyrimidines as has previously been observed for A-DNA structures [42]. As mentioned earlier the furanose conformations, as deduced from the δ values, are all C3′-*endo*. In accord with this the intra-strand phosphorus–phosphorus distances range from 5.3 Å to 6.3 Å with an average value of 5.8 Å. There is no gross distortion in the furanose–phosphate backbone conformation that can be attributed to the location of the mispair sites.

Base pair stacking patterns and local geometry

In addition to observing how the mispairs interact with the adjacent base pairs, we also see the detailed geometry of some base pair steps involving Watson–Crick pairs that have not previously been observed in crystal structures of duplex RNA. In the r(CGCGAAUUAGCG) duplex there are 11 base pair steps, 7 of which involve only Watson–Crick base pairs and 4 of which involve the mispairs. Our numbering scheme assigns the dinucleotide step 5′C1pG2 3′ (interacting with 5′C23:G24 3′) as step 1 which can be indicated as type CpG(CpG), 5′G2pC3 3′ with 5′G22:C23 3′ as step 2, type GpC(GpC) *etc.* Geometrical parameters used to define local geometry of base pair steps and of the base pairs themselves are presented in Table 2.

The four steps which involve bases of the mispairs are steps 3 and 9 of type CpG(ApG) and steps 4 and 8 of type GpA(UpA). There is consistency between these steps in that intra-strand purine–purine stacking interactions appear to dominate. In each case good stacking of the five-membered ring of a purine onto the six-membered ring of the adjacent purine is observed. There is also a small degree of intra-strand pyrimidine–purine stacking and of cross-strand stacking of purines. These four steps have a reduced twist (average of 28.5°) compared with the average value of 36.4° for the remaining seven steps. Steps 3 and 9 also have the highest rise values.

The remaining base pair steps can be grouped into the following categories. Steps 1 and 11 are type CpG(CpG), 2 and 10 are GpC(GpC), 5 and 7 are ApA(UpU), and step 6 is ApU(ApU). The stacking patterns observed here are similiar to those observed in $[U(UA_6)A]_2$ [3] and fibre models [1,26] and are determined by whether the bases involved are purines or pyrimidines. For the purpose of identifying a pattern it does not appear important to consider which type of purine or pyrimidine is involved. The patterns can be summarized as follows. The purine–pyrimidine steps, 2, 6 and 10, show pronounced intra-strand stacking of the pyrimidine on the six-membered ring of the purine. No cross-strand interactions are observed. The pyrimidine–purine steps, 1 and 11, display good cross-strand stacking of purines on each other. The purine–purine steps, 5 and 7, are arranged so that the five-membered ring on one purine stacks on the six-membered ring of the adjacent one. There is very little interaction between the pyrimidines on the other strand. This is very similar to the type of stacking observed in GpG(= CpC) steps in A-form DNA [42]. All of the steps at which some degree of cross-strand purine–purine interaction is observed have higher slide values (average 1.55 Å) than the other steps (average 1.03 Å). The lowest slide value is observed for step 5 where there is the strong intra-strand stacking of the purines.

A feature of the r(CGCGAAUUAGCG) duplex is the propeller twist of the A·U base pairs. Two of these base pairs (A6·U19 and A18·U7) have extremely high propeller twist (Table 3) with magnitudes of the same order as the A·U base pairs found in the structure of $U(UA)_6A$ [3] and the A·T base pairs in poly(A)·poly(T) stretches of B-DNA [43]. Calculation of idealized positions for hydrogen atoms in the r(CGCGAAUUAGCG) duplex, as detailed previously, and an analysis of possible hydrogen-bonding geometry as defined by Fritsch and Westhof [40] (Table 4) suggests that both of these base pairs may be involved in three-centre, cross-strand hydrogen bonding with the adjacent 5′ A·U base pair (Fig. 7) similar in nature to that found in B-DNA [43]. To the best of our knowledge this is the first time such

interactions have been proposed in duplex RNA and may be useful in molecular modelling of complex RNA structures.

The base pairs with the next highest values for propeller twist are at G2·C23 and C11·G14 with values of $-20.1°$ and $-22.6°$ respectively. These positions are one base pair in from the end of the duplex. As discussed earlier, the crystal lattice is in part stabilized by a base-stacking type interaction involving the terminal base pairs. The base steps (1 and 11) have the largest values for slide and step 11 has the largest value for twist. The intermolecular interactions may influence the G·C pairs one position in from the end of the duplex to promote a high propeller twist and the base-stacking pattern allows the pyrimidine room to flex away from the plane of the purine. The observation of high propeller twisted G·C pairs is in direct contradiction to suggestions, based on DNA structures, that there are two requirements for a base pair to exhibit a large propeller twist [44]. These requriements are that the base pair must be A·T and it has to be involved in major groove three-centre hydrogen bonding with an adjacent base pair. There are numerous examples of G·C pairs in DNA dodecamers with high propeller twist values in excess of $-15°$ and this is further discussed by Leonard and Hunter [45]. The high values for two of the G·C pairs in this RNA duplex are at the high end of the range observed in a survey of nucleic acid structural parameters [46], but there is precedent in A-form DNA. In the structure of d(ICCGG) [47], a propeller twist of $-21.4°$ is evident for a G·C pair and we also note that in the crystal structures of tRNA molecules high propeller twist values can be obtained for G·C pairings [29].

Hydration of the r(CGCGAAUUAGCG) duplex

Two views of the RNA duplex wth associated solvent molecules are presented in Fig. 8. Of the 97 water molecules included in the final model, 65 (67%) form direct hydrogen bonds to the RNA double helix and represent the first hydration shell. Of these first shell solvent molecules, 19 form hydrogen bonds to the major groove of the RNA, 14 interact directly with atoms in the minor groove and 32 form hydrogen bonds to the sugar–phosphate backbone. Of the latter solvent molecules, 11 interact with O2′ hydroxyl groups on the ribose moieties. There are no striking patterns of hydration in the major groove of the RNA duplex. For the backbone, we observe only a couple of well defined solvent molecules that bridge adjacent phosphate groups in a manner expected on the basis of Saenger and co-workers' study of hydration [48].

On the minor groove side of the duplex we observe a number of bridging solvent molecules, and examples are shown in Fig. 9. These involve 7 of the 14 solvents in this groove bridging atoms on the base edges and the O2′ hydroxyl groups of the riboses. This is reminiscent of the hydration pattern of the wide sections of the minor grooves of B-DNA duplexes [12] where solvent molecules also form bridges between the bases and the sugar–phosphate backbone and suggests that such a linking motif may be conserved in both DNA and RNA.

The O2′ hydroxyl group plays an important part in the stabilization of RNA secondary structure [1]. Modelling has suggested that this group can take part in three types of interactions [49]. Firstly, a weak hydrogen bond may form to O4′ of the neighbouring nucleotide unit. Secondly, the positioning of O2′ could promote the location of bridging solvent molecules linking up with O3′ of the phosphate group, or, thirdly the group could link up with the minor groove of the bases as already discussed. There are four examples of waters bridging the hydroxyl group and the phosphate O3′. We observe an additional interaction where, at two positions, a solvent link is formed with the O4′ of the adjacent ribofuranose. In both tRNA and the r[U(UA)$_6$A] structures the presence of similar bridging solvent molecules has been noted [3,29].

At numerous positions in the electron density maps there are indications of additional solvent molecules that could be included in the model. However, the quality of the peaks did not fit in with our selection criteria and we judged it best to leave them out. We feel that we have made a realistic interpretation of the solvent structure based on consideration of data quality.

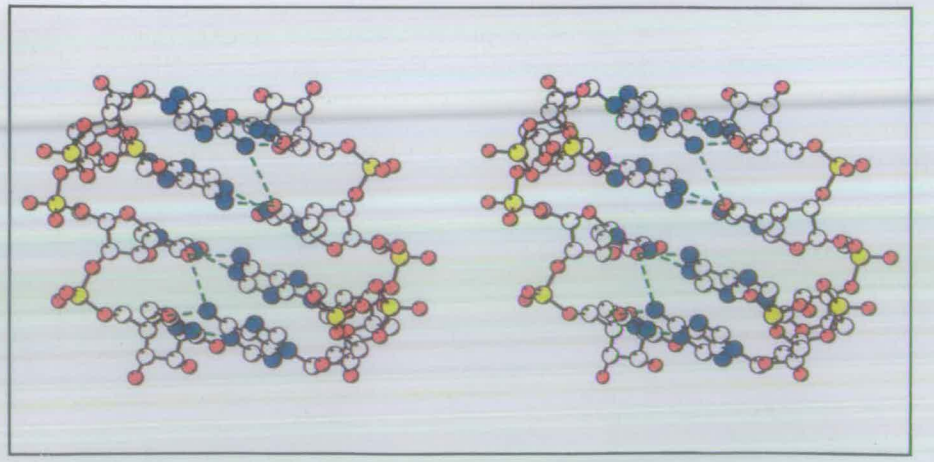


Fig. 7. A stereo ball-and-stick representation of the r(AAUU)$_2$ middle section of the RNA duplex. This view highlights the large propeller twist of the A6·U19 and A18·U7 base pairs and the major groove, cross-strand three-centre hydrogen bonds (green dashed lines) that could form as a result of this. Atoms are coloured as follows; phosphorous yellow, oxygen red, nitrogen blue, carbon white.

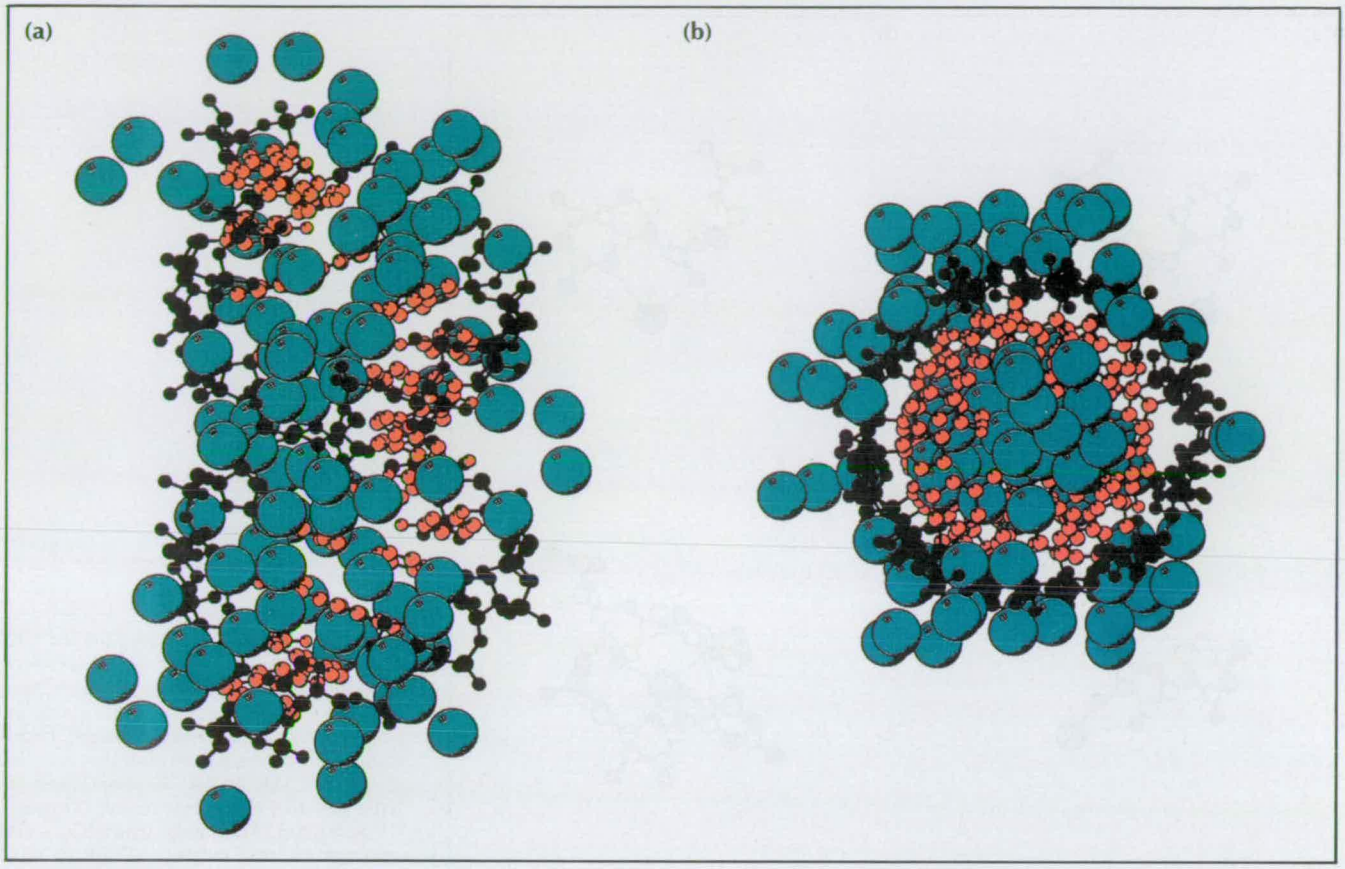


Fig. 8. Two views of the duplex (ball-and-stick model, red for base atoms, black for sugar–phosphate backbone atoms) with associated solvent positions depicted as cyan spheres. **(a)** Side view perpendicular to the duplex axis. **(b)** View down the helix axis.

Biological implications

The structures of RNA molecules are likely to be important in processes as diverse as RNA transport out of the nucleus, splicing, translation, RNA-mediated catalysis, viral assembly, and target binding in RNA-based drug candidates. Many RNA sequences have been determined, yet there is only a limited amount of information on the structures of RNA molecules. Analysis by NMR, molecular modelling [50] and X-ray crystallography (as in this study) is now beginning to give insight into the structure–function relationship of various RNAs, and should improve our understanding of the way that RNA interacts with other molecules, such as proteins.

In many RNA molecules it is clear that non-Watson–Crick base-pair associations are important in determining the three-dimensional structure of the molecule. Of the many examples found in tRNA, G·A mismatches appear most important, being commonly identified in sequence-based modelling and NMR studies. Our study of a duplex containing two such mismatches has allowed us to determine in detail the conformation that this mismatched pair adopts in a synthetic RNA dodecamer, and to probe the reasons for its stability. We find that both G·A pairs adopt the *anti-anti* conformation, and make two clear hydrogen bonds (N6 of adenine to O6 of guanine and N1 of guanine to N1 of adenine). We propose that the stability and lack of propeller twist seen in this structure may be partly due to weak, unconventional three-center hydrogen bonds to adjacent base pairs.

The data from our study provide a glimpse of the hydration patterns, geometry, and hydrogen bonding involved in interactions between RNA molecules. This may aid molecular modelling studies of higher-order RNA structures in cases where experimentally-derived structural detail is lacking.

Materials and methods

RNA synthesis and crystallization

The oligoribonucleotide was synthesized on a preparative scale using an ABI 320B automatic DNA synthesizer and phosphoramidites with 1-(2-fluorophenyl)-4-methoxypiperidin-4-yl (FPMP) protected 2′-hydroxyl functions. The crude product,

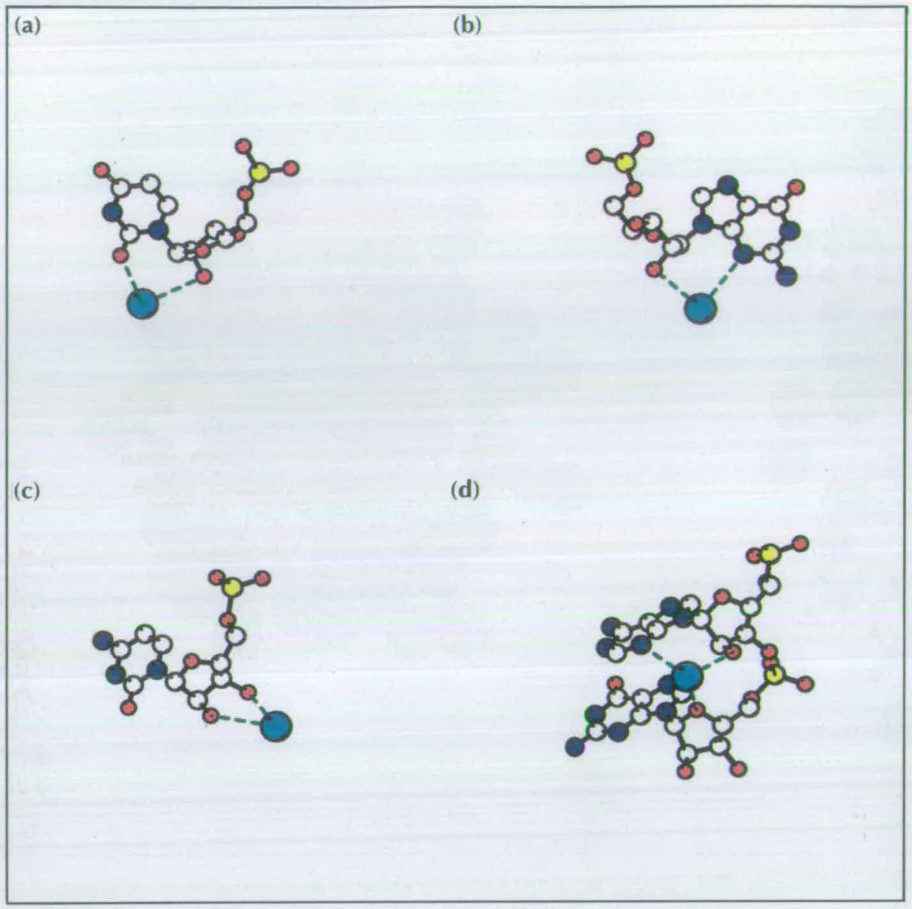


Fig. 9. Four specific examples of O2′ hydroxyl hydration. The colour scheme is; phosphorous yellow, oxygen red, nitrogen blue, carbon white, solvent molecules cyan spheres. Hydrogen bonds are green dashed lines. **(a)** The link between O2′U(8) to the O2 mediated by HOH89. **(b)** O2′G(16) to N3 through HOH40. **(c)** The link from O2′C(11) through HOH52 to the O3′. **(d)** HOH35 interacts with O2′A(21), N3A(21) and O4′G(22).

with the FPMP and 5′-trityl protecting groups still attached, was purified by reversed-phase high pressure liquid chromatography (HPLC) using a Brownlee aquapore octyl column (25 × 10 mm) and a gradient of 0 % to 50 % acetonitrile. This step was carried out over a period of 30 minutes in a buffer of 0.1 M ammonium acetate at pH 7.5. The sample was subsequently desalted using a Sephadex G-10 column (30 × 2.5 cm). Protecting groups were removed with aqueous acetic acid at pH 2.0–2.5 and the oligonucleotide was re-purified by anion-exchange, with a Partisil 10 SAX column (25 × 8 mm) using a gradient of 0.04–0.67 M potassium phosphate in the presence of 20 % acetonitrile (pH 6.4). The reversed-phase HPLC step (above) was repeated with a modified gradient of 0–20 % acetonitrile and the desalting stage was also repeated to produce the final sample. In our experience, sound laboratory practice has been sufficient to prevent nuclease degradation of this synthetic RNA both at the synthetic stage and during crystallization. Glassware was autoclaved, MilliQ purified (or doubly distilled) water, and high purity chemicals were used throughout.

Crystals were grown from sitting drops containing 1 mM oligonucleotide, 12.5 mM sodium cacodylate (pH 6.5), 50–60 mM magnesium chloride, 6.2 % (v/v) 2-methyl-2,4-pentanediol (MPD) and 0.9 mM spermine tetrahydrochloride equilibrated against an external reservoir of neat MPD. Single crystals appeared within 1–2 days and achieved final dimensions after 1–2 weeks. The largest crystals were grown at temperatures ranging from 20–25°C.

X-ray data collection and structure determination

A needle of dimensions 1.2 × 0.2 × 0.1 mm was used for data collection on an R-AXIS II c image plate system mounted on a Rigaku RU200 rotating anode generator operating at 50 KV and 100 mA. Data were measured at 24°C with graphite-monochromated CuK_α radiation ($\lambda = 1.5418$ Å, 0.5 mm focal spot). Unit cell determination was achieved by auto-indexing reflections on three still photographs taken at 45° intervals with 10 minute exposure times. This resulted in a triclinic unit cell of dimensions a = 32.1 Å, b = 27.1 Å, c = 27.0 Å, $\alpha = 79.5°$, $\beta = 118.1°$, $\gamma = 118.1°$ which transforms to a C-centred monoclinic cell with dimensions a = 41.69(1) Å, b = 34.62(1) Å, c = 32.13(3) Å, $\beta = 127.6(1)°$ after post-refinement. Initially the image plate was set at a distance of 84 mm (corresponding to a data collection limit of 1.8 Å) from the crystal and the data collected on 25 × 5° oscillation frames with an exposure time of 20 minutes per frame. Some of the low to medium resolution data were saturating the detector with that length of exposure. A second data set was collected with the image plate at a distance of 138 mm, data collection limit 2.5 Å. These data were collected on 18 × 10° oscillation frames with exposure times of 10 minutes and enabled the most intense reflections to be measured.

The length of the c-axis (32.13 Å) suggested that the duplexes pack parallel (or approximately so) to this axis. The cross-section of the cell perpendicular to the c-axis is of dimensions 34.6 Å by $a\sin\beta = 33.0$ Å and is appropriate for the close-packing, in a C-centred motif, of A-form nucleic acid duplexes with diameter of about 20 Å. Close inspection of the data indicated that the unit cell is primitive although the majority of reflections with h + k odd are very weak and the diffraction pattern is pseudo C-centred. The assumption of one duplex per unit cell leads to a calculated volume per base pair of 1527 $Å^3$ which is within the usual range for A-form duplexes [2]. We therefore anticipated a solution to the structure in which the RNA duplexes adopted a C-centred type packing as shown in Fig. 4.

For the first data set a total of 16956 reflections yielded 3447 unique data with $I \geq \sigma(I)$ ($R_{sym} = 4.3\%$). For the second set, 8535 reflections produced 1570 unique data with $I \geq \sigma(I)$ ($R_{sym} = 5.3\%$). Scaling the two data sets together and merging resulted in a data set of 3863 reflections with $R_{merge} = 5.7\%$ to a resolution of 1.8 Å. Auto-indexing of the still photographs and data measurement and reduction from the oscillation frames were performed with the software package PROCESS [51]. Scaling and merging together of the two individual data sets were carried out with the programs ANSC and AGROVATA from the CCP4 software package [SERC (UK) Collaborative Computer Project 4, Daresbury Laboratory, UK, 1979].

Structure solution in the two possible primitive space groups P2 and $P2_1$ was attempted using a canonical RNA duplex of sequence r(CGCGAAUUAGCG)·r(CGCUAAUUCGCG) generated using the SYBYL software package (SYBYL software, Tripos Associates, St. Louis, MO, USA) as a search model. The ULTIMA program [21] was used to find possible solutions based on a five-dimensional R-factor search using data in the resolution range 18.0–10.0 Å. In the space group P2, all possible solutions resulted in R-factors of greater than 60% when refined as rigid bodies against the same data. These were not pursued. In the space group $P2_1$, however, several equivalent possible solutions resulted in residuals of less than 40%. One of these was examined on an Evans and Sutherland ESV/30 graphics workstation and was found to have excellent crystal packing. The arrangement of the RNA double helices closely mimics a C-centred packing motif and given that the X-ray data had revealed the unit cell to be pseudo C-centred we considered this solution to be viable for further refinement.

Structure refinement

Preliminary refinement of the model was carried out using rigid-body techniques with a modified version of SHELX [22]. This was immediately followed by a round of refinement using the X-PLOR software package [23] which consisted of energy minimization of the model, a simulated annealing protocol [24] and 120 cycles of Powell minimization using data with $F \geq 2\sigma(F)$ in the resolution range 8.0–2.6 Å. The simulated annealing protocol consisted of a slow-cooling procedure in which the temperature of the system was reduced from 2000 K to 300 K in steps of 25 K with a time-step of 0.2 fs being employed for the molecular dynamics calculations. This resulted in an R-factor of 24.7% for 1646 reflections. Electron density ($2F_o - F_c$, α_{calc}) and difference density ($F_o - F_c$, α_{calc}) maps were then calculated and examined on the graphics workstation [52]. These confirmed that the solution was correct.

Our initial model was a duplex consisting of 12 Watson–Crick base pairs and the electron density maps calculated at this stage were unequivocal in suggesting that it should be rebuilt so that it contained two G(*anti*)·A(*anti*) base pairs. This was coupled with alterations of the sugar–phosphate backbone in order that it made a better fit with the density. The refinement was then continued with a further round of simulated annealing refinement with the data extended to 1.8 Å. This resulted in R = 30.7% for the 3766 $2\sigma(F)$ data. Further rounds of restrained least-squares refinement incorporating individual restrained isotropic temperature factors and the conservative addition of solvent molecules were then used to complete the refinement. Initially the X-PLOR software was used but in the latter stages the refinement was performed with the program NUCLSQ [25,27] in order to improve the geometry of our model. The final model has a crystallographic residual of 19.0% for 3147 reflections with $F \geq 3\sigma(F)$ in the resolution range 7.0–1.8 Å. This represents 46.3% of the total theoretically available data within the resolution range stated, 37.9% in the range 2.06 Å to 1.92 Å, 19.7% from 1.92 to 1.8 Å. These figures are a consequence of the pseudo C-centering (as those reflections with h + k odd are weak) but are also a reflection on the use of a thin crystal for data collection. When all processed data are included [$F \geq 1\sigma(F)$] in the structure factor calculations, the final residual is 21.2% for 3717 reflections. When the data used have $F \geq 6\sigma(F)$ the final R-factor is 17% for 2244 reflections. We have been able to carry out a resolution test on station PX9.6 at the Synchrotron Radiation Source at Daresbury. A crystal of approximate thickness 0.07 mm showed diffraction beyond 1.5 Å resolution. It should therefore be possible to improve the accuracy of this structure at some later date by both improving the completeness of and the resolution limit of the data that is incorporated into the refinement.

Structure factors and coordinates have been deposited with the Brookhaven Protein Data Bank.

Acknowledgements: We thank the Wellcome trust, the United Kingdom Science and Engineering Research Council for financial support, R. Beddoes for helpful discussions, M Papiz and P Rizkallah at Daresbury for support on PX9.6.

References

1. Saenger, W. (1984). *The Principles of Nucleic Acid Structure.* Springer-Verlag, New York.
2. Kennard, O. & Hunter, W.N. (1991). Single-crystal X-ray diffraction studies of oligonucleotides and oligonucleotide–drug complexes. *Angew. Chem.* **30**, 1254–1277.
3. Dock-Bregon, *et al.*, & Moras, D. (1989). Crystallographic structure of an RNA helix: $[U(UA_6)A]_2$. *J. Mol. Biol.* **209**, 459–474.
4. Holbrook, S.R., Cheong, C., Tinoco, I. & Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating non-Watson-Crick base pairs. *Nature* **353**, 579–581.
5. Hendrickson, W.A. & Wüthrich, K. (eds) (1992). *Macromolecular Structures 1992.* pp. 286–287, Current Biology Ltd, London.
6. Wang. A. H-J., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boeckle, S.A. & Rich, A. (1982). Molecular structure of r(GCG)d(TATACGC): A DNA–RNA hybrid helix joined to double helical DNA. *Nature* **299**, 601–604.
7. Egli, M., Usman, N., Zhang, S. & Rich, A. (1992). Crystal structure of an Okazaki fragment at 2 Å resolution. *Proc. Natl. Acad. Sci. USA* **89**, 534–538.
8. Egli, M., Usman, N. & Rich A. (1993). Conformational influence of the ribose 2′-hydroxyl group: crystal structures of DNA–RNA chimeric duplexes. *Biochemistry* **32**, 3221–3237.
9. Lorenz, S., *et al.*, & Erdmann, V.A. (1993). Crystallisation and preliminary diffraction studies of the chemically synthesized domain A of *Thermus flavus* 5S rRNA: an RNA dodecamer double helix. *Acta. Crystallogr. D* **49**, 418–420.
10. Gesteland, R.F. & Atkins, J.F. (eds) (1993). *The RNA World.* Cold Spring Harbor Laboratory Press, New York.
11. Brown, T., Leonard, G.A., Booth, E.D. & Kneale, G.G. (1990). Influence of pH on the conformation and stability of mismatch base pairs in DNA. *J. Mol. Biol.* **212**, 437–440.
12. Privé, G.G., Heinemann, U., Chandrasegaran, S., Kan, L-S., Kopka, M.L. & Dickerson, R.E. (1987). Helix geometry, hydration and G-A mismatch in a B-DNA decamer. *Science* **38**, 498–504.
13. Leonard, G.A., Booth, E.D. & Brown, T. (1990). Structural and thermodynamic studies on the adenine guanine mismatch in B-DNA. *Nucleic Acids Res* **18**, 5617–5623.
14. Heus, H.A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191–194.
15. Turner, D.H. & Bevilacqua, P.C. (1993). Thermodynamic considerations for evolution by RNA. In *The RNA World.* (Gesteland, R.A. & Atkins, J.F. eds) pp. 447–464, Cold Spring Harbor Laboratory Press, New York.
16 Gao, X. & Patel, D.J. (1988). G(syn)·A(anti) mismatch formation in DNA dodecamers at acidic pH: pH-Dependent conformational transition of GA mispairs detected by proton NMR. *J. Am. Chem. Soc* **110**, 5178–5182.
17. SantaLucia, J. Jr. (1991). The role of hydrogen bonding in the thermodynamics and structure of mismatches in RNA oligonucleotides. [PhD thesis]. University of Rochester, New York.

18. Katahira, M., Sato, H., Uesugi, S. & Fujii, S. (1993). NMR studies of G:A mismatches in oligodeoxyribonucleotide duplexes modelled after ribozymes. *Nucleic Acids Res.* **21**, 5418–5424.
19. Orita, M., Nishikawa, F., Shimayama, T., Taira, K., Endo, Y & Nishikawa, S. (1993). High-resolution NMR study of a synthetic oligoribonucleotide with a tetranucleotide GAGA loop that is a substrate for the cytotoxic protein, ricin. *Nucleic Acids Res.* **21**, 5670–5678.
20. SantaLucia, J. & Turner, D.H. (1993). Structure of r(GGCGA GCC)$_2$ in solution from NMR and restrained molecular dynamics. *Biochemistry* **32**, 12612–12623.
21. Rabinovitch, D. & Shakked, Z. (1984). A new approach to structure determination of large molecules by multidimensional search methods. *Acta Crystallogr. A* **40**, 195–200.
22. Sheldrick, G.M. (1976). *SHELX76 System of Computer Programs*, University of Cambridge, UK.
23. Brünger, A.T., (1990). *X-PLOR (Version 2.2) Manual*, Howard Hughes Medical Institute, Yale University, CT.
24. Brünger, A.T., Krukowski, A. & Erickson, J.W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A* **46**, 585–593.
25. Hendrickson, W.A. & Konnert, J.H. (1980). In *Computing and Crystallography*. (Diamond, R., Ramaseshan, S. & Venkatesan, K. eds) pp. 13.01–13.23. The Indian Academy of Sciences, Bangalore.
26. Arnott, S., Hukins, D.W.L., Dover, S.D., Fuller, W. & Hodgson, A.R. (1973). Structures of synthetic polynucleotides in the A-RNA and A′-RNA conformations: X-ray diffraction studies of the molecular conformations of polyadenylic acid polyuridylic acid and polyinosinicacid polycytilylic acid. *J. Mol. Biol.* **81**, 107–122.
27. Westhof, E., Dumas, P. & Moras, D. (1985). Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* **184**, 119–145.
28. Holbrook, S.R., Sussman, J.L., Warrant, R.W. & Kim, S.-H. (1978). Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.* **123**, 631–660.
29. Westhof, E. & Sundaralingam, M. (1986). Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA. Temperature factors and dynamics, coordinated waters and base pair propeller twists. *Biochemistry* **25**, 4868–4878.
30. Pyle, A.M. & Cech, T.R. (1991). Ribozyme recognition of RNA by tertiary interactions with specific ribose 2′-OH groups. *Nature* **350**, 628–631.
31. Brown, T., Hunter, W.N., Kneale, G. & Kennard, O. (1986). Molecular structure of the GA base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Natl. Acad. Sci. USA* **83**, 2402–2406.
32. Kan, L.-S., Chandrasegaran, S., Pulford, S.M. & Miller, P.S. (1983). Detection of guanine adenine base pair in a decadeoxyribonucleotide by proton magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. USA* **80**, 4263–4265.
33. Crick, F.H.C. (1966). Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555.
34. Umeyama, H. & Morokuma, K.J. (1977). The origin of hydrogen bonding. An energy decomposition study. *J. Am. Chem. Soc.* **9**, 1316–1332.
35. Steiner, T. & Saenger, W. (1993). Role of C-HO hydrogen bonds in the coordination of water molecules. Analysis of neutron diffraction data. *J. Am. Chem. Soc.* **115**, 4540–4547.
36. Howard, S.T., Hursthouse, M.B., Lehmann, C.W., Mallinson, P.R. & Frampton, C.S. (1992). Experimental and theoretical study of the charge density in 2-methyl-4-nitroaniline. *J. Chem. Phys.* **97**, 5616–5630.
37. Leonard, G.A., Guy, A., Brown, T., Téoule, R. & Hunter, W.N. (1992). Conformation of the guanine 8-oxoadenine base pairs in the crystal structure of d(CGCGAATT(O8A)GCG). *Biochemistry* **31**, 8415–8420.
38. Jeffrey, G.A. (1989). Hydrogen bonding in crystal structures of nucleic acid components: purines, pyrimidines, nucleosides and nucleotides. In *Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology. Group VII, Volume 1, Nucleic Acids. Subvolume b, Crystallographic and Structural Data II.* (Saenger, W. ed.) pp. 277–342, Springer-Verlag, Berlin.
39. Taylor, R., Kennard, O. & Verischel, W. (1983). Geometry of the N-H OC hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **105**, 5761–5766.
40. Fritsch, V. & Westhof, E. (1991). Three-center hydrogen bonds in DNA: molecular dynamics of poly(dA) poly(dT). *J. Am. Chem. Soc.* **113**, 8271–8277.
41. Carbonnaux, C., van der Marel, G.A., van Boom, J.H., Guschlbauer, W. & Fazakerly, G.V. (1991). Solution structure of an oncogenic DNA duplex containing a G.A mismatch. *Biochemistry* **30**, 5449–5458.
42. McCall, M.J., Brown, T. & Kennard, O. (1985). The Crystal Structure of d(G-G-G-G-C-C-C-C) a model for poly(dG) poly(dC). *J. Mol. Biol.* **183**, 385–396.
43. Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987). The structure of an oligo(dA) (dT) tract and its biological implications. *Nature* **330**, 221–226.
44. Yanagi, K., Prive, G.G. & Dickerson, R.E. (1991). Analysis of the local helix geometry in 3 B-DNA decamers and 8 dodecamers. *J. Mol. Biol.* **217**, 202–214.
45. Leonard, G.A. & Hunter, W.N. (1993). Crystal and molecular structure of d(CGTAGATCTACG). *J. Mol. Biol.* **234**, 198–208.
46. Kennard, O. & Hunter, W.N. (1988). Crystal structures of oligonucleotides. In *Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology. Group VII, Volume 1, Nucleic Acids. Subvolume a, Crystallographic and Sstructural Data II.* (Saenger, W. ed.), pp. 255–360. Springer-Verlag, Berlin.
47. Conner, B.N., Yoon, C., Dickerson, J.L. & Dickerson, R.E. (1984). Helix geometry and hydration in an A-DNA tetramer ICCGG. *J. Mol. Biol.* **174**, 663–695.
48. Saenger, W., Hunter, W.N. & Kennard, O. (1986). DNA conformation is determined by economics in the hydration of phosphate groups. *Nature* **324**, 385–388.
49. Kim, K. & Jhon, M.S. (1979). Theoretical study of hydration of RNA. *Biochim. Biophys. Acta* **565**, 131–147.
50. Westhof, E., Romby, P., Romaniuk, P.J., Ebel, J.-P., Ehresmann, C. & Ehresmann, B. (1989). Computer modelling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5S RNAs. *J. Mol. Biol.* **207**, 417–431.
51. Higashi, T. (1990). *PROCESS: a program for indexing and processing R-AXIS II image plate data.* Rigaku Corporation, Japan.
52. Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
53. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.

Received: **31 Jan 1994**; revisions requested: **23 Feb 1994**; revisions received: **13 Apr 1994.** Accepted: **14 Apr 1994.**