UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Chemistry

Propargylamino Modified Bases for DNA Duplex Stabilisation

by

James Alexander Booth

Thesis for the degree of Doctor of Philosophy

September 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF CHEMISTRY

Doctor of Philosophy

PROPARGYLAMINO MODIFIED BASES FOR DNA DUPLEX STABILISATION

by James Alexander Booth

Inherited diversity in the human genome leads to phenotypic variations that underlie an individual's response to their environment. By typing these differences, referred to as single nucleotide polymorphisms, on a whole genome basis, it is hoped that the complex genetic basis of susceptibility to common diseases can be elucidated.

Herein is described the basis of a potential assay for the determination of single nucleotide polymorphisms on a whole genome basis. Initial studies were focused on the ability of a resin bead to support oligonucleotide and peptide synthesis, oligonucleotide duplex formation and identification of a peptide by mass spectrometry from a minimum number of beads. Poor hybridisation kinetics led to a synthetic investigation into the ability to modify the properties of the oligonucleotides to improve the kinetics and thermodynamics of duplex association. Hereinafter it is shown that the propargylamino or 3-aminoprop-1-yne group stabilises DNA duplexes when placed at the C-5 position in pyrimidines and C-7 position in 7-deazapurines.

At physiological pH the cationic propargylamine group when introduced into an oligonucleotide helps to reduce the repulsion between the anions on the phosphodiester backbone and therefore reduces the salt dependence of duplex formation. The alkyne moiety increases the geometrical overlap and therefore the dispersion forces in single and double stranded DNA. The novel heterocyclic base 7-aminoprop-1-ynyl-7-deaza-2,6-diaminopurine when base paired with dT has a similar thermodynamic stability to a G.C base pair. This novel adenine analogue may therefore be useful in applications where T_m harmonisation would be an advantage.

LIST OF CONTENTS

AUTHOR'S DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any University or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

ACKNOWLEDGEMENTS

ABBREVIATIONS

λ_{max}	Wavelength of maximum absorption
A	7-(3-Aminopropyn-1-yl)-7-deaza-2'-deoxyadenosine
A	Adenine
Ac	Acetyl
ADP	Adenosine diphosphate
alpha	α-Cyano-4-hydroxycinnamic acid
A_M	The absorption coefficient of modified base M
AMP	Adenosine monophosphate
APS	Adenosine 5'-phosphosulfate
ATP	Adenosine triphosphate
Bps	Base pairs
C	5-(3-Aminoprop-1-ynyl)-2'-deoxycytidine
C	Cytosine
COSY	Correlated Spectroscopy
CPG	Controlled pore glass
CY5	Cyanine 5
D	7-(3-Aminopropyn-1-yl)-7-deaza-2'-deoxy-2-aminoadenosine
DCM	Dichloromethane
ddNTPs	Dideoxyribonucleotide triphosphates
DEPT	Distortionless enhancement by polarizaton transfer
DHB	2,5-Dihydroxybenzoic acid
DIC	Diisopropylcarbodiimide
DIPEA	Diisopropylethylamine
dmC	5-(3-*N*,*N*-dimethylaminoprop-1-ynyl)-2'-deoxycytidine
DMF	Dimethylformamide
DMS	Dimethylsulfate
DMS	Dimethylsulfate
DMSO	Dimethylsulphoxide
DMTr	4,4'-Dimethoxytrityl
dmU	5-(3-N,N-Dimethylaminoprop-1-ynyl)-2'-deoxyuridine

DNA	2'-Deoxyribonucleic acid
DNAzyme	DNA enzyme
dNDP	2'-Deoxyribonucleotide diphosphate
dNMP	2'-Deoxyribonucleotide monophosphate
dNTPs	2'-Deoxyribonucleotide triphosphates
DSC	Differential scanning calorimetry
dsDNA	Double-stranded 2'-deoxyribonucleic acid
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ES	Electrospray
Exo I	Exonuclease I
FACS	Fluorescence activated cell sorter
FAM	6-Carboxyfluorescein
FddNTPs	Fluorescent dideoxyribonucleotide triphosphates
FEN	Flap endonucleases
Fmoc	9-Fluorenylmethoxycarbonyl
FRET	Fluorescence resonance energy transfer
FT	Fourier transform
FT-ICR	Fourier-transform ion cyclotron resonance
G	7-(3-Aminopropyn-1-yl)-7-deaza-2'-deoxyguanosine
G	Gaunine
h	Hours
hν	Electromagnetic radiation
HMQC	Heteronuclear multiple-quantum correlation
HOBt	1-Hydroxybenzotriazole
HPLC	High pressure liquid chromatography
HRCA	Hyperbranched rolling circle amplification
HRMS	High resolution mass spectrometry
ht-SNP	Halotype tag-single nucleotide polymorphism
I	7-(3-Aminopropyn-1-yl)-7-deaza-2'deoxyinosine
IR	Infra-red
ITC	Isothermal titration calorimetry
J	Coupling constant

KDa	Kilodaltons
MALDI	Matrix assisted laser desorption ionisation
*m*CPBA	m-Chloroperoxybenzoic acid
MDA	Multiple displacement amplification
min	Minutes
Mod.	Modifications
MS	Mass spectroscopy
NIS	*N*-Iodosuccinimide
NMR	Nuclear magnetic resonance
No.	Number
N_x	The number of x bases in an oligonucleotide (where x = A, T, C, G)
O_{MAC}	Oligonucleotide molar absorption coefficient
PCR	Polymerase chain reaction
pH	Potential of hydrogen
PP_i	Pyrophosphate ($P_2O_7^{4-}$)
ppm	Parts per million
Py	Pyridine
RB	Round bottomed
RCA	Rolling circle amplification
R_f	Retention factor
RNA	Ribonucleic acid
rNTPs	Ribonucleotide triphosphates
ROX	6-Carboxy-X-rhodamine
RT	Reverse transcriptase or Room temperature
SAP	Shrimp alkaline phosphatase
SBCE	Single base chain extension
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded 2'-deoxyribonucleic acid
T	Thymine
TAMRA	*N,N,N',N'*-Tetramethyl-6-carboxyrhodamine
Taq	*Thermus aquaticus*
TDA-1	Tris[2-(2-methoxyethoxy)ethyl]amine
TEA	Triethylamine
TEBA	Triethyl Benzyl Ammonium Chloride

THF	Tetrahydrofuran
TLC	Thin layer chromatography
T_m	Melting temperature
TOF	Time of flight
U	Uracil
U	5-(3-Aminoprop-1-ynyl)-2'-deoxyuridine
UV	Ultra-violet
UV/vis	Ultra-violet / visable
WGA	Whole genome amplification
ΔG_T	The change in Gibbs free energy at a temperature T
ΔH	The change in enthalpy
ΔS	The change in entropy

GENERAL AIMS AND OBJECTIVES

The initial objective of this research was to produce an assay, capable of typing hundreds of thousands of single nucleotide polymorphisms (SNPs). The assay was envisaged to be a bead-based test, in which, each bead types one SNP of interest. The bead's surface was to support oligonucleotides that would hybridise to regions of the genome that code for SNP sites.

It was anticipated that a single base chain extension reaction would then be carried out using fluorescently labelled dideoxynucleotide triphosphates. On a macroscopic level, the beads that had positively identified a SNP would thus appear fluorescent. The single base extension reaction can only occur if the probe is exactly complementary to the genomic sequence. The beads would appear fluorescent, as the whole surface would have been supporting the oligonucleotides onto which the fluorescent nucleotide would have been added. It was envisaged that a fluorescence activated cell sorter would thereafter isolate the fluorescent beads. The beads would need to be isolated from the vast numbers of negative beads so that they could be individually identified. The beads were to be encoded, in such a way, that a mass tag could be removed and identified by mass spectrometry and therefore enabling the identification of the SNP.

However, during preliminary work, it was found that the kinetics and thermodynamics of oligonucleotide hybridisation needed optimisation. The optimisation was found to be necessary when it proved difficult to observe fluorescent oligonucleotides hybridising to complementary oligonucleotides attached to the surface of a bead. The objective of the research thus became, to find ways to improve the efficiency of the hybridisation through the modification of the heterocyclic bases in DNA. This led to an investigation into several propargylamino / 3-aminoprop-1-yne modified purine and pyrimidine nucleosides.

The modification introduces both an alkyne moiety and a positive charge, at physiological pH, that is spatially located in the region of the anionic phosphate backbone. The alkyne group increases the dispersion forces between the bases and the

amino group can potentially form a salt bridge so reducing the number of sodium ions that need to associate upon duplex formation.

Subsequently, a further objective arose. This was to investigate in detail the thermodynamic properties of the synthesised heterocyclic bases by UV-melting. The basis of any stabilisation or destabilisation was determined to improve future design of stabilising modified DNA bases.

1.   GENERAL INTRODUCTION

A general overview of nucleic acids is given with particular reference to their primary and secondary structures. This is followed by a summary of solid-phase oligonucleotide synthesis. A background to the occurrence of mutations and single nucleotide polymorphisms (SNPs) is included along with an overview of different SNP detection technologies. This is to illustrate the profound effect SNP detection and analysis is having on areas such as pharmacogenomics and pathogen detection and identification. Subsequently areas specific to this work are covered including: the modification of nucleic acid bases, the calculation of the melting point (T_m) and thermodynamics of DNA duplexes and the factors that influence DNA duplex formation.

1.1   DEOXYRIBONUCLEIC ACID BACKGROUND

1.1.1   Nucleic Acid Structure

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are polymers comprising of units called nucleotides. Phoebus Levene who found that nucleic acids contained either one of two pentose monosaccharides. Initially he discovered D-ribose, the sugar constituent of RNA and later he found D-2'-deoxyribose, the sugar component of DNA. In nucleic acids, both D-2'-deoxyribose and D-ribose are in the furanose form. Levene described how all the components of DNA are combined to form nucleotides and how these are combined into chains. In addition to the sugar moiety, nucleotides also contain a nitrogen-containing heterocyclic base and phosphate moiety. Compounds containing just the monosaccharide, glycon, and base moieties, aglycon, are referred to as nucleosides or N-glycosides. The monosaccharides discovered by Levene are optically active compounds and can exist in both D and L forms although most monosaccharides found in nature are in the D form. There are predominantly five heterocyclic bases found in DNA and RNA (Fig. 1), but only four are used by either DNA or RNA. In RNA uracil replaces the DNA base thymine (5-methyl uracil).
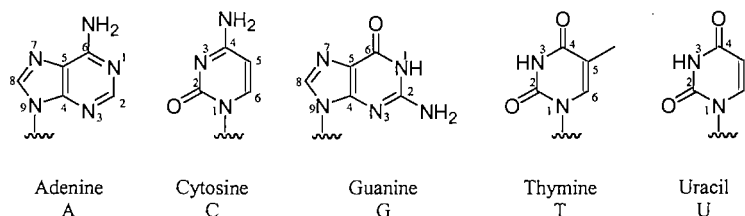
Figure 1. The five most common heterocyclic bases found in DNA and RNA. The bases are shown with their names, abbreviations and numbering system.

The heterocyclic bases are classified as either purines, a bicyclic structure, or pyrimidines, a monocyclic base. Adenine and guanine are functionalised purines while cytosine, thymine and uracil are functionalised pyrimidines (Fig. 1). The nucleosides are linked to form a nucleic acid chain via phosphodiester linkages at the 3' and 5' positions on the sugar. This gives the oligonucleotide its directionality. At physiological pH the phosphates are deprotonated. Hence, nucleic acids are highly negatively charged and exist as salts. The structure of the sugar-phosphate backbone is shown in Fig. 2.
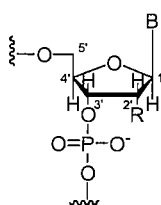
Figure 2. The sugar-phosphate backbone where, B = heterocyclic base, R = OH (RNA), H (DNA). The numbering system for the β-D-ribofuranose ring is shown.

In the nucleoside structure, the ribofuranose ring twists out of plane to minimise steric and electronic repulsion between the atoms. This 'puckering' is described as either C2'-*endo* (Southern, S-type) or C3'-*endo* (Northern, N-type). The equilibrium between these two extremes of conformation is influenced by the preference of electronegative groups at C2' and C3' for an axial conformation. DNA adopts an S-type and RNA an N-type conformation (Fig. 3). The *endo* conformation occurs when the out of plane atom (C2' or C3') is on the same side of the plane formed by the other four atoms as the 5' substituent. When on opposite sides the conformation is described as *exo*.

C3'-*endo* (N-type) C2'-*endo* (S-type)
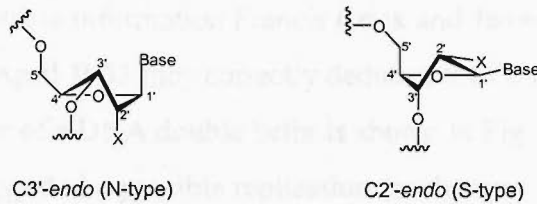
Figure 3. The β-D-ribofuranose ring conformations in DNA (X = H, S-type) and RNA (X = OH, N-type). In both examples the out of plane atom is on the same side of the plane, created by the other four atoms, as the 5' substituent.

In DNA and RNA the substituent at the anomeric carbon (C1') is in the β-form with the aglycon above the plane of the D-ribofuranose ring. In the α-form, which is much less common in nature, the 1' substituent is below the plane of the ring. The β-anomer is shown in both Fig. 2 and 3.

The secondary structure of DNA was identified following discoveries in various laboratories throughout the world. In 1951, Charagaff used UV-spectroscopy to deduce that the proportion of purines always equals the proportion of pyrimidines in cells although the sequence of these bases varies widely.[1,2] Maurice Wilkins and Rosalind Franklin used X-ray diffraction of DNA fibres to show that DNA forms an "A-form" at low humidity and a "B-form" at high humidity. Both forms were shown to be highly

Sugar-Phosphate Backbone

Minor Groove Major Groove
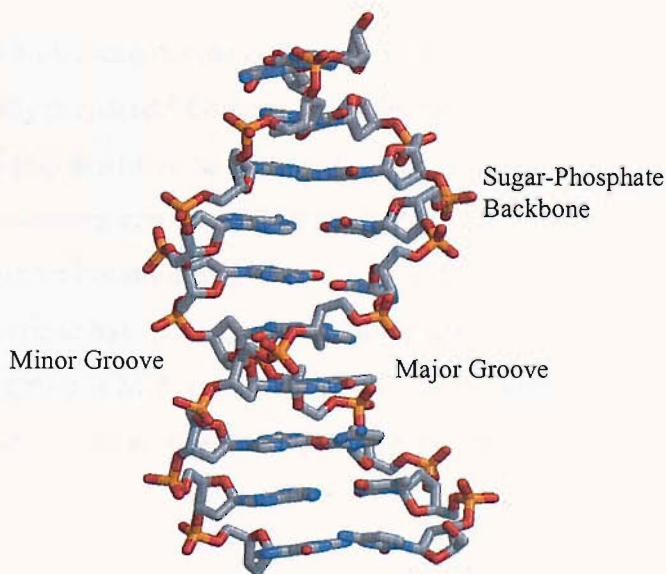
Figure 4. The secondary structure of DNA. A crystal structure of 5'-d(CTCTCGAGAG) obtained by x-ray crystal diffraction with a resolution of 1.70 Å. The sugar-phosphate backbone is on the outside exposed to the aqueous environment and the hydrophobic bases stacked internally in the centre of the helix.[3]

crystalline and helical in structure. This led them to believe that the phosphate groups must be on the outside exposed to water and the bases were on the inside of the helix.[4,5]

16

Using this and other available information Francis Crick and James Watson constructed a series of models, and in April 1953 they correctly deduced the double helical structure of DNA.[6] A crystal structure of a DNA double helix is shown in Fig. 4.

This structure suggested a possible replication mechanism for the genetic material.[7] The replication mechanism became evident because in building their structure they found that the bases, when in their keto/lactam, form were complementary. Adenine binds to thymine forming two hydrogen bonds and guanine to cytosine forming three hydrogen bonds (Fig. 5).

A : T   G : C
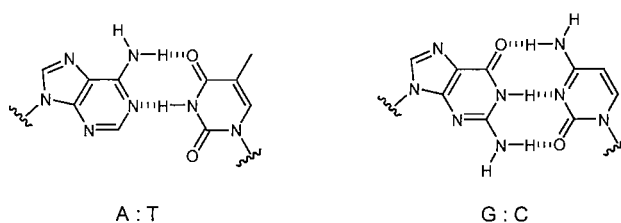
Figure 5. The complementary base pairs found in DNA. Each base pair consists of one purine and one pyrimidine. The adenine-thymine base pair has two hydrogen bonds and the guanine-cytosine base pair has three hydrogen bonds.

1.1.2 Oligonucleotide Synthesis

An oligonucleotide is a macromolecule composed of short sequences of nucleotides that are usually synthetically prepared.[§] Oligonucleotides can be synthesised by solid-phase methods. They are highly sensitive to a wide range of chemical reactions and only the mildest of reaction conditions can be used in the chemical synthesis of oligonucleotides.[8] The monomer units (phosphoramidites) must have the 5'-hydroxyl group protected with 4,4'-dimethoxytriphenylmethyl. This is commonly referred to as the 4,4'-dimethoxytrityl group (DMTr).[9] This group is highly acid labile and can be added regioselectively to the 5'-position in the presence of the 3'OH group due to its steric bulk.

[§] Glossary of terms used in bioinorganic chemistry – (IUPAC, 1997).
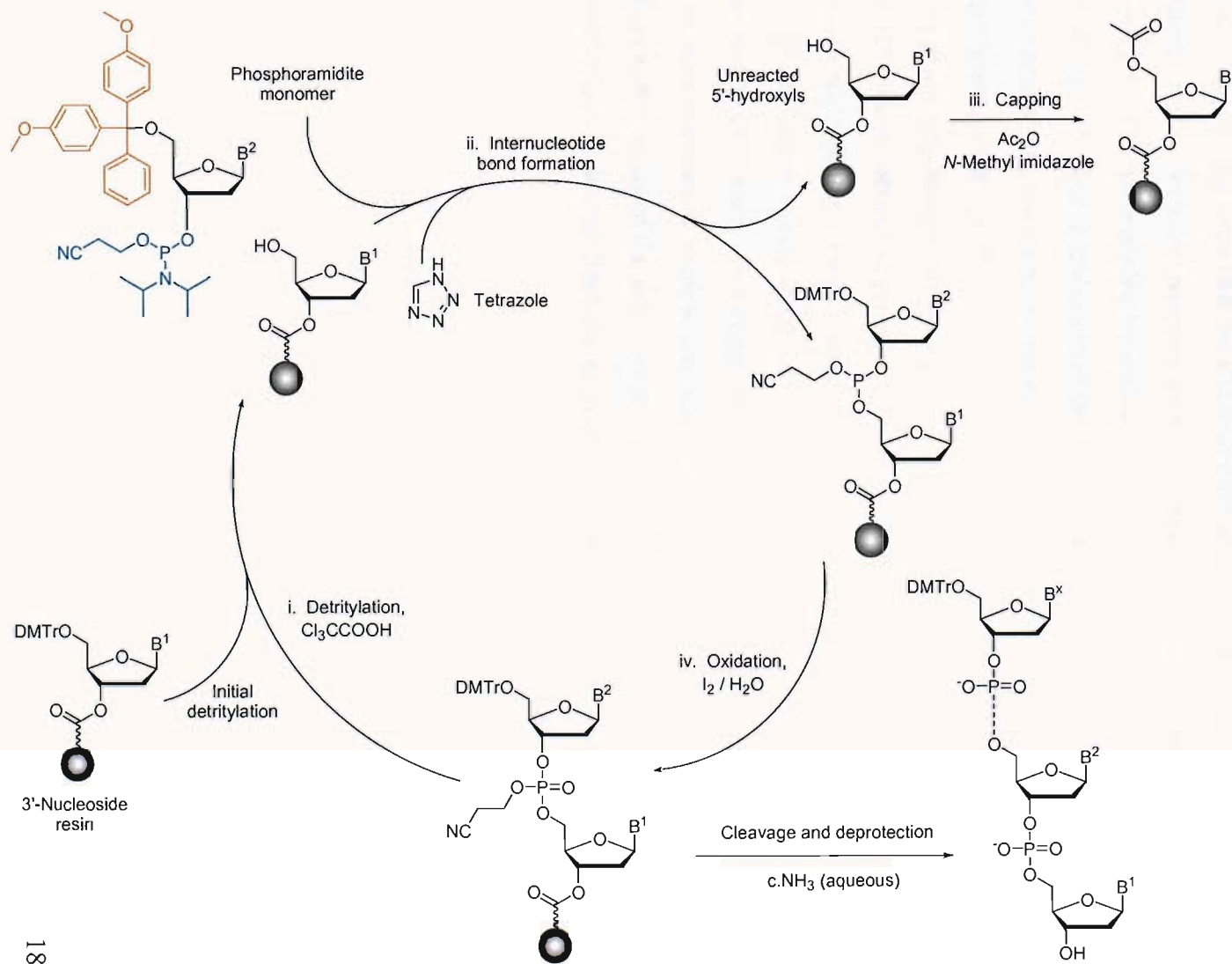
Figure 5. Assembly of an oligonucleotide on an automated DNA synthesizer. The 4,4'-dimethoxytrtiyl (DMTr) protecting group is removed from the 5' by trichloroacetic acid to reveal the free hydroxyl group. The incoming phosphoramidite is then mixed with tetrazole which protonates and therefore activates the *N,N*-diisopropylamine group on the phosphoramidite. This activated group is then displaced by the 5'-hydroxyl to form the new internucleotide bond. Any remaining unreacted hydroxyls are capped with activated acetic anhydride to prevent further reaction at these sites. The phosphorus is then oxidised in aqueous iodine to give the phosphodiester group. This cycle is repeated to add further nucleotides. The cleavage of the oligonucleotide from the resin, deprotection of the heterocyclic base protecting groups and the β-elimination of the cyanoethyl groups from the phosphates are all carried out by aqueous ammonia. The trityl group (4,4'-dimethoxytrtiyl) is highlighted in orange and the phosphoramidite group in blue as part of the incoming phosphoramidite monomer.

The dimethoxytrityl cation, liberated by acidic cleavage, absorbs strongly at 495 nm and can be used to monitor the coupling yield reactions during solid-phase oligonucleotide assembly. However, the yield of the reactions is normally monitored using the trityl ion conductivity. The structure of the 4,4'-dimethoxytrityl group is shown in orange in Fig. 5.

The exocyclic amino groups of the heterocyclic bases adenine, cytosine and guanine need to be protected during oligonucleotide synthesis. This is usually achieved with acyl protecting groups that are stable to mild acid and base, but are cleaved by treatment with concentrated aqueous ammonia. The protected nucleoside is then converted in a phosphoramidite by reaction of the 3'-hydroxyl group with 2-cyanoethoxy-N,N-diisopropylamine chlorophosphine.[10-12] Phosphoramidites are relatively stable at neutral pH, however, protonation at the nitrogen turns it into a highly reactive phosphitylating agent. This reactivity results in extremely high coupling yields of >99% in solid-phase oligonucleotide synthesis which is essential if long sequences are to be made. The phosphoramidite group is highlighted in blue in Fig. 5. The standard oligonucleotide synthesis cycle is depicted in Fig. 5.

The synthesis is carried out on a solid surface, controlled pore glass (CPG). Solid phase synthesis has many advantages, two of which are; large excesses of reagents can be used to force reactions to completion, and excess reagents can easily be removed by filtration and washing of the solid phase. The use of an ester as a linker means that the oligonucleotide is cleaved from the support when exposed to conc. NH_3 (aqueous).

1.2 SINGLE NUCLEOTIDE POLYMORPHISMS

1.2.1 The Formation of Single Nucleotide Polymorphisms

The theory described below is the Watson-Crick and Topal-Fresco model of base mispairing during DNA replication.[7,13] These models gave rise to the following predictions: transition mutations should occur primarily through G-T mispairs, transversions should occur primarily through purine-purine mispairs, and transitions should be more frequent than transversions at a given site. These predictions were then confirmed experimentally.[14] It should be noted that mutations also arise in different ways, for example when the DNA bases react with certain chemicals or high-energy radiation.

The accuracy of DNA replication relies on correct hydrogen-bonding interactions between the incoming base and the complementary base on the template strand. However the bases can exist in different tautomeric forms. For instance dG, under physiological conditions, exists mostly in its keto, lactam, tautomer and can momentarily switch to its enol, lactim, tautomer. This can result in the formation of a dG: dT base pair instead of the correct dG: dC base pairing (Fig. 6). The similarity in shape between the GT mismatch in Fig. 6 and the GC base pair means that polymerases and proofreading enzymes are unable to recognise the mismatch.

                    G : C   G : T
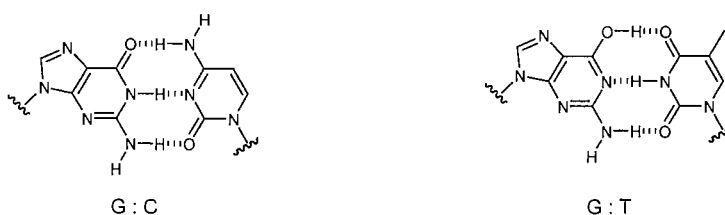
Figure 6. The rare enol tautomeric form of dG (right) can result in the formation of a dG: dT mismatch.

The frequency of occurence of the unusual tautomeric forms of the DNA bases mean that errors would be expected to occur every 10^4 to 10^5 nucleotides replicated. However, measurements of spontaneous mutation rates indicate that the rate is only 1 in 10^9 nucleotides. This discrepancy is due to the 'proof-reading' ability of DNA polymerases. After the insertion of a base, the enzyme pauses and checks to make sure the base pairing is correct. If incorrect base pairing results in a change in conformation in the enzyme, this then inactivates the polymerase activity and activates a 3'→5' exonuclease activity. The exonuclease then hydrolytically cleaves the previous

nucleotide. The probability that the base is in the rare tautomeric form during both insertion and during proof-reading is about once every 10^8 to 10^9 bases inserted which agrees with experimental observation of mutation rates.[15]

Mutations only become fixed in the genome if the individual survives to reproduce. The mutations therefore have to be benign, beneficial, or occur late in life to enter the gene pool permanently. However, mildly damaging mutations can exist as long as reproduction can take place. Nevertheless, the lack of evolutionary advantage means that these will probably disappear.

Single-nucleotide polymorphisms (SNPs) are the most prevalent form of sequence variation among individuals. SNPs account for about 90% of all human genetic variation. The remaining 10% is attributable to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements. SNPs occur when a single nucleotide (A, C, G or T) in the genome is changed to a different base. For one of these variations to be considered a SNP it must occur in at least 1% of the population. SNPs occur on average once every kilobase in the 3-billion-base human genome. The most common SNP is the mutation of a C to a T due to the stability of the G.T mismatch. The low rate of recurrent mutation makes the SNPs ideal as stable indicators of human variation.[16]

1.2.2   Linkage Studies

SNPs are important as they describe an individual's phenotypic variation, anthropometric characteristics, disease resistance and response to the environment, amongst many other characteristics. The traditional method for linking SNPs to disease susceptibility is through association studies. An individual can be genotyped to discover which allele of a SNP s/he has. This process is carried out with two large populations, one of which is diseased and the other healthy. If one of the alleles detected is more common in the diseased population, it is likely that this confers susceptibility to the disease or that the other allele found in the healthy population confers resistance.

This technique is problematic because in order to look at the three million SNPs in each individual requires as many typing assays. As large populations are needed to produce statistically significant results, millions of assays are required for each disease state. This problem can, in part, be overcome by a phenomenon called linkage disequilibrium. Linkage disequilibrium reflects the observation that many SNP tend to occur together. Therefore, if one SNP is found, many other associated SNPs are also

likely to be found. This is because these SNPs are located near to each other in the genome and so are rarely separated during recombination.

These collections of SNPs are known as a haplotypes. Haplotypes appear to be around 60 000 bp long and thus they usually carry around 60 SNPs. Recombination seems to be targeted to regions between haplotypes, so conserving them and also protecting genes from damage due to recombination. There are around 200 000 or so haplotype blocks. These appear to be the same in populations all over the world. This restricted set of haplotypes may be due to a lack of mixing such as the geographical isolation of certain populations. By linking a particular SNP to a haplotype block (thus unambiguously defining that haplotype block), the number of SNPs needing to be assayed can be reduced to around 300 000 – 500 000. Once the haplotype tag SNPs (ht-SNPs) have been identified they can be used to elucidate the genes underlying complex, multifactorial disorders. The ht-SNPs need to be detected cheaply and quickly.[17]

1.3   SINGLE NUCLEOTIDE POLYMORPHISM DETECTION

1.3.1   The Polymerase Chain Reaction (PCR)

In order to detect SNPs from only a few copies of DNA, it is necessary to be able to copy the regions of interest or copy the entire genome millions of times. This amplification is necessary in order that a detectable signal can be produced to indicate the presence of a SNP. The polymerase chain reaction is an *in vitro* method of amplification that uses the evolved biological machinery of organisms to carry out the replication.

In order to carry out the reaction, purified DNA, the four deoxynucleotide triphosphates (dNTPs), a DNA polymerase and primers (two short oligonucleotides) are mixed together in a suitable buffer. This mixture is then heated to 95°C, to separate the DNA strands, and then cooled to 55°C to allow the primers to bind to their complementary sites on the DNA template molecules. The primers are designed to bind at a specific site on the template DNA so that when the polymerase progresses along the molecule it copies the region of interest that will be downstream of the primer in the 3' direction. The primers are necessary to define the region of amplification, and as the polymerase will only bind to double stranded DNA they need to be placed 5' to the region of interest, as the polymerase only moves in a 3'→5' direction. Once the primers have annealed the polymerase will then bind.

The DNA polymerase most frequently employed is a thermostable enzyme isolated from T*hermus aquaticus* known as *Taq*. The enzyme is most efficient at 72°C so the polymerisation is carried out at this temperature and add dNTP's to form the new strand (Fig. 7).

The PCR is cycled through 95-55-72°C and repeated so that by cycle 3 the fragment of DNA to be amplified has been fully isolated. By the end of around 30 cycles the fragment will make up >99.9% of the DNA molecules contained in the sample. During each cycle one DNA molecule becomes two leading to an exponential increase in the number of product molecules.
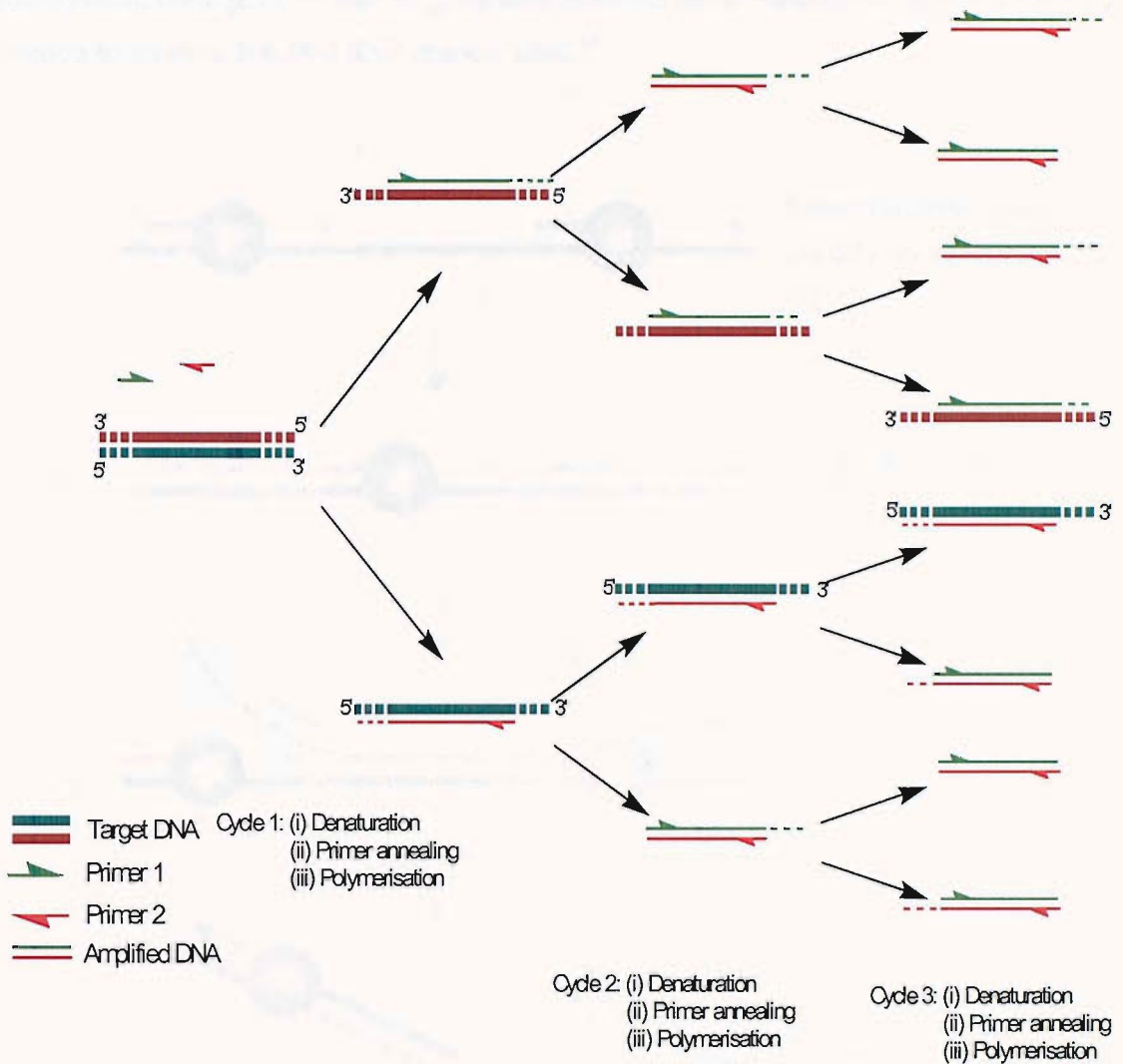
Figure 7. The polymerase chain reaction. In each of the three cycles shown the mixture is cycled through 95-55-72°C to denature the strands, allow the primers to bind and allow the thermostable polymerase to processes along the strand. The region of interest is only totally isolated after three rounds to cycling.

1.3.2 Whole Genome Amplification (WGA)

If hundreds of thousands of SNP sites are to be analysed to detect whole genome disequilibrium patterns, large (~1 µg) amounts of genomic DNA are required. To produce this amount of DNA from small quantities (~1 ng) a whole genome amplification system (WGA) is needed. One such system - "multiple displacement amplification" (MDA) - an isothermal method that uses random hexamers containing phosphorothioate-modified nucleotides as primers and Φ29 DNA polymerase (Fig. 8). This enzyme is used because of its high processivity, high fidelity and its strand displacement ability. Using

24

this system, enough DNA can be generated from ~3 ng of material to carry out the PCRs needed to analyse 300,000 SNP marker sites.[18]
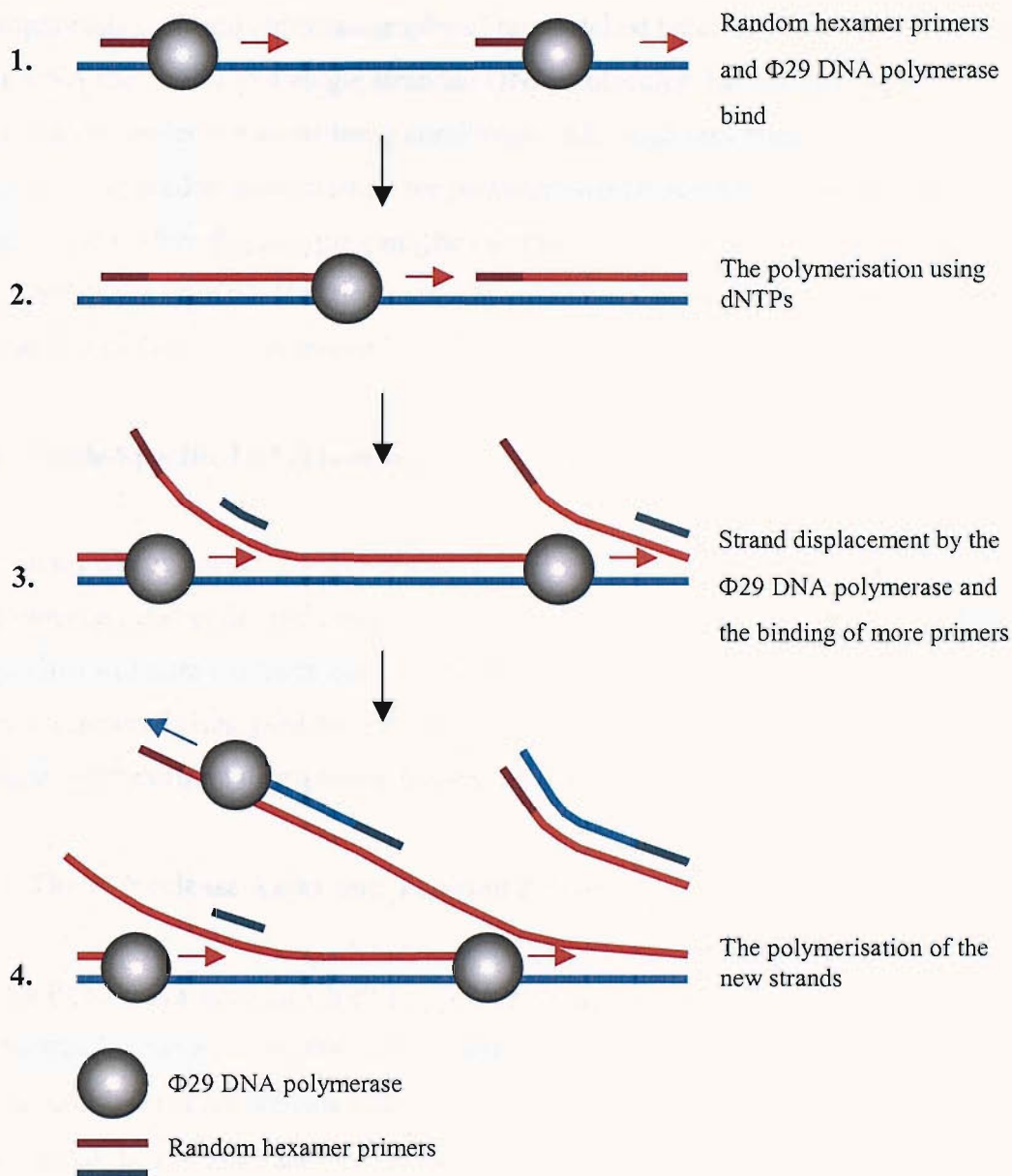
Figure 8. A whole genome amplification system. (1) Linear DNA is amplified using random hexamer primers that bind to random positions throughout the target DNA following heat denaturation of the target DNA. (2) Once bound the other components of the reaction are added and the Φ29 DNA polymerase will then process along incorporating the relevant dNTPs. (3-4) When the enzyme encounters another strand it displaces this strand so forming a new single strand that can be primed and copied.

1.3.3 Methods of SNP Detection[19,20]

SNPs can be detected in either a sequence-specific or non-specific way. Sequence non-specific detection is based on the capture, cleavage, or mobility change during electrophoresis or liquid chromatography of mismatched heteroduplexes formed between allelic DNA molecules or a single stranded DNA molecules that assume slightly different conformations under non-denaturing conditions. Although sequence non-specific detection is the predominant method for polymorphism discovery its non-specific nature means that the inferred genotyping maybe inaccurate.

Sequence-specific detection methods are more suitable for genotyping. These can be divided into four general groups.

1.3.4 Allele-Specific Hybridisation

In this assay a probe is designed to bind at a SNP site. Only when the probe is exactly complementary and under the assay conditions will the probe anneal. Any single base mismatches will alter the thermodynamics of binding so that the probe will not bind. The probes are generally designed so that the SNP site is placed centrally so any mismatches will have the maximum disruptive effect on the duplex.

1.3.5 The 5' Nuclease Assay and TaqMan® Probes[21,22]

This is a PCR-based assay in which the DNA polymerase, *Taq*, as well as acting as a polymerase also has an endogenous 5' nuclease activity. It displaces and cleaves oligonucleotides that are annealed downstream at the 5' end of the template. TaqMan® probes exploit this 5' nuclease activity by annealing to the template downstream of the primer so it is cleaved when the polymerase extends.

TaqMan® probes consist of a fluorophore at the 5' end and a fluorescence resonance energy transfer (FRET) quencher at the 3' end and they only bind when they are complementary to a SNP site. The intact probe fluoresces at a higher wavelength than the cleaved probe as the quencher acts on the fluorophore. However, if the probe binds, the *Taq* polymerase will cleave the probe so separating the fluorophore and the quencher. This then results in a lower wavelength fluorescent signal being observed (Fig. 9).
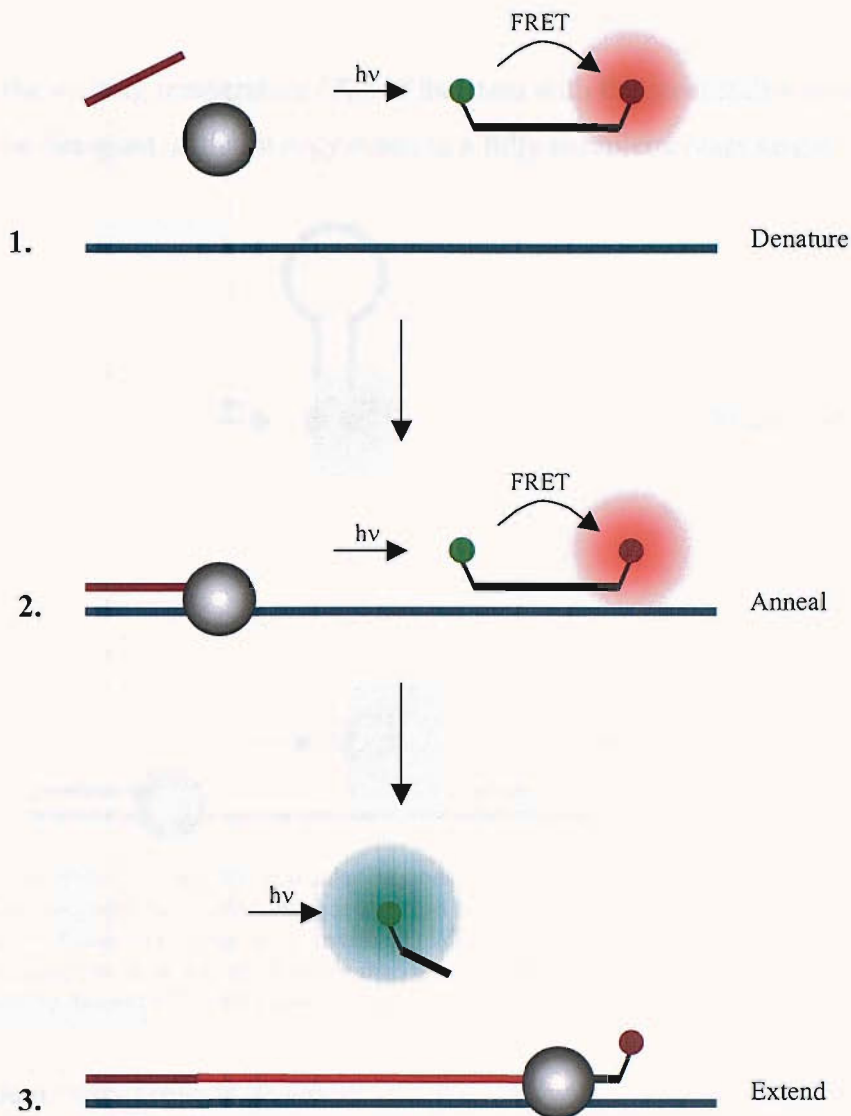
Figure 9. TaqMan® assay. (1) When denatured the TaqMan® probes fluorophore is quenched by FRET to the quencher (2) Following thermal denaturation, if the TaqMan® probe is complementary, it will bind to the target during the annealing phase of the PCR. (3) During the polymerisation phase the *Taq* polymerase will process along the strand displacing and cleaving the TaqMan® probe so releasing free fluorophore into solution that can then be detected. The fluorophore is represented by the green circle, the quencher by the red circle and the incoming visible electromagnetic radiation by hv.

1.3.6 Molecular Beacons.[23]

Molecular beacons are oligonucleotides that have a stem-loop structure in the absence of a complementary target. They have a fluorophore at one end and a collisional quencher at the other so that when in the stem-loop structure, no fluorescence is observed. When a complementary target is present the longer loop sequence will anneal to it, breaking the stem structure and therefore separating the fluorophore and quencher (Fig. 10). This results in a fluorescent signal in the presence of a fully complementary target. By

27

balancing the melting temperature (T_m) of the stem with the most stable mismatch the probe can be designed so that it only binds to a fully complementary target.
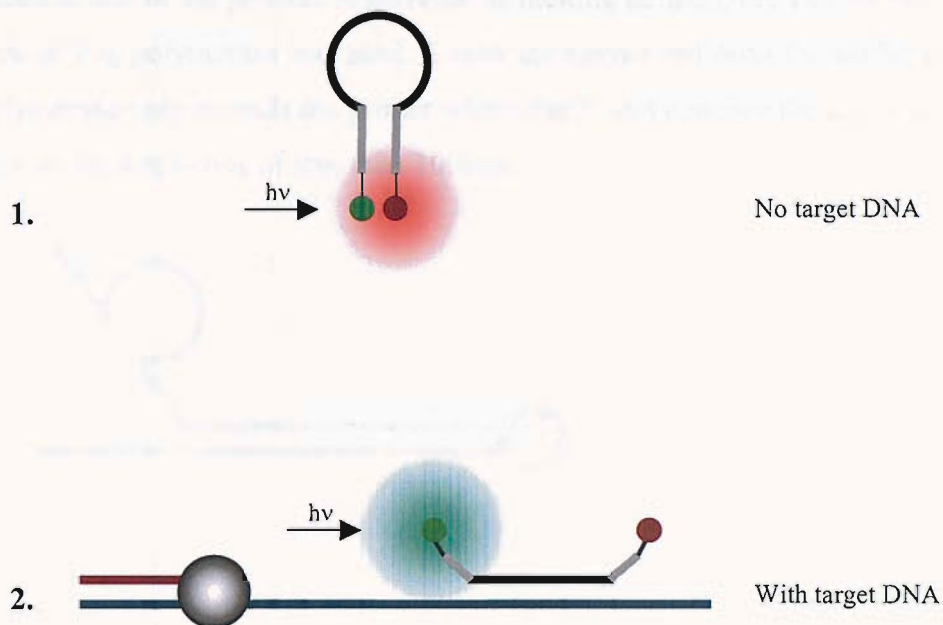
Figure 10. The mode of action for molecular beacons. (**1**) In the 'dark' state, with the stem loop structure, the fluorophore, in green, and collisional quencher, in red, are in close proximity so the fluorophore is quenched and no fluorescent signal is observed. (**2**) When a complementary target is present, the loop region of the probe binds, as it is more stable than the stem-loop structure so separating the collisional quencher from the fluorophore. The fluorophore is then free to fluoresce.

The complementary target is generated by PCR. As with TaqMan®, this format is combined with PCR to produce a closed-tube assay in which the fluorescence can be monitored in real time as the PCR product accumulates.

Different formats have been developed in an attempt to increase signal strength and accuracy, e.g. Sunrise® primers[24], HybProbe assay[25], Scorpion primers[26,27] and duplex scorpion primers.[28]

1.3.7 Allele-Specific PCR

A functional PCR requires the primers to be annealed to the target. In allele-specific PCR, if there is a mismatch between the primer and the target sequence, under the correct conditions the primer will not bind and the PCR will not occur. However, if the primer and target are fully complementary, the primer will bind and the PCR will occur.

Originally this method relied on gel electrophoresis, however false-negative results cannot be distinguished from negative results so undermining this methodology.

To improve this method an intercalating dye was used to detect double stranded DNA, real-time melt curve analysis was performed on the PCR products, a GC-rich sequence was added to one of the primers to increase its melting temperature and the stoffel fragment of *Taq* polymerase was used. Under the assay conditions the stoffel fragment of *Taq* polymerase only extends the primer where the 3' end matches the target sequence and only yields amplicons of less than 100bps.
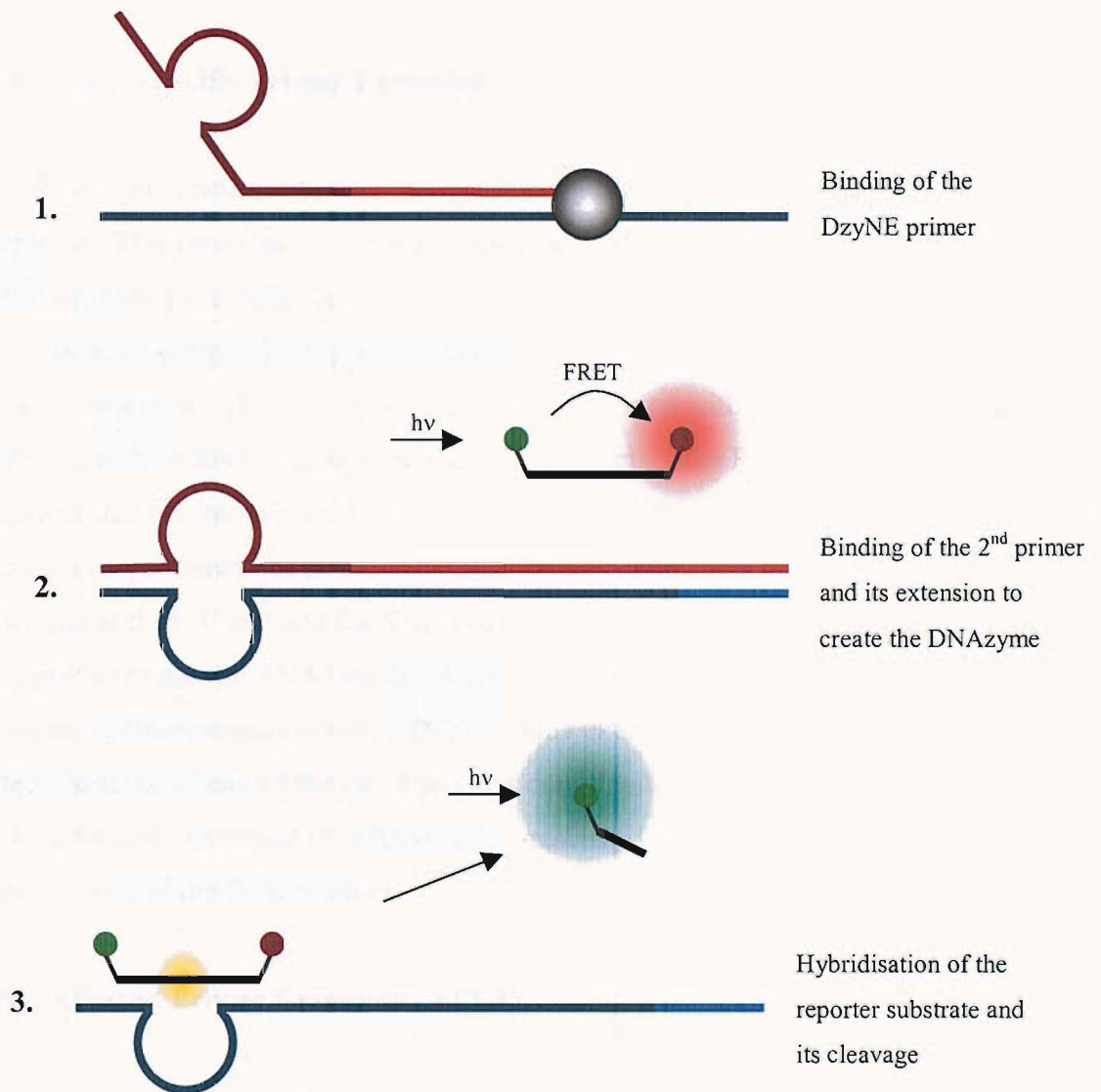
Figure 11. (1) The DzyNA primer (red line) containing the antisense DNAzyme at its 5' end (brown line) binds to the target DNA and is extended by a polymerase. (2) The complementary primer (light blue) binds and is extended, this time the sense DNAzyme is synthesised (dark blue). (3) The DNAzyme binds the reporter substrate and cleaves the phosphodiester bond between the two ribonucleotides in the centre of the reporter substrate (yellow circle). The fluorophore is then released into solution (green circle).

An alternative methodology based on the same principle uses a 10-23 DNAzyme that is capable of cleaving the phosphodiester bond between the ribonucleotides G and U

(Fig. 11). A primer that contains an antisense DNAzyme placed at the 5' end of a primer is used so that when this is copied a sense copy of the DNAzyme is synthesised. The active DNAzyme binds to the reporter substrate that contains the ribonucleotides G and U in the centre of the reporter. The DNAzyme then cleaves the chimeric reporter and the fluorophore is released into solution. In its uncleaved form, the FRET quencher prevents the reporter from fluorescing. If the primer cannot bind, the sense DNAzyme will not be generated and so no cleavage of the reporter will take place.[29]

1.3.8 Allele-specific Primer Extension

In allele-specific primer extension the genotyping is carried out after the PCR has been completed. This provides two levels of specificity, one during the PCR reaction and the other during the post-PCR genotyping.

In one example, PCR products are synthesised that contain a T7 RNA polymerase promoter sequence. These PCR products are then mixed with a T7 RNA polymerase and rNTPs to produce RNA targets from the PCR products. The RNA targets anneal to probe sequences that are immobilized on a surface (Fig. 12). The probes correspond to both the polymorphic combinations possible at the SNP site. The DNA probes have the SNP nucleotide at their 3' end and the SNP site is centrally located in the RNA target so the DNA probes prime the RNA targets. A reverse transcriptase is then used to extend the primer using fluorescently labelled dNTPs. This extension can only occur if the nucleotide at the 3' end of the DNA probe is complementary to the RNA target. The spots on the array therefore only become fluorescent if they correspond to the polymorphism in the PCR product.[30]

1.3.9 Arrayed Primer Extension (APEX)

In this example of SNP analysis, the primers for an extension assay are arrayed on a surface, in the case reviewed here a gel matrix on a surface. The primer's 3'-end is positioned so that it is one nucleotide short of the polymorphic site.
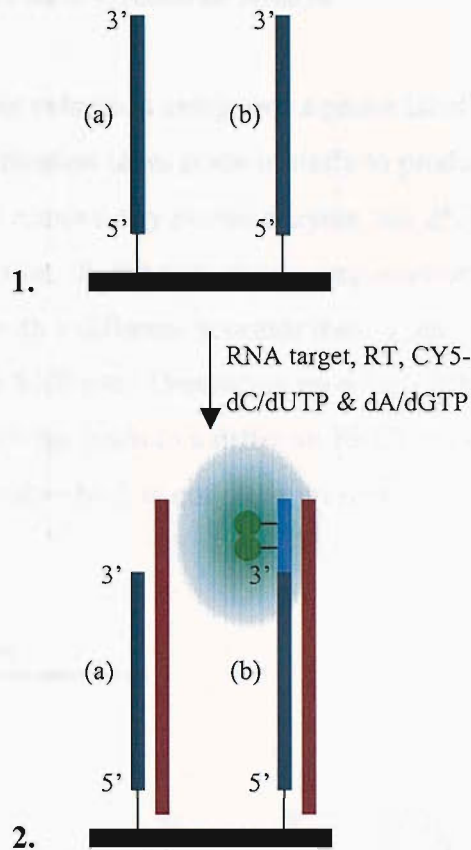
Figure 12. Allele-specific primer extension. A multiplex PCR is initially carried out that incorporates a T7 RNA polymerase promoter into PCR products. The PCR products are then mixed with a T7 RNA polymerase and rNTPs to produce RNA targets. The RNA PCR products are then exposed to an array of immobilised primers (**1**). The immobilised DNA primers are designed to bind to the RNA targets with the SNP nucleotide at their 3'-end. In this case (**a**) has the incorrect genotype at it 3'-end and (**b**) is complementary. If the 3' nucleotide is not complementary then the subsequent reverse transcriptase (RT) reaction with the fluorescent dNTPs will not occur and the spot will remain non-fluorescent **2(a)** however, if they are complementary then the fluorescent dNTPs will be incorporated and will be observed as a fluorescent spot on the array **2(b)**.

The PCR is carried out in a separate step and the products are exposed to the array (Fig. 12). Once hybridisation has occurred, a mini-sequencing reaction is carried out. Mini-sequencing is where the primer is extended by only one base using a polymerase and dideoxynucleotide triphosphates (ddNTPs) that act as terminators for the polymerase reaction. If each of the ddNTPs is labelled with a different fluorophore the colour of the spot will correspond to the base that has been inserted at the SNP site. The presence of a fluorescent spot indicates that a PCR product has been generated.[31]

1.3.10 Homogeneous Primer Extension Assays

The homogeneous primer extension assay uses a probe labelled with a donor dye that acts as a primer. PCR amplification takes place initially to produce the PCR products. The sample is then treated to remove any excess enzyme and dNTPs that might interfere with the minisequencing reaction. In the minisequencing reaction, two fluorescent ddNTPs are used, each labelled with a different acceptor dye. These two ddNTPs correspond to a different genotype at the SNP site. Depending on which ddNTP is incorporated (and this depends on the sequence) this leads to a different FRET signal from the acceptor dye (Fig. 13). This then reveals which genotype is present.[32]
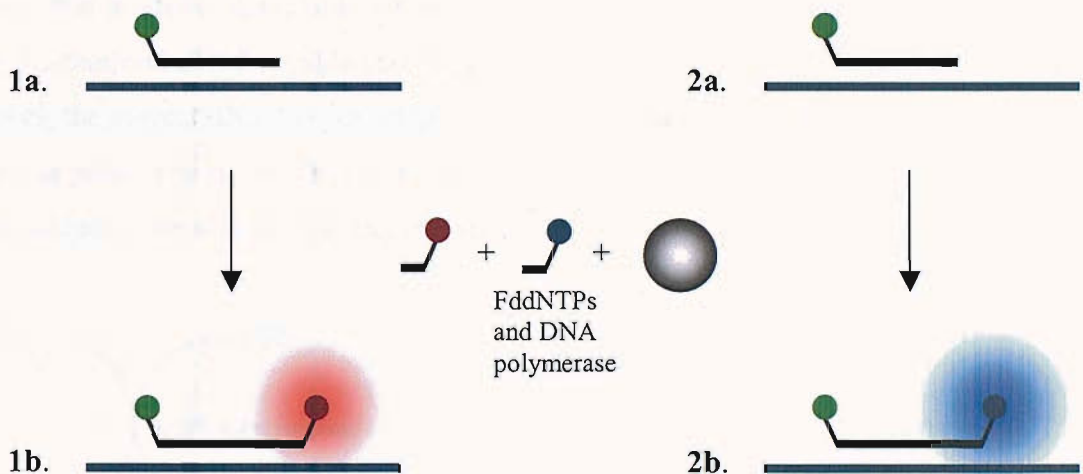
Figure 13. The template-directed dye-terminator incorporation assay. The region of interest is amplified using the PCR. The excess dNTPs are removed using shrimp alkaline phosphatase and exonuclease I and in a second step heat is used to deactivate the enzymes. (**1 & 2a**) A fluorescently labelled primer is then added together with labelled ddNTPs and a DNA polymerase. (**1 & 2b**) The ddNTP incorporated will depend on the base at the SNP site, therefore the fluorescent signature after the reaction will depend on the genotype.

1.3.11 Primer Extension with Detection by Mass Spectrometry

Many of the methods already mentioned can be adapted so that the detection is carried out using mass spectrometry instead of using fluorescence as the elucidating technique. Mass spectrometry has become a widely used detection technology because of the advent of matrix assisted laser desorption / ionisation time-of-flight mass spectroscopy (MALDI-TOF MS) and to a lesser extent electrospray (ES) MS.

Both of these techniques allow for the detection of non-volatile, thermally labile, intact molecules with a molecular weight greater than 2 kDa such as oligonucleotides. Many different assays have used MS as the method of detection and these assays have been reviewed.[33-37]

1.3.12 Pyrosequencing

The extension of a primer can be detected based on the detection of pyrophosphate (PP_i, $P_2O_7^{4-}$), which is a by-product of the polymerisation of the dNTPs. By using a combination of up to seven enzymes and other specialized reagents, the pyrophosphate can be converted into adenosine triphosphate (ATP), which is then used in a luciferase reaction that produces detectable visible radiation. If the dNTPs are added individually, when the required dNTP is added by the polymerase, light will be observed (Fig. 14). If, however, the correct dNTP is not added then no light will be observed. An enzyme, apyrase is present to remove any unreacted dNTP and excess ATP so that another dNTP can be added to see if it will be incorporated.[38]
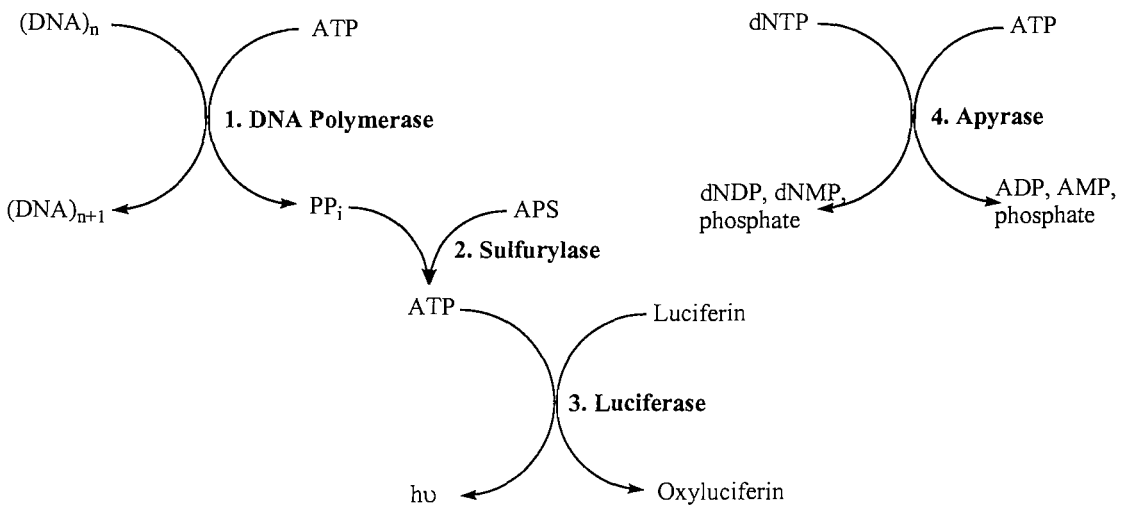
Figure 14. Pyrosequencing. During DNA synthesis by a DNA polymerase the dNTPs are added individually. If the incorrect base is added, the apyrase (**4**) present in the reaction will degrade the triphosphate. If the correct dNTP is added, the nucleotide is added yielding pyrophosphate as a by-product of the reaction (**1**). This is then converted into ATP by a sulfurylase enzyme (**2**), some of this is consumed by the apyrase but the remaining is utilised by a luciferase enzyme (**3**) in converting luciferin to oxyluciferin in a reaction that produces visible radiation. This is detected and the time it occurs is linked to the triphosphate that was present at the time in order to work out the nucleotide inserted. ADP and AMP, are adensoine di and monophosphate. dNDP and dNMP are 2'-deoxynucleotide di and monophosphate. APS is adenosine 5'-phosphosulfate.

1.3.13 Multiplex Primer Extension Sorted on Genetic Arrays

This format involves the PCR of the SNP sites, subsequently the excess PCR primers and dNTPs are degraded by a combination of shrimp alkaline phosphatase (SAP) and *E.coli* exonuclease I (Fig. 15). The combined PCR products are then mixed with capture probes. The capture probes are oligonucleotides that are complementary to a region adjacent to a SNP site and act as a primer for a single base chain extension (SBCE) reaction. The complementary region stops one base short of the polymorphic
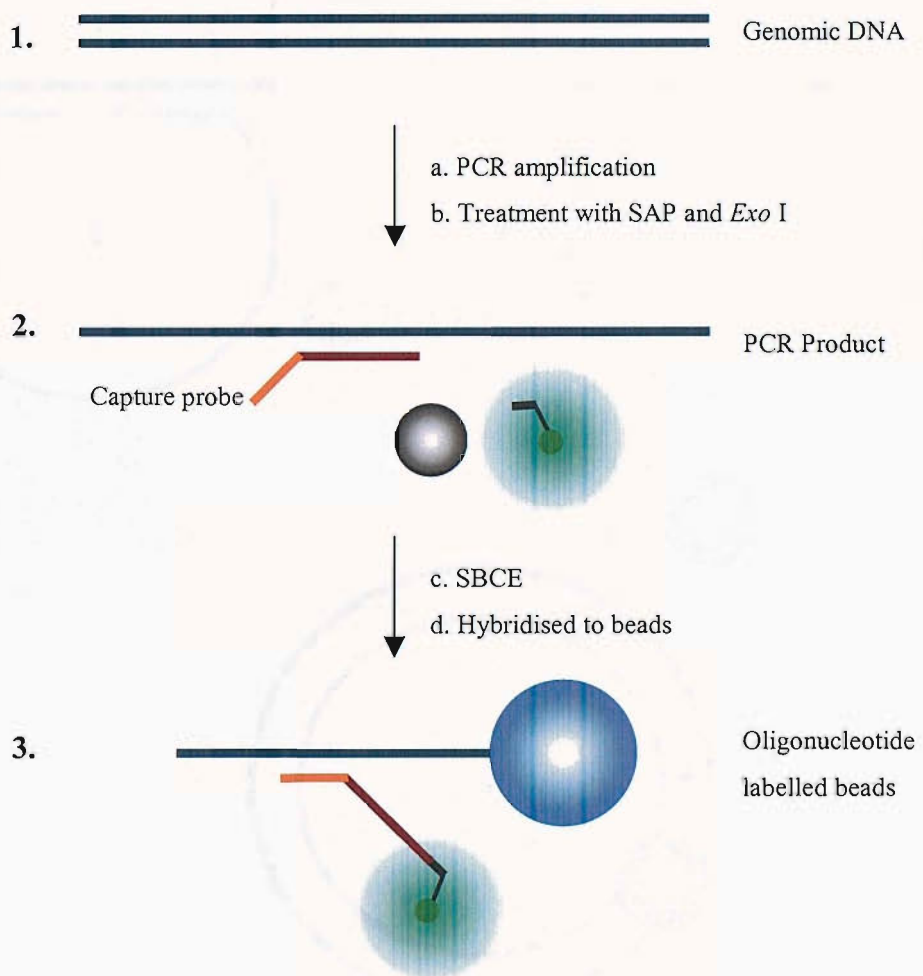
Figure 15. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension.[39] (**1a**) The region containing the SNP is amplified by the PCR. (**1b**) The excess primers and dNTPs are degraded. (**2**) The capture probe (brown) then binds to the PCR product and (**2c**) a SBCE reaction is carried out using FddNTPs. (**3**) The capture probe now hybridises to an oligonucleotide immobilised on a fluorescently labelled bead via another region referred to as the ZipCode (orange). The beads are then sorted by a FACS to determine if they have the labelled probe and if so which SNP does it correspond to.

site so the SBCE reaction can be carried out. The minisequencing reaction is carried out using one fluorescently labelled ddNTP that corresponds to a particular nucleotide being present at the SNP site and the other three ddNTPs are unlabelled. The capture probes from the SBCE reaction are then annealed to oligonucleotides that are attached to fluorescently labelled beads. The beads that emit a fluorescent signature from the fluorescently labelled ddNTP incorporated in the SBCE reaction are isolated by a fluorescently activated cell sorter (FACS). The fluorescent label of the bead is then read and this corresponds to the SNP that had been identified.[39]

The basic principles are demonstrated in this example and have been used to produce other bead based genotyping systems.[40-42]
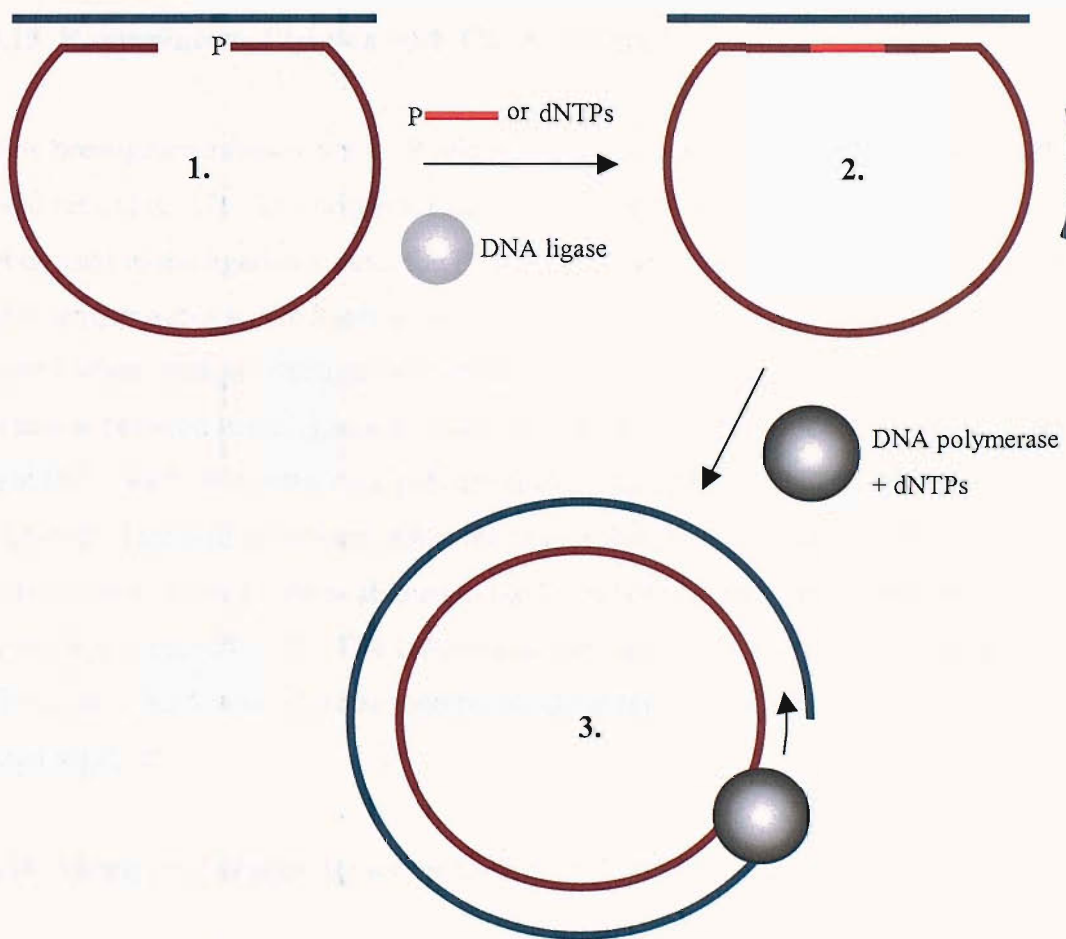
Figure 16. Circularisation of ssDNA and isothermal rolling-circle amplification. (**1**) The single stranded DNA target and the single stranded probe bind in such a way to place the 5' and 3' ends of the probe sequence close together. (**2**) The gap is then filled with dNTP's or a gap oligonucleotide and joined using a DNA ligase. A rolling circle primer can then be bound to the circularised DNA for amplification. (**3**) The circlularised DNA is then copied using a displacing DNA polymerase and dNTPs.

1.3.14 Ligation with Rolling Circle Amplification

Rolling circle amplification (RCA) is an isothermal DNA polymerase mediated reaction (Fig. 16). A single-stranded DNA target that needs amplification can be circularised by using a DNA ligase. Once circularised a primer and a displacing DNA polymerase are required to form a continuous polymerase reaction. A hyperbranched-RCA (HRCA) reaction can also occur with the use of another primer that binds to the displaced strand so starting off further polymerase reactions. This in turn means that the original primer can bind again to set off even more polymerisation reactions. SNP analysis can then be carried out in a variety of ways to determine if an SNP is present or not in the circularised DNA.[43]

1.3.15 Homogeneous Ligation with FRET detection

In this homogeneous assay the PCR and ligation reaction are carried out in the same closed tube (Fig. 17). The primers used for the PCR are longer than the oligonucleotide probes used in the ligation so that the PCR can be carried out independently by cycling at higher temperatures so the ligation probes cannot bind. The temperature can then be lowered when there are enough PCR products to carry out the ligation. The ligation reaction is between two oligonucleotides one with a dye at the 3' end and the other with a dye at the 5' end. When the two dyes are close in space FRET will occur and this can be monitored. This will only occur when the two probes bind to a complementary target so that the 5' end of one probe is aligned to the 3' end of the other and if there are no mismatches at the SNP site. The DNA ligase can ligate the two probes together.[44] The SNP site is at the 5' end of the second probe at the junction between the two probes to be ligated together.

1.3.16 Multiplex Ligation Reaction Sorted on Genetic Arrays

This method is very similar to the multiplex primer extension sorted on genetic arrays, however, instead of carrying out an SBCE with a fluorescent ddNTP a fluorescently labelled 5'-phosphate oligonucleotide is ligated to the capture oligonucleotide.[45]
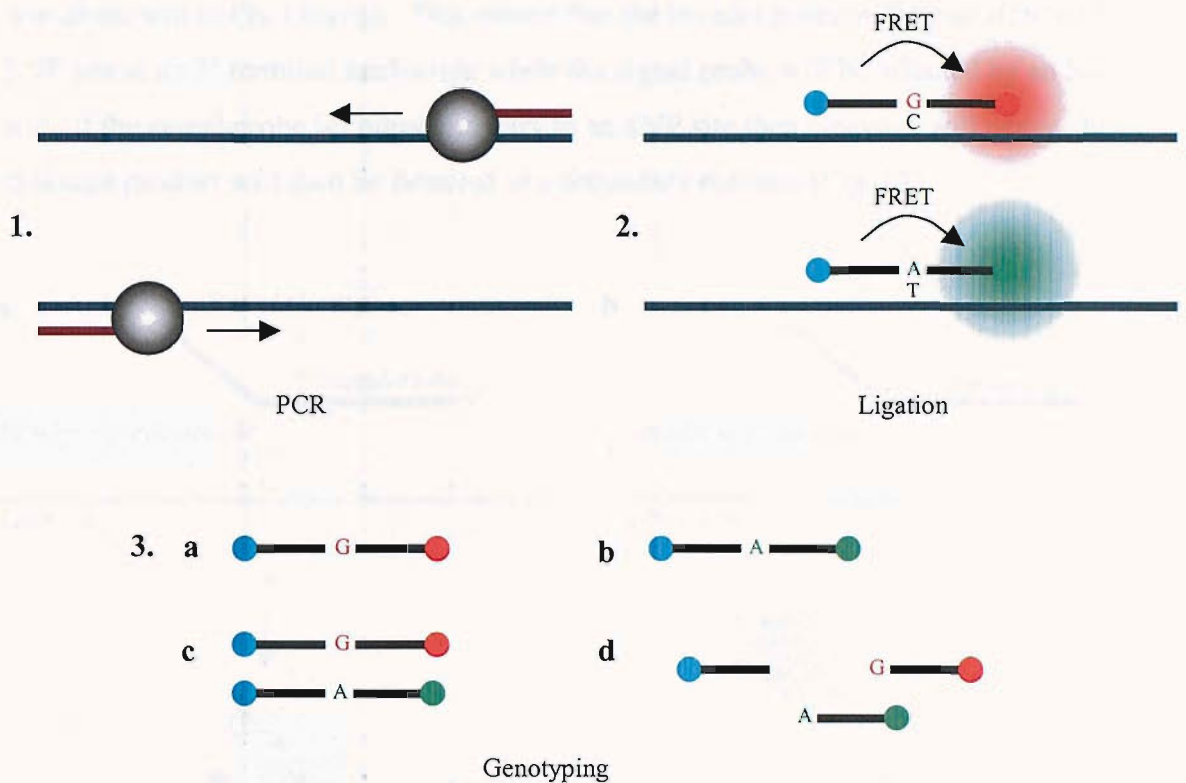
Figure 17. (**1**) Since the primers anneal at a higher temperature than the probes, the PCR can be conducted at these higher temperatures without the probes interfering with the reaction. (**2**) The probes can only bind when lower temperatures are used once PCR product has been generated. There are three probes one labelled with fluorescein, in blue, that is complementary to the region next to a SNP site. The other two probes, one labelled with TAMRA, in red, the other with ROX, in green, correspond to the genotype at the SNP site, in this case T or C. The ligation of the FAM probe to another will only take place if both probes are complementary. (**3c**) If the sample is a heterozygote then emission will be seen from both TAMRA and ROX when FAM is excited. (**3a** or **b**) If the sample is a homozygote then either TAMRA or ROX will be seen. (**3d**) If the probes are not complementary then only FAM emission will be observed.

1.3.17 Homogeneous Invasive Cleavage, the Invader Assay

The Invader Assay is based on the ability of a flap endonuclease (FEN), isolated from archaea to recognise and cleave a structure formed when two overlapping oligonucleotides hybridise to a target DNA strand.[46,47] Archaea are single-celled prokaryotic microorganisms. The enzyme cleaves at the 5' end of the downstream oligonucleotide, the precise site of cleavage is dependent on the amount of overlap. In order for cleavage to occur, the 5' end of the signal probe must be complementary to the target sequence in a region where the 3' end of the invader probe is. That is to say, the 3' terminal nucleotide of the invader probe does not need to be complementary to the target DNA but it needs to overlap with one or more 5' nucleotides of the signal probe that are complementary. This needs only be by one base, however an unpaired 5' signal probe

37

arm alone will not be cleaved. This means that the invader probe will be unaffected by an SNP site at its 3' terminal nucleotide while the signal probe will be affected by an SNP site. If the signal probe is complementary to an SNP site then cleavage will occur, the cleavage product will then be detected in a secondary reaction (Fig. 18).
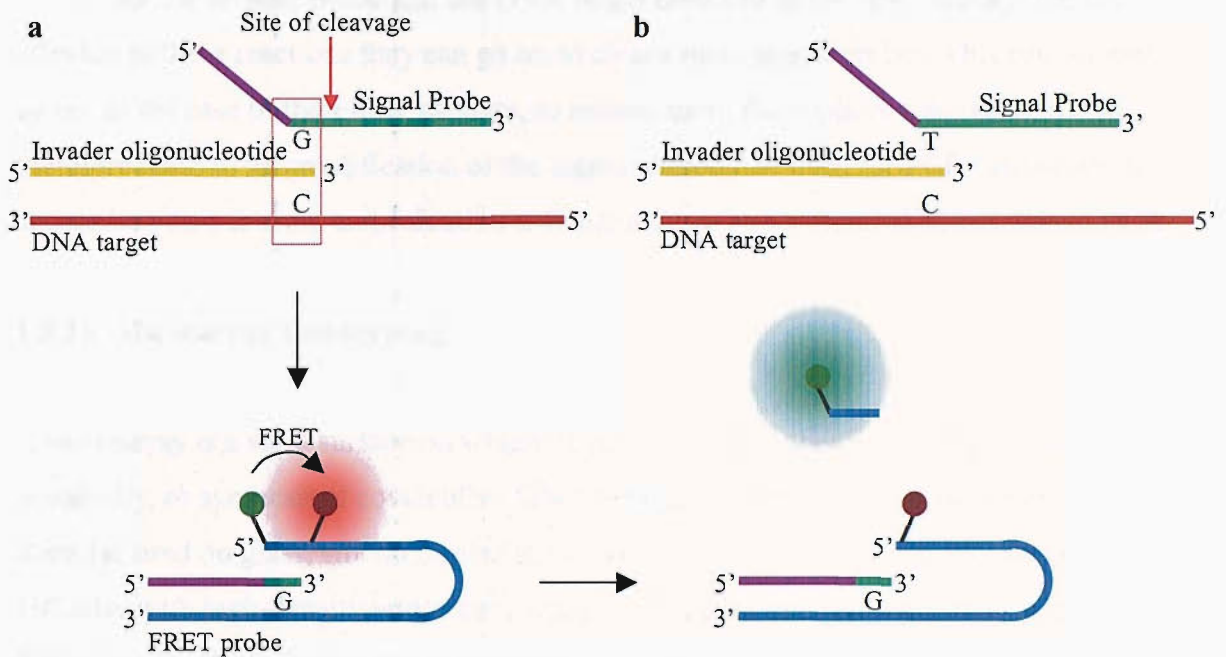
Figure 18. The invader assay. The target DNA in red contains a SNP site in this case a C base. An invader oligonucleotide, in yellow, binds to the target DNA. At the same time a signal probe binds downstream of the invader oligonucleotide. The signal oligonucleotide contains two regions; the 3' end is complementary to the DNA target, in green, with the 5' nucleotide of this region complementary to either SNP genotype. The two diagrams above denote this. The 5' end of the signal probe, in purple, is uncomplementary to any of the DNA target but is complementary to the 3' end of a FRET probe shown above in blue. (**a**) When the invader probe overlaps the polymorphic site but is not necessarily complementary to the polymorphism in the DNA target and the 5' end of the signal probes complementary region is complementary to the genotype present in the target DNA then a structure is formed that is recognised by the flap endonucleases (FENs). This enzyme on recognising this structure will cleave the downstream, signal probe, oligonucleotide on the 3' side of the overlapping nucleotide. This means that the released oligonucleotide contains all the uncomplementary sequence in purple, plus the nucleotide that was complementary to the polymorphic site in green. This sequence then acts as an invader probe in the FRET probe and so causes the release of a fluorophore into solution. (**b**) If the 5' end of the complmentary region of the signal probe is not complementary to the polymorphic site in the DNA target then no cleavage will take place.

In the secondary reaction, the cleavage product binds to a template longer than itself and acts as a primer so that a DNA polymerase can bind and incorporate fluorescent dNTPs. The fluorescent polymerase products are then separated and visualised using gel-electrophoresis. Several developments to the assay have been made since its inception. One of the improvements has been to the secondary signalling phase. The modified

procedure has a FRET cassette oligonucleotide where a fluorophore is placed adjacent to a FRET dye.[48] If the signal probe has been cleaved, it will bind to the FRET cassette and a second cleavage will take place between the two dyes. This means that FRET can no longer take place and will alter the spectral qualities of the solution (Fig. 18). This signalling technique has also been adapted for use in a FACS using streptavidin beads.[49]

As the invader probe and the DNA target involved in the first cleavage are not affected in these reactions they can go on to cleave more signal probes. This can, in turn, go on, in the case of the FRET cassette, to release more fluorophore into solution. This therefore leads to an amplification of the signal without the need for PCR. However, it should be noted that the amplification is linear and not exponential as in the case to PCR.

1.3.18 Microarray Genotyping

A microarray is a solid surface on which oligonucleotide probes can be placed non-covalently, or synthesised covalently. DNA hybridisation microarrays are generally manufactured on glass, silicon or plastic. Low-density microarrays can contain as few as 100 sites with high-density arrays containing >100K sites that can range in size from 10-500 microns. High-density arrays typically have a surface area of 1-2 cm^2. Positive results are detected *via* the reporter on the probe, often a fluorophore by a scanner (Fig. 19).

Micorarrays still face many technical challenges. The probes for instance all have different thermodynamic properties that can lead to false negative and positive results. To some extent, this can be resolved through the use of modified bases. For example, deoxyadenosine and deoxythymidine have been replaced with 2'-*O*-methyl-2, 6-diaminopurine and 2'-*O*-methyl-5-thymidine respectively to increase the signal intensity of A/T-rich probe sequences by increasing the melting temperature (T_m) of the probe-target duplex. The photolithographic process, by which probes are synthesised on the surface, has a stepwise yield of 95% so limiting the size of the probes.[50]
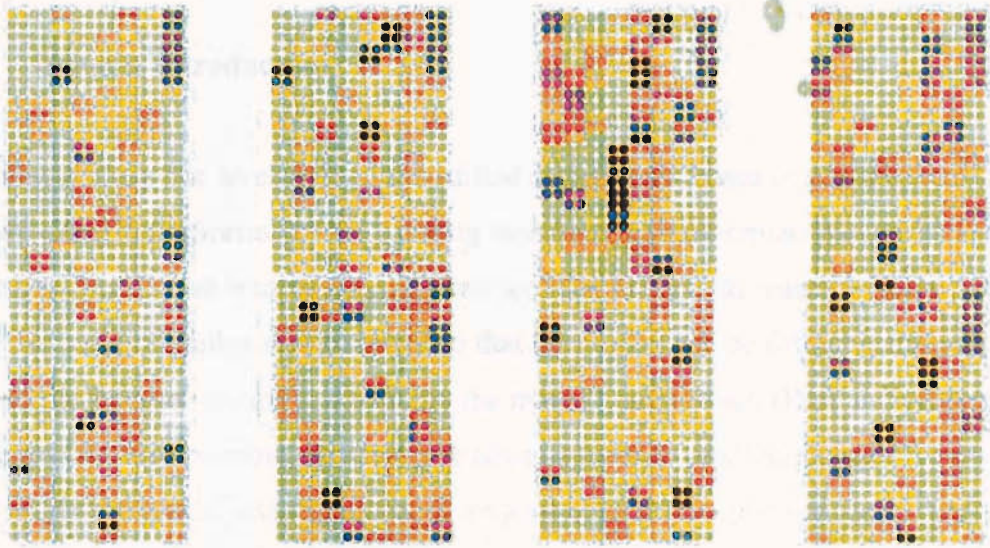
Figure 19. The results of a typical genotyping microarray. Each circle or spot is an area where an oligonucleotide has been attached or placed. The colour of the spot corresponds to whether a complementary oligonucleotide carrying a fluorescent label has hybridised or not. The tests are done in duplicate with the white or yellow/green spots corresponding to a negative test and all others a positive result.

Microarray to microarray variation is still high because of problems with reproducibility and testing, which means that microarrays are yet to make it into the clinic[51].

1.4   MODIFICATION OF OLIGONUCLEOTIDES

1.4.1   General Introduction

Oligonucleotides can be chemically modified at the sugar, bases or phosphates.[52] This can involve either modification of the existing structure or entire replacement with a different structure. It is desirable to modify oligonucleotides for several reasons. These may be to study the basis of duplex stabilisation, so that the duplex can be further stabilised, to incorporate reporter groups and to study the interaction between DNA and proteins. These reasons are desirable, because it is advantageous to predictably alter the properties of DNA for biological investigation and diagnostic or therapeutic purposes. For instance, in diagnostics it is important to be able to distinguish single base changes in a DNA sequence. These challenges greatly increase when trying to use oligonucleotides as therapeutic agents, due to the complexities of living cells.

1.4.2   Modified Heterocyclic Bases

In modifying the heterocyclic bases the electronic distribution is altered over the bases. The altered electronic structure of the bases is the foundation of their altered properties. The alteration of their properties can be conceptualised in many different ways. Conceptualisation can be aided, for example, by supercomputing and *ab initio* quantum chemical calculations to give a realistic insight into the effect of potential modifications on electronic distribution. However, in order to demonstrate these predicted properties or in order to test a structure the base needs to be synthesised.

The most suitable sites for substitution are those that are exposed to the solvents in the major groove of a DNA duplex so that substituents are less likely to cause general steric interference or interfere with base pairing. These criteria mean that the four and 5-positions in pyrimidines and the six and 7-positions in purines are the most appropriate sites for functionalisation. The numbering system for three heterocyclic bases are shown in Fig. 20.

The introduction of propyne increases the stability of many heterocyclic bases in DNA duplexes. Therefore, the Sonogashira reaction[53] as adapted by Hobbs[54] is a very important reaction. The palladium (0) cross-coupling reaction allows the coupling of many different alkynes into any of the four major DNA bases. The reaction is carried out

41

Pyrrolo[2,3-*d*]pyrimidine   Purine   Pyrimidine
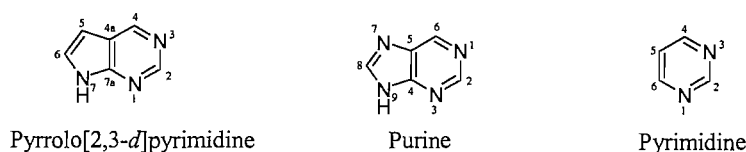
Figure 20. The numbering system for the heterocycles pyrrolo[2,3-*d*]pyrimidine, purine and pyrimidine.

under mild room temperature conditions and is often high yielding with few if any side reactions. An iodine atom is easily introduced at the five position of pyrimidines and the seven position of purines to produce the aryl halide needed for the reaction.

1.4.2.1 Pyrimidines

Pyrimidines modified in the 5-position are found in nature, for instance the bacteriophage XP-12 *Xanthomonas oryzae* has all its cytosines replaced by 5-methylcytosine this results in very thermally stable DNA. The effect of 5-methylpyrimidines appears to be a complex process as their incorporation has led to both positive and negative influences on duplexes. The introduction of alkynes at the 5-position has a pronounced influence on duplex stability. 5-Propynyl-2'-deoxyuridine increased duplex stability due to the increased hydrophobicity and the extended geometrical overlap that resulted in favourable vertical stacking interactions[55-57]. It has been demonstrated that propynylated pyrimidines enhance mismatch penalties of DNA.RNA duplexes and that long-range cooperativity is present in oligonucleotides containing multiple, sequential, propynylated pyrimidines.[56-58] Phosphorothioate oligodeoxynucleotides containing both 5-propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine were potent antisense inhibitors.[59,60] 5-Propynyl-2'-deoxycytidine was shown to stabilize duplexes in one of the above examples[57]. 5-Thiazolyl-uridine and 5-methylthiazolyl-uridine also increased the stability of duplexes in a similar fashion. These bases are shown in Fig. 21.
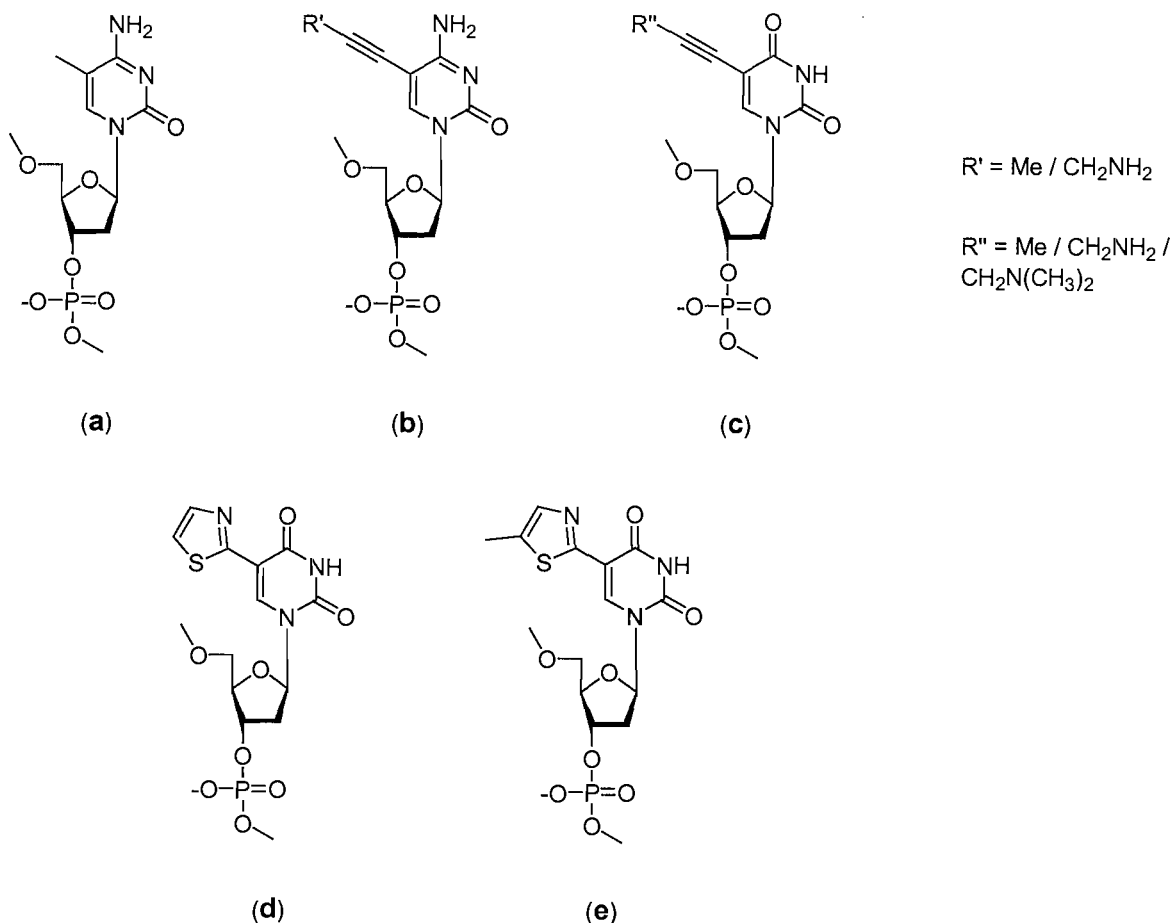
Figure 21. The structure of the 5-modified pyrimidines: (a) 5-methyl, (b) 5-propynyl, 5-(3-aminoprop-1-ynyl)cytosine, (c) 5-propynyl, 5-(3-aminoprop-1-ynyl), 5-(3-dimethylaminoprop-1-ynyl)uridine and (d) & (e) thiazolyl-substituted deoxyuridines.

The addition of the methyl has a positive effect over and above the effect of the thiazolyl group when incorporated into DNA and when it is forming DNA-RNA hybrid duplexes. In order to further increase the stability of 5-substituted pyrimidines an amino group was introduced to give 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine.[61] It was predicted that the amino group would form a salt bridge with the non-bridging major groove oxygen on the phosphate of the 5'-nucleotide.[62] 5-(3-Aminoprop-1-ynyl)-2'-deoxyuridine has found many uses that utilise its: nucleophilicity for use in chemical labelling,[54,63,64] thermal stability in triplexes[65,66] and duplexes,[62,67,68] and positive charge for *in vitro* selection,[69,70] induction of curvature in DNA,[71] attachment to a silver colloid surface for surface enhanced resonance Raman scattering (SERRS)[72] and mass spectrometry[73]. This modification was also introduced into cytosine to give 5-(3-aminoprop-1-ynyl)-2'-deoxycytosine and this base was used in some of the above applications.[54,64,68,73] In order to prevent the amino group from becoming charged the dimethylamino functional group

was used to give 5-(3-dimethylaminoprop-1-ynyl)-2'-deoxyuridine and it was shown that the 5-(3-dimethylaminoprop-1-ynyl)-2'-deoxyuridine was not as stabilising as 5-propynyl-2'-deoxyuridine.[67]

1.4.2.2 Purines and their analogues

Many of these modifications have been introduced into the heterocycle pyrrolo[2,3-d]pyrimidine, an analogue of purine. The N-7 in purines is replaced by a methinyl group in pyrrolo[2,3-d]pyrimidines. The methinyl group can then be functionalised, often *via* iodination.
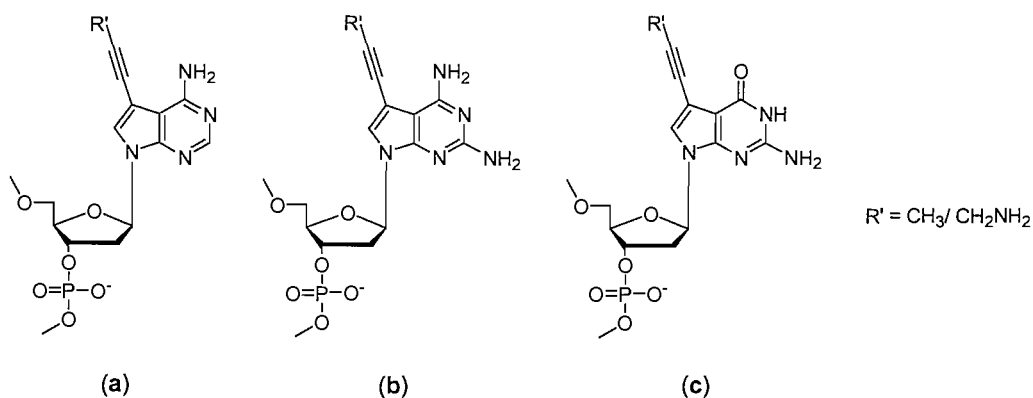
Figure 22. The structure of 7-modified pyrrolo[2,3-d]pyrimidines based on (a) adenine, (b) 2-amino adenine and (c) guanosine.

7-Propynyl-7-deaza-2'-deoxyadenine and 7-propynyl-7-deaza-2'-deoxyguanosine (Fig. 22) were synthesised and both were shown to have a stabilising effect when bound to RNA and increased the potency of oligonucleotides containing these analogues in an antisense assay.[74]

7-Propynyl-7-deaza-2'-deoxyadenine has also been used in *in vitro* selection studies due to its positive charge.[75] Similar effects have been observed with 7-propynyl-7-deaza-2-amino-2'-deoxyadenine with an increase in T_m of 3-4°C for a singly substituted oligonucleotide against an RNA target. Oligonucleotides containing 3 or 4 substitutions showed a 2-3-fold increase in potency over unmodified controls in an antisense assay.[76]

7-(3-Aminoprop-1-ynyl)-7-deaza-2'-deoxyadenine was shown to be more stabilising than many other 7-substituted-7-deazaadenines. However, the high density of charged residues in homomers impaired duplex formation due to counter-ion

condensation.[68,77,78] This modified base has been used to incorporate: methyl red into oligonucleotides[79], various reporter groups into 2'-deoxyribonucleoside-5'-triphosphates[64], positive charges into oligonucleotides for use in mass spectrometry.[73] 7-(3-Aminoprop-1-ynyl)-7-deaza-2'-deoxyguanosine has been used in two of the above applications.[64,73] It was shown to be a stabilising modification in DNA and DNA/RNA duplexes by 2-3°C per modification,[68,80] however it has also been shown that it can have a negligible effect on T_m.[81] The structure of these bases can be seen in Fig. 22.

One of the most stabilising adenine analogues is 8-aza-7-deaza-7-propynyl-2-amino-2'-deoxyadenosine that is based on the heterocycle pyrazolo[3,4-d]pyrimidine (Fig. 23). Each addition of the base gives rise to approximately a 5.5°C increase in T_m.

Pyrazolo[3,4-d]pyrimidine
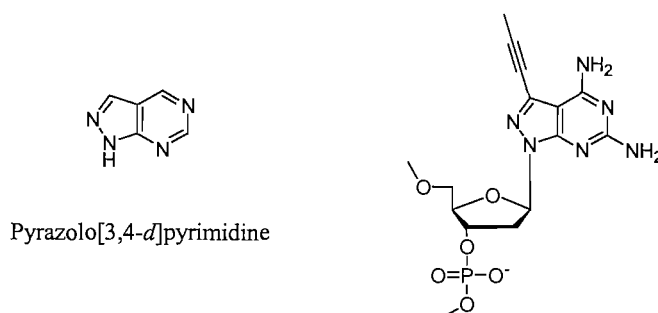
Figure 23. The structure of pyrazolo[3,4-d]pyrimidine and 8-aza-7-deaza-7-propynyl-2-amino-2'-deoxyadenosine

This increase in T_m is due in part to the different ring structure as it increases the proton donor properties of the two amino groups. The propyne group will also increase the hydrophobicity of the base as well as increasing its ability to base stack so stabilising a duplex further.[82] Halogenation with bromine or iodine at the site of propynylation also produces very stable duplexes by increasing the T_m by around 4.5°C.[83]

These are some of the most relevant modifications in relation to this work. However, many more base modifications have been synthesised and have been reviewed[84], as have those for sugar and backbone modifications that have been synthesised in order to increase nucleic acid duplex stability.[52]

1.5   ULTRA-VIOLET MELTING

1.5.1   General Introduction

Ultra-violet (UV) melting is the most commonly used technique for determining the thermodynamics of binding of nucleic acid duplexes and triplexes. Other well-established methods for determining the thermodynamics of duplex melting and formation are isothermal titration calorimetry (ITC) and differential scanning calorimetry (DSC).

1.5.2   DNA and UV Melting

The extinction co-efficient of dsDNA at 260 nm is about 40% less than that of ssDNA as the extinction co-efficient is dependent on the local electronic environment. This effect is known as hypochromicity and can be used as an indicator of the base pairing between two oligonucleotides. The hypochromic shift is due to electronic interactions between neighbouring bases. When the chromophore is excited a dipole forms as a result of the electronic transition. In turn this dipole induces opposing dipoles in neighbouring chromophores. This effect is much greater in a highly ordered, rigid, base stacked duplex structure than in a more disordered single strand. The hypochromicity is proportional to the inverse cube of the base separation. These induced dipoles oppose the inducing dipole and create an overall reduction in the transition dipole and this results in a lowering of the extinction coefficient.

Neither the sugar nor the phosphate components of nucleotides show any significant UV absorption above 230 nm. This means that nucleosides and nucleotides have UV absorption profiles similar to those of their constituent bases and absorb strongly with λ_{max} close to 260 nm. UV melting is observed as a change in absorbance of a buffered solution containing two complementary DNA strands. A definition of absorbance can be found in appendix 1.

UV Melting Curve and 1ˢᵗ Derivative
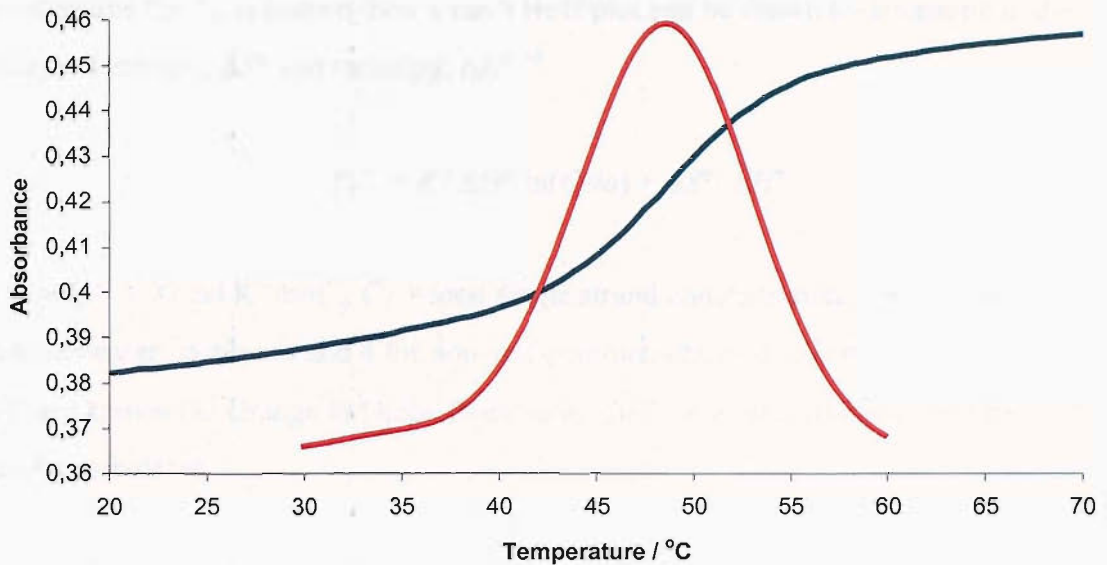
Figure 24. A UV melting curve of a DNA duplex from 20°C to 70°C. The duplex, which is present at 20°C, absorbs less UV light at 260 nm than the single strands present at 70°C. The smoothed first derivative is shown to show how the melting temperature, T_m, is determined, the point at which half the helical structure of the DNA duplex is lost. The **smoothing** (Savitzky-Golay)[85] accounts for the loss of data either side of the derivative. The derivative values as shown by the red line have been uniformly altered so that they overlap with the melting curve in blue.

The melting temperature (T_m) is described as the temperature at which half of the helical structure is lost. This is defined as the maximum turning point of the first derivative of a plot of absorbance versus temperature i.e. the steepest point of the melt curve.[86] A typical melting curve and its first derivative are shown in Fig. 24. The T_m can also be calculated using extrapolated upper and lower baselines, however, this method is less accurate when either of the baselines cannot be determined.[87] The abruptness of the transition indicates that the DNA double helix is a highly co-operative structure. In order to obtain the derivative the data first has to be smoothed and this is carried out using the Savitzky-Golay smoothing technique.[85] The method requires the data to be equidistant therefore an interval is required that determines the distance between each data point in the created equidistant data set. A filter size, odd integers between 5-101, is required that determines the level of smoothing. As the first and last data points are lost during the smoothing process the more highly smoothed the data the more points are lost and the narrower the range over which the derivative is calculated.

47

1.5.3 Thermodynamic Analysis

Once the melting temperature of a duplex and the concentration of the single strands used to determine the T_m is known then a van't Hoff plot can be drawn to determine to the change in entropy, $\Delta S°$ and enthalpy, $\Delta H°$.[88]

$$T_m^{-1} = R \:/\: \Delta H° \ \ln(C_T/a) + \Delta S°/\: \Delta H°$$

Where R = 1.99 cal $K^{-1}mol^{-1}$, C_T = total single strand concentration, a = 1 for self-complementary duplexes and 4 for non-self-complementary duplexes. Since $\Delta S°$ and $\Delta H°$ are known the change in Gibbs free energy, $\Delta G°$, at a particular temperature, T, can then be calculated.

$$\Delta G° = \Delta H° - T \: \Delta S°$$

For the regression line analysis to be significant a minimum of six points were required over a 10-fold range of concentrations.
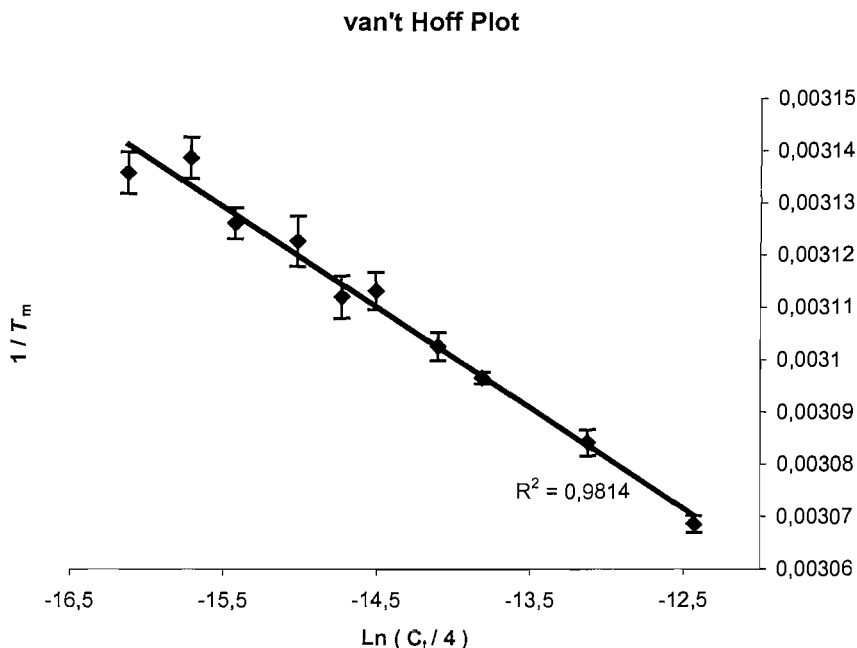
van't Hoff Plot

Figure 25. A van't Hoff plot analysis made up of 10 T_m determinations over a 40-fold range of concentrations. The temperatures are in Kelvin and the concentrations are in moles. The errors in the T_m are shown to a 95 % confidence level. The R^2 value gives an indication of the level of fit of the points to the linear regression line shown from which $\Delta S°$ and $\Delta H°$ are calculated. As can be seen from the graph the lower the concentration of single strands the higher the error in the readings.

1.6   STABILISATION OF DNA DUPLEXES

1.6.1   Introduction

1.6.1.1 Unmodified duplexes

The entropy of helix formation is always unfavourable and all stabilisation comes from favourable enthalpic interactions in unmodified duplexes. The stabilisation for duplex formation is the product of inter and intra-stand stacking interactions, burial of polar and non-polar surfaces and hydrogen bonding.

The contribution of each of these different interactions depends on the temperature at which duplex formation is observed at because of the changing heat capacity of the process discussed in 1.6.4. It has been reported for a 14mer duplex at 40°C that, 41% of the stabilisation comes from the intra-stand base stacking in the single strands, together with the burial of non-polar and polar surfaces. A further 65% of stabilisation comes from the docking of the two ordered single strands which comes from the formation of the hydrogen bonds, inter-strand stacking and a further increase in the burial of polar and non-polar surfaces. The intra-strand stacking accounts for ~73% of the total stacking interactions as shown by the hypochromism in the single strands when compared to the duplex. The remaining ~27% comes from inter-strand stacking. This has also been accounted for by molecular modelling in that ~27% of ring overlap comes from inter-strand base overlap. The final folding process to take into account is that of unfolding structures in the single strands. This requires energy i.e. the process has a positive ΔH and so there is a 6% penalty for unfolding the hairpins in the single strands.[89]

1.6.1.2 Modified duplexes

The stabilisation of DNA duplexes is an important area of study. The stabilisation must be carried out in a specific manner such that the most stable mismatches are not stabilised to any greater extent than the complementary base. In this work the stabilisation is brought about by the modification of the heterocyclic bases, however, DNA duplexes can be stabilised in other ways. Modifications of the backbone or the 2'-deoxyribose sugar can lead to increases in stability.[90] Stabilisation can also be brought about by the binding

of other compounds or chemicals that can bind non-specifically or can bind in the major or minor grooves of the helix again either specifically or non-specifically.

The heterocyclic bases are modified in an attempt to either increase the entropy i.e. make it less negative or decrease the enthalpy i.e. make it more negative relative to oligonucleotides containing unmodified bases. One of the ways of doing this is by trying to increase the levels of pre-organisation in the single strands. In an ideal situation the single strand would be rigid and in the conformation it adopts when it is in the double helix. In this form any mismatches would be extremely destabilising as the lack of flexibility would prevent the mismatched base pair from attempting to find a suitable conformation to minimise the unfavourable interactions.

One of the most effective ways to encourage pre-organisation when modifying the base pairs is to increase the base stacking interactions between the bases. The most recent *ab initio* quantum chemical calculations describe base stacking as a composition of three components: dispersion attractions (induced dipole-induced dipole interactions), short-range exchange repulsions and electrostatic interactions. The dispersion attractions are primarily responsible for the stabilisation in base stacking. The attractions are rather isotropic, that is independent of direction and proportional to the geometrical overlap of the bases. The short-range exchange repulsions, the correction in the coulomb repulsion between two electrons in different orbitals having the same spins, is responsible for the vertical distance between the two stacking bases. The electrostatic interaction dictates the mutual orientation and displacement of the bases. The results from the calculations show that a simple empirical potential relying on the use of atom-centred point charges is qualitatively sufficient to describe base stacking.[91] This can be seen in Fig. 26 where, within the constraints of the duplex, the atoms of one base appear to avoid lying immediately above or below atoms of an adjacent base.
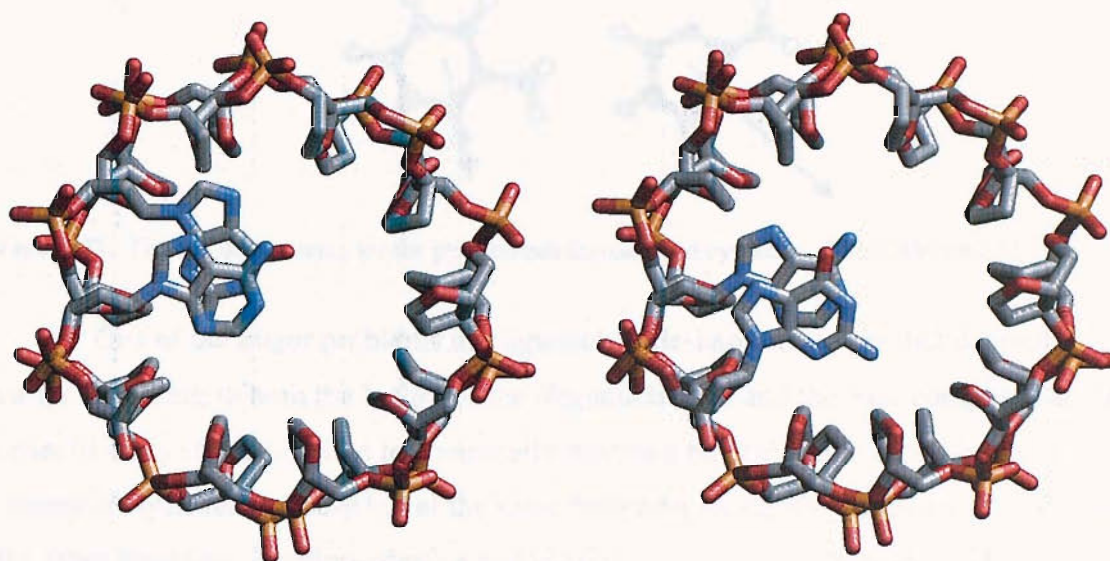
Figure 26. The stacking of an Adenine with a Guanine base above and below in a DNA duplex. The backbone is shown to demonstrate that the view is down the centre of the duplex and as such gives a fair representation of the base stacking of the adenine in this crystal structure.

Since dispersion attractions are effectively dependent only on geometrical overlap increasing this overlap should lead to increased base stacking and therefore stabilise duplexes. By using non-polar groups this will also have the added advantage that it increases the hydrophobicity of the bases. The increased hydrophobicity forces water molecules around the base into a more ordered hydrogen-bonding pattern. Once the base becomes stacked this water is then released into the random bulk of the water, an entropically favourable process.

Another way to increase the stability of duplexes is to introduce or strengthen hydrogen bonds. Hydrogen bonds are electrostatic interactions and as such are governed by the electronic distribution of the bases. It is not just the number of hydrogen bonds that is important but also the strength of the dipole moment of the base. This explains why G.C base pairs are almost twice as stable as A.T base pairs when a G.C base pair only has 50% more hydrogen bonds. The rest of the contributing stabilisation comes from the strong dipole moments in G and C compared with A and T as can be seen in Fig. 27 and 28.
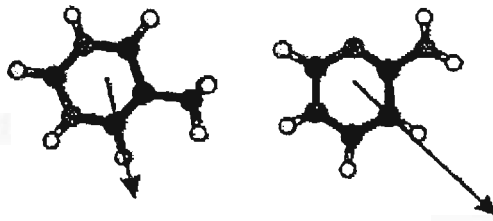
Figure 27. The dipole moments for the pyrimidines thymine and cytosine. [From reference 81.]

One of the major problems in oligonucleotide-based assays is that the probe T_m varies according to both the length of the oligonucleotide and the base composition. In order to study if it is possible to chemically modify a nucleobase so that it binds to the correct complementary base but at the same time alter its stability so that it is as stable as the other base pair, 2-amino-adenine and inosine were used in this work. 2-Amino-adenine, diaminopurine, was used as it can form three hydrogen bonds and may have the same stability as a G.C base pair when it binds to T. Inosine (6-oxopurine) was used as it only forms two hydrogen bonds with C and could potentially reduce the stability to that of an A.T base pair.
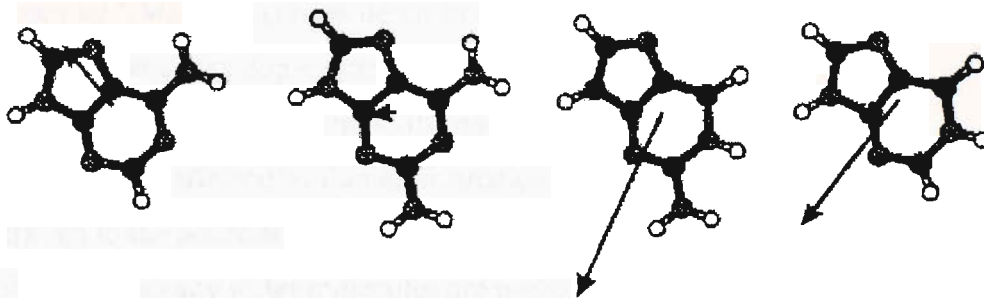
Figure 28. The dipole moments for the purines, from left to right: adenine, 2-amino-adenine, guanine and inosine. [From reference 81.]

One of the drawbacks of using 2-amino-adenine, despite its extra hydrogen bond, is its weak dipole moment as can be seen in Fig. 28. In order to increase its stability 1-aminoprop-2-yne, propargylamine, was added. The alkyne will increase the geometrical overlap between the base it is on and the base of the 5' nucleotide so increasing base stacking interactions. The primary amino group will introduce a salt bridge between itself, which should be ionised at physiological pH, and the non-bridging major groove oxygen on the phosphate of the 5' nucleotide.

1.6.2 The Hydration of DNA.

DNA duplex stability is influenced by the availability of water molecules, known as water activity, and sodium ions. There are many water molecules associated with a DNA duplex. However, the rate at which they exchange with the bulk medium varies a great deal. It has been demonstrated that there are about 17 ordered water molecules per base pair. It is the stability of these 17 water molecules in the duplex structure that makes them significant when considering which waters molecules are likely to contribute to the conformation and stability of DNA structures. The number of water molecules appears to be independent of the type of base pair as guanine and adenine are similarly hydrated as are cytosine and thymine. These ordered water molecules are approximately distributed as follows: five around the bases, four associated with the phosphates and two in the sugar region.[92]

X-ray crystal structures have shown that the water hydrogen bonding to the bases is bound the most specifically and therefore is probably the least exchanged. The least exchanged of all water appears to be that bound in the minor groove of A_nT_n tracts as shown by NMR. It has been determined that four water molecules per base pair will dissociate when the duplex melts into its single strands. These uniquely bound waters are probably those that are bound to the base pairs as the sugar and phosphate associated waters are not effected by duplex formation to the same extent as they are always exposed to the aqueous solvent. The bases in the single strand are exposed to the bulk solvent and so any water molecules are probably in rapid exchange and will have a chemical potential very similar to that of the bulk solvent. This is different to the uniquely bound water molecules, as their chemical potential will be affected by the DNA duplex to which they are bound. This effect is one of the entropic penalties of duplex formation. An important conclusion is that any changes in water activity will alter the conformation of a DNA duplex through the influence of the bound water molecules.[92]

1.6.3 Sodium Ions Associated with DNA

Sodium ions are released during the melting of DNA duplexes. This is because in the duplex form the DNA has a higher charge density than when there are two single strands. The high charge density combined with the rigidity of duplex DNA results in the condensation of ions along the surface of the DNA with approximately 0.75 sodium ions

per phosphate. The lower charge density of the single strand and its increased flexibility means that fewer sodium ions are required to neutralise the charge with approximately 0.38 sodium ions per phosphate. The number of sodium ions bound by the single strand is sequence dependent. This is due to stacking interactions between purines that rigidify the single strand and so increase the charge density resulting in a larger number of sodium ions associating. Any factor that affects the charge density will therefore lead to a change to the numbers shown here for ion release or association. The difference in the numbers leads to a release of approximately 0.37 sodium ions per phosphate.[92]

1.6.4 The Change in Heat Capacity for Duplex Formation.

Any change in the heat capacity (ΔC^{o}_{obs}) is important as it is assumed to be zero during duplex formation. If there is a significant decrease in C^{o}_{obs} then as the temperature increases enthalpy will become more negative. A decrease in C^{o}_{obs} would be observed if more nonpolar surface groups were to become buried in the duplex than polar surface groups. This is due to the ordered network of low-density water that surrounds a hydrophobic surface. A large negative ΔC^{o}_{obs} would also be observed if there were temperature dependent changes in the thermodynamic state of the single strands.
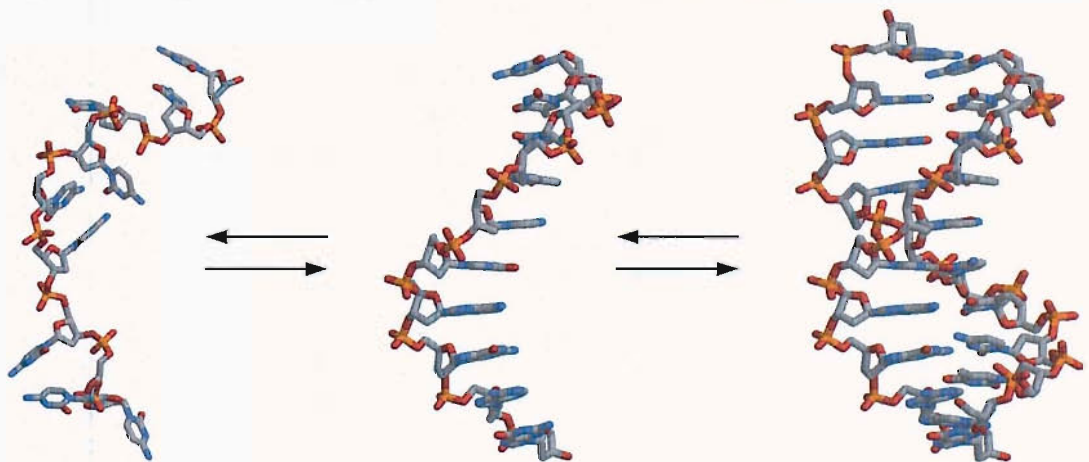
Figure 29. Two of the folding processes that are involved in duplex formation. On the left is a disordered single strand that can become an ordered single strand primarily due to base stacking interactions. Two of these ordered single strands can then dock to form the duplex seen on the right.

The surface area changes during various discrete folding processes have been modelled. In all processes from a disordered ssDNA to a helical ssDNA for both strands and for the docking of those ssDNA oligomers to form dsDNA and for the processes of going from two disordered single strands to a dsDNA the removal of nonpolar groups

from solution is compensated for by the removal of polar groups. The folding processes are shown in Fig. 29. This therefore does not lead to any change in C°_{obs}.

This leads to the conclusion that the ΔC°_{obs} is due to temperature dependent changes in the thermodynamic state of the single strands. These temperature-dependent coupled conformational transitions in the unpaired strands are responsible as predicted and shown experimentally by changing enthalpies at different temperatures. The two discrete processes involved are the unfolding of residual hairpin structures in the single strand (not shown) and the formation of a helix in the single strand (the left hand transition in Fig. 29).[89]

2. BEAD-BASED MULTIPLEX PRIMER EXTENSION

2.1 INTRODUCTION TO ASSAY

The International SNP map working group has identified 1.42 million SNPs. It is estimated 60 K of these fall within exon coding and untranslated regions and 85% of exons are within 5 kb of a SNP. If a whole genome analysis of gene expression is to be conducted these SNPs will need to be genotyped, and the exact number will rise as further SNPs are discovered. However, the total number that will have to be detected will probably fall when a haplotype map, and hence ht-SNPs become available.[16] Although presently available DNA microarrays can be used to analyse a few hundred thousand SNPs they are expensive in capital equipment, operating expenses, sample preparation, and reagent costs.[93] In addition, the results from microarrays can vary because of variations in the efficiency of different hybridisation probes and cross-hybridisation between target sequences and mismatched probes.[93]

Microbeads are particles on which chemical and biological reactions can be conducted. They vary in size 2-300 μm and at the diameter of a typical bacterium (3 μm) 1 ml of a microbead suspension (20% solids) contains around 10^{10} beads.[93] The beads provide a surface onto which oligonucleotides can be synthesised and used as immobilised probes. Since each bead is a discrete unit and one SNP probe can be placed on each bead, each individual bead becomes the platform for a single SNP assay. In this way many millions of SNPs can potentially be detected in one assay. Bead production and oligonucleotide chemistry are both inexpensive and optimised technologies. However, if a bead-supported primer extension assay is to be used, the immobilised oligonucleotides will have to be synthesised from 5' to 3' (opposite to the normal direction). This requires a less well developed chemistry[94] than the traditional 3' to 5' method.

A whole genome amplification system such as the MDA method can be used to generate the target DNA[18] and this needs to be to be combined with random cleavage by restriction endonucleases so that the fragments are small enough to interact specifically and efficiently with the oligonucleotides on the beads. Once the resin bound oligonucleotide and a complementary ssDNA fragment generated by the WGA are present they can hybridise together. The resin bound probe can be designed to bind to the

region one base short of the polymorphic site of an SNP. A single base chain extension (SBCE) reaction[95-97] can then be carried out using a DNA polymerase and four fluorescently labelled dideoxynucleotide triphosphates (FddNTPs), A, G, C and T, each bearing a different fluorophore, to probe the ht-SNP or SNP site, with the inserted fluorescent nucleotide indicating the genotype of the SNP. On a macroscopic level, this would be observed by measuring the fluorescence of the bead on which the reaction has been carried out. In this detection method the oligonucleotides have to be present only on the surface of the resin. This is because the light from the laser cannot penetrate the core of the resin and any fluorophore inside the core is not detected. As with microarrays, however, a significant technical challenge is to match the T_ms of all the probe-target duplexes so that all the reactions can be carried out at one convenient temperature.

In the proposed method (Fig. 30), a fluorescence activated cell sorter/scanner (FACS) can be used to detect the fluorescent beads[40]. This instrument was initially designed for sorting cells and can also be used on fluorescent beads. The particular advantage of this instrument is the speed at which it can sort through the samples (100,000 particles sec^{-1})[93]. At this speed it would take around a minute to assess the full complement of ht-SNPs using 10 replicates for each SNP. The final stage is identification of the sequence on the fluorescent bead. This requires encoding of the bead so that the sequence of the oligonucleotide can be determined and therefore the SNP analysed.

(1)   (2)   (3)
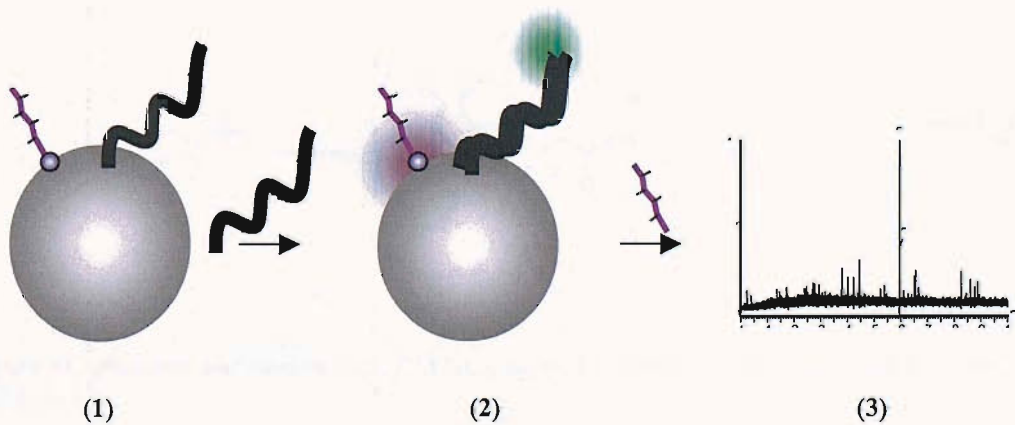
Figure 30. The proposed assay. (1). The resin beads have an oligonucleotide probe on the surface, (solid black line) as well as an encoding moiety (purple) that is peptide-based. Following a whole genome amplification and random endonuclease cleavage the PCR products are mixed with the functionalised resin. (2) If the mixture contains a product that is exactly complementary to the probe sequence on the resin the DNA strands will hybridise. Following hybridisation, a single base extension reaction is carried out (green). As a result the beads become fluorescent. The fluorescent beads are then identified by FACS. (3) The active beads are individually treated to cleave the encoding peptide moiety that is identified by MS.

Direct fluorescent labelling of the beads[39] or addition of fluorescent particles to the surface of the beads during oligonucleotide synthesis can be used for encoding.[93,98,99] Another method involves the synthesis of an encoding chemical moiety on the bead in parallel with oligonucleotide synthesis to record the sequence of the oligonucleotide as it is synthesised[100]. This requires orthogonality between the chemistry of oligonucleotide synthesis and the method of attachment of the encoding moiety. Furthermore, both these chemistries have to be compatible with the properties of the bead. In our work it was envisaged that the encoding moiety would be detected using mass spectrometry.

2.2   BACKGROUND STUDIES

2.2.1   Oligonucleotide Synthesis

In order to assess a range of polymer resin beads for their ability to support oligonucleotide synthesis the resins were functionalised with 12-(4,4'-dimethoxytrityloxy)dodecanoic acid (Fig. 31). A non-cleavable attachment was used as it is only necessary to observe the trityl coupling yields and the oligonucleotide is not required once the synthesis is complete. The attachment method is a one step procedure using readily available starting materials.
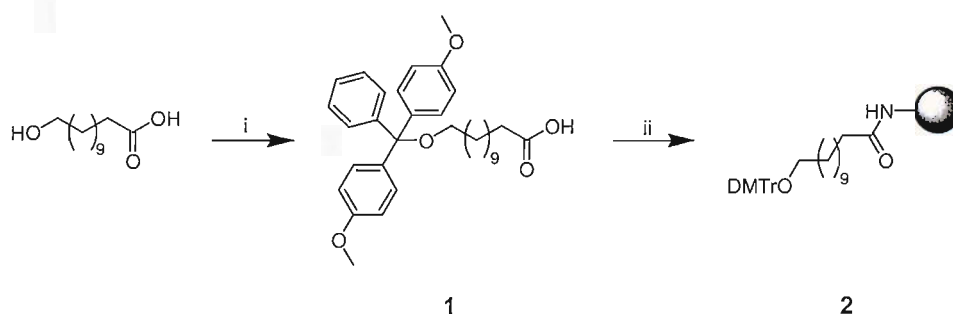
Figure 31. *Reagents and conditions*: i, DMTrCl, py, rt, 18 h, 48%; ii, DIC, HOBt, resin, 1% DIPEA, DCM, rt, 1 h.

Amine-functionalised beads were chosen as they could be coupled with 12-(4,4'-dimethoxytrityloxy)dodecanoic acid. Several types of beads were purchased in order to evaluate a wide range of bead properties to determine which is likely to be the most suitable. The beads selected and their properties are shown in Table 1.

Table 1. A summary of the properties of the beads under investigation.

Resin Code	Resin	Supplier	Polymer Core	Particle Size (μm) Pore size (Å)	Stated Loading (mmolg^{-1})	Functionalisation
R1	HypoGel® 200	Rapp Polymere	PS / 1% DVB	110-150	0.92	Grafted amine functionalised penta-ethylene glycol
R2	HypoGel® 400	Rapp Polymere	PS / 1% DVB	110-150	0.76	Grafted amine functionalised deca-ethylene glycol
R3	ArgoPore	Argonaut technologies	MP PS / DVB	100-250	0.74	Amino-methylated
R4	TentaGel N	Rapp Polymere	PS / PEG	90	0.23	Amine functionalised PEG
R5	Amino PS	Glen Research	PS	–	–	Amino-methylated
R6	30 HL	Nycomed Amersham	–	–	0.16	–
R7	PL-PEGA	Polymer Laboratories	PS / PEG	300-500	0.2	Amine functionalsied PEG
R8	PL-PEGA	Polymer Laboratories	PS / PEG	150-300	0.4	Amine functionalsied PEG
R9	PL-AMS	Polymer Laboratories	PS / 1% DVB	75-150	0.91	Amino-methylated
R10	PL-AMS	Polymer Laboratories	PS / 1% DVB	75-150	1.79	Amino-methylated
R11	PEG-AMPS	Biosearch technologies	PS / 1% DVB	40-75	0.4	Amino-methylated and PEG
R12	Biomac	Biosearch technologies	Polymethacrylate	50-100 (1000 Å)	0.14	Amino substituted diethylene glycol
R13	Biomac	Biosearch technologies	Polymethacrylate	30-60 (1800 Å)	0.15	Amino substituted diethylene glycol
R14	Biomac	Biosearch technologies	Polymethacrylate	50-100 (300 Å)	0.19	Amino substituted diethylene glycol
R15	TentaGel N	Rapp Polymere	PS / PEG	90	0.21	Amine functionalised PEG
R16	PS-PTFE T$_{38}$ + Ext	E.B.Hirschfeld	PTFE	–	0.16	Amino-methyl PS
R17	PS-PTFE T$_{38}$	E.B.Hirschfeld	PTFE	–	0.18	Amino-methyl PS
R18	PS-PTFE T$_{28}$	E.B.Hirschfeld	PTFE	–	0.07	Amino-methyl PS
R19	PS-PTFE T$_{23D}$	E.B.Hirschfeld	PTFE	–	0.16	Amino-methyl PS
R20	PS-PTFE T$_{44}$	E.B.Hirschfeld	PTFE	–	0.08	Amino-methyl PS
R21	PL-Rink	Polymer Laboratories	PL-AMS	75-150	0.7	Fmoc-protected Rink linker
R22	PL-Rink	Polymer Laboratories	PL-AMS	150-300	1	Fmoc-protected Rink linker
R23	CPG	Link technologies	CPG	–	–	Amine

Table 2. The results of oligonucleotide synthesis on various resins defined in Table 1.

Resin Code	Quantative trityl loading [a] / μmolg⁻¹	Average stepwise yield [b] / %	Normalised trityl conductivities			
			10bp	30bp	50bp	80bp
R1	44.6	97.4	0.62	0.15	0.10	0.12
R2	16.2	97.7	0.93	0.39	0.21	0.16
R2	8.6	98.9	0.97	0.78	0.57	0.40
R2	16.4	97.6	0.79	0.38	0.19	0.14
R2	23.9	97.2	0.74	0.31	0.12	0.11
R2	11.9	-	-	-	-	-
R3	38.5	-	-	-	-	-
R4	55.5	98.3	0.88	0.60	-	-
R5	6.8	101.1	1.03	1.40	-	-
R6	13.2	-	-	-	-	-
R7	13.2	95.4	0.70	0.24	-	-
R8	22.1	100.5	1.12	1.17	-	-
R9	2.1	0	0	0	-	-
R10	17.8	94.8	0.49	0.20	-	-
R11	21.3	101.1	1.28	1.40	-	-
R12	6.3	99.4	0.75	0.83	-	-
R13	27.5	96.0	0.61	0.29	-	-
R14	25.2	98.3	0.93	0.60	-	-
R15	35.4	98.1	0.81	0.56	-	-
R16	6.2	99.4	0.96	0.86	0.73	-
R17	56.6	100.0	0.96	0.89	1.02	-
R18	33.6	99.5	1.06	0.92	0.76	-

$$Y_c = Y_s^N$$

$$\log Y_s = \log Y_c / N$$

$$Y_s = 10^{(\log Y_c / N)}$$

Average stepwise yield % = Y_s x 100

Where, $\quad Y_c$ = The cumulative trityl yield
Y_s = The mean stepwise yield
N = The number of steps

[a] The initial loading with 12-(4,4'-dimethoxytrityloxy)dodecanoic acid.

[b] The final normalised cumulative trityl yield is used to calculate the mean stepwise yield as shown by the equations in the insert.

The required bead has to support both oligonucleotide and peptide synthesis and be biocompatible so that the SBCE reaction can be performed. The beads must also be compatible with aqueous buffers for the FACS sorting to be carried out.

DMTr dodecanoic acid was added to the amine-functionalised resins using standard diisopropyl carbodimide (DIC), 1-hydroxy-benzotriazole (HOBt), coupling conditions to produce a loading of around 30 μmolg^{-1}. The loading of the beads, which was determined by a quantitative trityl analysis, is an important factor that needs to be carefully controlled.
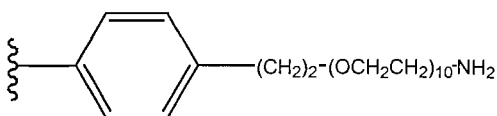
Figure 32. The functional group of the HypoGel® 400 PS resin. The 400 refers to the approximate molecular weight of the ethylene glycols units, 10 are required in HypoGel® 400.

If the resin is functionalised at too high a density then the reaction efficiencies during oligonucleotide synthesis will rapidly fall off due to the sterically encumbered surface. However, opposing that, it is desirable for the surface to have as many oligonucleotides on it as possible so that a large number of fluorophores can be added during the SBCE reaction. This, in turn, will make the fluorescence emission of the beads more intense and therefore make the beads easier to identify by FACS. Beyond a certain level of loading fluorescence quenching is likely to occur.

One of the most promising resins identified according to the trityl yields was HypoGel® 400, a hydrophilic polystyrene gel-type resin. The resin is based on a low cross-linked (1% divinylbenzene, DVB) polystyrene matrix on which decethylene glycol units are grafted to form a high loaded hydrophilic resin. The reactive amino groups are located at the terminus of the glycol spacers (Fig. 32).

Once a potential resin was identified according to the trityl conductivity yields, different loading levels were synthesised to determine the optimal loading for oligonucleotide synthesis. The three HypoGel® 400 resin samples were used in an ABI 394 automated DNA synthesiser and the yields from 80 sequential coupling reactions were monitored using the trityl conductivities. It was found that the lowest

The Effect of Loading on HypoGel® 400
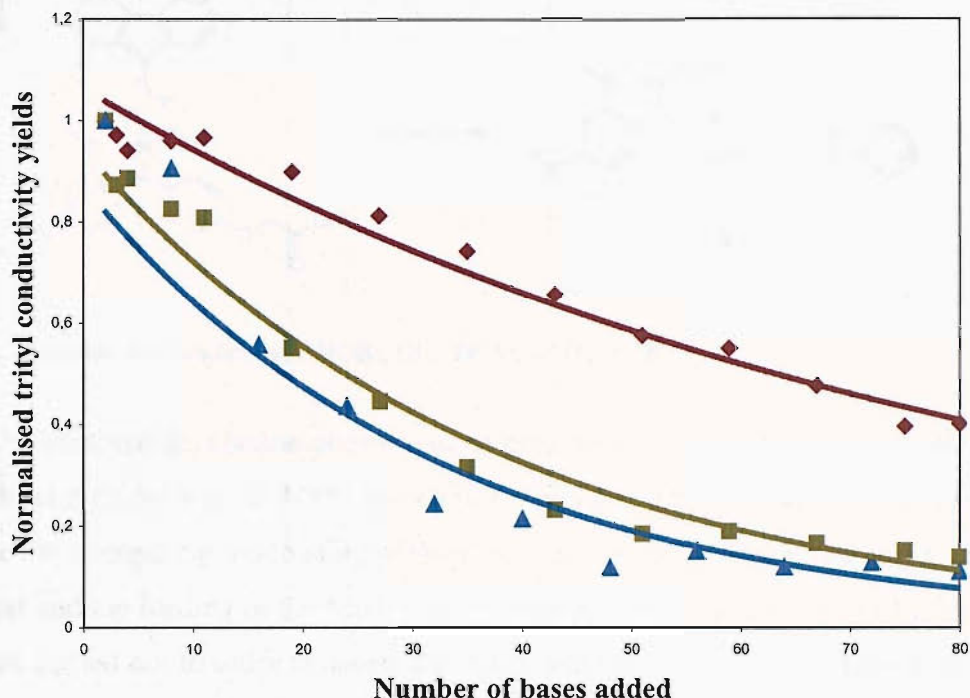
Figure 33. Effect of varying the loading of DMT dodecanoic acid on the efficiency of oligonucleotide synthesis on HypoGel® 400 resin. Brown = loading of 8 μmolg^{-1} (mean stepwise yield of 98.9%), green = 17 μmolg^{-1} (mean stepwise yield of 97.6%) and blue = 27 μmolg^{-1} (mean stepwise yield of 97.3%). The trityl conductivities are normalised against the first coupling yield and give an indication of the coupling efficiencies.

loading (8 μmolg^{-1})gave optimal stepwise trityl yields and a mean stepwise yield of 98.9% (Fig. 33 and Table 2). CPG, the most commonly used support in oligonucleotide synthesis has mean stepwise yields of >99%.[101] At this stage it was not known whether a loading of 8 μmolg^{-1} would be sufficient for the assay with respect to the SBCE reaction and the fluorescence detection. In later studies, it was shown that TentaGel, **R4** and **R15** and some PTFE resins **R16-18** were also promising candidates for oligonucleotide synthesis with trityl yields of >98% and >99.3% respectively.[101,102]

2.2.2 Peptide Synthesis

As peptides were to be used for the encoding moiety, the resins must be able to support peptide chemistry. The same resins that were examined for their ability to support oligonucleotide synthesis were derivatised with the acid cleavable Rink linker.[103] This linker was used as conditions required to cleave it are orthogonal to the basic conditions

used in the Fmoc method of peptide synthesis.[104] The Rink linker was attached to the resins using standard DIC, HOBt chemistry (Fig. 34).
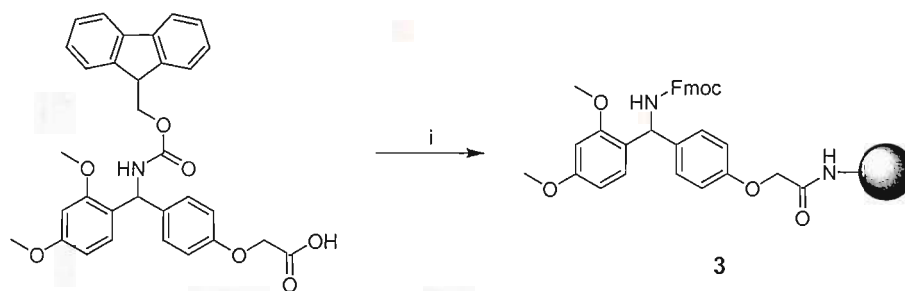
Figure 34. *Reagents and conditions*: i, HOBt, DIC, DCM, DMF, rt, 36 h.

The tetrapeptide, alanine-phenylalanine-glycine-alanine (AFGA), was synthesised using manual methodology and DIC and HOBT coupling. The synthesis efficiency was determined by comparing the loading of the Rink linker as determined by a quantative FMOC test and the loading of the final peptide using several different resins (Table 3). HPLC was carried out In order to assess the purity and integrity of the peptides produced, and the products were analysed by mass spectrometry.

The amount of available peptide is critical for the MS detection. The greater the number of peptide molecules synthesised the greater the MS signal strength. The results from the peptide synthesis experiments showed that **R5, 11, 12, 16, 19-21** gave total yields of less than 10%. Resins **R1, 3** and **10** had linker yields of equal to or less than 30% and the stated loadings of resins **R6, 7** and **12-20** were less than or equal to 0.2 mmol g^{-1}. These unsuitable resins are highlighted in blue in Table 3. The most appropriate resins for use in the assay are highlighted in Table 3 by the underlining of the resin code number.

The origin of the anomalous yield results is unknown. However, incomplete capping of the resin following incomplete amino acid coupling will lead to over estimates for loading values, one inaccurate loading result will effect all results related to it.

63

Table 3. Rink linker loading and final peptide loading for the synthesis of the peptide AFGA on various resins. The yields of the linker loading and both final loadings are given. Values in blue indicate a stated loading of equal to or less than 0.2 mmolg[-1], a linker yield of equal to or less than 30% and a final yield of peptide of less than 10%. The underlined resin codes numbers indicate those resins that do not fall into any of these undesirable criteria. Resin R23 is ruled out as CPG is not suitable for use in a FACS as it is too brittle.

Resin Code No	Stated Loading (mmolg[-1])	Rink linker loading [a]/ mmolg[-1]	Linker Yield / %	Final quantative fmoc loading [b]/ mmolg[-1]	Final Yield as determined by the fmoc loading (%)	Final quantative ninhydrin loading [c]/ mmolg[-1]	Final Yield as determined by the ninhydrin loading (%)
R1	0.92	0.13	14	0.55	423	0.17	131
R2	0.76	0.62	82	0.5	81	0.08	13
R3	0.74	0.16	22	0.39	244	0.11	69
R4	0.23	0.21	91	0.17	81	0.1	48
R5	-	0.16	-	n.d.	-	0.008	5
R6	0.16	0.19	119	n.d.	-	0.06	32
R7	0.2	0.31	155	n.d.	-	0.09	29
R8	0.4	0.27	68	0.51	189	0.08	30
R9	0.91	0.63	69	0.54	86	0.31	49
R10	1.79	0.54	30	0.63	117	0.35	65
R11	0.4	0.88	220	0.28	32	0.05	6
R12	0.15	0.81	540	n.d.	-	0.05	6
R13	0.19	0.22	116	n.d.	-	0.1	45
R14	0.14	0.14	100	n.d.	-	0.04	29
R16	0.16	0.28	175	n.d.	-	0.01	4
R17	0.18	0.14	78	0.26	186	0.05	36
R18	0.07	0.07	100	0.06	86	0.04	57
R19	0.16	0.21	131	n.d.	-	0.005	2
R20	0.08	0.09	113	n.d.	-	0.0007	1
R21	0.7	0.7	100	0.45	64	0.03	7
R22	1	1	100	0.62	62	0.21	21
R23	-	0.09	-	n.d.	-	0.03	33

[a] A quantative fmoc test was used in this case.
[b] Some of the values were to low to be determined (n.d. not determined).
[c] In some examples it proved difficult to remove the adduct from the resin resulting in underestimates of the loading.

2.2.3 Fluorescence Melting

Initial studies by fluorescence and UV-melting showed that the resin bearing a complementary oligonucleotide did not remove the fluorescent oligonucleotide from solution. The oligonucleotides **O1** and **O2** that were used in this assay are shown below. Oligonucleotide **O1** was fluorescently labelled at the 5' end with fluorescein (FAM). Resin **R2** was loaded with 12-(4,4'-dimethoxytrityloxy)dodecanoic acid at 11.9 μmolg^{-1}. The underlined section indicates the regions where the two oligonucleotides are complementary.

O1   5'd-**FAM**GGG<u>CTGAAAAGCTCCCG</u>AT-3'

O2   5'd-CTACGTCTGC<u>CGGGAGCTTTTCAGC</u>TTGAAGCT**R2**

The assay involved adding the fluorescent oligonucleotide to a buffered solution (100 mM NaCl, 1 mM EDTA and 100mM Na-phosphate pH 7.0) at 25°C containing 1-2 mg of the resin and observing any decrease in fluorescence. However, no such decrease was observed. In order to force the fluorescent oligonucleotide to bind to the oligonucleotide on the resin surface it was necessary to concentrate the solution and leave it overnight. After this process had been carried out, the oligonucleotide could be melted off the resin by addition to buffer (100 mM NaCl, 1 mM EDTA and 100mM Na-phosphate pH 7.0) and heating 25-75°C - a process that could be monitored using either fluorescence or UV-light (Fig. 35). These experiments were only carried out on resin **R2** (HypoGel® 400).
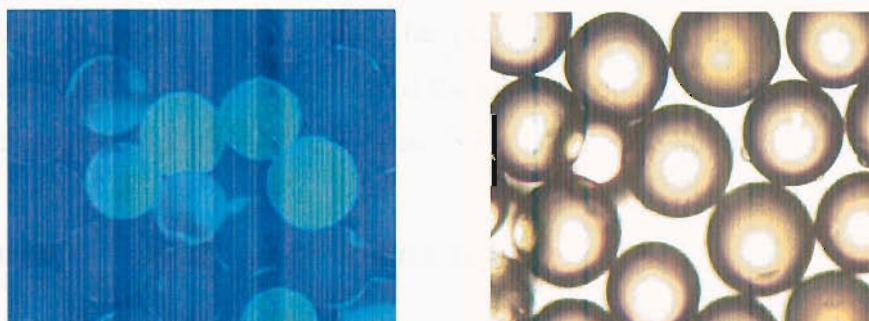
Figure 35. Hybridisation of fluorescent oligonucleotide **O1** to the resin bound oligonucleotide **O2** (a) fluorescence microscopy, (b) white light.

The problem of poor hybridisation kinetics is a fundamental drawback of this assay, so in order to improve hybridisation modified nucleosides were synthesised. It was hoped that introducing positive charges into the oligonucleotides would help overcome any charge repulsion problems between the immobilised oligonucleotide probes and the target DNA. It might also be possible to harmonise the T_ms of the duplexes by altering the thermodynamic stabilities of the different base pairs so that A.T=G.C=T.A=C.G.
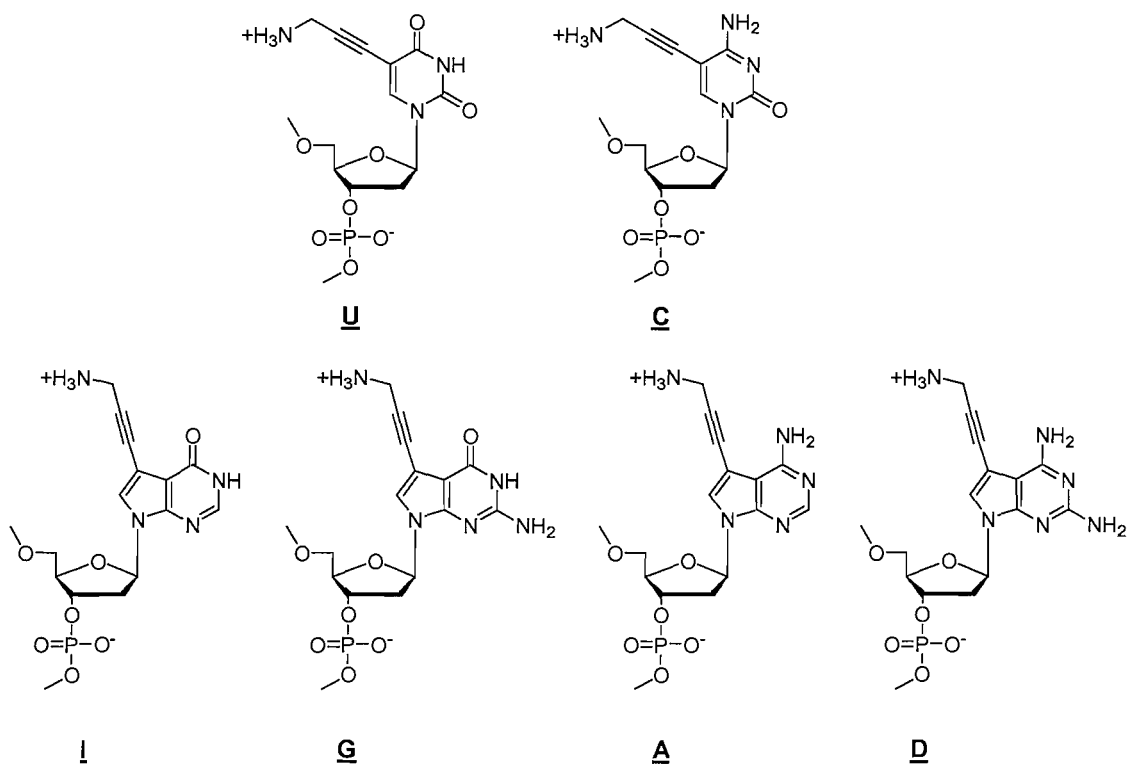
Figure 36. The structures of the incorporated modified nucleotides containing the propargylamino functionality at physiological pH.

The 2-aminoprop-1-yne (propargylamino) functionality was chosen as it is positively charged at physiological pH, and the alkyne moiety has been shown be to be effective in stabilising duplexes.[62] All bases in Fig. 36 were to be modified and evaluated.

2.2.4 Mass Spectrometry Analysis of the Peptide

In order to be able to detect the encoding moiety, in this case a peptide, it is necessary to be able to cleave the peptide from a single resin bead and then detect it.[105,106] To evaluate this possibility, some of the beads used in the peptide synthesis compatibility experiment

were set aside with the peptide still attached. Several **R2** beads were exposed to TFA to determine if the cleaved peptide could be detected by MADLI-TOF MS. The minimum number of beads required to achieve detection was found to be five. Single bead detection at this stage was not achieved. However, it has already been demonstrated that single bead detection is possible[105,106] and can be enhanced by isotopically labelling the analyte.[105]

2.2.5 Conclusions

It has been demonstrated that HypoGel® 400 resin can support both oligonucleotide and peptide synthesis. Before optimisation, the resin could be used to support the synthesis of a sufficient quantity of peptide for five beads to yield a detectable signal. The major drawback with this resin appears to be the inefficient hybridisation of the fluorescent oligonucleotide in solution to the oligonucleotide bound to the surface of the resin.

It was anticipated that the kinetics of binding of a solution phase DNA strand to the oligonucleotide bound to the resin bead surface may be improved by reducing the overall negative charge on the resin-bound oligonucleotide backbone. Therefore it was decided to introduce positive charges onto the bases by the addition of a primary amine[62] and simultaneously introduce a duplex-stabilising propynyl group. In this manner, it might be possible to harmonise the T_ms of duplexes by using a mixture of modified and unmodified bases. This would then make the analysis of the beads far more precise by eliminating false positive and negative results arising from variation in the duplex T_ms. These modified bases might also find applications in many other oligonucleotide-based applications that would benefit from harmonisation of duplex T_ms.

3.1   INTRODUCTION

When added to the C-5 position of dC and dT or the C-7 position of 7-deaza-dA, dG or diaminopurine, the 1-aminoprop-2-yne functional group (Fig. 36) increases the geometrical overlap, and therefore the dispersion interactions between the bases in DNA duplexes. It protrudes into the major groove and may affect hydration by displacing some of the strongly bound water molecules, thereby increasing duplex stability. The alkynyl moiety and methylene group increase hydrophobicity and may cause the release of ordered water molecules into bulk solution. The primary amino group lies in the proximity of the phosphate groups of the DNA duplex and at physiological pH will mask charge repulsion between the two anionic phosphodiester backbones by forming a salt bridge with the non-bridging major groove oxygen on the phosphate of the 5'-nucleotide.[62] (the pK$_a$ of 1-aminoprop-2-yne is 8.15).[62]

3.2   SYNTHESIS OF THE AMINO PROTECTED 1-AMINOPROP-2-YNE

Two different base labile protecting groups were evaluated to protect 1-aminoprop-2-yne. Initially the trifluoroacetamide protecting group was used (Fig. 37).[61]
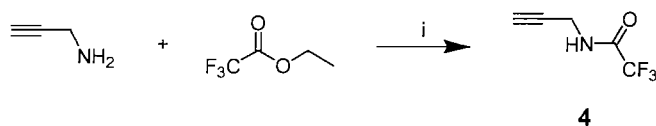
Figure 37. *Reagents and conditions*: i, MeOH, TEA, rt, 16 h, 78%.

However, it became apparent that oligonucleotides synthesised using a TFA protected 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine phosphoramidite gave multiple products on HPLC analysis, and the greater the number of 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine nucleosides added the more complex the HPLC chromatogram. Seela reported a similar problem and interpreted the cause as partial exchange of acetyl for the TFA protecting group during the capping step of oligonucleotide synthesis. The resulting acetyl-protected amine is not hydrolysed during the standard ammonia treatment.[68] However, Seela refers to a study on 5-trifluoracetamidouracil, not 5-[3-(trifluoroacetamido)prop-1-ynyl]-2'-deoxyuridine.[107] Difficulties with TFA-protected 5-

(3-aminoprop-1-ynyl)-dU led Seela to evaluate the phthalimido protecting group that can be removed by treatment with concentrated aqueous ammonia. In accordance with Seela the phthalimido protecting group was adopted. The synthesis of 3-phthalimidoprop-1-yne is shown in Fig. 38.
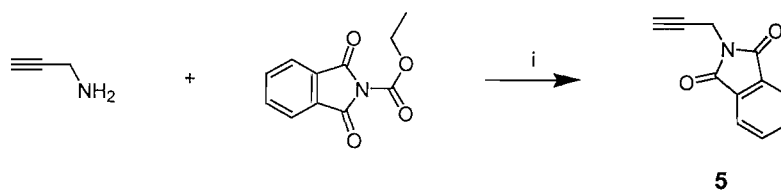
Figure 38. *Reagents and conditions*: i, DCM, TEA, rt, 0.5 h, 77%.

Subsequently, careful mass spectrometric analysis (analysis by Ian van Delft) of the different peaks isolated from the HPLC purification (HPLC analysis by Dr Dorcas Brown) of oligonucleotides containing 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine indicated the addition of acrylonitrile (cyanoethylene-adducts) (Fig. 39). Acrylonitrile is the by-product of the β-elimination of the cyanoethyl protecting groups from the phosphotriesters during the ammonia deprotection at the end of the oligonucleotide synthesis. It was found (by Professor Tom Brown) that addition of phenol as a scavenger to 10% aqueous methylamine (2.5 mg/ml) solves this problem. This method may be generally applicable as this side reaction also occurs with thymidine and uridine. However, it is a much greater problem with propynyl dU derivatives, possibly because of the higher pK$_a$ and nucleophilic character of the imide nitrogen.
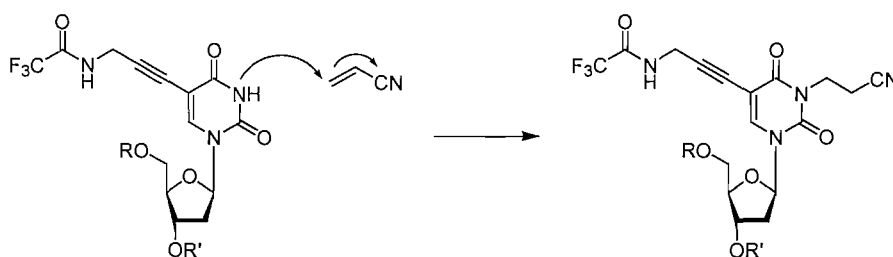
Figure 39. The addition of acrylonitrile to the imide nitrogen in 5-[3-(trifluoroacetamido)prop-1-ynyl]-2'-deoxyuridine.

69

3.3   1-AMINOPROP-2-YNE MODIFIED PYRIMIDINES

3.3.1   5-(3-Aminoprop-1-ynyl)-2'-deoxyuridine

The thermodynamic effects of both C-5 3-aminoprop-1-ynyl and 3- dimethylaminoprop-1-ynyl 2'-deoxyuridine have previously been studied.[62,67,68] The structures of the incorporated bases are shown in Fig. 40. The phosphoramidite monomer was synthesised in four steps from the commercially available 5-iodo-2'-deoxyuridine. This short synthesis means that it is a suitable model for studying the effects of the 1-aminoprop-2-yne modification on duplexes and triplexes.[65]

In order to determine the thermodynamic basis of duplex stability, UV-melting studies were carried out over a range of salt concentrations. The 2- dimethylaminoprop-1-yne dU phosphoramidite (synthesised by Dr Oliver Lack) was studied in parallel.

(a)   (b)
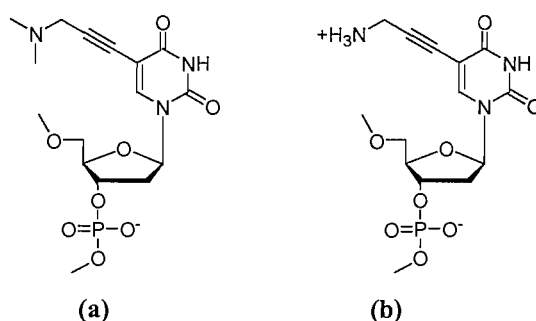
Figure 40. (a) 5-(3-Dimethylaminoprop-1-ynyl)-dU and (b) 5-(3-aminoprop-1-ynyl)-dU.

3.3.2   5-(3-Aminoprop-1-ynyl)-dU Phosphoramidite, 8

5'-O-(4,4'-Dimethoxytrityl)-5-[3-N-(2,2,2-trifluoroacetyl)amino-1-propynyl)-3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidyl)-2'-deoxyuridine was synthesised in four steps (Fig. 41) with the fourth step being the protection of 1-aminoprop-2-yne (Fig. 37). 5-Iodo-2'-deoxyuridine was tritylated using standard procedures to give 5-iodo-5'-O-(4,4'-dimethoyxtrityl)-2'-deoxyuridine, 6. A Sonogashira reaction[53] adapted by Hobbs[54] was then carried out with 2,2,2-trifluoro-N-(prop-2-yne)acetamide to yield nucleoside 7. The 3'-hydroxyl group of the nucleoside was then phosphitylated using standard procedures to produce the phosphoramidite 8.
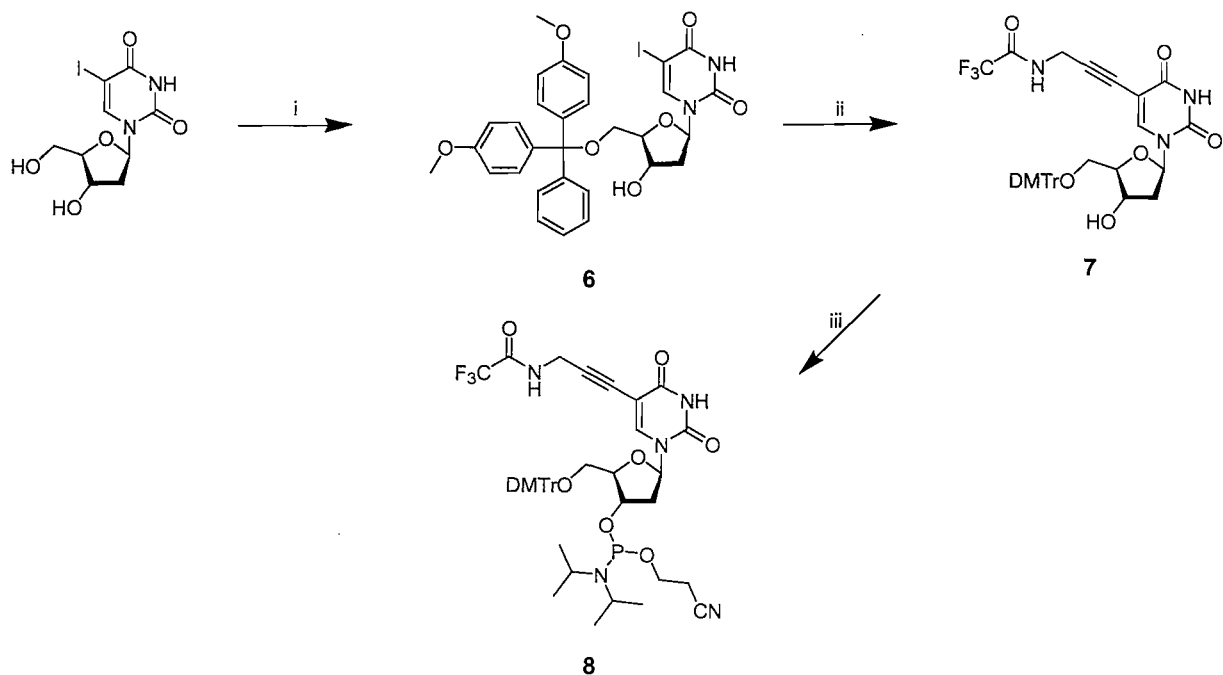
Figure 41. *Reagents and conditions*: i, DMTrCl, py, rt, 12 h, 47%, ii, 2,2,2-trifluoro-*N*-(prop-2-ynyl)acetamide, tetrakis(triphenylphosphine)palladium, CuI, TEA, DMF, rt, 12 h, 79%, iii, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 52%.

An analogue of **8** was also synthesised using phthalimide-protected 1-aminoprop-2-yne (section 3.2).

Both 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine and 2-cyanoethyltetraisopropylphosphorodiamidite / tetrazole diisopropylammonium salt were evaluated as phosphitylating agents for compound **7**. Longer reaction times were required in the diamidite approach and both methods gave a similar yield and levels of purity. It was therefore concluded that in this case the chlorophosphine is the preferred reagent.

The phosphoramidite **8** was then used on an ABI 394 automated DNA synthesiser and an Expedite™ Nucleic Acid Synthesis System for incorporation into three 12mer oligonucleotides. The sequences of the oligonucleotides are shown below where the **U** represents 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine. This modified nucleoside was incorporated in one, two or three non-adjacent positions (**O5-O7**) so that its cumulative effect could be observed. The unmodified sequence and unmodified complement were also synthesised (**O3 & O4**).

3'-dCGATAGATAGAC-5' **O3**

5'-dGCTATCTATCTG-3' **O4**

5'-dGCTATC**U**ATCTG-3' **O5**

5'-dGCTATC<u>U</u>A<u>U</u>CTG-3' **O6**

5'-dGCTA<u>U</u>C<u>U</u>A<u>U</u>CTG-3' **O7**

5-(3-*N,N*-Dimethylaminoprop-1-ynyl)-2'-deoxyuridine **dmU** was also incorporated using the same oligonucleotide sequence in the same positions as <u>U</u> in (**O5-O7**) in order to compare the modifications (**O8-O10**). These oligonucleotides were then used in UV-melting studies.

3.3.3   Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

Ultraviolet melting curves were carried out at 10 different oligonucleotide concentrations from 0.2 μM + 0.2 μM to 8 μM + 8 μM in 100 mM NaCl, 1 mM EDTA and 10 mM Na-phosphate (pH 7.0). Initial data processing was carried out using the melting temperatures (T_ms) generated by the Cary WinUV Thermal melting application software. These were calculated using the Savitzky-Golay smoothing technique[85] to derive an equidistant data set from which the derivative is taken. These T_m values were then used in van't Hoff plots to calculate the thermodynamic parameters for duplex melting (enthalpy, entropy and Gibbs free energy).

This data processing technique failed to give consistent results for enthalpy and entropy, so additional data were collected at higher oligonucleotide concentrations. However this failed to resolve the discrepancies in the data.

T_m values were then determined manually from the raw data by selecting the maximum value in the raw derivative data and correlating this to the temperature at which it occurred. This value was adjusted for the difference in value between the data points either side and the maximum in an attempt to find the true maximum. Each melting curve was also processed using the Meltwin 3.5 software.[108] This was done for two reasons; firstly the program calculates the T_m by a different method, and secondly it also calculates the enthalpy and entropy. Unfortunately these alternative T_m analysis methods when used in a van't Hoff plot also gave inconsistent values for enthalpy and entropy.

In an attempt to resolve this problem, the mean enthalpy, entropy and Gibbs free energy from each individual melting curve for each unique duplex was determined using the Meltwin 3.5 software. Sixty individual melting curves were available as each melting analysis was carried out in triplicate to give three melting and three annealing curves at 10 different salt concentrations. It was important to use both melting and annealing values to

eliminate errors due to the inherent hysteresis caused by delayed heating and cooling in the individual cuvettes. Each melting curve was checked for the presence of a proper upper and lower baseline. Melt curves containing an obviously anomalous data point had the point removed and a value at the mid point of the values immediately above and below the anomalous point inserted. Data collected below 20°C were noisy and were therefore eliminated before processing. The melt curves were processed in this way and the mean was calculated for each duplex. The errors were propagated as shown in the errors section.

The results of this analysis gave rise to logical trends and these values have been used throughout this work.

Table 4. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing <u>U</u> and <u>dmU</u>.

Code Number	Mod. Type	No. of Mod.	$-\Delta H$ kcal mol^{-1}	+/−	$-\Delta S$ cal K^{-1} mol^{-1}	+/−	$-\Delta G_{310}$ kcal mol^{-1}	+/−	$T_{m\,(2\,\mu M)}{}^{a}$	$\Delta T_m{}^{b}$
O4	Non	0	84.8	1.7	243.9	5.7	9.1	0.07	37.6	0.0
O5	<u>U</u>	1	93.3	1.1	267.3	3.4	10.4	0.04	41.6	4.0
O6	<u>U</u>	2	94.6	1.4	270.3	4.3	10.8	0.08	42.5	2.5
O7	<u>U</u>	3	96.5	0.7	274.2	2.2	11.5	0.04	44.6	2.3
O8	<u>dmU</u>	1	91.2	0.2	261.0	0.7	10.2	0.01	41.5	3.8
O9	<u>dmU</u>	2	86.1	0.3	244.3	1.0	10.3	0.01	41.9	2.2
O10	<u>dmU</u>	3	85.0	1.0	240.6	3.2	10.4	0.02	43.3	1.9

[a] $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 μM + 1 μM.

[b] ΔT_m is the difference in melting temperature observed between the unmodified oligonucleotide and the modified oligonucleotide divided by the number of modifications in the oligonucleotide.

These results (Table 4) demonstrate that both 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine and 5-(3-N,N-dimethylaminoprop-1-ynyl)-2'-deoxyuridine stabilise duplexes relative to the unmodified duplex. The stabilising effect is cumulative, but addition of the first modified nucleoside brings about the greatest increase in ΔG_{310} and T_m.

Previous thermodynamic studies of 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine indicated that a 12-mer oligonucleotide incorporating two modifications was stabilised by 3°C or by 0.9 kcal mol^{-1} at 298 K.[68] (100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na-cacodylate (pH 7.0), 5 μM single stranded DNA concentration) Another report on a 12mer duplex showed that one and four modifications gave increases in T_m of 4.5 and

11.9°C respectively.[62] (100 mM NaCl, 0.1 mM EDTA, 10 mM Na-phosphate (pH 7.0), 5 µM single strand concentration). Both reports show that 5-(3-aminoprop-1-ynyl)-dU stabilises DNA duplexes and this is consistent with our findings.

The Fidelity of Base Pairing
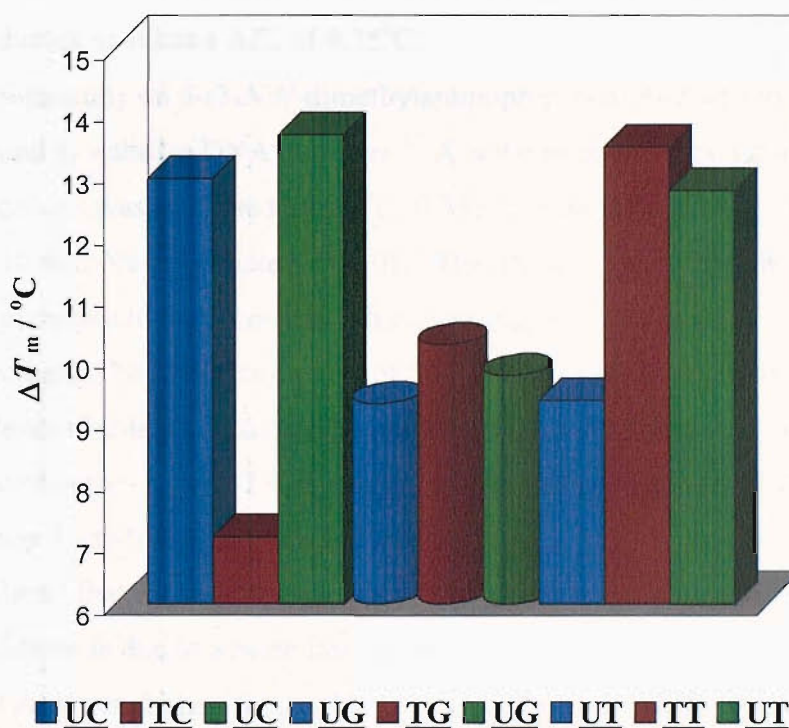
Figure 42. The ΔT_m is the difference in melting temperature between the fully complementary oligonucleotide duplexes and duplexes containing a single mismatch. (the lower the ΔT_m the more stable the mismatch). The blue bars represents the data gathered for this study, the red bars are for unmodified oligonucleotides as specified in the literature and the green bars are for data on singly modified oligonucleotides in the literature.[62] $\underline{U}$ = 5-(3-Aminoprop-1-ynyl)-2'-deoxyuridine

In order to establish that the fidelity of base pairing was not compromised by introduction of the stabilising nucleoside, the singly modified oligonucleotide with 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine was paired against all three possible mismatches (**O11-O13**). This was compared to mismatch data in the literature for both a singly modified oligonucleotide duplex and unmodified duplexes.[62] The results (Fig. 42) are consistent with existing literature.[62]

The lower salt concentration used in the literature accentuates any differences in melting temperature as low salt conditions destabilise the unmodified oligonucleotides more than oligonucleotides containing 1-aminoprop-2-yne modified bases. The results

indicate that the most stable mismatch is **UG** in both the literature case and the one presented here, with the difference in T_m between the complementary modified oligonucleotide and the mismatched oligonucleotide being 9.25°C and 9.7°C. The data appears to indicate that the **UT** mismatch from our study is stabilised with a ΔT_m of only 9.3°C. However, this may be due to the differences in sequence and buffering conditions used in the literature study. These differences can make any comparisons difficult. Despite this, the most stable mismatch should be easily identifiable from a complementary duplex as it has a ΔT_m of 9.25°C.

In a previous study on 5-(3-*N,N*-dimethylaminoprop-1-ynyl)-2'-deoxyuridine this analogue was found to stabilise DNA duplexes.[67] A self-complementary 12-mer duplex with two modifications was stabilised by 0.7°C, 0.35°C per modification (1.0 M NaCl, 1 mM EDTA and 10 mM Na-phosphate (pH 7.0)). This value is lower than the one obtained in this study due to the increased salt concentration.[67] Increased salt concentration decreases the stabilising effect of the modifications in comparison to unmodified duplexes (Table 5). It should be noted that the salt studies were only carried out on the 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine base and not on the 5-(3-*N,N*-dimethylaminoprop-1-ynyl)-2'-deoxyuridine base.

Table 4 shows that the stabilisation from the 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine additions is due to a more favourable enthalpic term. At the same time, entropy becomes increasingly less favourable but not at the same rate at which enthalpy becomes favourable (i.e. ΔG_{310} for duplex formation is more favourable for this analogue than for thymidine). However, with increasing numbers of 5-(3-*N,N*-dimethylaminoprop-1-ynyl)-2'-deoxyuridine additions the enthalpy becomes less favourable while the entropy becomes increasingly favourable. One potential explanation for this could be that the extra steric bulk of the methyl groups in the major groove displaces some of the ordered water molecules.[92] These water molecules are gained with an entropic penalty during duplex formation, their loss due to the steric bulk of the methyl groups could lead to an increasingly favourable entropy of duplex formation with increasing numbers of modified bases.

3.3.4 Salt Dependent Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

The primary amine, which is ionised at physiological pH,[62] in the 1-aminoprop-2-yne modification should interact with the negatively charge backbone of the DNA.

Table 5. The salt dependent enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing U̲.

Code Number	Mod. Type	Salt Concentration (NaCl) / M	No. of Mod.	$-\Delta H$ kcal mol^{-1}	$-\Delta S$ cal K^{-1} mol^{-1}	$-\Delta G_{310}$ kcal mol^{-1}	$T_{m\,(2\,\mu M)}$[a]	ΔT_m[b]	Mean ΔT_m[c]
O2	Non	0	0	72.8	213.3	6.7	27.8	0.0	0.0
O3	U̲	0	1	89.1	265.0	6.9	30.0	2.2	
O4	U̲	0	2	81.6	238.9	7.5	31.6	1.9	2.2
O5	U̲	0	3	84.7	246.3	8.4	34.9	2.4	
O2	Non	0.025	0	94.1	279.5	7.4	32.1	0.0	0.0
O3	U̲	0.025	1	90.1	262.5	8.7	36.1	4.0	
O4	U̲	0.025	2	82.6	237.2	9.0	37.2	2.6	3.1
O5	U̲	0.025	3	89.1	255.1	9.9	40.5	2.8	
O2	Non	0.05	0	93.8	275.4	8.4	35.2	0.0	0.0
O3	U̲	0.05	1	83.2	238.3	9.3	38.5	3.3	
O4	U̲	0.05	2	84.5	241.0	9.7	40.0	2.4	2.8
O5	U̲	0.05	3	91.3	259.6	10.7	43.2	2.7	
O2	Non	0.1	0	86.7	249.5	9.3	38.3	0.0	0.0
O3	U̲	0.1	1	82.5	233.2	10.1	41.6	3.3	
O4	U̲	0.1	2	92.7	264.0	10.8	43.5	2.6	2.8
O5	U̲	0.1	3	100.7	286.4	11.8	46.1	2.6	
O2	Non	0.2	0	92.5	264.2	10.5	42.3	0.0	0.0
O3	U̲	0.2	1	87.4	246.1	11.1	44.7	2.3	
O4	U̲	0.2	2	96.7	274.3	11.7	46.0	1.8	2.1
O5	U̲	0.2	3	113.9	325.3	13.0	48.5	2.1	
O2	Non	1	0	93.0	261.2	12.0	47.6	0.0	0.0
O3	U̲	1	1	84.7	234.2	12.1	48.9	1.3	
O4	U̲	1	2	96.9	271.2	12.8	49.8	1.1	1.2
O5	U̲	1	3	98.4	274.8	13.2	50.9	1.1	

[a] $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 μM + 1 μM.

[b] ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

[c] The mean ΔT_m is the average increase in melting temperature per modified base at a particular salt concentration.

If this electrostatic component of stabilisation is present then oligonucleotide duplexes containing the modified base 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine should be less sensitive to salt concentration. The greater the number of modified bases the less sensitive the duplex.

Table 5 shows that over the range of salt concentration used (0-1 M) that the unmodified duplexes were more sensitive than the modified duplexes. The unmodified duplexes varied in T_m by 19.79°C, singly modified duplexes by 18.94°C, doubly modified duplexes by 18.19°C and the triply modified duplexes by 16.05°C. These data are consistent with data found in the literature.[62]

In a study where the salt concentration was varied from 50-500 mM, the unmodified duplex showed a variation of 13.6°C while a singly modified duplex had a variation of 11.5°C and a quadruply modified duplex had a variation of 7.8°C.[62]

3.3.5 5-(3-Aminoprop-1-ynyl)-2'-deoxycytidine

Initial studies on analogues similar to this compound and the other three 1-aminoprop-2-yne modified natural bases were to synthesise fluorescent dideoxynucleotide triphosphates (FddNTPs) for use in fluorescent gel-based sequencing.[109] A subsequent report has appeared in the literature with a synthesis and T_m data on 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine.[68] In this thermodynamic study two oligonucleotides were synthesised containing two 5-(3-aminoprop-1-ynyl)-2'-deoxycytidines.[68] The novel 5-(3-N,N-dimethylaminoprop-1-ynyl)-2'-deoxycytidine (synthesised by Sunil Vardia) was studied in parallel. Salt dependent studies were carried out to determine the electrostatic effect of the 1-aminoprop-2-yne amine. The incorporated structures of these bases are shown in Fig. 43.
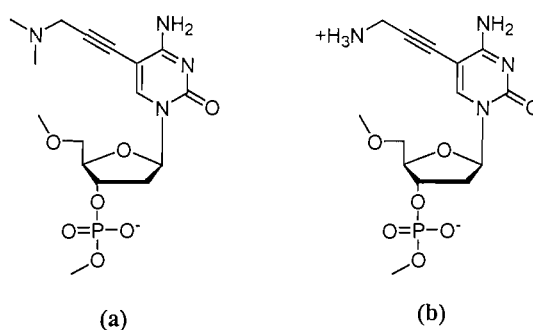
Figure 43. (a) 5-(3-N,N-Dimethylaminoprop-1-ynyl)-dC and (b) 5-(3-aminoprop-1-ynyl)-dC.

3.3.6 5-(3-Aminoprop-1-ynyl)-dC Phosphoramidite, 12 & 15
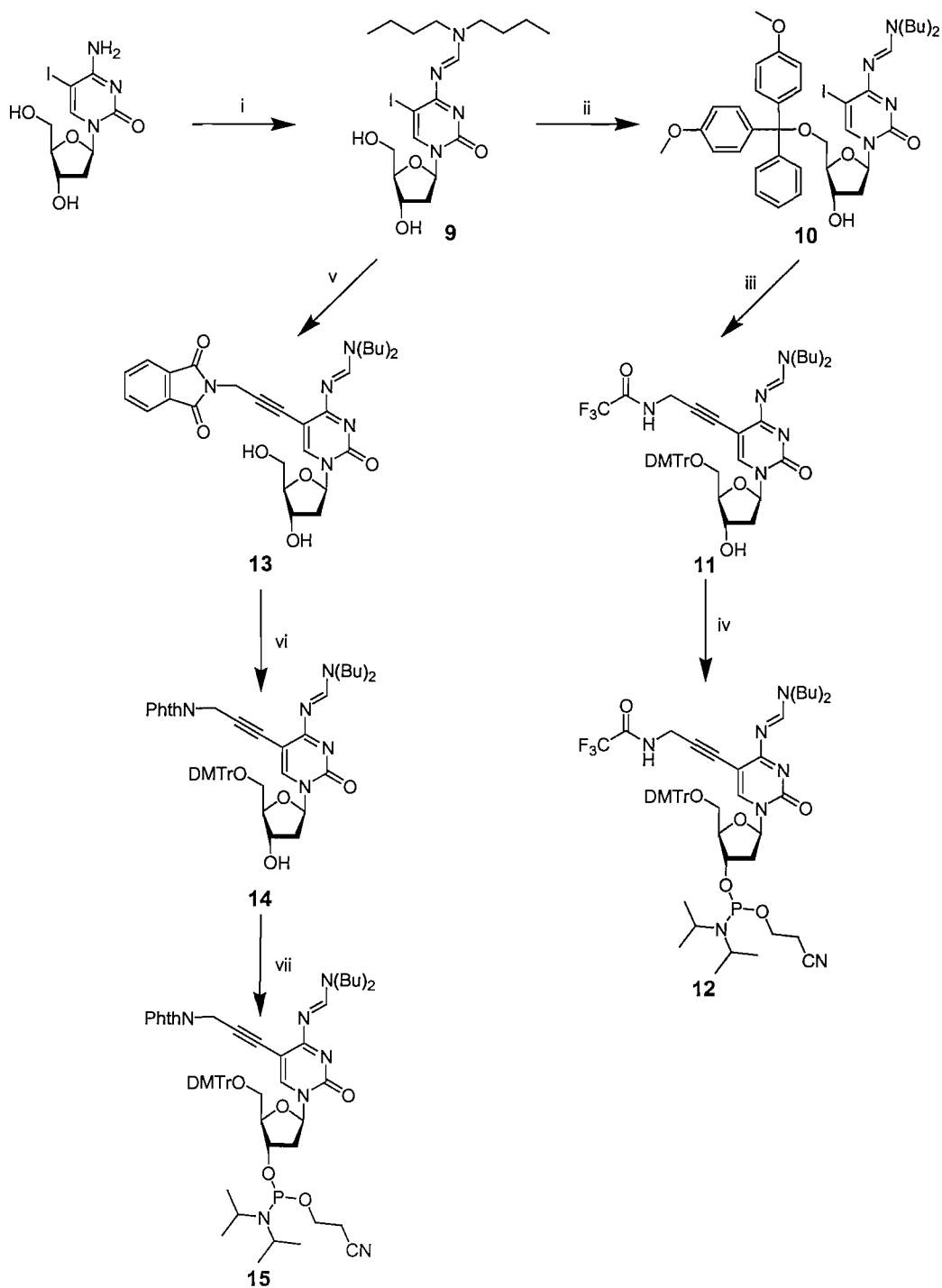
Figure 44. *Reagents and conditions* i, *N,N*-dimethylformamide diethyl acetal, DMF, rt, 16 h, 62% ii DMTrCl, py, rt, 16 h, 54%, iii, 2,2,2-trifluoro-*N*-(prop-1-ynyl)acetamide, tetrakis(triphenylphosphine) palladium, CuI, TEA, DMF, rt, 12 h, 83%, iv, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 58%, v, 3-phthalimido-1-propyne, tetrakis(triphenylphosphine) palladium, CuI, TEA, DMF, rt, 5 h, 84%, vi, DMTrCl, py, rt, 12 h, 62%, vii, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 79%

In this five step synthesis, shown in Fig. 44, the initial step was to protect the exocyclic amine of 5-iodo-2'-deoxycytidine with N,N-di-n-butylformamidine dimethyl acetal. The protecting group was synthesised from di-n-butylamine and N,N-dimethlyformamide dimethyl acetal and was used crude after distillation.[110] The 5'-hydroxyl of N4-[(di-n-butylamino)methylidene]-5-iodo-2'-deoxycytidine, **9**, was tritylated using standard conditions to yield **10**.[9] This was followed by a Hobbs modified Sonogashira reaction to introduce 2,2,2-trifluoro-N-(prop-2-ynyl)acetamide using a standard protocol.[54] The final step was to phosphitylate at the 3'-hydroxyl to yield 5-(3-amino-1-propynyl)-dC phosphoramidite **12**. The phthalimide protected 5-(3-amino-1-propynyl)-dC analogue was also synthesised from **9** *via* **13** and **14** to generate the phosphoramidite **15**.

The phosphoramidites **12** and **15** were then used on an ABI 394 automated DNA synthesiser and an Expedite™ Nucleic Acid Synthesis System for incorporation into three 12mer oligonucleotides. The sequences of the oligonucleotides are shown below where the **C** represents 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine. This modified nucleoside was incorporated in one, two or three non-adjacent positions (**O14-O16**) so that its cumulative effect could be observed. The unmodified sequence and unmodified complement were also synthesised (**O3 & O4**).

<div align="center">

5'-dGCTAT<u>C</u>TATCTG-3' **O14**

5'-dGCTAT<u>C</u>TAT<u>C</u>TG-3' **O15**

5'-dG<u>C</u>TAT<u>C</u>TAT<u>C</u>TG-3' **O16**

</div>

5-(3-N,N-Dimethylaminoprop-1-ynyl)-2'-deoxycytidine **dmC** was also incorporated into the same oligonucleotide sequence (**O17-O19**) in the same position as **C** in (**O14-O16**) in order to compare the modifications. These oligonucleotides were then used in UV-melting studies.

To calculate the absorption coefficient at 260 nm of oligonucleotides that incorporate either **12** or **15**, the absorption coefficient of the nucleoside was determined $(5000 \text{ mol}^{-1}\text{cm}^{-1})$. As protecting groups will interfere with this, they must first be removed. The phthalimide group was removed by 10% methylamine in water (Fig. 45). The absorption coefficient of the 5-(3-N,N-dimethylaminoprop-1-ynyl)-2'-deoxycytidine nucleoside was also determined $(5900 \text{ mol}^{-1}\text{cm}^{-1})$.
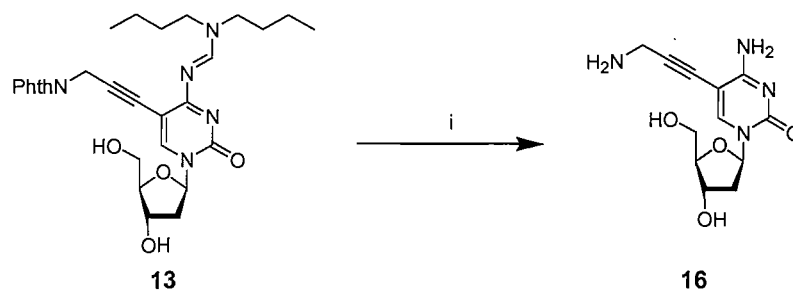
Figure 45. *Reagents and conditions* i, methylamine, water, rt, 24 h, 39%

3.3.7 Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

Table 6. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing $\underline{C}$ and $\underline{dmC}$.

Code Number	Mod. Type	No. of Mod.	$-\Delta H$ kcal mol^{-1}	+/-	$-\Delta S$ cal K^{-1} mol^{-1}	+/-	$-\Delta G_{310}$ kcal mol^{-1}	+/-	$T_{m\,(2\,\mu M)}{}^{a}$	$\Delta T_m{}^{b}$
O4	Non	0	84.8	1.7	243.9	5.7	9.11	0.07	37.6	0.0
O14	$\underline{C}$	1	89.3	0.7	255.4	2.1	10.08	0.03	40.8	3.2
O15	$\underline{C}$	2	91.0	0.7	259.0	2.1	10.69	0.03	43.0	2.7
O16	$\underline{C}$	3	91.8	0.9	258.3	2.9	11.65	0.04	46.7	3.0
O17	$\underline{dmC}$	1	95.0	1.1	273.2	3.5	10.24	0.03	43.4	5.8
O18	$\underline{dmC}$	2	91.5	1.2	259.2	3.7	11.14	0.06	44.2	3.3
O19	$\underline{dmC}$	3	89.3	1.3	251.8	3.9	11.22	0.05	46.7	3.0

[a] $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 µM + 1 µM.
[b] ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

Table 6 reveals that both the modified cytidine analogues are stabilising relative to the unmodified oligonucleotide. In a literature report, a doubly modified 12mer oligonucleotide increased the melting temperature of the duplex by 6°C, 3°C per modification and a quadruply modified 12mer by 9°C, 2.3°C per modification.[68] (5 µM + 5 µM single stranded oligonucleotide in 100 mM NaCl, 10 mM MgCl$_2$ and 10 mM Na-cacodylate (pH 7.0)).

Base-on-base increases in T_m here are around 3°C per modification. This is different to the uridine analogue that showed a decline in influence with each additional base.

The thermodynamic analysis here shows an increasingly favourable enthalpy with the addition of more modified bases whereas the literature results indicate an increasingly favourable entropic term.

The 5-(3-*N*,*N*-dimethylaminoprop-1-ynyl)-2'-deoxycytidine base is stabilising with per modification increases in T_m of 5.8, 3.3 and 3.0°C. In this case, there is a reducing influence for each additional modified base. As with the equivalent uridine analogue after the addition of the first base subsequent additions are primarily stabilised by more favourable entropy.

3.3.8 Salt Dependent Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

Oligonucleotide duplexes containing 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine should not be as thermodynamically influenced by salt concentration as an unmodified oligonucleotide duplex. The results in Table 7 reveal that the T_m of the unmodified duplex varies by 19.8°C over the six salt concentrations (0-1 M).

The singly modified duplex varies by 19.5°C the doubly by 19.4°C and the triply by 17.3°C. These ranges are greater than those for oligonucleotides containing 5-(3-aminoprop-1-ynyl)-2'-deoxyuridines, this indicates that there is less of an electrostatic effect with the 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine bases.

The reduced electrostatic effect can be seen in the ΔT_m values, as, at low salt concentrations' the 5-(3-aminoprop-1-ynyl)-2'-deoxyuridines have more of a stabilising effect. However, at high concentrations of salt, the ΔT_m values demonstrate that the 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine is more stabilising. This therefore indicates the 5-(3-aminoprop-1-ynyl)-2'-deoxycytidines gain more of their stability from intermolecular forces other than the electrostatic bridge formed between the 1-aminoprop-2-yne and the phosphates. Another observation shows that the addition of new bases can lead to greater increases in stability. Table 7 indicates that the inclusion of three modifications in the oligonucleotide is the most stabilising format when dividing the thermodynamic gains between the three bases. The effect of two additions is mixed but generally the least stable configuration. This may indicate some form of co-operativity between the bases with the triply modified oligonucleotide.

Table 7. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing $\underline{C}$.

Code Number	Mod. Type	Salt Concentration (NaCl) / M	No. of Mod.	$-\Delta H$ kcal mol^{-1}	$-\Delta S$ cal K^{-1} mol^{-1}	$-\Delta G_{310}$ kcal mol^{-1}	$T_{m\,(2\,\mu M)}$[a]	ΔT_m[b]	Mean ΔT_m[c]
O4	Non	0	0	72.8	213.3	6.68	27.8	0.0	0.0
O14	$\underline{C}$	0	1	88.0	262.2	6.73	29.4	1.6	
O15	$\underline{C}$	0	2	86.2	254.3	7.33	31.3	1.8	2.0
O16	$\underline{C}$	0	3	85.1	246.8	8.54	35.5	2.6	
O4	Non	0.025	0	94.1	279.5	7.43	32.1	0.0	0.0
O14	$\underline{C}$	0.025	1	88.3	257.6	8.40	35.1	3.0	
O15	$\underline{C}$	0.025	2	80.7	231.3	8.94	37.0	2.5	2.9
O16	$\underline{C}$	0.025	3	83.8	237.3	10.16	41.6	3.2	
O4	Non	0.05	0	93.8	275.4	8.39	35.2	0.0	0.0
O14	$\underline{C}$	0.05	1	81.6	233.5	9.17	37.9	2.7	
O15	$\underline{C}$	0.05	2	85.8	245.0	9.82	40.2	2.5	2.7
O16	$\underline{C}$	0.05	3	88.4	249.8	10.94	44.2	3.0	
O4	Non	0.1	0	86.7	249.5	9.31	38.3	0.0	0.0
O14	$\underline{C}$	0.1	1	83.3	236.1	10.04	41.1	2.8	
O15	$\underline{C}$	0.1	2	88.9	252.0	10.73	43.4	2.5	2.7
O16	$\underline{C}$	0.1	3	91.4	256.5	11.79	47.0	2.9	
O4	Non	0.2	0	92.5	264.2	10.50	42.3	0.0	0.0
O14	$\underline{C}$	0.2	1	90.1	254.8	11.06	44.5	2.2	
O15	$\underline{C}$	0.2	2	92.0	259.3	11.61	46.3	2.0	2.2
O16	$\underline{C}$	0.2	3	94.9	265.2	12.66	49.7	2.4	
O4	Non	1	0	93.0	261.2	12.02	47.6	0.0	0.0
O14	$\underline{C}$	1	1	88.7	246.5	12.22	48.9	1.3	
O15	$\underline{C}$	1	2	92.0	255.2	12.83	50.7	1.5	1.5
O16	$\underline{C}$	1	3	91.2	251.0	13.36	52.8	1.7	

[a] $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 μM + 1 μM.

[b] ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

[c] The mean ΔT_m is the average increase in melting temperature per modified base at a particular salt concentration.

The third addition could increase stability more than the other two additions as it can stack with a purine, with superior geometrical overlap, and a pyrimidine whereas the first two additions are stacked between two pyrimidines. The last observation reveals the greatest difference in T_m between the modified and unmodified oligonucleotides is at 0.025 M NaCl. For salt concentrations lower than this, 0 M NaCl, the low sodium activity affects the modified oligonucleotides more than the unmodified oligonucleotides so narrowing the ΔT_m more than at any other salt concentration apart from at 1 M NaCl.

3.4   1-AMINOPROP-2-YNE MODIFIED PURINES

3.4.1   1-Chloro-2-deoxy-3,5-di-*O-p*-toluoyl-α-*erythro*-pentofuranose, 19

The lack of commercially available precursors meant that the purine nucleosides had to be synthesised. The sugar component was synthesised in three steps from 2'-deoxyribose.[111,112] The standard Hoffer[111] procedure for this important sugar has been superseded by the more synthetically convenient synthesis produced by Lhomme *et al.*[112] The Lhomme synthesis followed is illustrated in Fig. 46. The sugar is first methylated at the anomeric carbon using acidified methanol to give an anomeric mixture of the methyl pentofuranosides **17** in quantitative yield. The 5' and 3' hydroxyls of the glycoside were then protected with toluoyl chloride in pyridine to give **18**. The stereoselective chlorination was carried out using an HCl saturated solution generated *in situ* by the addition of acetyl chloride to a mixture of water and acetic acid to give **19**.
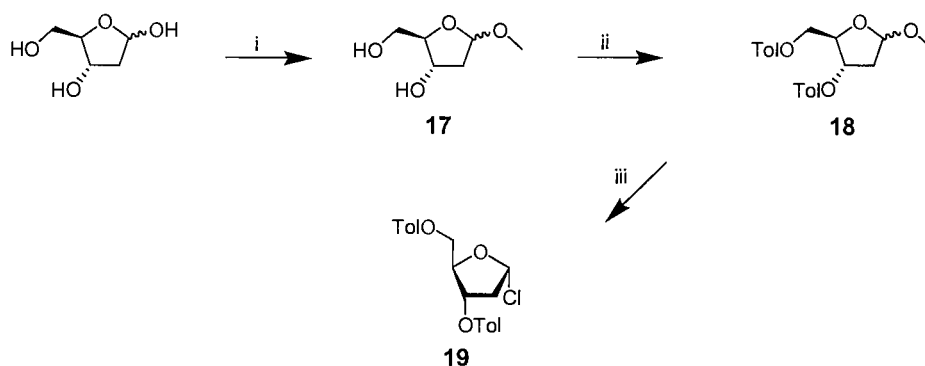
Figure 46. *Reagents and conditions* i, acetyl chloride, MeOH, rt, 2 h, 98%, ii, toluoyl chloride, py, 0°C - rt, 12 h, 54%, iii, acetyl chloride, acetic acid, 0°C - rt, 0.5 h, 83%

3.4.1   7-(3-Aminoprop-1-ynyl)-7-deaza-2'-deoxy-adenine

The synthesis and thermodynamic effects of this compound have already been reported.[68,77,78] In the most comprehensive study it was found the base, **A**, shown in Fig. 47 stabilised duplexes by 1°C per modification.[77] The phosphoramidites **34** and **35** were synthesised in 17 or 16 steps respectively depending on the protection of the exocyclic amine at the C-6 position. This convergent synthesis requires 6 steps for the heterocyclic

base 25, 3 steps for the sugar 19, 1 step for the alkyne 5 and 1 for the formamidine to yield 34 and 6 steps to assemble and convert 25, 19 and 5 into the phosphoramidites.

Oligonucleotides incorporating 34, 35 were synthesised. UV-melting was used to examine the thermodynamic effect of incorporated 34 and 35 on duplex association and disassociation relative to an unmodified duplex of identical sequence.
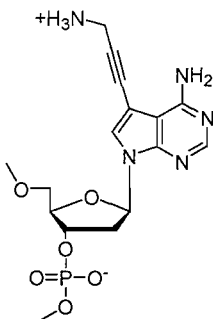
Figure 47. 7-(3-Aminoprop-1-ynyl)-7-deaza-2'-deoxyadenosine, <u>A</u>.

3.4.2   7-(3-Aminoprop-1-ynyl)-7-deaza-dA Phosphoramidite, 34 & 35

The purine numbering convention will be used throughout the discussion of the pyrollo[2,3-d]pyrimidines. The first five steps of this synthesis were originally reported by Davol.[113] The scheme followed can be seen in Fig. 48. In the first step potassium carbonate deprotonates between the nitrile and the ester to form an enolate. The enolate then attacks and displaces the iodine that exchanged with the bromine of bromoacetaldehyde to form a carbon-carbon bond and 20. The reaction is activated with a catalytic quantity of sodium iodide. The iodine exchanges with the bromine to form sodium bromide and hence activates the acetal by forming iodoacetaldehyde. The cyclisation is carried out with thiourea in the presence of sodium ethoxide that is present in order to remove the protons generated in the reaction so preventing the deprotection of the acetal to yield 21.[113,114] The thiol is removed by reduction with Raney® nickel to give 22 and a second cyclisation is carried out, this time in an acidic solution, to form the pyrrolo[2,3-d]pyrimidine ring system or 7-deaza-purine as seen in 23. These reactions were carried out in reverse order to the ones found in the literature.[113,114] Chlorination at the six position is achieved using phosphorous oxychloride to give 24[113] and then iodination at the seven position with N-iodosuccinimide gives 25.[75,115] It has also been reported that this reaction can be carried out using iodine monochloride after the glycosylation.[74] The base was then coupled to the sugar moiety, 19, in a stereoselective

manner with potassium hydroxide that deprotonates the base at the *N*-9-position. The TDA-1 stabilises this anion in solution that then act as a nucleophile attacking stereospecifically from above the plane of the 2'-deoxyribose ring so producing only the β-anomer, **27**.[116,117] A sodium salt glycosylation method is also available, that gives similar yields.[118,119]
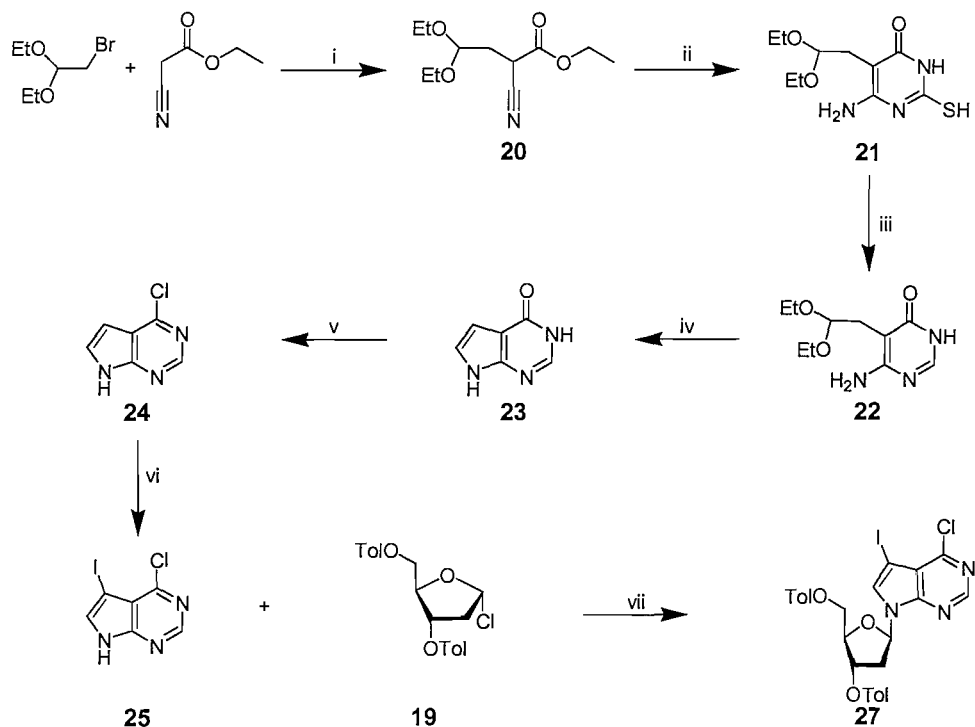
Figure 48. *Reagents and conditions* i, K_2CO_3, NaI, ethyl cyanoacetate, 150°C, 6 h, 42%, ii, thiourea, NaOEt, EtOH, reflux, 4 h, 69%, iii, Raney® 2800 nickel, H_2O, reflux, 1 h, 76%, iv, HCl(aq), H_2O, rt, 6 h, 93%, v, $POCl_3$, reflux, 1 h, 67%, vi, NIS, DMF, rt, 24 h, 96%, vii, KOH, TDA-1, MeCN, rt, 1 h, 77%

The final transformations are shown in Fig. 49. A Parr general-purpose acid digestion bomb (125 ml) was used to carry out the simultaneous deprotection of the toluoyl protecting groups and the aromatic nucleophilic displacement of the chloride by ammonia.[74] Another method is available.[117] The phthalimide protected 1-aminoprop-2-yne group was then added in high yield at the 7-position of the pyrrolo[2,3-*d*]pyrimidine using a Sonogashira reaction adapted by Hobbs.[54] A protocol using the alkyne, 2,2,2-trifluoro-*N*-(prop-2-ynyl)acetamide, can be found in the literature.[77] The exocyclic amine was then protected using *N*,*N*-di-*n*-butylformamidine dimethyl acetal in DMF[110] although the reaction can also be carried out in methanol.[77,120] The 5'-primary alcohol was then protected with dimethoxytrityl chloride using standard procedures to give **32** and then the 3'-secondary alcohol was phosphitylated to yield the target phosphoramidite, **34**. The

phosphitylation can be carried out in DCM as described in the experimental section here or THF as described elsewhere.[77]
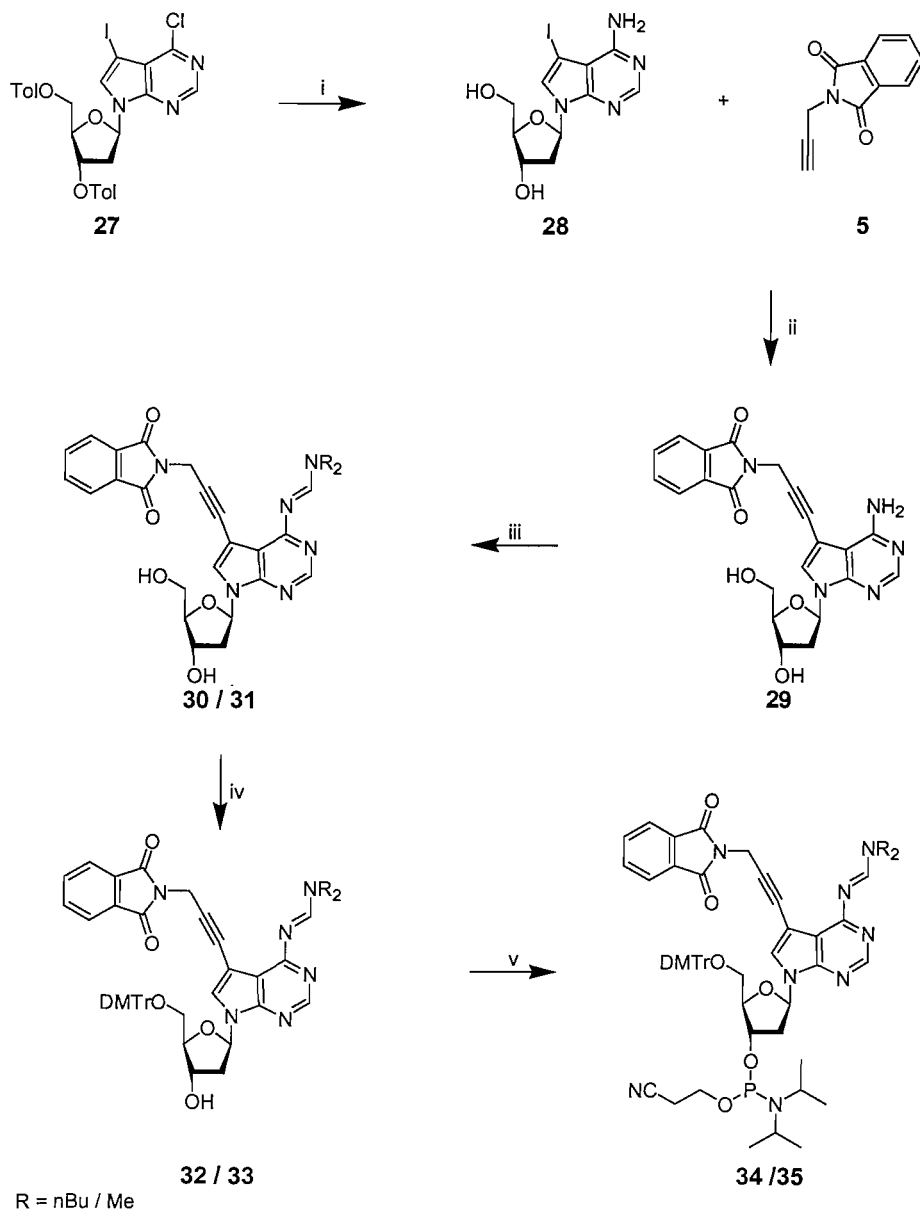
Figure 49. *Reagents and conditions* i, NH$_3$, MeOH, 150°C, 24 h, 54%, ii, tetrakis(triphenylphosphine) palladium, CuI, TEA, DMF, rt, 12 h, 94%, iii, **30**, *N,N*-di-*n*-butylformamide dimethyl acetal, DMF, rt, 12 h, 81%, **31**, *N,N*-dimethylformamide dimethyl acetal, DMF, 40°C, 2 h, 76 %, iv, **32**, DMTrCl, py, rt, 16 h, 38 %, **33**, DMTrCl, py, rt, 4 h, 54 %, v, **34**, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 72%, **35**, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 61%

The exocyclic amine was also protected using *N,N*-dimethylformamidine dimethyl acetal so that both protection strategies could be tested during oligonucleotide deprotection at the end of the oligonucleotide synthesis. This resulted in phosphoramidite **35** *via* the intermediates **31** and **33.** There was no detectable difference in the protection strategies as shown by HPLC.

The phosphoramidites **34** and **35** were incorporated into oligonucleotides using an Applied Biosystems 394 automated DNA synthesiser. The cleaved and deprotected oligonucleotides were purified by HPLC.

The two 12mer oligonucleotides contained one or two non-adjacent modifications (**O22** & **O23**). These oligonucleotides are complementary to **O3**. Due to the constraints of the sequence only one or two modifications could be incorporated. Included in Table 8 are the results for the unmodified oligonucleotides. These oligonucleotides have a G.C base pair, highlighted in blue, incorporated at the same position as the **A**.T base pairs in the modified oligonucleotides. The unmodified oligonucleotides aided in the determination of the stability of **A**.T *versus* G.C for duplex T_m harmonisation.

<div align="center">

5'-dGCTATCT**A**TCTG-3' **O22**

5'-dGCT**A**TCT**A**TCTG-3' **O23**

5'-dGCTATCTGTCTG-3' **O24**

5'-dGCTGTCTGTCTG-3' **O25**

5'-dCAGACAGATAGC-3' **O26**

5'-dCAGACAGACAGC-3' **O27**

</div>

The molar absorption coefficient was determined. This requires the nucleoside to be free of any protecting groups that will alter the absorption spectrum at 260 nm. These protecting groups are removed using 10% methylamine in water and this was done with **29** to yield the free nucleoside **36** (Fig. 50).

The value obtained, 1700 moles^{-1}cm^{-1}, varied significantly from the literature value (6100 moles^{-1}cm^{-1}).[77] A Job plot was carried out to determine if the stoichiometry of binding was 1:1 when using the value determined here.

29   **36**
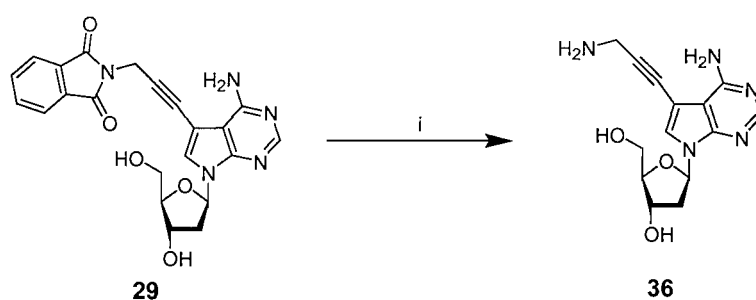
Figure 50. *Reagents and conditions* i, Methylamine, water, rt, 24 h, 43%

Stoichiometry of binding
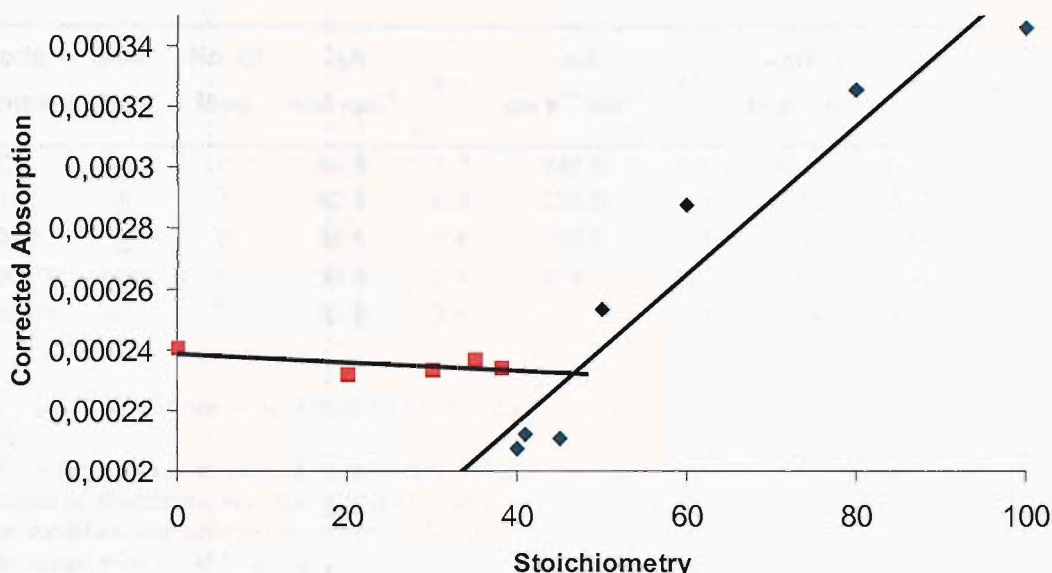
Figure 51. The Job plot was carried out with the bis-<u>A</u>-modified oligonucleotide, O23, and its unmodified complement. The absorption readings are corrected for the absorption coefficients of the single strands. A stoichiometry of 0 or 100 is equal to 0:1 or 1:0 and 50 to 1:1. The data set for each regression line was taken by using the data sets that gave the closest correlations as determined by the Pearson product moment correlation (R^2).

Fig. 51 shows that the stoichiometry is around 40:60 to 45:55. One of the causes for this deviation from a 1:1 stoichiometry could be the low value for the absorption coefficient. An incorrect value for the absorption coefficient however would not lead to a large error in the oligonucleotide concentration as can be seen from the equation below.

$$O_{MAC} = 0.9(15.4N_A + 11.7N_G + 7.3N_C + 8.8N_T + A_M N_M)$$

O_{MAC} = Oligonucleotide molar absorption coefficient. The 0.9 is a correction used for any base stacking in the single stranded DNA at room temperature, this can be ignored if the absorption readings are carried out at high temperatures, >80°C. The other numerical values represent the absorption coefficients of the individual bases and N_A, N_G, N_C and N_T represent the number of As, Gs, Cs and Ts present in the oligonucleotide. Any modifications can be included with A_M representing the absorption coefficient of the modified base and N_M the number of modified bases. Further arguments can be added if more modifications are present in the oligonucleotide.

In a 12mer the total absorption coefficient is made up of the 12 individual absorption coefficients of the individual bases. Therefore any error would only affect one or two bases in 12.

3.4.3 Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

Table 8. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing $\underline{A}$.

Code Number	Mod. Type	No. of Mod.	$-\Delta H$ kcal mol^{-1}	+/-	$-\Delta S$ cal K^{-1} mol^{-1}	+/-	$-\Delta G_{310}$ kcal mol^{-1}	+/-	$T_{m\,(2\,\mu M)}{}^{a}$	$\Delta T_m{}^{b}$
O4	Non	0	84.8	1.7	243.9	5.7	9.11	0.07	37.6	0.0
O22	$\underline{A}$	1	82.9	0.9	235.6	3.0	9.83	0.02	40.6	3.0
O23	$\underline{A}$	2	88.6	1.4	250.2	4.3	11.03	0.04	44.3	3.4
O26/28	Non	1^{c}	89.6	5.4	254.5	16.6	10.64	0.23	43.0	5.4
O27/29	Non	2^{c}	89.9	2.4	251.3	7.5	11.96	0.08	47.8	5.1

^{a} $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 μM + 1 μM.

^{b} ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

^{c} The modifications referred to in the unmodified oligonucleotides are where the $\underline{A}$.T base pair has been substituted with a G.C base pair.

The higher errors associated with the unmodified oligonucleotide duplexes (**O26/28 & O27/29**) are due to the reduced number of melts that were recorded to produce the above mean values. Typically, 60 melting curves were averaged to get the modified oligonucleotide data while the unmodified data is the result of six melt curves.

Table 8 shows that the replacement of 2'-deoxyadenosine with 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyadenosine results in more thermodynamically stable duplexes relative to the unmodified duplex (**O3/O4**). Two reports by Seela have also shown that the adenosine analogue, $\underline{A}$, is stabilising.[77,78] In the first publication a single self-complementary oligonucleotide was synthesised which included 12 adenosine analogue bases. An increase in T_m of 17°C, 1.4°C per modification was observed. (8 μM + 8 μM oligomer concentration, 100 mM MgCl$_2$ and 100 mM NaCl, 60 mM Na-cacodylate (pH 7.1)).[78] In a later publication the same oligonucleotide was shown to increase the T_m by 20°C, 1.7°C per modification.[77] (5 μM + 5 μM single-stranded concentration, 0.1 mM EDTA, 1 M NaCl, 10 mM Na-phosphate (pH 7.0)). Using a non-self complementary oligonucleotide a stabilisation of 1°C per modified base was observed.[77] (5 μM + 5 μM single-stranded concentration, 10 mM MgCl$_2$, 0.1 M NaCl, 10 mM Na-cacodylate (pH 7.0)) The same conditions in the latest publication resulted in a stabilisation of 3°C and

1.3°C per modification for oligonucleotides containing two and six modifications respectively.[68]

The duplexes here show increases in T_m of 3.0 and 3.4°C per modification. The increase in stabilisation for the first base is associated with a favourable increase in entropy while the introduction of the second base is due to a decrease in enthalpy. This increase in stability is not great enough to equal the replacement of $\underline{A}$.T base pairs with G.C base pairs at the same positions.

3.4.4 7-(3-Aminoprop-1-ynyl)-7-deaza-2'-deoxy-2-aminoadenosine

This novel nucleoside as seen in Fig. 52, is structurally similar to other nucleosides that have been synthesised.[76,82,83,121] Its potential to form three hydrogen bonds means it intuitively should have a similar stability to that of a G.C base pair. The addition of the 1-aminoprop-2-ynyl group means that any loss of stabilisation due to the alternating hydrogen bond donors and acceptors could potentially be overcome.
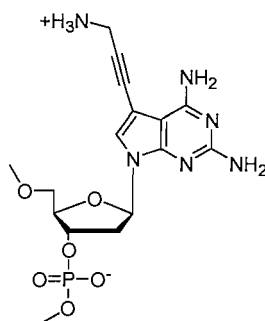
Figure 52. 7-(3-Aminoprop-1-ynyl –7-deaza-2-amino-2'-deoxyadenosine, $\underline{D}$

The convergent synthesis of phosphoramidite **46** was carried out in 16 steps. Oligonucleotides incorporating **46** were synthesised. UV-melting was used to examine the thermodynamic effect of $\underline{D}$ on duplex association and disassociation relative to an unmodified duplex of identical sequence and duplexes in which the $\underline{D}$.T base pairs had been replaced with G.C base pairs.

3.4.5   7-(3-Aminoprop-1-ynyl)-7-deaza-dD Phosphoramidite, 46

The synthesis diverges from that of the 7-(3-aminoprop-1-ynyl)-7-deaza-dA phosphoramidite when the thiol introduced by the addition of thiourea is methylated. This is carried out with dimethyl sulphate in the presence of sodium hydroxide.[113,122] By leaving this reaction overnight it was found that the second ring formed to yield the pyrrolo[2,3-*d*]pyrimidine ring structure to give **37** as seen in Fig. 53. A different protocol was necessary for the chlorination at the 6-position as the use of only phosphorous oxychloride did not result in a high yielding transformation. The alternative protocol uses acetonitrile as the solvent with *N*,*N*-dimethylaniline and the phase transfer catalyst benzyl triethyl ammonium chloride to yield **38**.[123] The heterocycle is then iodinated using the same procedure as used in the synthesis of the 7-(3-aminoprop-1-ynyl) -7-deaza-dA phosphoramidite using *N*-iodosuccinimide in DMF producing **39**.[75,115] A solid-liquid phase-transfer glycosylation using the aprotic solvent acetonitrile with powdered potassium hydroxide and the cryptand tris[2-(2-methoxyethoxy)ethyl]amine (TDA-1) was then carried out to form the protected nucleoside **41**.[116,117] The sulfide was oxidised by *m*CPBA to form a mixture of the sulfoxide and the sulfone using a literature protocol.[74] This mixture is partially purified to remove any polar material and then the mixture is heated in dioxane that has been saturated with NH_3 at 10°C in a sealed Parr general-purpose acid digestion bomb (125 ml). The ammonia displaces both the sulfoxide and the sulfone to leave an amine at the 2-position. Again the product can be reacted on without purification in methanolic ammonia again in a Parr general-purpose acid digestion bomb (125 ml). The ammonia in this case displaces the toluoyl protecting groups to leave the free primary and secondary hydroxyls. The extended reaction time is required for the nucleophilic displacement of the chloride at the 4-position to yield an amine at this position, **42**, and the requisite diamino configuration. Protocols from the literature were followed for v, a) b) & c).[74] 3-Phthalimido-1-propyne was then coupled to the 7-deaza purine at the 5-position using the Hobbs modified Sonogashira reaction to complete the modification of the 7-deaza diaminopurine.[109]
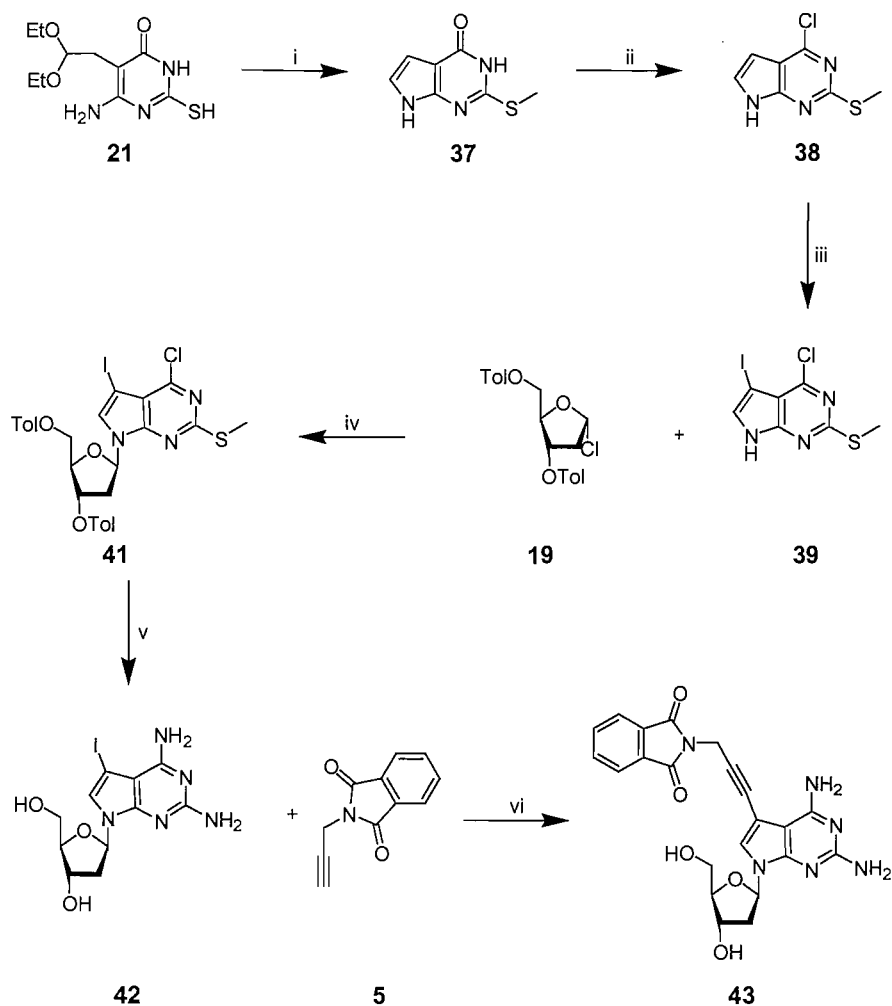
Figure 53. *Reagents and conditions* i, sodium hydroxide, DMS, water, rt, 12 h, 94%, ii, phosphorus oxychloride, *N,N*-dimethylaniline, TEBA, acetonitrile, reflux, 1 h, 65%, iii, NIS, DMF, rt, 16 h, 78%, iv, potassium hydroxide, TDA-1, acetonitrile, rt, 1 h, 77%, v, a) *m*CPBA, DCM, 0°C – rt, 2 h, b) NH₃, dioxane, 140°C, 9 h, c) NH₃, MeOH, 150°C, 24 h, 43% over 3 steps, vi, tetrakis(triphenylphosphine) palladium, CuI, TEA, DMF, 12 h, 94%

The exocyclic amines were then protected using *N,N*-di-*n*-butyl formamidine dimethyl acetal [120] and this was followed by tritylation of the 5'-hydroxyl as can seen in Fig. 54. Upon purification it was noted that the *N,N*-di-*n*-butyl formamidine protecting group at the 2-position partly degrades to the formamide. This reaction was also observed by Seela who observed full deprotection during the basic treatment at the end of the synthesis.[121] The phosphoramidite was formed using standard conditions and the compound was prepared for use on the DNA synthesiser.
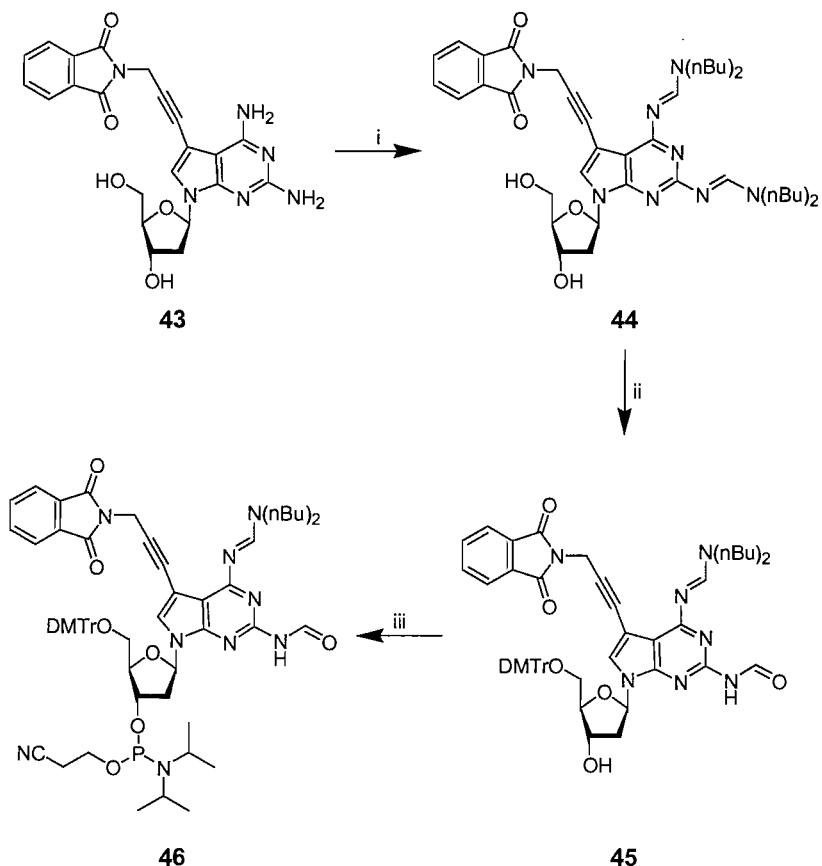
Figure 54. *Reagents and conditions* i, *N,N*-di-*n*-butylforamidine dimethyl acetal, DMF, rt, 12 h, 37%, ii, DMTrCl, py, rt, 16 h, 42% iii, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 47%

The phosphoramidite **46** was incorporated into oligonucleotides using an Applied Biosystems 394 automated DNA synthesiser. The oligonucleotides were purified by HPLC following cleavage from the resin and deprotection. The two 12mers contained one or two non-adjacent modifications (**O28** & **O29**). Due to the constraints of the sequence only one or two modifications could be incorporated. Included in Table 9 are the results for the unmodified oligonucleotides (**O24-O27**). The unmodified oligonucleotides aided in the determination of the stability of **D**.T *versus* G.C for duplex T_m harmonisation.

<div align="center">

5'-dGCTATCT<u>D</u>TCTG-3'   **O28**

5'-dGCT<u>D</u>TCT<u>D</u>TCTG-3'   **O29**

</div>

To determine the absorption coefficient of oligonucleotides containing 7-(3-aminoprop-1-ynyl –7-deaza-2-amino-2'-deoxyadenosine the absorption coefficient of the

nucleoside was determined. As with the previous nucleoside, the intermediate **43** was deprotected in 10% methylamine to give the free nucleoside **47** (Fig. 55).
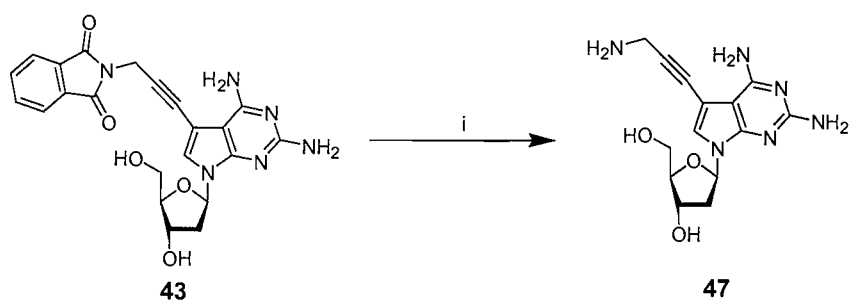
Figure 55. *Reagents and conditions* i, Methylamine, water, rt, 24 h, 56%

The absorption coefficient was determined in water and was calculated to be 5900 moles^{-1}cm^{-1}.

3.4.6 Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

Table 9. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing **D**.

Code Number	Mod. Type	No. of Mod.	$-\Delta H$ kcal mol^{-1}	+/−	$-\Delta S$ cal K^{-1} mol^{-1}	+/−	$-\Delta G_{310}$ kcal mol^{-1}	+/−	$T_{m\,(2\,\mu M)}$[a]	ΔT_m[b]
O4	Non	0	84.8	1.7	243.9	5.7	9.11	0.07	37.6	0.0
O28	**D**	1	87.0	1.3	247.1	4.1	10.31	0.06	42.3	4.6
O29	**D**	2	90.5	1.3	253.5	3.9	11.89	0.05	47.7	5.1
O26/28	Non	1[c]	89.6	5.4	254.5	16.6	10.64	0.23	43.0	5.4
O27/29	Non	2[c]	89.9	2.4	251.3	7.5	11.96	0.08	47.8	5.1

[a] $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 µM + 1 µM.

[b] ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

[c] The modifications referred to in the unmodified oligonucleotides are where the **D**.T base pair at the position where the modified base has been used is substituted with a G.C base pair.

The higher errors associated with the unmodified oligonucleotides are due to the reduced number of melts that were used to produce the above means. Typically 60 melting curves were average to get the modified oligonucleotide data while the unmodified data is the result of six melt curves.

The data shows that this 2'-deoxyadenosine analogue, **D**, is strongly stabilising with ΔT_ms of 4.6°C and 5.1°C for the singly **O28** and doubly **O29** modified oligonucleotides. This corresponds to a stabilisation in the Gibbs free energy of 1.2 and 1.4 kcal mol^{-1} per modification for **O28** and **O29** respectively (Table 9). According to the

data in Table 9 the stabilisation is due to a decrease in enthalpy (2.2 kcal mol^{-1} per modification for **O28**, 2.9 kcal mol^{-1} per modification for **O29**) with a modest entropic penalty (3.16 cal K^{-1} mol^{-1} per modification for **O28**, 4.81 cal K^{-1} mol^{-1} per modification for **O29**).

Close structural analogues in the literature and 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyadenoisne give an interesting insight into the effect of individual functional groups (Table 10).[76,82]

Table 10. The highlighted groups are those that are under inspection with this analysis. The red groups are present in the 7-(3-aminoprop-1-ynyl)-7-deaza-2-amino-2'-deoxyadenosine and the blue aza group shows any changes to the afore mentioned nucleoside.
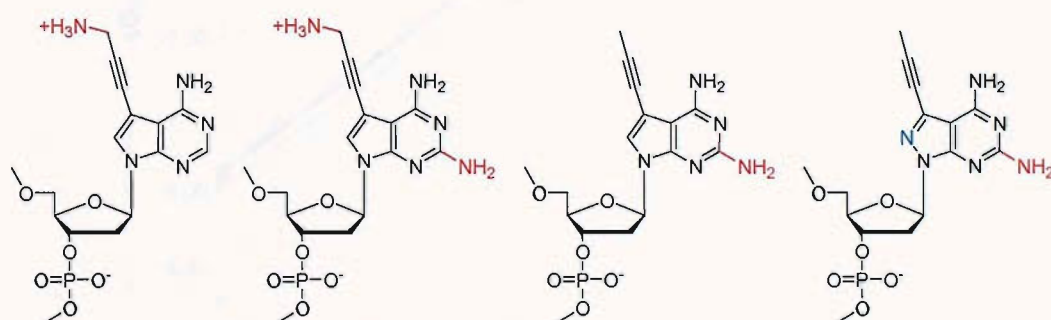
No. Mod	1	2	1	2	1[a]	4[b]	1[c]	2[c]
ΔT_m [d]	2.95	3.35	4.64	5.06	3.86	1.54	6	5.5
Δ-ΔG_{310} [e]	0.72	0.96	1.20	1.39	nd	nd	1.0	1.75

[a] The modification was incorporated into an 18-mer and were melted using 100 mM NaCl, 0.1 mM EDTA, 10 mM Na-phosphate buffer (pH 7.1)

[b] The four modifications were incorporated into a 17-mer using the conditions in [a]

[c] The modifications were incorporated into a 12-mer and were melted using 100 mM NaCl, 10 mM MgCl$_2$ and 10 mM Na-cacodylate (pH 7.0) and 5 μM of each single stranded oligonucleotide

[d] ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

[e] Δ-ΔG_{310} is the change in Gibbs free energy per modification / kcal mol^{-1}.

nd = not determined.

The 2-amino group contributes 1.7°C or -0.45 kcal mol^{-1} in both cases. The amine on the 1-aminoprop-2-yne group adds 0.8°C to the stabilisation of duplexes containing a single modification. The addition of the 8-aza group makes a contribution of 2.1°C, however, the conditions of the melting could also contribute to this difference. This is in comparison to the 7-propynyl-7-deaza-2-amino-2'-deoxyadenosines.

As with the 7-propynyl-8-aza-7-deaza-2-amino-2'-deoxyadenosine nucleoside[82], 7-(3-aminoprop-1-ynyl)-7-deaza-2-amino-2'-deoxyadenosine harmonises the T_m of A.T

base pairs to that of G.C base pairs. A direct comparison of the Gibbs free energies at 310 K is shown in Fig. 56.

D:T vs G:C
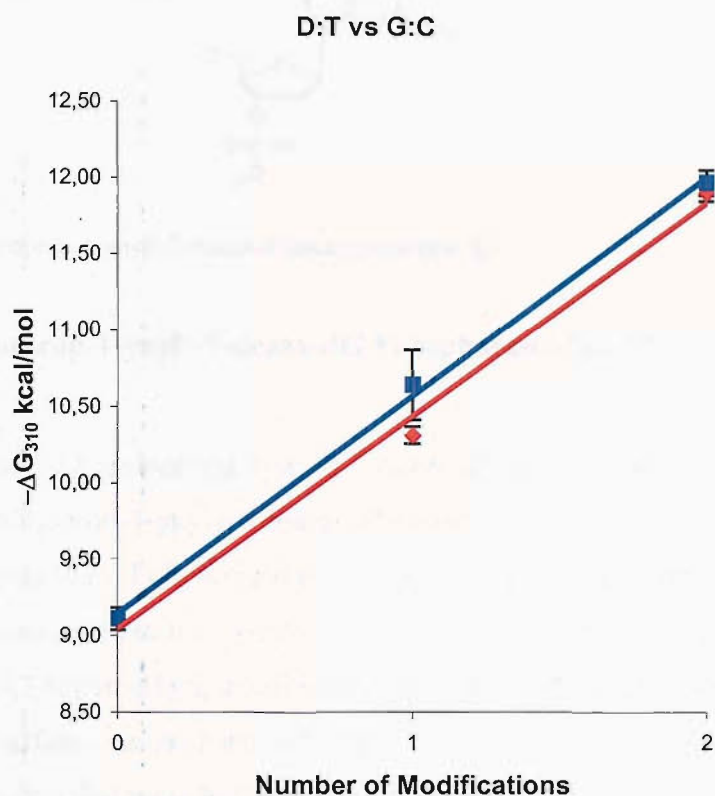
Figure 56. A comparison of the ΔG (kcal/mol) values at 310 K between oligonucleotides that have had their A.T base pairs replaced with either **D**.T base pairs or G.C base pairs. The number of replacements is referred to as the number of modifications. The y-axis error-bars are shown as calculated by the Meltwin 3.5 software and the propagation of random errors were necessary. The blue line corresponds to replacement with G.C and the red line to **D**.T.

3.4.7   7-(3-Aminoprop-1-ynyl)-7-deaza-2'deoxyguanosine

Two reports have been published where 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyguanosine (Fig. 57) has been incorporated into duplexes.[80,81]

The convergent synthesis of phosphoramidite **53** was carried out in 16 steps. Oligonucleotides incorporating **53** were synthesised. UV-melting was used to examine the thermodynamic effect of **G** on duplex association and disassociation relative to an unmodified duplex of identical sequence.
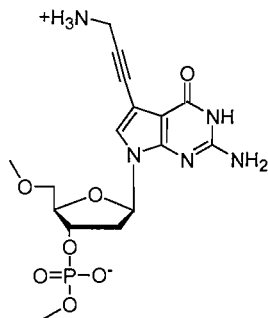
Figure 57. 7-(3-Aminoprop-1-ynyl)-7-deaza-2'deoxyguanosine, <u>G</u>.

3.4.8   7-(3-Aminoprop-1-ynyl)-7-deaza-dG Phosphoramidite, 53

The synthesis of the 7-(3-aminoprop-1-ynyl)-7-deaza-dG phosphoramidite is identical to that for the 7-(3-aminoprop-1-ynyl)-7-deaza-dD phosphoramidite up to and including the glycosylation, **41** (Fig. 53). Following the glycosylation the chloro-pyrrolo[2,3-*d*]pyrimidine was converted to the pyrrolo[2,3-*d*]pyrimidone with *syn*-2-pyridine aldoxime and 1,1,3,3-tetramethylguanidine with a yield of 95% to give **48** (Fig. 58).[74] Subsequently the sulfide was oxidised with *m*CPBA to produce a mixture of the sulphoxide and sulfone that were both displaced by ammonia-saturated dioxane to give an amine at the 2-position. The toluoyl groups were then removed by methanolic ammonia to yield the free nucleoside.[74] The nucleoside was then functionalised at the 7-position with 3-phthalimido-1-propyne, **5**, using the Pd (0) catalysed Sonogashira reaction as adapted by Hobbs.[80,109] The exocyclic amine was then protected to form the dimethylformamidine as in a similar synthesis by Froehler *et al.*[74] The reagent *N,N*-dimethylformamide diethyl acetal was used in order to prevent methylation at the *N*-3 position.[120] The synthesis was then completed by the tritylation of the 5'-hydroxyl, **53**, and then phosphitylation of the 3'-hydroxyl to yield the completed monomer ready for use on the DNA synthesiser, **54**.

Phosphoramidite **54** was incorporated into 12mer oligonucleotides. As the modified nucleoside is an analogue of 2'-deoxy-guanosine a new sequence was required. This is because the previous sequence had only two 2'deoxy-guanosines and these were restricted to the 3' and 5' ends of the oligonucleotide. Instead an inverted sequence of the complementary strand was used. The base was inserted in 1, 2 or 3 non-adjacent positions (**O31-O33**). The positions of the 7-(3-aminoprop-1-
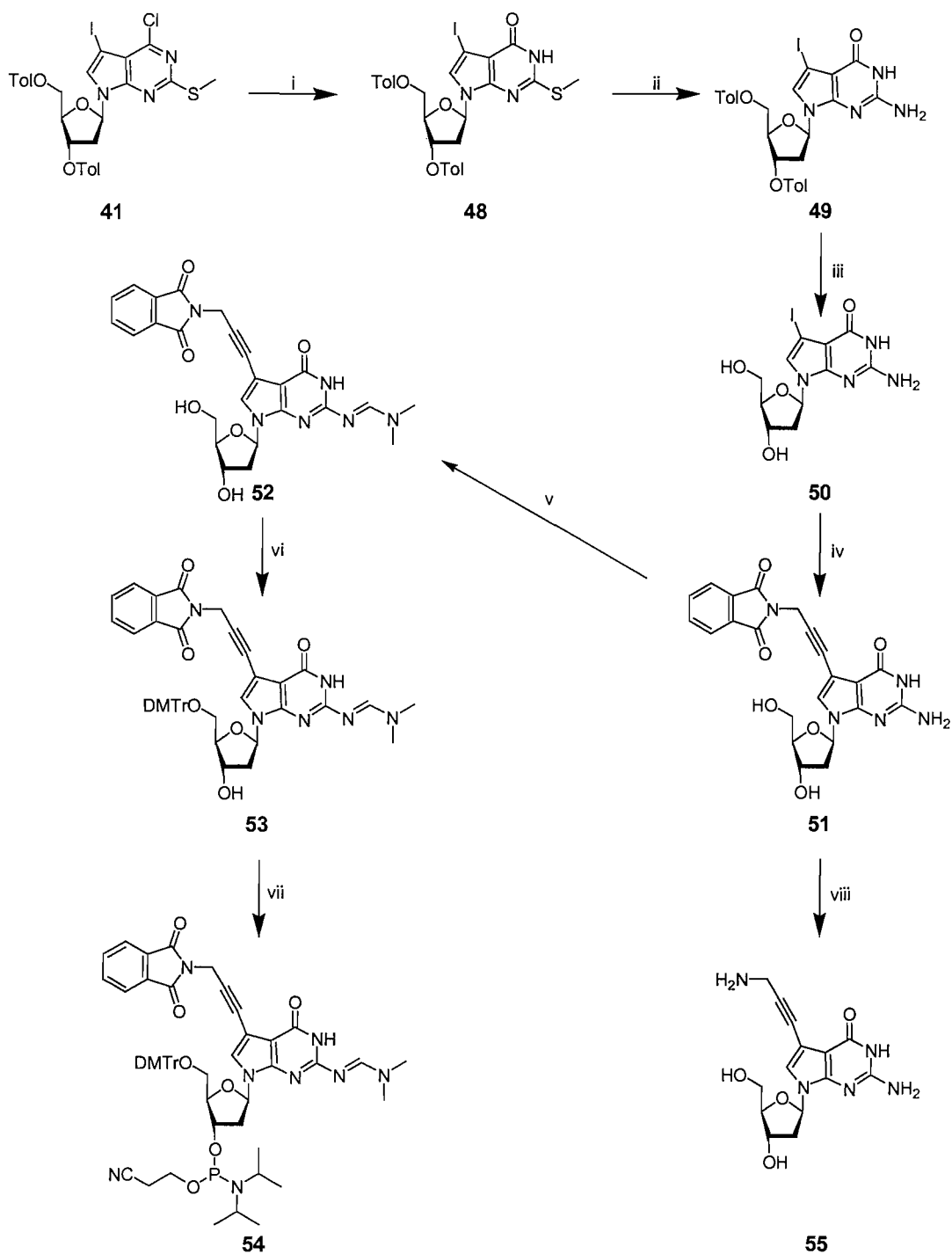
Figure 58. *Reagents and conditions* i, *syn*-2-pyridinealdoxime, 1,1,3,3-tetramethylguanidine, DMF, dioxane, rt, 24 h, 95%, ii, a) *m*CPBA, DCM, 0°C – rt, 2 h, b) NH₃, dioxane, 140°C, 9 h, 90% over 2 steps iii, NH₃, MeOH, 140°C, 2 h, 34%, iv, tetrakis(triphenylphosphine)palladium, CuI, TEA, DMF, rt, 12 h, 72%, v, *N,N*-dimethylformamide diethyl acetal, EtOH, DMF, rt, 24 h, 40%, vi, DMTrCl, py, rt, 16 h, 51%, vii, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 75%, viii, methylamine, water, rt, 24 h, 39%

ynyl)-7-deaza-2'-deoxyguanosines are in bold and underlined, **G**. The unmodified complement **O30** was also synthesised.

5'-dGTCTATCTATCG-3'   **O30**

5'-dCGATA<u>G</u>ATAGAC-3'   **O31**

5'-dCGATA<u>G</u>ATA<u>G</u>AC-3'   **O32**

5'-dC<u>G</u>ATA<u>G</u>ATA<u>G</u>AC-3'   **O33**
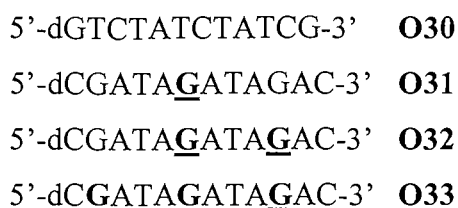
The absorption coefficient of the free nucleoside was determined (4600 moles⁻¹ cm⁻¹). Using this value the concentration of oligonucleotides containing the nucleoside could be calculated. Exposing **47** to 10% methylamine in water generated the free nucleoside **55** (Fig. 58).

3.4.9   Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

Table 11. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing <u>G</u>.

Code Number	Mod. Type	No. of Mod.	$-\Delta H$ kcal mol⁻¹	$+/-$	$-\Delta S$ cal K⁻¹ mol⁻¹	$+/-$	$-\Delta G_{310}$ kcal mol⁻¹	$+/-$	$T_{m\,(2\,\mu M)}{}^{a}$	$\Delta T_m{}^{b}$
O4	Non	0	84.8	1.7	243.9	5.7	9.11	0.07	37.6	0.0
O31	<u>G</u>	1	88.0	0.8	250.9	2.6	10.14	0.03	41.2	3.6
O32	<u>G</u>	2	84.2	1.3	237.1	4.0	10.60	0.04	43.5	3.0
O33	<u>G</u>	3	89.0	1.1	249.7	3.4	11.51	0.04	46.3	2.9

^{a} $T_{m\,(2\,\mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 μM + 1 μM.

^{b} ΔT_m is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

The anomalous result in the ΔH and ΔS data for the bis-modified oligonucleotide **O32** could not be traced to anomalous individual melting curves. However as the errors in ΔH and ΔS, to an extent cancel each other, the values of ΔG follow an understandable trend.

The base is stabilising when compared to the unmodified oligonucleotide duplex. In the initial literature paper an increases in T_m of 1°C per modified base was observed.[81] In the subsequent literature paper however an increase of between 2.0-2.5°C (0.60-0.48 kcal mol⁻¹) per modification was recorded for duplexes containing two and four modifications respectively.[80] (100 mM NaCl, 10 mM MgCl₂, 10 mM Na-cacodylate (pH 7) with 5 + 5 μM single strand concentrations).

The results in Table 11 show T_m increases of 2.9-3.6°C per modified base and increases in the Gibbs free energy of 0.75-1.03 kcal mol^{-1} per modification.

3.4.10 7-(3-Aminoprop-1-ynyl)-7-deaza-2'-deoxyinosine

2'-Deoxyinosine is a universal nucleoside that shows ambiguous base pairing with the four natural bases of DNA.[124,125] 7-(3-Aminoprop-1-ynyl)-7-deaza-2'deoxyinosine is a novel analogue of 2'-deoxyinosine and its incorporated structure is shown in Fig. 59. Its synthesis and pairing against 2'-deoxycytidine gave some insight into the thermodynamic effect of the 2-amino group in 7-(3-aminoprop-1-ynyl)-7-deaza-2'deoxyguanosine.

The convergent synthesis of phosphoramidite **60** was carried out in 15 steps. Oligonucleotides incorporating **60** were synthesised. UV-melting was used to examine the thermodynamic effect of **I** on duplex association and disassociation relative to an unmodified duplex of identical sequence and duplexes in which the **I**.C base pairs had been replaced with A.T base pairs.
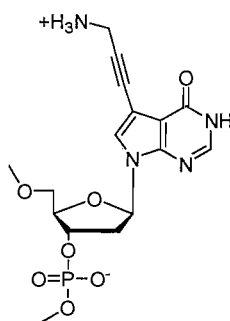
Figure 59. 7-(3-Aminoprop-1-ynyl)-7-deaza-2'deoxyinosine, **I**.

3.4.11 7-(3-Aminoprop-1-ynyl)-7-deaza-dI Phosphoramidite, 60

The synthesis of the 7-(3-aminoprop-1-ynyl)-7-deaza-dI phosphoramidite is identical to the 7-(3-aminoprop-1-ynyl)-7-deaza-dA phosphoramidite up to and including the glycosylation, **27** (Fig. 48). The conversion of the chloro-pyrrolo[2,3-*d*]pyrimidine to the pyrrolo[2,3-*d*]pyrimidone was carried out with *syn*-2-pyridine aldoxime and 1,1,3,3-tetramethylguanidine with a yield of 95% as in the synthesis of the 7-(3-aminoprop-1-ynyl)-7-deaza-dG phosphoramidite to yield **56** (Fig. 60).[74] Next the toluoyl protecting groups were cleaved using methanolic ammonia[74] and then 3-phthalimido-1-propyne, **5**, was added to the 7-position using a Pd(0) catalysed Sonogashira reaction adapted by

Hobbs to give **58**.[109] The nucleoside was then tritylated and phosphitylated at the 5' and 3' hydroxyls respectively using standard protocols to yield the phosphoramidite **60**.
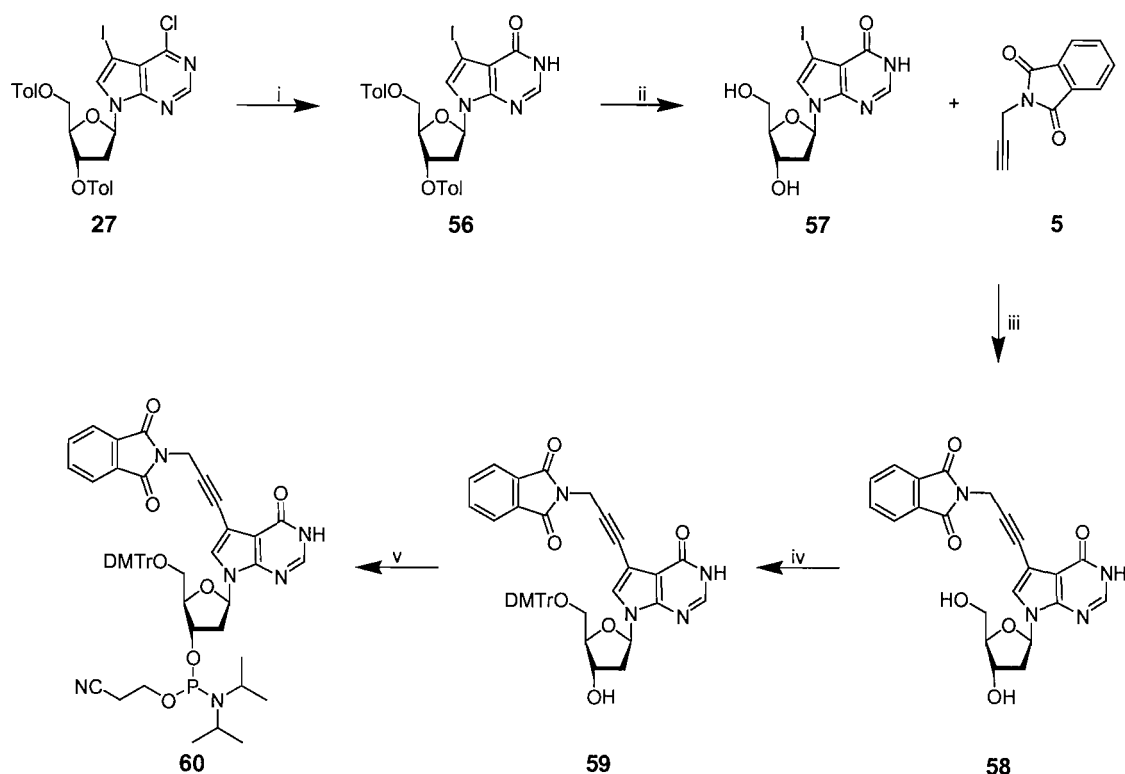
Figure 60. *Reagents and conditions* i, *syn*-2-pyridinealdoxime, 1,1,3,3-tetramethylguanidine, DMF, dioxane, rt, 24 h, 95%, ii, NH$_3$, MeOH, 140°C, 2 h, 76%, iii, tetrakis(triphenylphosphine) palladium, CuI, TEA, DMF, rt, 12 h, 56%, iv, DMTrCl, py, rt, 16 h, 41%, v, 2-cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM, rt, 1 h, 45%

Phosphoramidite **60** was incorporated into three 12mer oligonucleotides. The base was inserted in 1, 2 or 3 non-adjacent positions (**O34-O36**) that were complementary to **O30**. The sequence was the same as that used for the 7-(3-aminoprop-1-ynyl)-7-deaza-dG studies. The positions of the 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyinosines are in bold and underlined, **I**.

5'-dCGATA**I**ATAGAC-3'	**O34**
5'-dCGATA**I**ATA**I**AC-3'	**O35**
5'-dC**I**ATA**I**ATA**I**AC-3'	**O36**
5'-dCGATAAATAGAC-3'	**O37**
5'-dCGATAAATAAAC-3'	**O38**
5'-dCAATAAATAAAC-3'	**O39**
5'-dGTCTATTTATCG-3'	**O40**

101
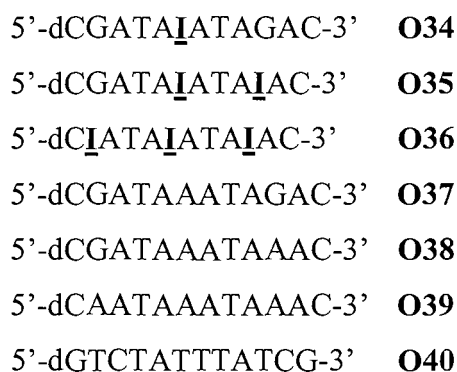
5'-dGTTTATTTATCG-3' **O41**

5'-dGTTTATTTATTG-3' **O42**

The absorption coefficient of the free nucleoside was determined (6900 moles⁻¹cm⁻¹). Using this value the concentration of oligonucleotides containing the nucleoside could be calculated. Exposing **58** to 10% methylamine in water generated the free nucleoside **61** (Fig. 61).
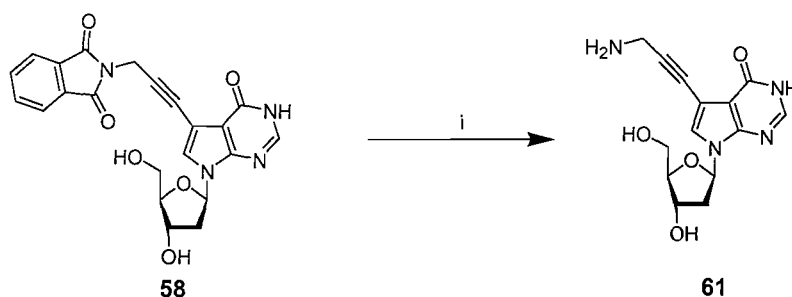
Figure 61. *Reagents and conditions* i, Methylamine, water, rt, 24 h, 29%

3.4.12 Ultraviolet Melting and Thermodynamic Analysis of Duplex Dissociation

As expected the 7-(3-aminoprop-1-ynyl)-7-deaza-dI destabilises the duplex with respect to G.C base pairs (Table 12).

Each addition of 7-(3-aminoprop-1-ynyl)-7-deaza-dI led to a destabilisation when compared to the unmodified duplex and each extra addition led to an even greater disruption. The reduction in T_m per modification increases from 1.58°C for **O34** to 3.02 °C for **O36**. This corresponds to an increase in the Gibbs free energy at 310 K of 0.45 kcal mol⁻¹ for **O34** to 0.62 kcal mol⁻¹ for **O36**. The loss of stability is due to a more positive enthalpy with per base increases from 2.6-3.8 kcal/mol with some compensation coming from a more favourable entropic term with per base increases from 7-10.3 cal K⁻¹ mol⁻¹.

Table 12. The enthalpy, entropy and Gibbs free energy of duplex melting for oligonucleotides containing **I**.

Code Number	Mod. Type	No. of Mod.	$-\Delta H$ kcal mol^{-1}	+/-	$-\Delta S$ cal K^{-1} mol^{-1}	+/-	$-\Delta G_{310}$ kcal mol^{-1}	+/-	$T_{m\ (2\ \mu M)}$ [a]	ΔT_m [b]
O4	Non	0	84.8	1.7	243.9	5.7	9.11	0.07	37.6	0.0
O34	**I**	1	82.1	1.0	236.6	3.1	8.66	0.02	36.0	-1.6
O35	**I**	2	79.5	1.0	229.9	3.1	8.16	0.02	33.9	-1.9
O36	**I**	3	73.3	1.0	213.0	3.2	7.26	0.03	28.5	-3.0
O37/40	Non	1 [c]	84.3	0.6	244.6	2.0	8.4	0.056	35.0	-2.6
O38/41	Non	2 [c]	84.2	2.9	247.4	9.6	7.45	0.058	31.6	-3.0
O39/42	Non	3 [c]	74.0	1.1	217.8	3.4	6.47	0.033	27.0	-3.6

[a] $T_{m\ (2\ \mu M)}$ is the melting temperature at a total single stranded oligonucleotide concentration of 1 μM + 1 μM.

[b] ΔT_m Is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide.

[c] The modifications referred to in the unmodified oligonucleotides are where the **I**.C base **pairs have** been substituted with **A.T** base pairs.

The higher errors associated with the unmodified oligonucleotides are due to the reduced number of runs that were used to produce the above means. Typically 60 melting curves were average to get the modified oligonucleotide data while the unmodified data is the result of six melt curves.

To determine whether the **I**.C base pair could be of benefit when modifying the melting temperatures of duplexes it was compared to the A.T base pair (Fig. 62). In each position where an **I**.C base pair was placed an A.T was placed in a control oligonucleotide that contained no modified bases.
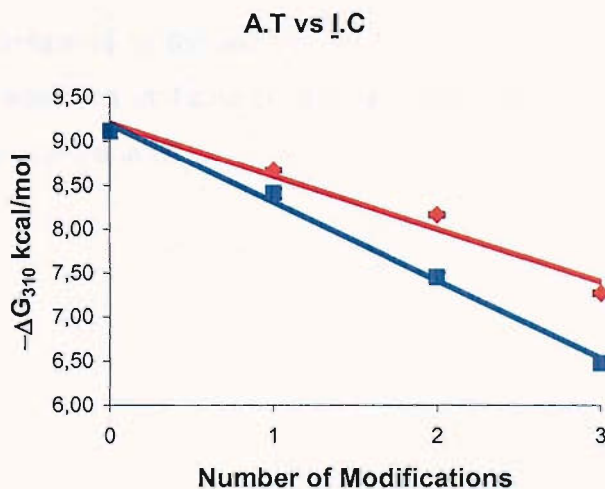
Figure 62. A comparison of the ΔG (kcal/mol) values at 310 K between oligonucleotides that have had their G.C base pairs replaced with either **I**.C base pairs or A.T base pairs. The number of replacements is referred to as the number of modifications. The y-axis error-bars are shown as calculated by the Meltwin 3.5 software and the propagation of random errors where necessary. The blue line is the A.T data whereas the red line represents the **I**.C data.

The data shows the two regression lines diverging with the introduction of A.T base pairs being more destabilising to the duplex than the introduction of **I**.C base pairs.

Table 13 shows that when the 2-amino group is placed next to another hydrogen bond donor it is a very stabilising group for duplexes. It contributes 5°C or 1.6 kcal mol^{-1} per addition to a duplex for low numbers of additions.

Table 13. A thermodynamic comparison of 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyguanosine and 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyinosine. The 2-amino group is highlighted in 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxyguanosine.
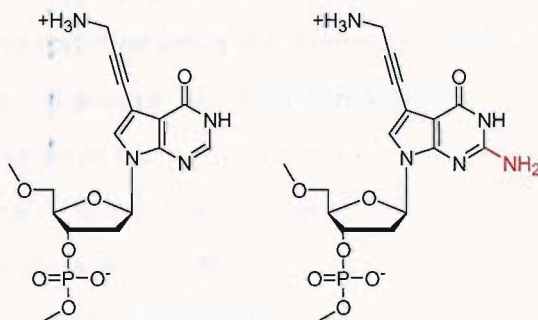
No. Mod	1	2	1	2
$\Delta T_m{}^a$	-1.6	-1.9	3.6	3.0
$\Delta\text{-}\Delta G_{310}{}^b$	-0.5	-0.5	1.0	0.7

^{a} ΔT_m Is the increase in melting temperature observed over the unmodified oligonucleotide divided by the number of modifications in the oligonucleotide / °C.

^{b} $\Delta\text{-}\Delta G_{310}$ is the change in Gibbs free energy per modification / kcal/mol.

This can be compared to the addition of the 2-amino group next to a hydrogen bond acceptor. This was seen in Table 10 and each addition only contributes 1.7°C or -0.45 kcal mol^{-1} in this configuration.

3.5   CONCLUSIONS

The 1-aminoprop-2-yne group when placed at the C-5 position of pyrimidines and the C-7 positions of 7-deazapurines is stabilising in all cases. The thermodynamic basis is a more favourable enthalpic term in almost all cases and normally accompanied by a less favourable entropic term.

In the case of the pyrimidine analogues the electrolytic nature of the 1-aminoprop-2-yne at physiological pH creates a duplex that is less sensitive to salt concentration. Comparatively 5-(3-aminoprop-1-ynyl)-2'-deoxyuridine is less sensitive to salt than 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine under the conditions used here.

The novel adenosine analogue 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxy-2-aminoadenosine when inserted in place of an adenosine in an oligonucleotide alters the T_m of a duplex to that of a duplex in which a G.C base pair has been placed at the same position. The fidelity of this analogue needs to be investigated to ensure that it does not stabilise mismatches to any greater extent than natural adensoine.

The novel inosine analogue 7-(3-aminoprop-1-ynyl)-7-deaza-2'-deoxy-inosine when paired with cytidine does not destabilise a duplex to the same extent that replacement with an A.T does. The introduction of inosine alone paired with cytidine may however, be destabilising enough to do this. The universal stability of inosine paired with the other natural base pairs may negate the need for this to be investigated in this context.

3.6 FUTURE DIRECTION

The research described within this thesis may be extended in a number of ways. This may be by examining the molecular basis of stabilisation or by reviewing different aspects of the overall assay design.

To identify if the origin of the stabilisation could be explained further, the orientation of the bases in the duplex could be examined by crystallising oligonucleotides that contain the modified bases. Of particular interest would be the position of the propargylamino modification and how this impacts the hydration of the major groove. The determination of the origin of the stabilisation is important in the design of future modifications. Thus, the most desirable aspects of these modifications could be combined in such a way as to further enhance stabilisation.

The results with the modified inosine base demonstrated that it did not destabilise a G.C base pair to that of an A.T base pair. However, despite the promiscuity of inosine as a base, it would be worthwhile to examine if inosine could be used to reduce the stability of a G.C base pair down to that of an A.T base pair. The harmonisation of the melting temperature of G.C and A.T base pairs is an important area of study in the generation of probes, which only depend on their length for their melting temperature. With harmonised melting temperatures, genetic arrays, for example, may become much more accurate.

The complete assay could be further developed by using the modified bases to explore if they do aid the kinetics and thermodynamics of DNA hybridisation by incorporating them into the oligonucleotide on the bead surface. MS analysis of isotopically labelled peptides, the single base extension reaction and the sorting of fluorescently labelled beads by a FACS could be carried out to develop the full assay. This may lead to the complete assay as envisaged at the outset of the research.

4   EXPERIMENTAL

4.1   SOLVENTS AND REAGENTS

Acetonitrile, dichloromethane, pyridine, triethylamine and diisopropylethylamine were distilled from calcium hydride. Ethanol and methanol were distilled from magnesium ethoxide or methoxide prepared by the addition of magnesium and iodine to the solvent.[126] Anhydrous dimethylformamide and dioxane were purchased from Aldrich. Other reagents were purchased from Sigma, Aldrich (subsidiaries of Sigma-Aldrich) or Avocado Organics (now part of Alfa Aesar® a Johnson Matthey company). 2-Cyanoethyl-*N,N*-diisopropylchlorophosphine, 5-iodo-2'-deoxyuridine, 5-iodo-2'-deoxycytidine and 4,4'-dimethoxytritylchloride[9] were purchased from Link Technologies Ltd. Resin beads were purchased or provided by: Rapp Polymere GmbH, Argonaut Technologies, Inc., Glen Research Corp., Amersham Biosciences (now part of GE Healthcare), Polymer Laboratories Ltd, Biosearch Technologies, Inc. and E. Birch-Hirschfeld.[101,102] All reactions were carried out under an argon atmosphere using oven dry glassware unless otherwise stated.

4.2   ANALYTICAL TECHNIQUES

4.2.1   Chromatography

Flash silica gel column chromatography[127] was carried out under pressure using Fisher Scientific DAVISIL 60A (35-70 micron) silica or 60 mesh silica gel (Merck or Apollo) or using a FlashMaster personal chromatography system with FlashPack modules (Argonaut Technologies, Inc.). Thin layer chromatography (TLC) was carried out on silica gel 60, F_{254}, 0.2mm layer aluminium sheets (Merck) or ISOLUTE silica TLC plates. Compounds were visualised on TLC by irradiation at 254nm, by heating compounds that contained a dimethoxytrityl group or by staining with a solution of:

- 4-Methoxybenzaldehyde (p-anisaldehyde): glacial acetic acid: conc. sulphuric acid: ethanol (5:1:1:50, v:v:v:v) – stains nucleophilic compounds, for example protected or unprotected 1,2-diols give a dark blue colour on heating.
- Ninhydrin: ethanol (1:100, wt:v) – stains compounds containing a primary amino group to give a dark brown colour on heating.

- Phosphomolybdic acid acid: ethanol (1:10, v:v) – stains most functional groups however functional groups can not be distinguished by the colour the spots. Generally the compounds will appear dark green on a light green background.

- Potassium permanganate: sodium carbonate: water (1:2:100, wt,wt,v) – stains easily oxidised compounds on contact and others on heating yellow with a pink background.

- Iodine – the iodine vapour stains unsaturated and aromatic compounds dark brown on a light brown background.

4.2.2 NMR Spectrometry

NMR spectra were assigned using standard ^{1}H, ^{13}C, Dept135, ^{31}P, COSY and HMQC experiments. These were performed on a Bruker AC300 or a Bruker DPX400 NMR spectrometer in the solvents indicated at 300K: all chemical shifts (δ) are in ppm, proton and carbon spectra were internally referenced to residual solvent peaks[128] and phosphorus spectra externally referenced to 85 % phosphoric acid in D_2O. J Values are given in hertz.

4.2.3 Mass Spectrometry

Low-resolution mass spectra were recorded either using electrospray technique on a Micromass platform II instrument in acetonitrile or a Waters ZMD quadrupole mass spectrometer in methanol or water. Low-resolution MALDI-TOF spectra were recorded on a ThermoBioAnalysis Dynamo Linear MALDI-TOF mass spectrometer. High-resolution mass spectra were recorded using electrospray using a Bruker APEX III FT-ICR mass spectrometer.

4.2.4 IR Spectroscopy

Infrared spectra were recorded on a BIORAD FT-IR using a Golden Gate adapter and BIORAD WIN-IR software or a Satellite FT-IR using a Golden Gate adapter and WIN FIRST-lite software. Absorptions are described as strong (s), medium (m), broad (b) or weak (w).

4.2.5 UV/vis Spectroscopy

UV-vis data was recorded on a Cary 400 UV/visible spectrometer using 1 ml cuvettes at $20^{\circ}C$. All absorption coefficients are quoted in ($cm^{-1}mol^{-1}$).

4.2.6 Elemental Analysis

C, H, N Elemental analysis was undertaken by MEDAC Ltd. Calculated and observed values are shown.

4.2.7 Melting Point Analysis

Melting points were determined using a Gallenkamp electrothermal melting point apparatus, values are quoted as a range of temperature ($^{\circ}C$) and are uncorrected.

4.3 LIST OF COMPOUNDS

1. 12- *O*- (4,4'-Dimethoxytrityl)dodecanoic acid

2. 12- *O*- (4,4'-Dimethoxytrityl)dodecanamide resins

3. 4-(2,4-Dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyacetamido resins

4. 2,2,2-Trifluoro-*N*-(prop-2-ynyl)acetamide[61]

5. 3-Phthalimidoprop-1-yne[63]

6. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-5-iodo-uridine[9,67]

7. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl]- uridine[62]

8. 3'-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl)-uridine[62]

9. 2'-Deoxy-*N*4-[(di-*n*-butylamino)methylidene]-5-iodo-cytidine

10. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-iodo-cytidine

11. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl)]-cytidine

12. 3'-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl)]-cytidine

13. 2'-Deoxy-*N*4-[(di-*n*-butylamino)methylidene]-5-(phthalimidoprop-1-ynyl)-cytidine

14. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-(phthalimidoprop-1-ynyl)-cytidine

15. 3'-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-(phthalimidoprop-1-ynyl)-cytidine

16. 5-(3-Aminoprop-1-ynyl)-2'-deoxycytidine

17. Methyl-2-deoxy-D-ribofuranoside[112]

18. Methyl-3,5-di-*O*-*p*-toluoyl-2-deoxy-D-ribofuranoside[112]

19. 1-Chloro-2-deoxy-3,5-di-*O*-*p*-toluoyl-α-*erythro*-pentofuranose[112]

20. Ethyl-2-cyano-4,4-diethoxybutanoate[113,114]

21. 6-Amino-5-(2,2-diethoxy-ethyl)-2-(mercapto)-3*H*-pyrimidin-4-one[113,122]

22. 6-Amino-5-(2,2-diethoxy-ethyl)-3*H*-pyrimidin-4-one[113]

23. 3,7-Dihydro-pyrrolo[2,3-*d*]pyrimidin-4-one[113]

24. 4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine[113]

25. 4-Chloro-5-iodo-7*H*-pyrrolo[2,3-*d*]pyrimidine[75]

26. 4-Chloro-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)pyrrolo[2,3-*d*]pyrimdine[75,116,119]

27. 4-Chloro-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-(iodo)pyrrolo[2,3-*d*]pyrimdine[74,116]

28. 4-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidine[74]

29. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-[phthalimidoprop-1-ynyl]-7*H*-pyrrolo-[2,3-*d*]pyrimidin-4-amine[74]

30. 7-[2'-Deoxy-β-D-*erythro*-pentofuranosyl]-*N*^{4}-[(di-*n*-butylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

31. 7-[2'-Deoxy-β-D-*erythro*-pentofuranosyl]-*N*^{4}-[(dimethylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[74]

32. 7-[2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*^{4}-[(di-*n*-butyl-amino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

33. 7-[2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*^{4}-[(dimethylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

34. 7-[3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*^{4}-[(di-*n*-butylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

35. 7-[3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*4-[(dimethylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

36. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine

37. 2-(Methylthio)pyrrolo[2,3-*d*]pyrimidin-4-one[113]

38. 4-Chloro-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine[123]

39. 4-Chloro-5-iodo-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine[75]

40. 4-Chloro-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine

41. 4-Chloro-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodo-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine[116]

42. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-2,4-diamine[74]

43. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine[80]

44. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)- N^4,N^2-[di-(di-*n*-butylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

45. 7-(2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)- N^4-[(di-*n*-butylamino)methylidene]-N^2-formyl-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

46. 7-(3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)- β-D -*erythro*-pentofuranosyl)-N^4-[(di-*n*-butylamino)methylidene]-N^2-formyl-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

47. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

48. 7-(2'-Deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodo-2-(methylthio)pyrrolo[2,3-*d*]pyrimidin-4-one[74]

49. 2-Amino-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-4-one[74]

50. 2-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-4-one[74]

51. 2-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[74]

52. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-2-([(*N,N*-dimethylamino)methylidene]amino)-5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[74]

53. 7-(2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)-2-{[(*N,N*-dimethylamino)methylidene]amino}-5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[80]

54. 7-(3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)-2-{[(*N,N*-dimethylamino)methylidene]amino}-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-one

55. 2-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)- 5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimid-4-one

56. 7-(2'-Deoxy-3',5'-di-*O*-*p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-4-one[74]

57. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimid-4-one

58. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[74]

59. 7-(2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)- 5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[80]

60. 7-(3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimid-4-one

61. 7-(2-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimid-4-one

4.4 CHEMICAL SYNTHESIS

1. 12- *O*- (4,4'-Dimethoxytrityl)dodecanoic acid

1
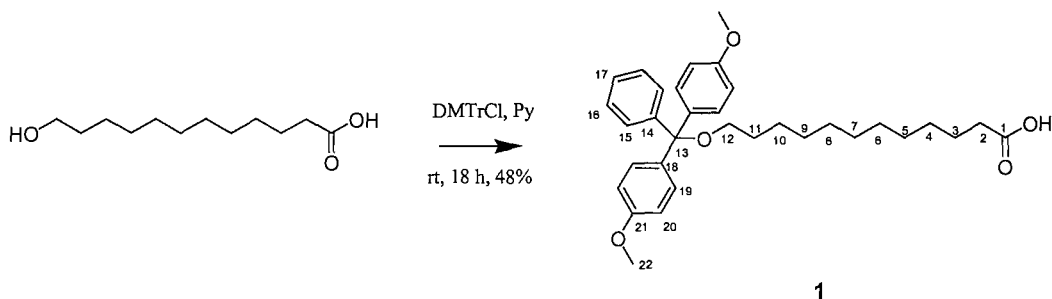
12-Hydroxydodecanoic acid (5 g, 23.1 mmol) was dissolved in pyridine (18 ml) and a solution of 4,4'-dimethoxytrityl chloride (8.61 g, 25.4 mmol) in pyridine (35 ml) was added drop-wise (45 min). The reaction was left stirring under argon (RT, 18 h). The mixture was concentrated *in vacuo*, extracted with DCM (200 ml), washed with saturated sodium bicarbonate solution (2 x 200 ml), saturated KCl solution (200 ml), dried (anhydrous MgSO$_4$) and evaporated to dryness. The resulting cloudy yellow/orange oil was purified by flash silica gel column chromatography, using silica pre-equilibrated with DCM:TEA (95:5) and eluted with DCM:methanol (0-10%) giving a light brown oil, **1**, (5.7 g, 48%), that was dried over P$_2$O$_5$ *in vacuo* (24 h).

R$_f$ 0.56 (DCM:MeOH:Et$_3$N, 9:1:1)

δ$_H$ (300 MHz; CDCl$_3$) 8.88 (1H, br. s, OH), 7.46-7.16 (9H, m, CH$^{15-17, 19}$), 6.81 (4H, dd, *J* = 11.8, 2.9 Hz, CH^{20}), 3.73 (6H, s, CH^{22}), 3.02 (2H, t, *J* = 6.6 Hz, CH^{12}), 2.22 (2H, t, *J* = 7.4 Hz, CH^{2}), 1.60 (4H, q, *J* = 6.6 Hz, CH$^{3, 11}$), 1.31-1.20 (14H, m, CH$^{4-10}$).

δ$_C$ (75.5 MHz; CDCl$_3$) 180.2 (C^{1}), 158.4 (C^{21}), 145.6 (C^{14}), 136.9 (C^{18}), 130.2 (C^{19}), 128.3 (C^{16}), 127.8 (C^{15}), 126.7 (C^{17}), 113.1 (C^{20}), 85.7 (C^{13}), 63.7 (C^{12}), 55.3 (C^{22}), 37.1 (C^{2}), 30.3 (C$^{5-9, 11}$), 29.8 (C^{4}), 26.5 (C^{10}), 26.3 (C^{3})

IR ν$_{max}$/cm^{-1}: 2923 (m), 2851 (m), 1709 (w), 1606 (m), 1577 (w), 1507 (m), 1246 (m), 1173 (m), 1032 (m)

ES$^-$/MS: 631.2 (M+TFA)$^-$, 553.2 & 555.3 (M+Cl)$^-$

2.   12- O- (4,4'-Dimethoxytrityl)dodecanamide resins
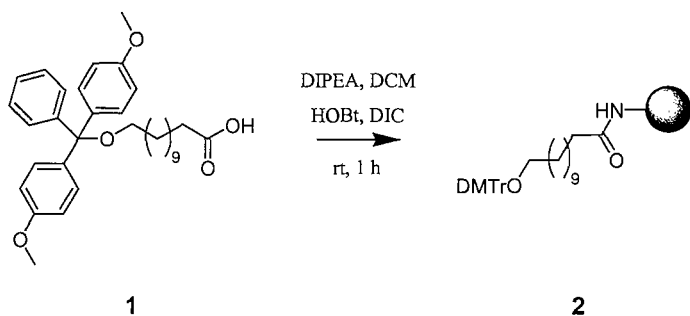
The resin was washed with anhydrous N,N-diisopropylethylamine in DCM (1%, 5 ml, x 5) and dried. **1** Was dissolved in a minimum of anhydrous N,N-diisopropylethylamine in DCM (1%) and to this was added HOBt (5eq) and DIC (5eq). The solution was added to the resin with sufficient N,N-diisopropylethylamine in DCM (1%) to cover the resin. The resin was agitated occasionally to mix the resin and solution. The time the resin was exposed to this mixture depended on the loading required and the resin type. Typical exposure times were 0.2-1 h. The resin was then filtered and washed sequentially with N,N-diisopropylethylamine in DCM (1%, 2 ml, x 5), DCM (2 ml, x 3), Et₂O (2 ml, x 5) and dried *in vacuo*. The resin was then soaked in a capping solution of acetic anhydride, 2,6-lutidine in THF: 1-methylimidazole in THF (1:1) for 1 h. The resin was then washed as previously described and dried to give **2**.

3.   4-(2,4-Dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyacetamido resins
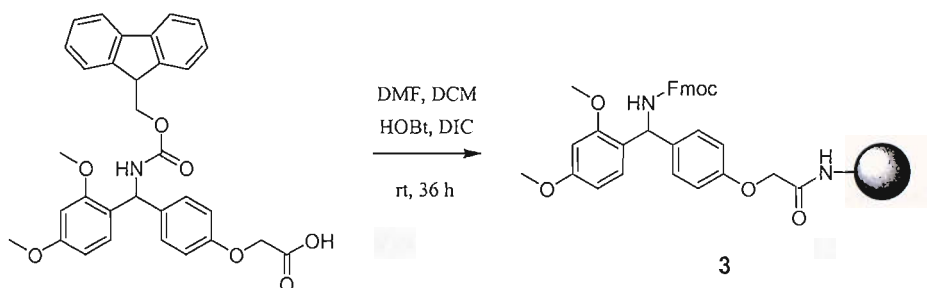
The resin was swollen in a minimum of DCM (30 min). 4-(2,4-Dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyacetic acid (1.5 eq) and HOBt (5 eq) were dissolved in a minimum of DCM with the addition of a few drops of DMF and stirred (RT, 10 min). DIC (5 eq) was added and stirred (RT, 10 min) before addition to the resin. The mixture was agitated occasionally over 36 h. The resin was washed DMF (1ml, x 3), DCM (1ml, x 3), MeOH (1ml, x 3), Et$_2$O (1ml, x 3) and dried *in vacuo* over P$_2$O$_5$.

4. 2,2,2-Trifluoro-*N*-(prop-2-ynyl)acetamide[61]

4
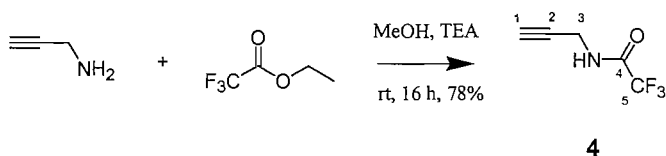
Propargylamine (20.54 ml, 300 mmol) was suspended in anhydrous methanol (150 ml) and triethylamine (125 ml, 900 mmol). Ethyl trifluoroacetate (39 ml, 330 mmol) was added drop-wise. The reaction was allowed to stir under argon overnight at RT. The reaction was evaporated to a yellow oil, this was dissolved in DCM (500 ml), and washed with sat. $NaHCO_3$ (1 x 150 ml), then sat. KCl (1 x 150 ml) and dried over sodium sulphate. Concentration to an oil was followed by purification by vacuum distillation 15 mmHg, 58°C (lit. 51°C, 2.5 torr.) to give a clear, colourless liquid **4** (35.48 g, 78%).

R$_f$ 0.46 (1% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 6.79 (1H, br. s, **NH**), 4.15 (2H, dd, J = 5.5, 2.6 Hz, **CH³**), 2.33 (1H, t, J = 2.6 Hz, **CH¹**)

δ_C (75.5 MHz; CDCl$_3$) 179.7 (**C⁴**), 113.6 (**CF₃**, q, J = 287.7 Hz, **C⁵**), 77.0 (**C²**), 73.2 (**C¹**), 29.7 (**C³**)

IR ν_{max}/cm⁻¹: 3304 (m), 2361 (w), 1706 (s), 1550 (m), 1429 (w), 1364 (w), 1335 (w), 1208 (m), 1154 (s), 1046 (w), 998 (w), 927 (w), 827 (w), 725 (m)

ES⁺/MS: 150.1 (M-H)⁻

5. 3-Phthalimidoprop-1-yne[63]

5
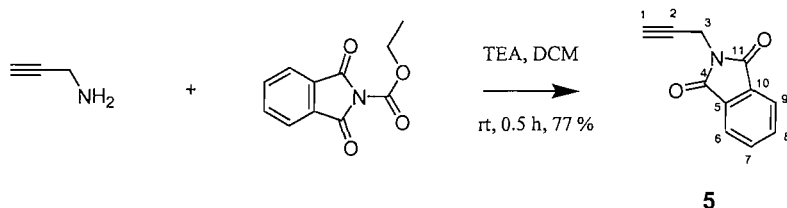
Into DCM (30 ml) were added, in order, triethylamine (5.57 ml, 39.94 mmol), 3-aminoprop-1-yne (2.49 ml, 36.31 mmol) and *N*-carboethoxyphthalimide (7.96 g, 36.31 mmol). After the mixture had been stirred at RT for 20 min DCM (60 ml) was added. The mixture was extracted with water (3 x 50 ml) and the combined aqueous was back-extracted with DCM (30 ml). The combined DCM extracts were decolourised with carbon, filtered through Celite®, and evaporated to dryness. The solid residue was crystallised from acetone. The white crystals were collected by filtration, washed with water and air-dried the give **5** (5.2 g, 77%).

R$_f$ 0.61 (1% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 7.87 (2H, dd, J = 5.5, 2.9 Hz, C**H**^{6}), 7.73 (2H, dd, J = 5.5, 2.9 Hz, C**H**^{7}), 4.44 (2H, d, J = 2.6 Hz, C**H**^{3}), 2.22 (1H, t, J = 2.6 Hz, C**H**^{1})

δ_C (75.5 MHz; CDCl$_3$) 166.9 (**C**$^{4, 11}$), 134.2 (**C**$^{7, 8}$), 131.9 (**C**$^{5, 10}$), 123.5 (**C**$^{6, 9}$), 77.1 (**C**^{2}), 71.4 (**C**^{1}), 26.9 (**C**^{3})

IR ν_{max}/cm^{-1}: 3465 (w), 3294 (m), 2967 (w), 2361 (w), 1769 (m), 1695 (s), 1613 (m), 1468 (m), 1428 (s), 1395 (s), 1352 (s), 1325 (s), 1190 (m), 1164 (w), 1118 (s), 1088 (m), 948 (m), 924 (w), 841 (w), 801 (w), 724 (s), 707 (s)

ES$^+$/MS: 208.1 (M+Na)$^+$

Mpt. 142-144°C (lit. 151-152°C)

6. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-5-iodo-uridine[9,67]
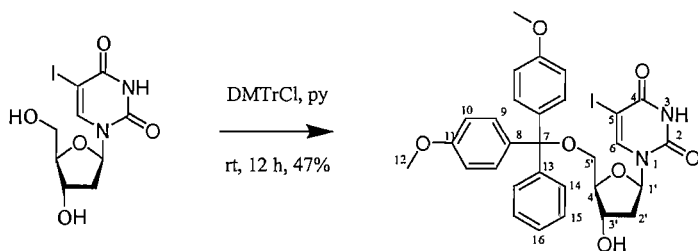
6

To a solution of 5-iodo-2'-deoxyuridine (10 g, 28.24 mmol) in dry pyridine (100 ml) was added 4,4'-dimethoxytrityl chloride (13.40 g, 39.54 mmol) in small portions over a period of 10 min. The reaction mixture was then stirred at RT for 4 h. Upon completion of the reaction MeOH (10 ml) was added and the solvents removed *in vacuo*. The residue was re-dissolved in DCM (500 ml), washed with sat. KCl (2 x 100 ml), CuSO₄ (2 x 50 ml) dried with Na₂SO₄, and the solvent removed *in vacuo*. The pyridine free product was purified by flash silica gel column chromatography eluting with DCM and then 2% MeOH/DCM using the flash master chromatography system to yield **6** as a white foam (8.5 g, 47%).

R$_f$ 0.18 (5% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 11.75 (1H, br, NH), 8.02 (1H, s, CH^6), 7.42-7.20 (9H, m, CH^{9,} $^{14\text{-}16}$), 6.90 (4H, d, J = 9.0 Hz, CH^{10}), 6.10 (1H, t, J = 6.8 Hz, CH$^{1'}$), 5.33 (1H, br, OH$^{3'}$), 4.22 (1H, m, CH$^{3'}$), 3.90 (1H, m, CH$^{4'}$), 3.74 (6H, s, CH^{12}), 3.21-3.14 (2H, m, CH$^{5'}$), 2.30-2.13 (2H, m, CH$^{2'}$)

δ_C (75.5 MHz; DMSO-d6) 160.5 (C^4), 158.0 (C^{11}), 150.0 (C^2), 144.7 (C^6), 135.4 (C^{13}), 129.7 (C^9), 127.9 (C^{15}), 127.6 (C^{14}), 126.7 (C^{16}), 113.2 (C^{10}), 85.8 (C$^{4'}$), 84.8 (C$^{1'}$), 70.4 (C$^{3'}$), 69.8 (C^5), 63.7 (C$^{5'}$), 55.0 (C^{12}), 40.3 (C$^{2'}$)

IR ν_{max}/cm^{-1}: 3401 (br w), 3057 (br w), 1675 (s), 1606 (s), 1507 (s), 1443 (m), 1351 (w), 1300 (w), 1247 (m), 1174 (s), 1147 (m), 1087 (m), 1059 (m), 1029 (s), 962 (m), 825 (s), 790 (m), 754 (m), 726 (m), 700 (m)

ES$^+$/MS: 679.09 (M+Na)$^+$, 655.09 (M-H)$^-$

7.   2'-Deoxy-5'-O-(4,4'-dimethoxytrityl)-5-[3-N-(2,2,2-trifluoroacetamidoprop-1-ynyl]- uridine[62]

2,2,2-Trifluoro-N-(prop-2-ynyl)acetamide,
tetrakis(triphenylphosphine)palladium,
CuI, TEA, DMF

→

rt, 12 h, 79%

6   **7**

To a solution of **6** (5.5 g, 8.38 mmol) and copper iodide(I) (0.32 g, 1.68 mmol) in anhydrous DMF (30 ml) was added triethylamine (5.84 ml, 41.89 mmol) and then **4** (1.39 g, 9.22 mmol). The reaction was stirred in the dark for 10 min and tetrakis(triphenylphosphine)palladium (0.97 g, 0.84 mmol) was added. The yellow reaction mixture was stirred in the absence of light for 4 h and then concentrated *in vacuo* to give a brown oil which was dissolved in DCM (500 ml). The organic phase was washed with 5% EDTA solution (100 ml), 5% NaHCO$_3$ (100 ml), sat. KCl (100 ml) and dried over anhydrous Na$_2$SO$_4$. Evaporation to dryness led to a brownish powder which purified by flash silica gel column chromatography eluting with DCM then 5% MeOH/DCM to give **7** as a yellow foam (4.5 g, 79%).

R$_f$ 0.14 (5% MeOH/DCM)

δ$_H$ (300 MHz; CDCl$_3$) 8.22 (1H, s, CH^6), 7.79 (1H, br. s, NH), 7.18-7.42 (9H, m, CH$^{9-12}$), 6.83 (4H, d, J = 9.2 Hz, CH^{13}), 6.36 (1H, dd, J = 6.3, 7.0 Hz, CH$^{1'}$), 4.60 (1H, m, CH$^{3'}$), 4.15 (1H, d, J = 2.2 Hz, CH$^{4'}$), 3.90 (2H, d, J = 6.3 Hz, CH$_2$), 3.76 (6H, s, 2 x OCH$_3$), 3.34 (2H, m, CH$^{5'}$), 2.55-2.62 (1H, m, CH$^{2'}$), 2.28-2.37 (1H, m, CH$^{2'}$)

δ$_C$ (75.5 MHz; CDCl$_3$) 162.9 (C^4), 158.5 (C^{14}), 156.9 (q, J = 37.3 Hz, C^{19}), 149.5 (C^2), 144.4 (C^8), 143.5 (C^6), 135.4 (C^{21}), 129.9 (C^{12}), 128.0 (C^{10}), 127.8 (C^9), 127.0 (C^{11}), 117.5 (q, J = 287.7 Hz, C^{20}), 113.3 (C^{13}), 99.0 (C^5), 87.5 (C^7), 87.0 (C^{16}), 87.0 (C$^{4'}$), 86.0 (C$^{1'}$), 75.2 (C^{17}), 72.3 (C$^{3'}$), 63.5 (C$^{5'}$), 55.2 (C^{15}), 55.2 (C^{15}), 30.4 (C^{18})

IR v$_{max}$/cm^{-1}: 3071 (w), 2935 (w), 2861 (w), 2359 (w), 1696 (s), 1608 (m), 1508 (m), 1461 (m), 1366 (m), 1282 (m), 1249 (s), 1209 (s), 1174 (s), 1091 (m), 1032 (m), 828 (m), 774 (m), 756 (m), 726 (m)

ES$^+$/MS: 702.3 (M+Na)$^+$

8. 3'-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl)-uridine[62]

Method 1

2-Cyanoethoxy-(*N*,*N*-diisopropylamino)chlorophosphine, DIPEA, DCM

rt, 1h, 77%

7   8
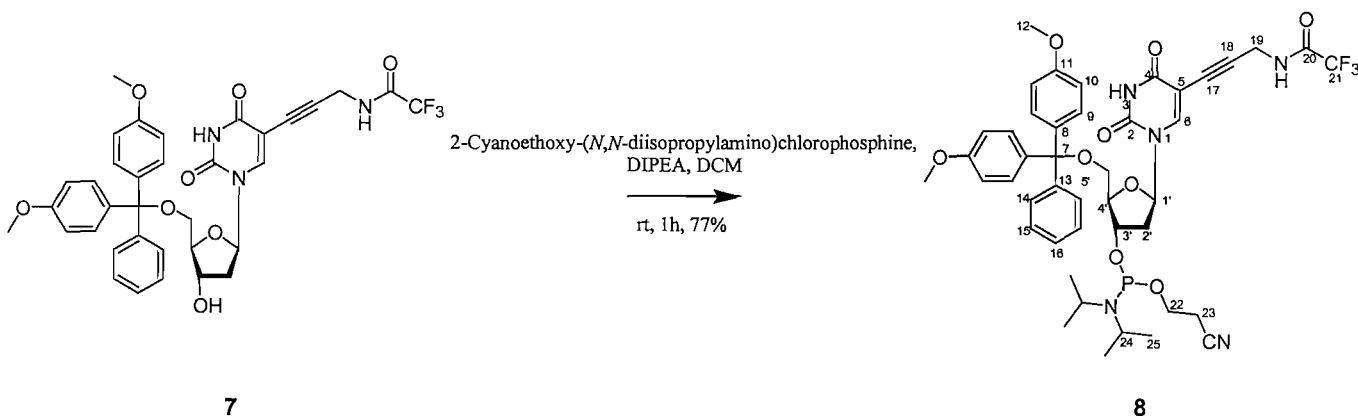
To a solution of **7** (1 g, 1.47 mmol) in anhydrous DCM (10 ml) were added diisopropylethylamine (1.03 ml, 5.89 mmol) and 2-cyanoethyl-*N*,*N*-diisopropyl chlorophosphine (0.48 ml, 2.06 mmol). This was stirred (RT, 1 h) before DCM (100 ml) was then added. This was washed with sat. KCl (100 ml), dried over Na₂SO₄ and the solvents were removed *in vacuo*. The white foam was purified using flash silica gel column chromatography (Flash master) 40% EtOAc/Hexane 10 g column to yield **8** (1 g, 77%).

Method 2

2-Cyanoethyl tetraisopropylphosphorodiamidite, DCM, tetrazole diisopropylammonium salt

rt, 24 h, 80%

7   8

121
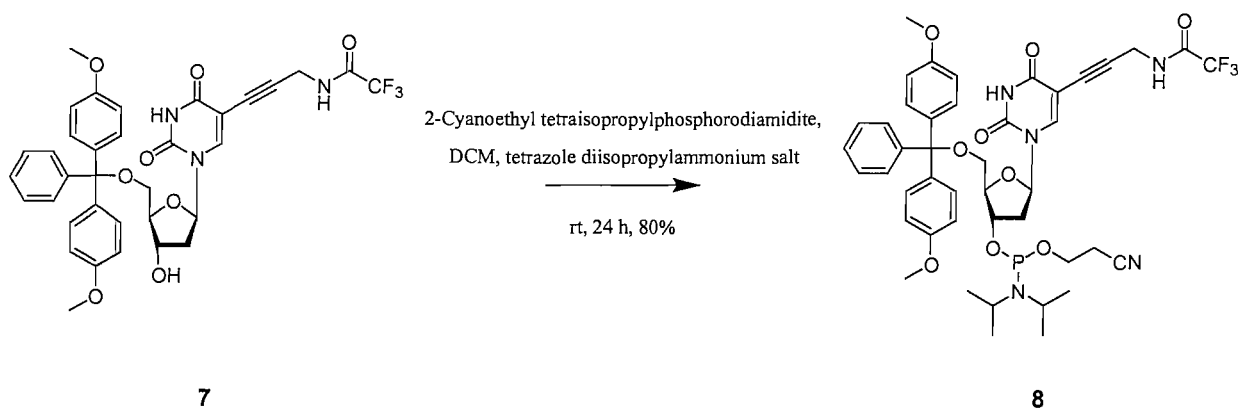
Compound **7** (1 g, 1.47 mmol) was azeotroped with MeCN. The resultant residue was dissolved in alcohol free, anhydrous DCM (7.5 ml) under argon. 2-Cyanoethyl tetraisopropylphosphorodiamidite (0.56 ml, 1.77 mmol) was added together with the tetrazole diisopropylammonium salt (0.13 g, 0.74 mmol), to form a suspension. The reaction was stirred overnight at RT under argon. Following completion, as shown by TLC, the solution was diluted with DCM (20 ml), washed (2 x 50 ml, Na$_2$CO$_3$ $_{(aq)}$, 1 x 50 ml, NaCl $_{(aq)}$), dried over anhydrous Na$_2$SO$_4$ and then the solvents were removed *in vacuo*. The resultant oil was dissolved in EtOAc (20 ml) and precipitated into hexane twice before passing through a small silica gel column (60% EtOAc / Hexane) to yield a white foam **8** (1.03 g, 80%).

R$_f$ 0.53 (70% EtOAc / Hexane)

δ_H (300 MHz; CDCl$_3$) 8.25 & 8.29 (1H, 2 s, CH^6), 7.21-7.46 (9H, m, CH$^{9,\ 14\text{-}16}$), 7.09 (1H, br. s, NH^3), 6.84-6.87 (4H, m, CH^{10}), 6.28-6.32 (1H, m, CH$^{1'}$), 4.65-4.71 (1H, m, CH$^{3'}$), 4.18 & 4.22 (1H, m, CH$^{4'}$), 3.90 (2H, br. s, CH^{19}), 3.80 (3H, s, CH^{12}), 3.79 (3H, s, CH^{12}), 3.53-3.78 (4H, m, CH^{24}, CH^{22}), 3.32-3.49 (2H, m, CH$^{5'}$), 2.57-2.69 (2H, m, CH$^{2'}$, CH^{23}), 2.36-2.47 (2H, m, CH$^{2'}$, CH^{23}), 1.19 (3H, d, J = 6.8 Hz, CH^{25}), 1.18 (6H, d, J = 6.8 Hz, CH^{25}), 1.08 (3H, d, J = 6.8 Hz, CH^{25})

δ_P (90 MHz; CDCl$_3$) 149.33, 149.57

ES$^+$/MS: 902.4 (M+Na)$^+$, 878.4(M-H)$^-$

9. 2'-Deoxy-*N*4-[(di-*n*-butylamino)methylidene]-5-iodo-cytidine
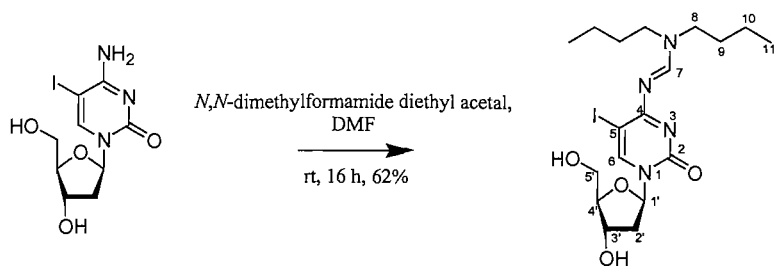
9

5-Iodo-2'deoxycytidine (4.95 g, 14.02 mmol) was dissolved in anhydrous DMF (50 ml). *N,N*-Di-*n*-butylformamide dimethyl acetal (3.95 g, 19.43 mmol) was added and the reaction stirred (16 h) at RT. The solution was evaporated to an oil, re-dissolved in CH_2Cl_2, washed with $NaHCO_3$, dried over Na_2SO_4 and purified by flash silica gel column chromatography eluting with MeOH (0-5%)/DCM. This yielded **9** as a yellow solid (4.3 g, 62%).

R_f 0.06 (5% MeOH/DCM)

δ_H (300 MHz; $CDCl_3$) 8.64 (1H, s, CH^7), 8.32 (1H, s, CH^6), 6.09 (1H, t, J = 6.1 Hz, $CH^{1'}$), 4.53 (2H, br. s, 2 x OH), 4.20 (1H, m, $CH^{3'}$), 4.02 (1H, m, $CH^{4'}$), 3.87 (2H, m, $CH^{5'}$), 3.57 (2H, t, J = 7.5 Hz, CH^8), 3.35 (2H, t, J = 7.2 Hz, CH^8), 2.29-2.50 (2H, m, $CH^{2'}$), 1.54-1.70 (4H, m, CH^9), 1.24-1.42 (4H, m, CH^{10}), 0.94 (6H, q, J = 7.2 Hz, CH^{11})

δ_C (75.5 MHz; $CDCl_3$) 168.8 (C^4), 158.3 (C^7), 155.8 (C^2), 147.9 (C^6), 88.0 ($C^{4'}$), 87.5 ($C^{1'}$), 70.2 (C^5), 70.0 ($C^{3'}$), 61.5 ($C^{5'}$), 52.6 (C^8), 46.4 (C^8), 40.7 ($C^{2'}$), 30.9 (C^9), 28.9 (C^9), 20.2 (C^{10}), 19.7 (C^{10}), 13.8 (C^{11}), 13.7 (C^{11})

IR v_{max}/cm^{-1}: 3400 (w), 3289 (w), 2957 (w), 2931 (w), 2873 (w), 2358 (w), 1738 (w), 1631 (m), 1558 (s), 1478 (m), 1446 (m), 1394 (s), 1364 (s), 1298 (m), 1216 (m), 1173 (w), 1132 (w), 1097 (m), 1056 (s), 993 (m), 952 (m), 857 (w), 786 (m)

ES$^+$/MS: 493.1 (M+H)$^+$

Mpt. 161-163°C

10. 2'-Deoxy-5'-O-(4,4'-dimethoxytrityl)-N4-[(di-n-butylamino)methylidene]-5-iodo-cytidine

DMTrCl, py

rt, 16 h, 54%

9   **10**
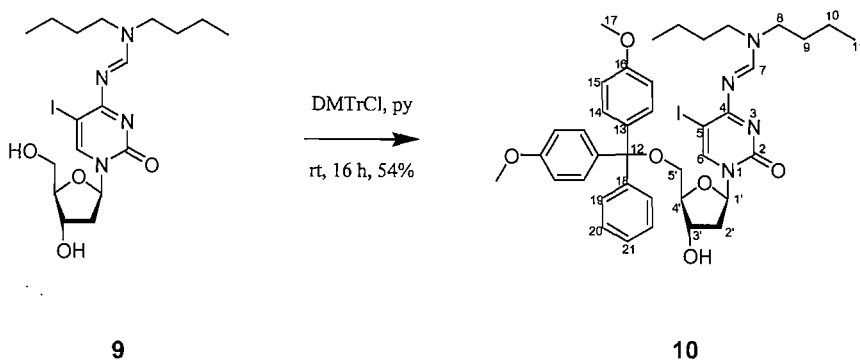
To a solution of **9** (1.95 g, 3.96 mmol) in dry pyridine (20 ml) was added 4,4'-dimethoxytrityl chloride (1.36 g, 4 mmol) in small portions over a period of 10 min, the reaction mixture was allowed to stir at RT overnight. Upon completion of the reaction, MeOH (10 ml) was added and the solvents removed *in vacuo*. The residue was re-dissolved in DCM (60 ml), washed with sat. KCl (2 x 40 ml), sat. CuSO$_4$ (aq) (2 x 40 ml) dried over Na$_2$SO$_4$, and the solvents removed *in vacuo*. The pyridine-free product was purified by flash silica gel column chromatography eluting with 5-10% MeOH / DCM using the flash master chromatography system to yield **10** as a white foam (1.7 g, 54%).
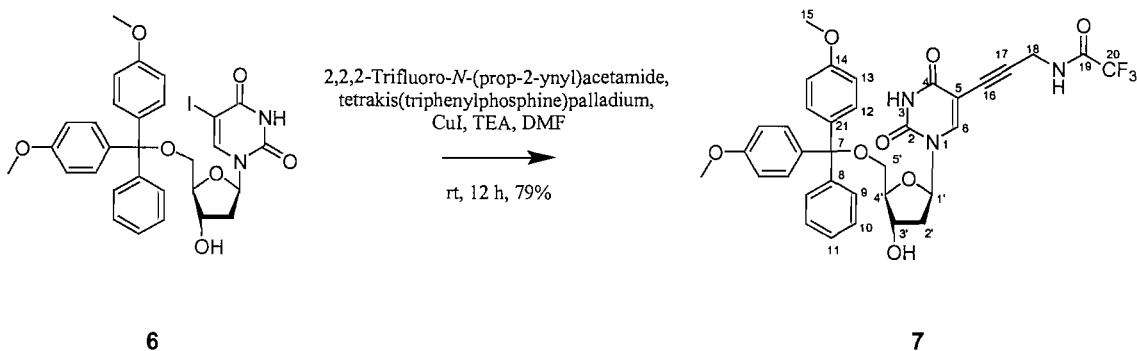
R$_f$ 0.18 (5% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 8.74 (1H, s, CH^6), 8.27 (1H, s, CH^7), 7.19-7.46 (9H, m, CH$^{14,\,19\text{-}21}$), 6.85 (4H, d, J = 8.1 Hz, CH^{15}), 6.36 (1H, dd, J = 5.5, 7.4 Hz, CH$^{1'}$), 4.51 (1H, m, CH$^{3'}$), 4.16 (1H, m, CH$^{4'}$), 3.80 (6H, s, CH^{17}), 3.58-3.63 (2H, m, CH^8), 3.33-3.38 (4H, m, CH^8 and CH$^{5'}$), 2.69-2.76 (1H, m, CH$^{2'}$), 2.15-2.24 (1H, m, CH$^{2'}$), 1.56-1.72 (4H, m, CH^9), 1.24-1.44 (4H, m, CH^{10}), 0.97 (3H, t, J = 7.4 Hz, CH^{11}), 0.95 (3H, t, J = 7.4, CH^{11})

δ_C (75.5 MHz; CDCl$_3$) 168.7 (C^4), 158.5 (C^2), 158.3 (C^7), 155.7 (C^{16}), 146.3 (C^6), 144.5 (C^{18}), 135.6 (C^{13}), 130.1 (C^{14}), 128.0 (C$^{19,\,20}$), 126.8 (C^{21}), 113.3 (C^{15}), 86.9 (C$^{4'}$), 86.7 (C^{12}), 86.3 (C$^{1'}$), 72.4 (C$^{3'}$), 70.2 (C^5), 63.6 (C$^{5'}$), 55.2 (C^{17}), 52.5 (C^8), 46.3 (C^8), 42.3 (C$^{2'}$), 31.0 (C^9), 28.9 (C^9), 20.2 (C^{10}), 19.8 (C^{10}), 13.8 (C^{11}), 13.7 (C^{11})

IR v$_{max}$/cm^{-1}: 2956 (w), 2933 (w), 2872 (w), 2362 (w), 1739 (w), 1645 (m), 1606 (m), 1568 (s), 1508 (m), 1475 (m), 1401 (s), 1365 (m), 1248 (s), 1175 (m), 1094 (m), 1033 (m), 985 (m), 903 (w), 826 (m), 780 (m), 752 (m), 726 (m)

ES$^+$/MS: 795.2 (M+H)$^+$

11. 2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl)]-cytidine
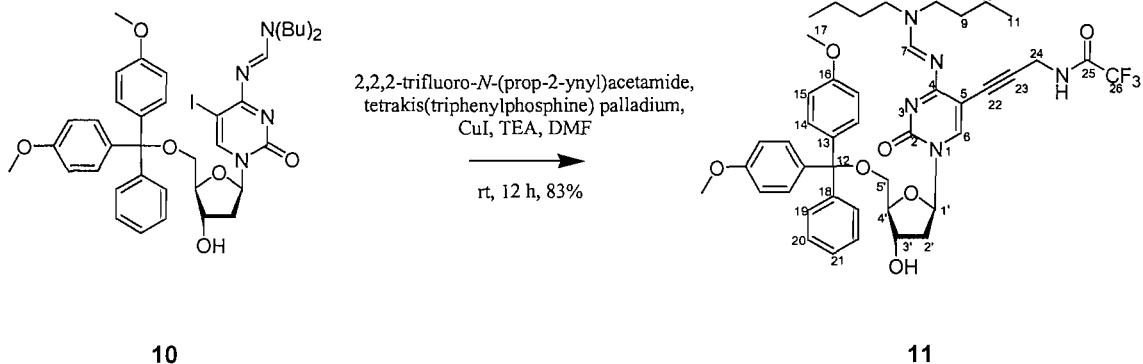
To a solution of **10** (1.4 g, 1.76 mmol) and copper iodide(I) (0.07 g, 0.35 mmol) in anhydrous DMF (25 ml) was added triethylamine (1.23 ml, 8.81 mmol) and then **4** (0.29 g, 1.94 mmol). The reaction was stirred in the dark for 10 min and tetrakis(triphenylphosphine)palladium (0.20 g, 0.18 mmol) was added. The yellow reaction mixture was stirred in the absence of light for 2 h and then concentrated *in vacuo* to give a brown oil which was dissolved in DCM (70 ml). The organic phase was washed with 5% EDTA solution (70 ml), 5% NaHCO₃ (70 ml), sat. KCl (70 ml) and dried over anhydrous Na₂SO₄. Evaporation to dryness led to a brownish powder which purified by flash silica gel column chromatography (flash master) eluting with DCM then 0.5% MeOH/DCM to give **11** (1.2 g, 83%) as a yellow foam.

R$_f$ 0.22 (5% MeOH / DCM)

δ_H (300 MHz; CDCl₃) 8.81 (1H, s, CH⁶), 8.38 (1H, s, CH⁷), 7.18-7.47 (9H, m, CH¹⁴, ¹⁹⁻²¹), 6.85 (4H, d, *J* = 8.8 Hz, CH¹⁵), 6.47 (1H, br. s, NH), 6.37 (1H, t, *J* = 6.4 Hz, CH¹'), 4.53 (1H, br. s, OH³'), 4.16 (1H, m, CH³'), 3.95-4.00 (2H, m, CH²⁴), 3.78 (6H, m, CH¹⁷), 3.59 (2H, t, *J* = 7.4 Hz, CH⁸), 3.28 (5H, m, CH⁴', CH⁵', CH⁸), 2.64-2.76 (1H, m, CH²'), 2.18-2.30 (1H, m, CH²'), 1.55-1.67 (4H, m, CH⁹), 1.26-1.39 (4H, m, CH¹⁰), 0.94 (6H, t, *J* = 7.4 Hz, CH¹¹)

δ_C (75.5 MHz; CDCl₃) 170.6 (C⁴), 158.4 (C¹⁶), 158.3 (C⁷), 154.8 (C²), 145.4 (C⁶), 144.6 (C¹⁸), 135.9 (C¹³), 135.7 (C¹³), 130.0 (C²⁰), 129.9 (C¹⁹), 127.9 (C¹⁴), 126.8 (C²¹), 113.3 (C¹⁵), 98.3 (C⁵), 86.9 (C¹²), 86.7 (C⁴'), 86.3 (C¹'), 85.3 (C²²), 78.6 (C²³), 71.8 (C³'), 63.5 (C⁵'), 55.2 (C¹⁷), 52.4 (C⁸), 45.8 (C⁸), 42.5 (C²'), 30.9 (C⁹), 30.5 (C²⁴), 28.9 (C⁹), 20.0 (C¹⁰), 19.8 (C¹⁰), 13.7 (C¹¹), 13.6 (C¹¹)

IR v_{max}/cm^{-1}: 2959 (w), 2935 (w), 2874 (w), 2342 (w), 1722 (m), 1651 (m), 1583 (s), 1508 (m), 1480 (m), 1398 (s), 1367 (m), 1333 (m), 1249 (m), 1206 (m), 1175 (s), 1111 (m), 1034 (m), 828 (m), 788 (m), 755 (m), 726 (m)

ES$^+$/MS: 818.6 (M+H)$^+$

12. 3'-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-[3-*N*-(2,2,2-trifluoroacetamidoprop-1-ynyl)]-cytidine

11 → **12**

2-Cyanoethoxy-(*N*,*N*-diisopropylamino)chlorophosphine, DIPEA, DCM

rt, 1 h, 58%
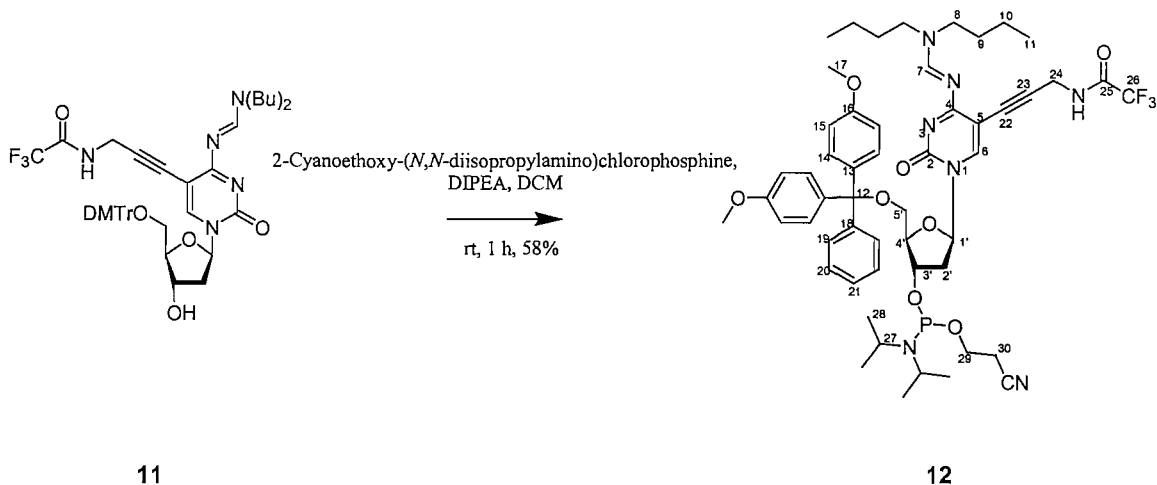
To a solution of **11** (0.55 g, 0.67 mmol) in anhydrous DCM (6 ml) were added diisopropylethylamine (0.47 ml, 2.69 mmol) and 2-cyanoethyl *N*,*N*-diisopropyl chlorophosphine (0.22 ml, 0.94 mmol). This was stirred for 1 h at RT, then EtOAc (60 ml) was added. This was washed with sat. KCl (60 ml), drying over Na_2SO_4 and removal of solvent *in vacuo*. The yellow foam was then purified using a flash master column (10 g) using 40% EtOAc/Hexane to yield **12** (0.4 g, 58%) as a pale yellow foam.

R_f 0.44, 0.37 (90% EtOAc / Hexane)

δ_H (300 MHz; $CDCl_3$) 8.84 (1H, s, CH^6), 8.42 (1H, s, CH^7), 7.19-7.49 (9H, m, $CH^{14,\ 19\text{-}21}$), 6.83-6.88 (4H, m, CH^{15}), 6.35 (1H, m, $CH^{1'}$), 4.55 (1H, m, $CH^{3'}$), 4.19 (1H, m, $CH^{4'}$), 3.79 (6H, s, CH^{17}), 3.25-3.62 (12H, m, $CH^{5'}$, CH^8, CH^{24}, CH^{29}, CH^{27}), 2.56-2.80 (2H, m, $CH^{2'}$, CH^{30}), 2.17-2.44 (2H, m, $CH^{2'}$, CH^{30}), 1.56-1.70 (4H, m, CH^9), 1.30-1.42 (4H, m, CH^{10}), 1.28 (3H, d, $J = 6.6$ Hz, CH^{28}), 1.15 (3H, d, $J = 6.6$ Hz, CH^{28}), 1.14 (3H, d, $J = 6.6$ Hz, CH^{28}), 1.04 (3H, d, $J = 7.0$ Hz, CH^{28}), 0.95 (6H, t, $J = 7.4$ Hz, CH^{11})

δ_P (121.5 MHz; $CDCl_3$) 149.52, 148.99

ES$^+$/MS: 1018.6 (M+H)$^+$, 1040.6 (M+Na)$^+$

13. 2'-Deoxy-*N*4-[(di-*n*-butylamino)methylidene]-5-(phthalimidoprop-1-ynyl)-cytidine

3-Phthalimido-1-propyne,
tetrakis(triphenylphosphine) palladium,
CuI, TEA, DMF
→
rt, 5 h, 84%

9   **13**
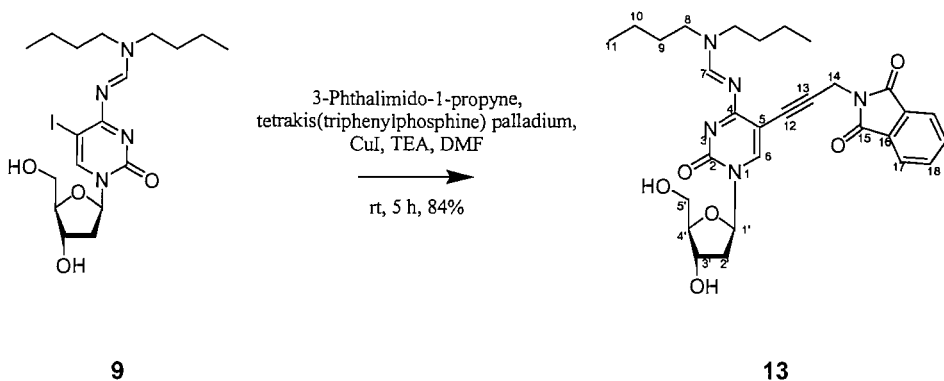
To a solution of **9** (1 g, 2.03 mmol) and copper iodide(I) (0.08 g, 0.41 mmol) in anhydrous DMF (20 ml) was added triethylamine (1.42 ml, 10.16 mmol) and then 3-phthalimido-1-propyne (0.41 g, 2.23 mmol). The reaction was stirred in the dark for 10 min and tetrakis(triphenylphosphine)palladium (0.24 g, 0.2 mmol) was added. The yellow reaction mixture was stirred in the absence of light (5 h) and then concentrated *in vacuo* to give a brown oil which was purified by flash silica gel column chromatography eluting with DCM then 5% MeOH/DCM to give **13** (0.94 g, 84%) as a yellow amorphous solid.

R$_f$ 0.28 (10% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 8.73 (1H, s, C**H**^{6}), 8.16 (1H, s, C**H**^{7}), 7.85 (2H, dd, J = 3.1, 5.5 Hz, C**H**^{17}), 7.73 (2H, dd, J = 3.1, 5.5 Hz, C**H**^{18}), 6.12 (1H, t, J = 6.3 Hz, C**H**$^{1'}$), 4.64 (2H, s, C**H**^{14}), 4.52-4.56 (1H, m, C**H**$^{3'}$), 4.02-4.04 (1H, m, C**H**$^{4'}$), 3.84-3.95 (2H, m, C**H**$^{5'}$), 3.57-3.62 (2H, m, C**H**^{8}), 3.33 (2H, t, J = 7.4 Hz, C**H**^{8}), 2.30-2.50 (2H, m, C**H**$^{2'}$), 1.55-1.63 (4H, m, C**H**^{9}), 1.26-1.37 (4H, m, C**H**^{10}), 0.93 (3H, t, J = 7.4 Hz, C**H**^{11}), 0.91 (3H, t, J = 7.4 Hz, C**H**^{11})

δ_C (75.5 MHz; CDCl$_3$) 167.1 (**C**^{15}), 162.2 (**C**^{4}), 158.1 (**C**^{7}), 154.8 (**C**^{2}), 146.5 (**C**^{6}), 134.1 (**C**^{18}), 132.0 (**C**^{16}), 123.4 (**C**^{17}), 98.5 (**C**^{5}), 88.2 (**C**$^{1'}$), 87.4 (**C**$^{4'}$), 85.7 (**C**^{12}), 77.1 (**C**^{13}), 70.5 (**C**$^{3'}$), 61.9 (**C**$^{5'}$), 52.3 (**C**^{8}), 45.8 (**C**^{8}), 40.9 (**C**$^{2'}$), 30.9 (**C**^{9}), 29.0 (**C**^{9}), 28.1 (**C**^{14}), 20.0 (**C**^{10}), 19.7 (**C**^{10}), 13.7 (**C**^{11}), 13.6 (**C**^{11})

IR ν$_{max}$/cm^{-1}: 3373 (br.), 2957 (w), 2932 (w), 2872 (w), 1771 (w), 1716 (s), 1650 (s), 1582 (s), 1477 (s), 1392 (s), 1344 (s), 1267 (m), 1209 (m), 1092 (s), 1055 (m), 989 (m), 941 (m), 870 (w), 788 (m), 725 (m), 624 (m)

ES$^+$/MS: 550.5 (M+H)$^+$

2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-(phthalimidoprop-1-ynyl)-cytidine

13   **14**
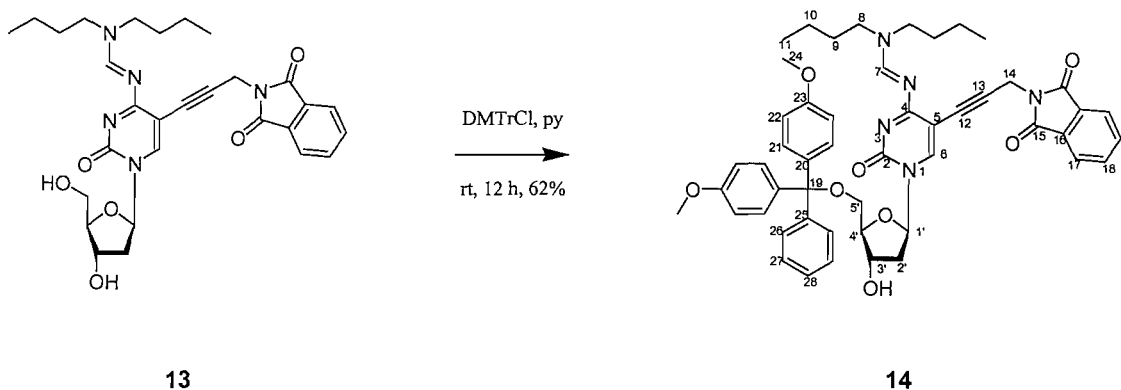
To compound **13** (0.94 g, 1.71 mmol) in dry pyridine (10 ml) was added 4,4'-dimethoxytrityl chloride (0.61 g, 1.80 mmol) in small portions over a period of 10 min, the reaction mixture was allowed to stir at RT overnight. Upon completion, MeOH (1 ml) was added and the solvents removed *in vacuo*. The residue was re-dissolved in DCM (60 ml), washed with sat. KCl (2 x 40 ml), $CuSO_4$ (2 x 40 ml) dried with Na_2SO_4, and the solvent removed *in vacuo*. The pyridine free product was purified by flash silica gel column chromatography eluting with MeOH (0-10%) / DCM to yield **14** as a white foam (0.9 g, 62%).

R$_f$ 0.17 (5% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 8.78 (1H, s, CH^6), 8.13 (1H, s, CH^7), 7.81 (2H, dd, J = 3.0, 5.3 Hz, CH^{17}), 7.71 (2H, dd, J = 3.0, 5.3 Hz, CH^{18}), 7.17-7.46 (9H, m, CH$^{21, 26-28}$), 6.86 (4H, d, J = 7.3 Hz, CH^{22}), 6.29 (1H, t, J = 6.3 Hz, CH$^{1'}$), 4.45 (3H, m, CH$^{3'}$, CH^{14}), 4.12 (1H, m, CH$^{4'}$), 3.80 (6H, s, CH^{24}), 3.58 (2H, m, CH^8), 3.30-3.43 (4H, m, CH$^{5'}$, CH^8), 2.65-2.71 (1H, m, CH$^{2'}$), 2.14-2.21 (1H, m, CH$^{2'}$), 1.52-1.62 (4H, m, CH^9), 1.23-1.36 (4H, m, CH^{10}), 0.94 (3H, t, J = 7.3 Hz, CH^{11}), 0.86 (3H, t, J = 7.3 Hz, CH^{11})

δ_C (75.5 MHz; CDCl$_3$) 170.8 (C^4), 166.8 (C^{15}), 158.5 (C^6), 158.1 (C^{23}), 154.8 (C^2), 144.7 (C^{25}), 144.6 (C^7), 135.7 (C^{20}), 133.9 (C^{18}), 132.1 (C^{16}), 130.0 (C$^{27, 21}$), 127.9 (C^{26}), 126.8 (C^{28}), 123.3 (C^{17}), 113.3 (C^{22}), 98.7 (C^5), 86.9 (C^{19}), 86.7 (C$^{1'}$), 86.1 (C$^{4'}$), 85.8 (C^{12}), 77.2 (C^{13}), 72.2 (C$^{3'}$), 63.6 (C$^{5'}$), 55.2 (C^{24}), 52.2 (C^8), 45.6 (C^8), 42.1 (C$^{2'}$), 30.9 (C^9), 29.1 (C^9), 27.9 (C^{14}), 19.9 (C^{10}), 19.8 (C^{10}), 13.7 (C^{11}), 13.6 (C^{11})

ES$^+$/MS: 852.4 (M+H)$^+$

15. 3'-*O*-(2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-*N*4-[(di-*n*-butylamino)methylidene]-5-(phthalimidoprop-1-ynyl)-cytidine

2-Cyanoethoxy-(*N*,*N*-diisopropylamino)chlorophosphine, DIPEA, DCM
→
rt, 1 h, 79%

14   **15**
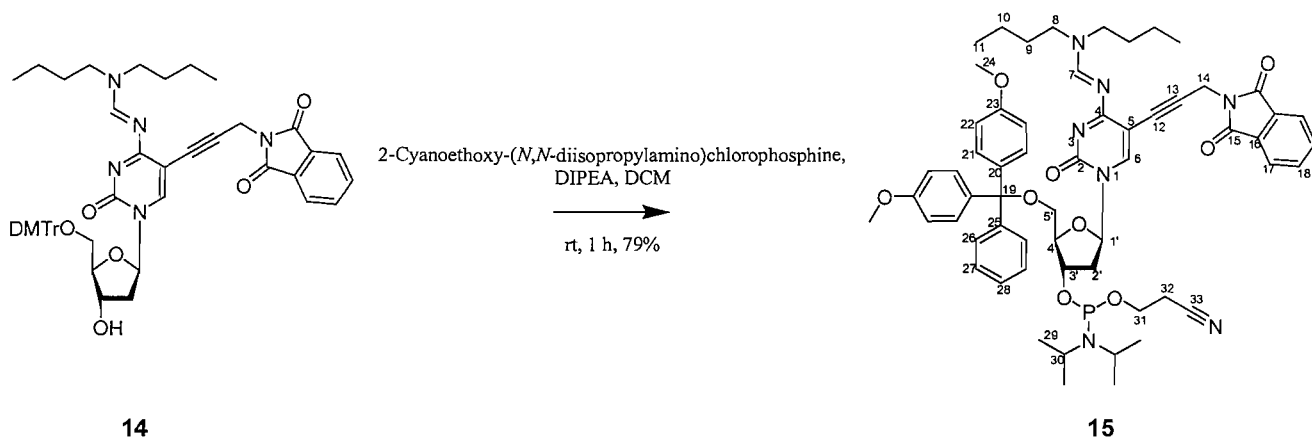
To compound **14** (0.82 g, 0.96 mmol) in anhydrous DCM (10 ml) were added diisopropylethylamine (0.42 ml, 2.40 mmol) and 2-cyanoethoxy-(*N*,*N*-diisopropylamino)chlorophosphine (0.24 ml, 1.06 mmol). Following completion (1 h), as shown by TLC, the solution was diluted with DCM (20 ml), washed (2 x 50 ml, Na_2CO_3 (aq), 50 ml, NaCl (aq), dried over anhydrous Na_2SO_4 and then the solvents were removed *in vacuo*. The resultant oil was dissolved in EtOAc (5 ml) and precipitated into hexane (250 ml) twice. The solid was then purified by flash silica gel column chromatography 90% EtOAc / hexane to yield **15** (0.8 g, 79%) as a white foam.

R_f 0.46, 0.40 (90% EtOAc/Hexane)

δ_H (300 MHz; $CDCl_3$) 8.72 (1H, s, CH^6), 8.05 (1H, s, CH^7), 7.73 (2H, dd, J = 5.5, 3.0 Hz, CH^{17}), 7.63 (2H, dd, J = 5.5, 3.0 Hz, CH^{18}), 7.11-7.40 (9H, m, $CH^{21, 26-28}$), 6.76-6.80 (4H, m, CH^{22}), 6.18-6.24 (1H, m, $CH^{1'}$), 4.32-4.46 (3H, m, $CH^{3'}$, CH^{14}), 4.12 (1H, m, $CH^{4'}$), 3.72 (7H, m, CH^{24}, CH^{31}), 3.43-3.59 (5H, m, CH^8, CH^{30}, CH^{31}), 3.22-3.33 (4H, m, $CH^{5'}$, CH^8), 2.58-2.70 (1H, m, $CH^{2'}$), 2.54 (1H, t, J = 6.5 Hz, CH^{32}), 2.36 (1H, t, J = 6.5 Hz, CH^{32}), 2.03-2.16 (1H, m, $CH^{2'}$), 1.43-1.55 (4H, m, CH^9), 1.14-1.29 (4H, m, CH^{10}), 1.09 (9H, m, CH^{29}), 0.99 (3H, d, J = 6.8 Hz, CH^{29}), 0.86 (3H, t, J = 7.3 Hz, CH^{11}), 0.77 (3H, dt, J = 7.3, 2.0 Hz, CH^{11})

δ_P (90 MHz; $CDCl_3$) 148.9, 149.4

ES⁺/MS: 1052.4 (M+H)⁺

16. 5-(3-Aminoprop-1-ynyl)-2'-deoxycytidine
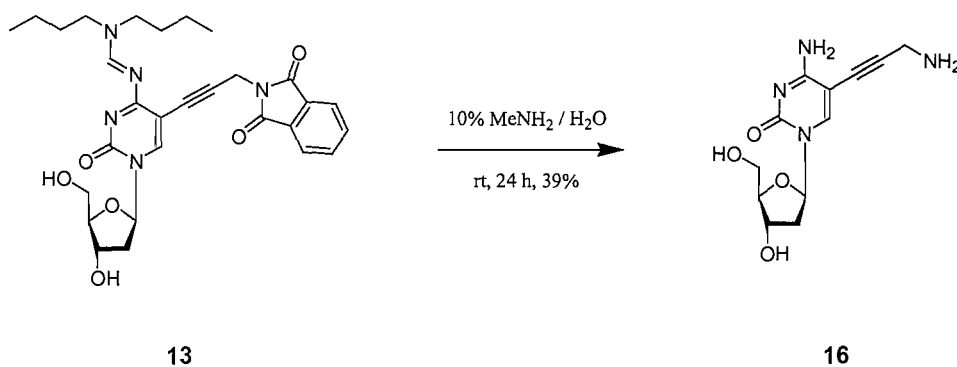
Compound **13** (0.05 g, 0.09 mmol) was dissolved in 10% $MeNH_2$ in water (2 ml) and left for 24 h. TLC analysis showed the reaction to be finished and the solvent was removed *in vacuo*. The residue was then purified by flash silica gel column chromatography (20-40% MeOH / DCM) to yield **16** as an amorphous solid (0.01 g, 39%).

R_f 0.07 (20% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 8.07 (1H, s, CH^6), 6.11 (1H, ap.t, $J = 6.5$ Hz, $CH^{1'}$), 5.19 (1H, br.s, $OH^{3'}$), 5.07 (1H, br.s, $OH^{5'}$), 4.20 (1H, m, $CH^{3'}$), 3.78 (1H, dt, $J = 3.4, 6.5$ Hz, $CH^{4'}$), 3.60 (1H, dd, $J = 3.4, 11.8$ Hz, $CH^{5'}$), 3.54 (1H, dd, $J = 3.4, 11.8$ Hz, $CH^{5'}$), 3.44 (2H, s, CH^{10}), 2.14 (1H, ddd, $J = 3.4, 5.9, 13.2$ Hz, $CH^{2'}$), 1.96 (1H, ddd, $J = 6.3, 6.5, 13.2$ Hz, $CH^{2'}$)

δ_C (75.5 MHz; DMSO-d6) 164.3 (C^4), 153.4 (C^2), 143.2 (C^6), 90.0 (C^8), 87.3 ($C^{4'}$), 85.2 ($C^{1'}$), 73.1 (C^7), 70.9 (C^5), 70.1 ($C^{3'}$), 61.0 ($C^{5'}$), 40.7 ($C^{2'}$), 31.5 (C^9)

UV/vis (H_2O): 260 nm (5.0 x 10^3 $cm^{-1}mol^{-1}$)

ES^+/MS: 303.1 $(M+Na)^+$, 583.3 $(2M+Na)^+$

17.   Methyl-2-deoxy-D-ribofuranoside[112]
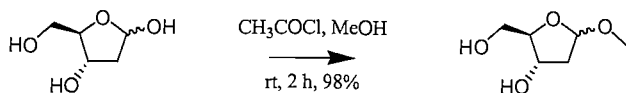
17

2-Deoxy-D-ribose (25 g, 186.38 mmol) was dissolved in anhydrous MeOH (300 ml) under argon then 0.7% (wt:wt) HCl/anhydrous MeOH (prepared by the cautious addition of acetyl chloride (1 ml) to anhydrous MeOH (60 ml) with stirring and external cooling with ice) was added and the reaction was stirred at RT for 2 h. TLC analysis after that time showed that the reaction was complete. The reaction mixture was neutralized by the portion-wise addition of NaHCO$_3$ (5 g), the resultant was filtered and concentrated *in vacuo*. The cloudy brown oil was dissolved in DCM and the resultant precipitate was removed by filtration, the filtrate was then concentrated *in vacuo* to afford **17** as a light brown oil that solidified after storage at −18°C for 24 h (27.04 g, 98%).

R$_f$ 0.19, 0.28 (5% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 5.08 (1H, m, **CH^1**), 4.46 & 4.14 (1H, 2 x m, **CH^3**), 4.08 & 4.00 (1H, 2 x m, **CH^4**), 3.69-3.58 (2H, 2 x m, **CH^5**), 3.37 & 3.36 (3H, 2 x s, **OCH$_3$**), 3.32 & 3.19 (1H, 2 x m, **OH^3**), 2.64 (1H, m, **OH^5**), 2.26-1.84 (2H, m, **CH^2**)

δ_C (75.5 MHz; CDCl$_3$) 105.9 & 105.8 (**C^1**), 87.7 & 87.5 (**C^4**), 73.0 & 72.3 (**C^3**), 63.1 & 63.3 (**C^5**), 55.8 & 55.4 (**OCH$_3$**), 42.6 & 41.8 (**C^2**)

IR ν_{max}/cm^{-1}: 3368 (br,m), 2991 (w), 2920 (w), 2836 (w), 2361 (w), 1648 (w), 1443 (w), 1364 (w), 1276 (w), 1206 (m), 1176 (w), 1083 (s), 1027 (s), 964 (m), 946 (m), 855 (m), 827 (m), 764 (m), 740 (m)

ES$^+$/MS: 171.2 (M+Na)$^+$

18. Methyl-3,5-di-*O*-*p*-toluoyl-2-deoxy-D-ribofuranoside[112]

17   **18**
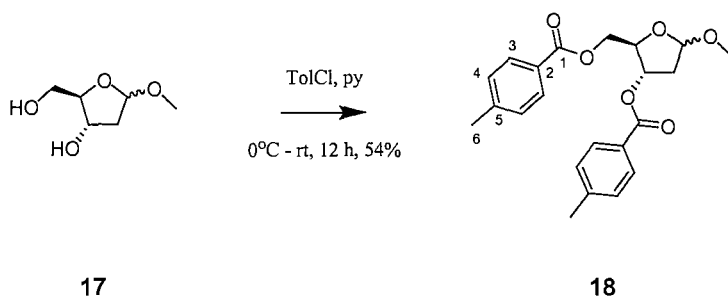
Compound **17** (27.04 g, 182.51 mmol) was co-evaporated with anhydrous pyridine (3 x 15 ml) and dissolved in anhydrous pyridine (100 ml). The reaction mixture was stirred under argon at 0°C with cooling over ice and *p*-toluoyl chloride (53.1 ml, 401.52 mmol) was added drop-wise, with stirring over 30 min. The mixture was allowed to warm to RT then heated to 65°C and stirred at that temperature for 4 h. The reaction mixture was allowed to cool to RT overnight with stirring. At that time TLC analysis showed the reaction to be complete. The reaction mixture was partitioned between water (150 ml) and DCM (100 ml) and the aqueous phase was extracted with further DCM (2 x 100 ml). The combined organic extracts were washed with 2 M $HCl_{(aq)}$ (2 x 100 ml) and then 6% $NaHCO_{3(aq)}$ (200 ml), dried over Na_2SO_4 and concentrated *in vacuo*. The resulting yellow oil was co-evaporated with toluene (3 x 15 ml) to remove residual pyridine, then with diethyl ether (3 x 15 ml) and dried *in vacuo* to constant mass. This afforded methyl **18** a 1:1 mixture of anomers as a yellow oil (38 g, 54%). The oil **18** was used without further purification. A small sample was purified to give the following analysis.

R_f 0.69 (1% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 7.98 (2H, d, $J = 8.1$ Hz, CH^3), 7.80 (2H, d, $J = 8.1$ Hz, CH^3), 7.25 (2H, d, $J = 8.1$ Hz, CH^4), 7.14 (2H, d, $J = 8.1$ Hz, CH^4), 5.65 (0.5H, dd, $J = 2.9$, 4.8 Hz, CH$^{3'}$), 5.62 (0.5H, dd, $J = 3.3$, 4.8 Hz, CH$^{3'}$), 5.50 (1H, m, CH$^{1'}$), 4.97 (1H, dd, $J = 2.6$, 2.6 Hz, CH$^{1'}$), 4.09 (1H, dd, $J = 1.7$, 12.9 Hz, CH$^{4'}$), 3.95 (2H, m, CH$^{5'}$), 3.42 (3H, s, CH$_3$), 2.42 (3H, s, CH^6), 2.35 (3H, s, CH^6), 2.37 (1H, m, CH$^{2'}$), 2.14 (1H, m, CH$^{2'}$)

δ_C (75.5 MHz; CDCl$_3$) 165.9 (C^1), 165.4 (C^1), 143.8 (C^5), 143.6 (C^5), 129.7 (C^3), 129.6 (C^3), 129.1 (C^4), 128.9 (C^4), 127.3 (C^2), 127.1 (C^2), 98.7 (C$^{3'}$), 68.3 (C$^{4'}$), 66.4 (C$^{4'}$), 60.9 (CH$_2$$^{5'}$), 55.1 (C^7), 31.3 (CH$_2$$^{2'}$), 21.6 (C^6), 21.6 (C^6)

IR ν_{max}/cm^{-1}: 2930 (w), 1711 (s), 1611 (m), 1578 (w), 1509 (w), 1451 (w), 1409 (w), 1372 (m), 1272 (s), 1210 (m), 1179 (m), 1102 (s), 1068 (s), 1044 (m), 1019 (m), 976 (m), 926 (m), 858 (m), 841 (m), 752 (s)

ES$^+$/MS: 407.1 (M+Na)$^+$

Mpt. 60-64°C

19.   1-Chloro-2-deoxy-3,5-di-*O*-*p*-toluoyl-α-*erythro*-pentofuranose[112]
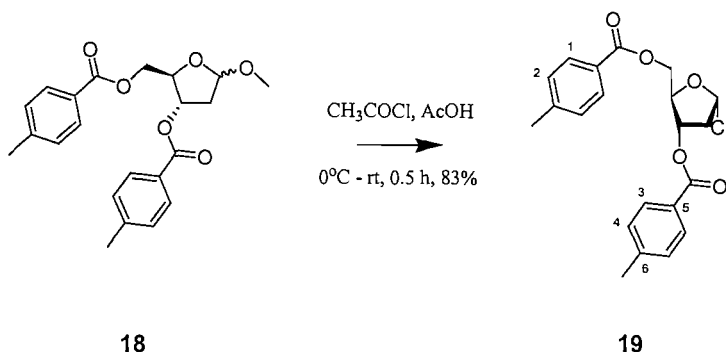
Crude **18** (36.08 g, 93.85 mmol) was dissolved in glacial acetic acid (42 ml) and stirred at RT. A pre-saturated solution of HCl in glacial acetic acid (70 ml) (prepared by adding acetyl chloride (16.3 ml) to a mixture of glacial AcOH (81.0 ml) and water (4 ml), at 0°C to RT) was added in a steady stream with vigorous stirring. Sufficient acetyl chloride (~20 ml) to cause a thick white precipitate was added drop-wise. The mixture was then cooled to 0°C on ice and shaken vigorously with diethyl ether (70 ml, 0°C). The white precipitate was obtained by filtration and washed with diethyl ether (2 x 70 ml, 0°C). The precipitate was quickly transferred to a desiccator and dried *in vacuo* over KOH affording **19** (30.12 g, 77.46 mmol, 83%), as a micro-crystalline white solid.

R$_f$ 0.21 (1% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 8.07 (2H, d, J = 8.1 Hz, ArH^3), 7.98 (2H, d, J = 8.1 Hz, ArH^1), 7.33 (4H, 2d, J = 8.1, 8.1 Hz, ArH^{2,4}), 6.55 (1H, d, J = 4.8 Hz, CH$^{1'}$), 5.64 (1H, ddd, J = 1.1, 2.9, 6.3 Hz, CH$^{3'}$), 4.93 (1H, dd, J = 3.3, 3.7 Hz, CH$^{4'}$), 4.76 (1H, dd, J = 3.3, 12.1 Hz, CH$^{5'}$), 4.67 (1H, dd, J = 3.7, 12.1 Hz, CH$^{5'}$), 2.99-2.90 (2H, m, CH$^{2'}$), 2.50 (3H, s, CH$_3$), 2.49 (3H, s, CH$_3$)

δ_C (75.5 MHz; CDCl$_3$) 166.4, 166.0 (C=O), 144.3, 144.0 (C^6), 130.1 (C^1), 129.8 (C^1), 129.4 (C^2), 129.2 (C^2), 126.9 (C^5), 126.8 (C^5), 95.5 (C$^{1'}$), 84.9 (C$^{4'}$), 73.7 (C$^{3'}$), 63.7 (C$^{5'}$), 44.7 (C$^{2'}$) 21.9 (2xCH$_3$)

IR v_{max}/cm^{-1}: 1710 (s), 1606 (m), 1452 (w), 1393 (w), 1280 (s), 1180 (s), 1094 (s), 1012 (m), 971 (m), 872 (m), 836 (m), 799 (m), 750 (s), 695 (m), 659 (m)

ES$^+$/MS: 799.0 & 801.3 (2M+Na)$^+$

Mpt. 108-110°C

20. Ethyl-2-cyano-4,4-diethoxybutanoate[113,114]
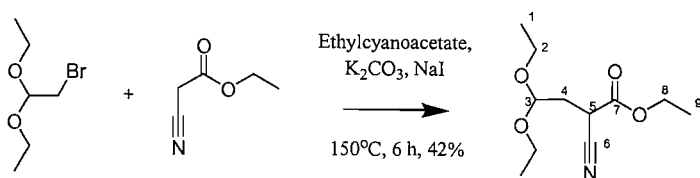
20

Bromoacetaldehyde diethyl acetal (53.5 ml, 0.388 mol), ethyl cyanoacetate (235 ml, 2.21 mol), anhydrous potassium carbonate (dried @ 1 mmHg / 160°C for 12 h) (53.6 g, 0.388 mol) and sodium iodide (2.9 g, 19.4 mmol) were heated to 150°C (oil bath) over 2 h, stirred for 6 h and allowed to cool overnight to RT. The reaction mixture was diluted with de-ionised water (600 ml), sufficient to dissolve the inorganic solids) and extracted with ether (4 x 200 ml). The combined organic extracts were dried over sodium sulphate (50 g) and concentrated *in vacuo* to give a dark brown oil. This was purified by double distillation under reduced pressure through a 15 cm Fenske column yielding **20** (62.07 g, 42%) as a colourless oil (115-125°C @ 2 mmHg).

R_f 0.35 (DCM)

δ_H (300 MHz; CDCl$_3$) 4.69 (1H, dd, J = 5.1, 6.3 Hz, C**H**^{3}), 4.25 (2H, q, J = 7.0 Hz, C**H**^{8}), 3.63-3.76 (3H, m, C**H**^{5}, C**H**^{2}), 3.48-3.59 (2H, m, C**H**^{2}), 2.15-2.33 (2H, m, C**H**^{4}), 1.32 (3H, t, J = 7.0 Hz, C**H**^{9}), 1.22 (3H, t, J = 7.0 Hz, C**H**^{1}), 1.20 (3H, t, J = 7.0 Hz, C**H**^{1})

δ_C (75.5 MHz; CDCl$_3$) 165.8 (**C**^{7}), 116.3 (**C**^{6}), 99.9 (**C**^{3}), 62.8 (**C**^{8}), 62.7 (**C**^{2}), 62.6 (**C**^{2}), 33.6 (**C**^{4}), 33.5 (**C**^{5}), 15.2 (**C**^{1}), 13.9 (**C**^{9})

IR ν_{max}/cm^{-1}: 2978 (m), 2933 (m), 2899 (m), 2251 (w), 1744 (s), 1444 (m), 1371 (m), 1334 (m), 1301 (m), 1259 (m), 1214 (m), 1188 (m), 116 (m), 1123 (s), 1056 (s), 855 (m), 815 (m), 763 (m)

ES$^+$/MS: 252.2 (M+Na)$^+$

21. 6-Amino-5-(2,2-diethoxy-ethyl)-2-(mercapto)-3*H*-pyrimidin-4-one[113,122]

20   **21**
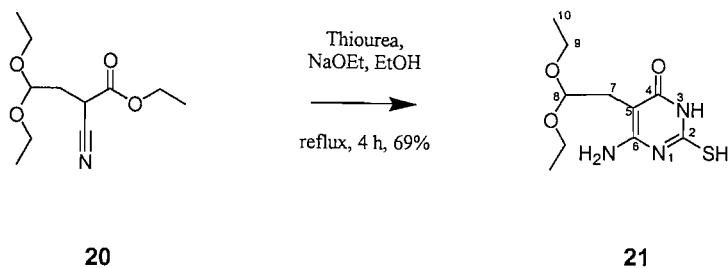
Compound **20** (130 g, 0.567 mol) was dissolved in 0.5 M NaOEt/EtOH (800 ml) and stirred at RT. A suspension of thiourea (44.3 g, 567 mmol) in 0.5 M NaOEt/EtOH (600 ml) was added portion-wise by decantation to the stirred reaction mixture. The reaction was heated at reflux (4 h) then allowed to cool and concentrated *in vacuo* to give a coloured solid. The solid was dissolved in water and glacial acetic acid added till pH 6. This resulted in a white precipitate that was obtained by filtration and washed by displacement with i-PrOH and then Et_2O and re-crystallised from MeOH. The solid was recovered by suction, washed with i-PrOH, then Et_2O and dried *in vacuo* over P_2O_5. This gave **21** (101 g, 69%) as a white crystalline solid.

R$_f$ 0.24 (5% MeOH / DCM)

δ_H (300 MHz; DMSO-d6) 11.77 (1H, s, **SH**), 11.46 (1H, s, **NH^3**), 6.08 (2H, s, **NH$_2$**), 4.50 (1H, t, J = 5.5 Hz, **CH^8**), 3.58 (2H, dq, J = 7.0, 9.6 Hz, **CH^9**), 3.39 (2H, dq, J = 7.0, 9.6 Hz, **CH^9**), 2.43 (2H, d, J = 5.5 Hz, **CH^7**), 1.06 (6H, t, J = 7.0 Hz, **CH^{10}**)

δ_C (75.5 MHz; DMSO-d6) 172.8 (**C^2**), 161.7 (**C^4**), 151.8 (**C^6**), 101.6 (**C^8**), 85.6 (**C^5**), 61.6 (**C^9**), 27.8 (**C^7**), 15.3 (**C^{10}**)

IR v_{max}/cm^{-1}: 3207 (br), 1633 (m), 1551 (s), 1474 (m), 1438 (m), 1370 (m), 1339 (m), 1370 (m), 1339 (m), 1295 (m), 1216 (m), 1178 (m), 1100 (m), 1041 (s), 981 (s), 830 (m), 782 (m), 605 (m)

ES$^+$/MS: 282.2 (M+Na)$^+$

22. 6-Amino-5-(2,2-diethoxy-ethyl)-3*H*-pyrimidin-4-one[113]

21   **22**
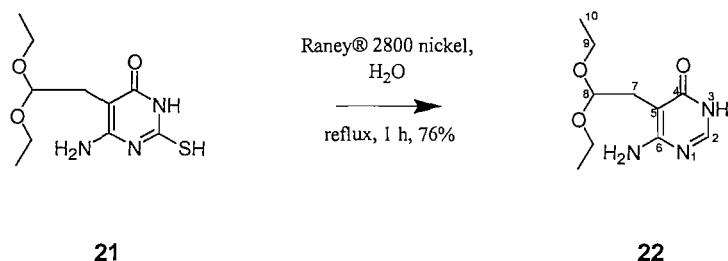
Compound **21** (1 g, 3.86 mmol) was dissolved in de-ionised water (50 ml) and c.NH₃(aq)
(4 ml) was added followed by Raney® 2800 Nickel (4 g). The resultant was heated at
reflux (1 h) and hot-filtered through a sintered glass funnel. The Raney® 2800 nickel
residue was washed by displacement with MeOH (4 x 5 ml) and the combined filtrate and
washings were concentrated *in vacuo*. The amorphous residue was crystallized from
MeOH (10 ml) affording **22** (0.67 g, 76%) as a white crystalline solid.

$\mathbf{R_f}$ 0.39 (5% MeOH / DCM)

δ_H (300 MHz; DMSO-d6) 11.54 (1H, s, $\mathbf{NH^3}$), 7.70 (1H, s, $\mathbf{CH^2}$), 6.11 (2H, s, $\mathbf{NH_2}$), 4.54
(1H, t, J = 5.5Hz, $\mathbf{CH^8}$), 3.59 (2H, dq, J = 9.6, 7.0 Hz, $\mathbf{CH^9}$), 3.39 (2H, dq, J = 9.6, 7.0
Hz, $\mathbf{CH^9}$), 2.51 (2H, d, J = 5.5 Hz, $\mathbf{CH^7}$), 1.04 (6H, t, J = 7.0 Hz, $\mathbf{CH^{10}}$)

δ_C (75.5 MHz; DMSO-d6) 161.7 ($\mathbf{C^4}$), 161.4 ($\mathbf{C^6}$), 147.0 ($\mathbf{C^2}$), 101.8 ($\mathbf{C^8}$), 93.1 ($\mathbf{C^5}$), 61.4
($\mathbf{C^9}$), 28.7 ($\mathbf{C^7}$), 15.3 ($\mathbf{C^{10}}$)

$\mathbf{ES^+/MS}$: 250.3 (M+Na)⁺

23.   3,7-Dihydro-pyrrolo[2,3-*d*]pyrimidin-4-one[113]
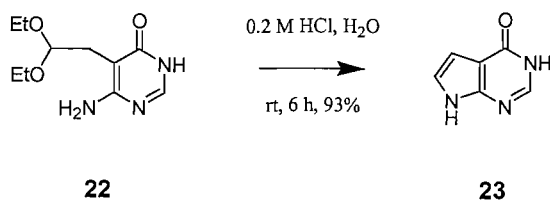
Compound **22** (20.01 g, 88.05 mmol) was suspended in 0.2 M HCl$_{(aq)}$ (880 ml) and agitated at RT for 6 h. The resulting solid was then filtered and rinsed by displacement with de-ionised water (2 x 40 ml) then ether (2 x 40 ml) and dried overnight *in vacuo* over P_2O_5 affording **23** (11.07 g, 93%) as an amorphous white solid.

R$_f$ 0.22 (10% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 11.84 (1H, br. s, N**H**), 11.76 (1H, br. s, N**H**), 7.83 (1H, s, C**H**^{2}), 7.03 (1H, dd, *J* = 2.7, 3.0 Hz, C**H**^{6}), 6.44 (1H, dd, *J* = 2.2, 3.0 Hz, C**H**^{5})

δ_C (75.5 MHz; DMSO-d6) 158.5 (**C**^{7a}), 148.1 (**C**^{4}), 143.2 (**C**^{2}), 120.4 (**C**^{6}), 107.6 (**C**^{4a}), 102.0 (**C**^{5})

IR ν_{max}/cm^{-1}: 2862 (br.), 1663 (m), 1571 (m), 1512 (m), 1474 (m), 1431 (m), 1373 (m), 1335 (m), 1304 (m), 1230 (m), 1150 (w), 1119 (m), 1093 (m), 939 (w), 916 (w), 883 (m), 861 (m), 728 (s), 705 (m), 693 (s), 614 (s)

ES$^+$/MS: 271.04 (2M+H)$^+$, 293.04 (2M+Na)$^+$

Mpt. Decomposition >290°C

24.   4-Chloro-7*H*-pyrrolo[2,3-*d*]pyrimidine[113]

23   **24**
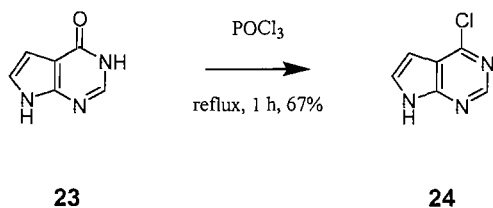
Compound **23** (11.07 g, 135.12 mmol) was suspended in POCl$_3$ (115 ml) and heated at reflux for 40 min. The reaction mixture was allowed to cool to RT and concentrated *in vacuo* to give an oil. The excess POCl$_3$ was quenched by pouring onto ice (250 ml), which was allowed to melt and sat. KCl$_{(aq)}$ (100 ml) was added. The resultant mixture was extracted with DCM (10 x 100 ml) and the combined organic extracts were dried over Na$_2$SO$_4$ (5 g) and concentrated *in vacuo* to give a residual solid. The residue was dried overnight *in vacuo* over P$_2$O$_5$ affording **24** (8.55 g, 67%) as a yellow amorphous solid.

R$_f$ 0.51 (10% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 12.58 (1H, br.s, N**H**), 8.58 (1H, s, C**H**^{2}), 7.68 (1H, dd, *J* = 2.6, 3.3 Hz, C**H**^{6}), 6.58 (1H, dd, *J* = 1.8, 3.3 Hz, C**H**^{5})

δ_C (75.5 MHz; DMSO-d6) 151.7 (**C**^{4}), 150.4 (**C**^{7a}), 150.2 (**C**^{2}), 129.3 (**C**^{6}), 116.5 (**C**^{4a}), 98.7 (**C**^{5})

IR v_{max}/cm^{-1}: 3102 (w), 2961 (w), 2807 (w), 1602 (m), 1547 (m), 1488 (w), 1434 (m), 1352 (m), 1266 (m), 1207 (m), 1103 (w), 1067 (w), 971 (s), 899 (m), 849 (s), 740 (s)

ES$^+$/MS: 152.2 (M-H)⁻ 154.2 (M-H)⁻

25. 4-Chloro-5-iodo-7*H*-pyrrolo[2,3-*d*]pyrimidine[75]
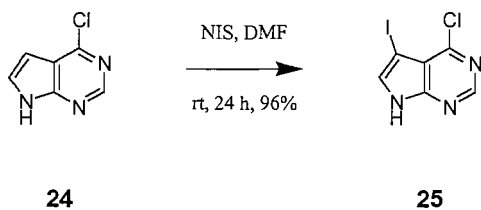
To a solution of **24** (3.6 g, 18.03 mmol) in DMF (72 ml), was added *N*-iodosuccinimide (4.26 g, 18.93 mmol). After stirring for 18 h at RT the solvent was evaporated. The residue was crystallized from MeOH to yield **25** (6.1 g, 96%) a brown amorphous solid.

R_f 0.41 (90% EtOAc/Hexane)

δ_H (300 MHz; DMSO-d6) 12.92 (1H, br.s, **NH**), 8.58 (1H, s, **CH²**), 7.92 (1H, s, **CH⁶**)

δ_C (75.5 MHz; DMSO-d6) 152.0 (**C⁴**), 151.2 (**C⁷ᵃ**), 150.9 (**C²**), 134.3 (**C⁶**), 116.2 (**C⁴ᵃ**), 52.1 (**C⁵**)

IR v_{max}/cm⁻¹: 3185 (m), 3122 (m), 3062 (m), 2953 (m), 2814 (m), 1910 (w), 1660 (m), 1592 (s), 1549 (s), 1437 (s), 1408 (m), 1335 (s), 1275 (m), 1233 (s), 1203 (m), 1158 (m), 1043 (w), 982 (w), 953 (m), 837 (s), 780 (s)

ES⁺/MS: 311.3, 313.3 (M+Na)⁺

26. 4-Chloro-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)pyrrolo[2,3-*d*]pyrimdine[75,116,119]

24   **19**   **26**
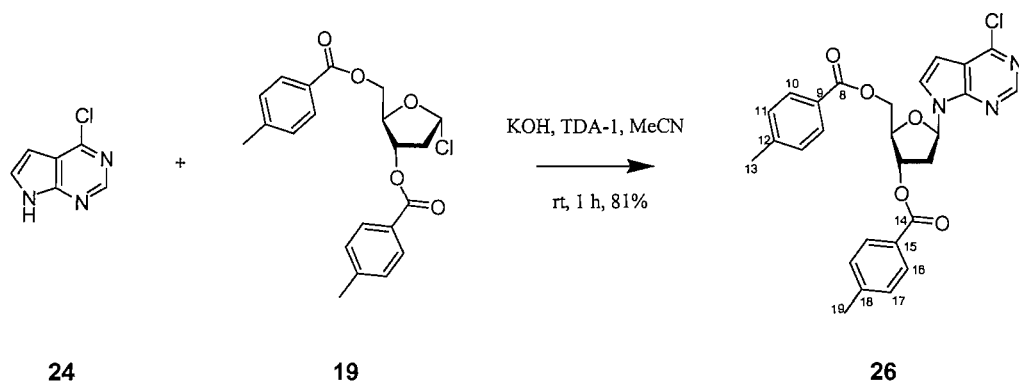
Powdered potassium hydroxide (8.77 g, 156.28 mmol), TDA-1 (1.67 ml, 5.21 mmol) and **24** (8 g, 52.09 mmol) were added to anhydrous acetonitrile (640 ml) over a period of 5 min. After an additional 5 min, **19** (20.26 g, 52.09 mmol) was added and the mixture stirred for 15 min. Insoluble material was filtered off, the residue washed with acetonitrile and the solution was evaporated *in vacuo*. The residue was purified by flash silica gel column chromatography (25% EtOAc / Hexane) to give a white foam **26** (20.27 g, 81%).

R$_f$ 0.14 (DCM)

δ_H (300 MHz; CDCl$_3$) 8.63 (1H, s, CH^2), 7.99 (2H, d, J = 8.1 Hz, CH^{16}), 7.93 (2H, d, J = 8.1 Hz, CH^{10}), 7.43 (1H, d, J = 3.8 Hz, CH^6), 7.29 (2H, d, J = 8.1 Hz, CH^{17}), 7.25 (2H, d, J = 8.1 Hz, CH^{11}), 6.82 (1H, dd, J = 5.8, 8.5 Hz, CH$^{1'}$), 6.60 (1H, d, J = 3.8 Hz, CH^5), 5.77 (1H, m, CH$^{3'}$), 4.77-4.60 (3H, m, CH$^{4', 5'}$), 2.96-2.86 (1H, ddd, J = 6.1, 8.6, 14.4 Hz, CH$^{2'}$), 2.82-2.74 (1H, ddd, J = 2.1, 5.7, 14.4 Hz, CH$^{2'}$), 2.45 (3H, s, CH^{19}), 2.43 (3H, s, CH^{13})

δ_C (75.5 MHz; CDCl$_3$) 166.1 (C^{14}), 166.0 (C^8), 151.0 (C^4), 150.8 (C^2), 144.5 (C^{18}), 144.2 (C^{12}), 129.8 (C^{16}), 129.6 (C^{10}), 129.2 (C$^{11, 17}$), 126.7 (C^{15}), 126.3 (C^9), 125.9 (C^6), 118.3 (C^{7a}), 101.0 (C^5), 84.4 (C$^{1'}$), 82.4 (C$^{4'}$), 75.0 (C$^{3'}$), 64.1 (C$^{5'}$), 38.1 (C$^{2'}$), 21.7 (C$^{13, 19}$)

IR v_{max}/cm^{-1}: 1719 (m), 1706 (m), 1611 (w), 1593 (w), 1506 (w), 1452 (w), 1420 (w), 1366 (w), 1266 (s), 1198 (m), 1108 (m), 1085 (s), 1017 (m), 922 (m), 840 (m), 754 (s), 731 (m), 691 (m)

ES$^+$/MS: 528.2, 530.2 (M+Na)$^+$

27. 4-Chloro-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-(iodo)pyrrolo[2,3-*d*]pyrimdine[74,116]

Method 1

26   **27**
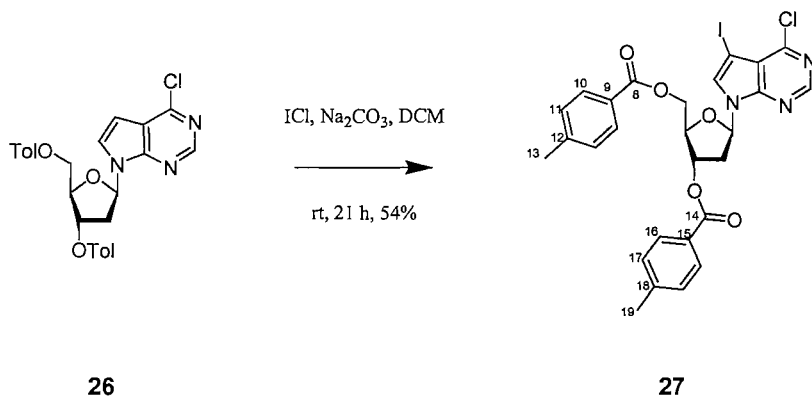
To a solution of **26** (18 g, 35.58 mmol) in DCM (100 ml) was added sodium carbonate (15.08 g, 0.142 mol) and iodine monochloride (3.57 ml, 71.15 mmol). The reaction was stirred for 21 h, washed with aqueous sodium hydrosulfite ($Na_2S_2O_4$), saturated aqueous $NaHCO_3$, dried over Na_2SO_4 and evaporated. Flash silica gel column chromatography (30% ethyl acetate / hexane) afforded (12.17 g, 54%) of **27** as a brown foam.

Method 2

25   **19**   **27**
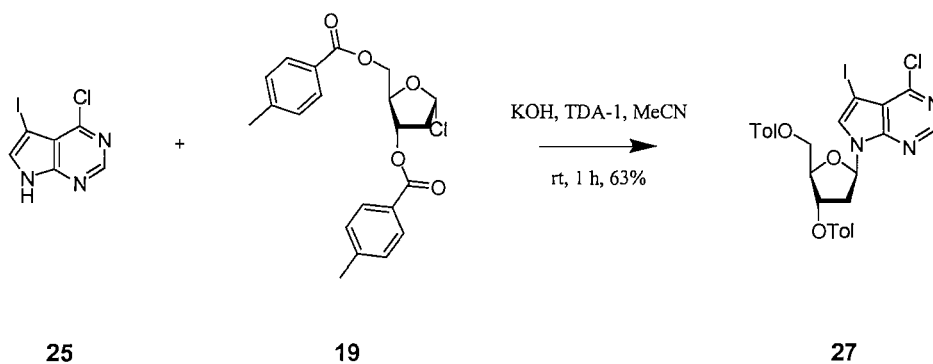
Powdered potassium hydroxide (3.5 g, 62.4 mmol), TDA-1 (0.67 ml, 2.08 mmol) and **25** (6.1 g, 21.83 mmol) were added to anhydrous acetonitrile (400 ml) within 5 min. After an additional 5 min, **19** (8.08 g, 20.79 mmol) was added and the mixture stirred for 10 min. Insoluble material was filtered off, the residue washed with hot acetone and the combined filtrates were evaporated *in vacuo*. The residue was purified by flash silica gel column

143

chromatography on a flash master (20% EtOAc / Hexane) to yield 27 as a yellow foam (8.3 g, 63%).

R$_f$ 0.24 (DCM)

δ$_H$ (300 MHz; CDCl$_3$) 8.60 (1H, s, **CH^2**), 7.98 (2H, d, J = 8.2 Hz, **CH^{16}**), 7.93 (2H, d, J = 8.2 Hz, **CH^{10}**), 7.57 (1H, s, **CH^6**), 7.29 (2H, d, J = 3.1 Hz, **CH^{17}**), 7.27 (2H, d, J = 3.1 Hz, **CH^{11}**), 6.79 (1H, apparent t, J = 7.0 Hz, **CH$^{1'}$**), 5.75 (1H, m, **CH$^{3'}$**), 4.79-4.59 (3H, m, **CH$^{3', 5'}$**), 2.78 (2H, m, **CH$^{2'}$**), 2.44 (3H, s, **CH^{19}**), 2.43 (3H, s, **CH^{13}**)

δ$_C$ (75.5 MHz; CDCl$_3$) 166.0 (**C$^{8, 14}$**), 151.0 (**C^2**), 150.6 (**C^4**), 144.6 (**C^{18}**), 144.3 (**C^{12}**), 131.5 (**C^6**), 129.8 (**C^{17}**), 129.6 (**C^{11}**), 129.5 (**C^{16}**), 129.3 (**C^{10}**), 126.5 (**C^{15}**), 126.0 (**C^9**), 84.4 (**C^5**), 82.9 (**C$^{1'}$**), 75.0 (**C$^{4'}$**), 63.9 (**C$^{5'}$**), 52.9 (**C$^{3'}$**), 38.5 (**C$^{2'}$**), 21.7 (**C$^{13, 19}$**)

IR ν$_{max}$/cm^{-1}: 1715 (m), 1606 (w), 1583 (w), 1538 (w), 1506 (w), 1448 (w), 1339 (w), 1262 (s), 1203 (m), 1180 (m), 1094 (s), 1017 (m), 940 (m), 840 (w), 759 (s), 695 (w)

ES$^+$/MS: 654.1, 656.1 (M+Na)$^+$

28. 4-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidine[74]
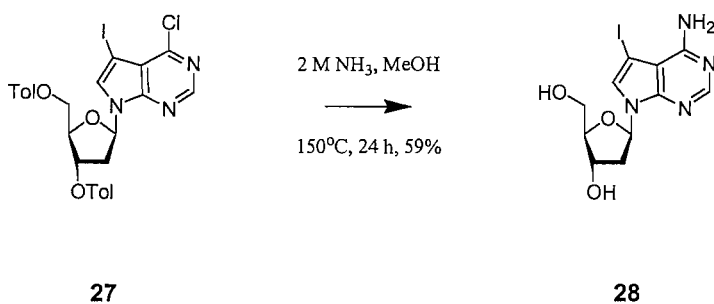
Compound **27** (4 g, 6.33 mmol) was dissolved in 2 M NH$_3$/MeOH (50 ml) in a Parr®
general-purpose acid digestion bomb (stainless steel 125 ml). The sealed vessel was then
heated (150°C, 24 h), cooled (3 h) and the contents then concentrated *in vacuo*. The
residue was dissolved in H$_2$O, washed with Et$_2$O and then the combined Et$_2$O layers back
extracted with H$_2$O. The combined aqueous phases were then concentrated *in vacuo*. The
product was then crystallized from a minimum hot MeOH. The solid was filtered,
washed with cold acetone and dried by vacuum desiccation to yield **28** (1.4 g, 59%).

R$_f$ 0.4 (20% MeOH / DCM)

δ$_H$ (300 MHz; DMSO-d6) 8.10 (1H, s, CH^2), 7.67 (1H, s, CH^6), 6.71 (2H, br. s, NH$_2$),
6.48 (1H, dd, *J* = 6.0, 8.1 Hz, CH$^{1'}$), 5.28 (1H, br. s, OH$^{5'}$), 5.07 (1H, br. s, OH$^{3'}$), 4.32
(1H, m, CH$^{3'}$), 3.81 (1H, m, CH$^{4'}$), 3.53 (2H, m, CH$^{5'}$), 2.49-2.41 (1H, m, CH$^{2'}$), 2.19-
2.11 (1H, ddd, *J* = 2.6, 5.9, 13.1 Hz, CH$^{2'}$)

δ$_C$ (75.5 MHz; DMSO-d6) 157.0 (C^4), 151.7 (C^2), 149.6 (C^{7a}), 126.8 (C^6), 103.0 (C$^{4a,\,5}$),
87.4 (C$^{4'}$), 82.9 (C$^{1'}$), 70.9 (C$^{3'}$), 61.8 (C$^{5'}$), 51.9 (C$^{2'}$)

IR v$_{max}$/cm^{-1}: 3450 (w), 3346 (m), 3137 (m), 2930 (m), 2361 (w), 1619 (s), 1586 (m),
1553 (s), 1481 (m), 1436 (m), 1335 (m), 1306 (m), 1253 (m), 1181 (m), 1103 (m), 1050
(s), 993 (m), 941 (s), 917 (m), 855 (m), 786 (m), 760 (m), 716 (m)

ES$^+$/MS: 376.9 (M+H)$^+$, 752.9 (2M+H)$^+$

29. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-[phthalimidoprop-1-ynyl]-7*H*-pyrrolo-[2,3-*d*]pyrimidin-4-amine[74]
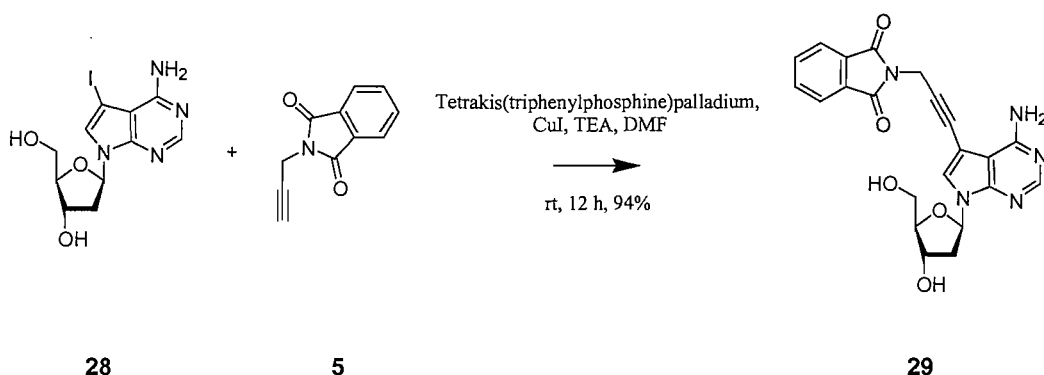
To a solution of **28** (0.1 g, 0.27 mmol) in anhydrous DMF (2 ml) under argon was added copper iodide(I) (10 mg, 0.05 mmol), anhydrous triethylamine (0.19 ml, 1.33 mmol) and **5** (50 mg, 0.29 mmol). The reaction was stirred in the dark (10 min), then tetrakis(triphenylphosphine)palladium (30 mg, 0.03 mmol) was added. The reaction was stirred overnight after which TLC showed the reaction to be incomplete. Copper iodide(I) (10 mg, 0.05 mmol), tetrakis(triphenylphosphine)palladium (30 mg, 0.03 mmol) and **5** (50 mg, 0.29 mmol) were added and after 5 h TLC showed the reaction to be complete. The reaction was then concentrated *in vacuo* to give a brown oil. The oil was purified by flash silica gel column chromatography (0-10% MeOH / DCM) to yield **29** (0.11 g, 94%) as a yellow amorphous solid.

R_f 0.51 (20% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 8.10 (1H, s, CH^2), 7.89 (4H, m, CH^{Ar}), 7.76 (1H, s, CH^6), 6.45 (1H, dd, J = 6.25, 8.09 Hz, $CH^{1'}$), 5.23 (1H, d, J = 2.9 Hz, $OH^{3'}$), 5.03 (1H, s, $OH^{5'}$), 4.68 (2H, s, CCH_2N), 4.31 (1H, m, $CH^{3'}$), 3.80 (1H, m, $CH^{4'}$), 3.52 (2H, m, $CH^{5'}$), 2.43 (1H, m, $CH_\beta^{2'}$), 2.16 (1H, m, $CH_\alpha^{2'}$)

δ_C (75.5 MHz; DMSO-d6) 166.99 ($C^{12, 19}$), 157.30 (C^4), 152.61 (C^2), 149.11 ($C^{13, 18}$), 134.68 ($C^{15, 16}$), 131.46 ($C^{13, 18}$), 126.64 (C^6), 123.31 ($C^{14, 17}$), 102.15 (C^{7a}), 93.95 ($C^{5, 4a}$), 87.47 ($C^{4'}$), 86.03 (C^9), 83.03 ($C^{1'}$), 75.90 (C^8), 70.87 ($C^{3'}$), 61.78 ($C^{5'}$), 39.73 ($C^{2'}$), 27.91 (C^{10})

IR v_{max}/cm^{-1}: 3358 (w), 2936 (w), 2763 (w), 2681 (w), 2360 (w), 2338 (w), 1768 (w), 1708 (s), 1626 (m), 1587 (m), 1571 (m), 1536 (w), 1466 (m), 1418 (m), 1393 (m), 1337

(m), 1308 (m), 1206 (m), 1168 (m), 1088 (m), 1035 (m), 940 (m), 797 (m), 725 (s), 708 (s)

ES$^+$/MS: 434.3 (M+H)$^+$

HRMS (ES$^+$) for $C_{22}H_{19}N_5O_5$ (M+H)$^+$; calculated 434.1459, found 434.1458

30. 7-[2'-Deoxy-β-D-*erythro*-pentofuranosyl]-*N*[4]-[(di-*n*-butylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

N,N-Di-*n*-butylformamide dimethyl acetal,
DMF

rt, 12 h, 81%

29   **30**
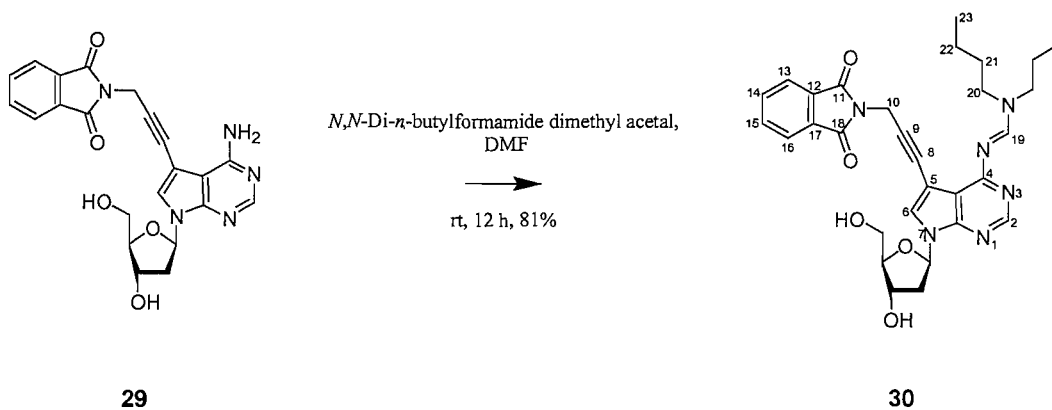
Compound **29** (0.5 g, 1.15 mmol) was dissolved in DMF (50 ml) and stirred at RT. *N,N*-Di-*n*-butylformamide dimethyl acetal (0.33 g, 1.62 mmol) freshly distilled was added drop-wise and the reaction was stirred at RT (48 h). TLC analysis after that time showed complete conversion to a single product. The reaction mixture was diluted with Et_2O (100 ml) and shaken with sat. KCl (3 x 30 ml). The organic phase was dried over Na_2SO_4 and concentrated *in vacuo* and the resulting oil/solid was purified by flash silica gel column chromatography (6% MeOH / DCM) to yield **30** (0.5 g, 81%) as a pale yellow amorphous solid.

R$_f$ 0.14 (5% MeOH/DCM)

δ$_H$ (300 MHz; CDCl$_3$) 8.76 (1H, s, CH^2), 8.33 (1H, s, CH^6), 7.85 (2H, dd, J = 3.0, 5.3 Hz, CH^{13,16}), 7.71 (2H, dd, J = 3.0, 5.3 Hz, CH^{14,15}), 7.26 (1H, s, CH^{19}), 6.15 (1H, dd, J = 5.5, 9.3 Hz, CH$^{1'}$), 4.69 (3H, m, CH$^{10,3'}$), 4.14 (1H, m, CH$^{4'}$), 3.94-3.67 (3H, m, CH$^{5',20}$), 3.37-3.16 (3H, m, CH^{20}), 2.98 (1H, ddd, J = 5.3, 9.3, 13.3, CH$^{2'}$), 2.18 (1H, dd, J = 5.5, 13.3, CH$^{2'}$), 1.62 (4H, s, CH^{21}), 1.33 (4H, m, CH^{22}), 0.93 (6H, m, CH^{23})

δ$_C$ (75.5 MHz; CDCl$_3$) 167.1 (C^{11,18}), 162.7 (C^4), 156.5 (C^2), 151.4 (C^6), 149.5 (C^{7a}), 134.1 (C^{13,16}), 132.1 (C^{12,17}), 131.1 (C^{19}), 123.4 (C^{14,15}), 112.6 (C^5), 96.1 (C^{4a}), 89.6 (C$^{1'}$), 89.0 (C$^{4'}$), 83.2 (C^9), 76.8 (C^8), 73.4 (C$^{3'}$), 63.4 (C$^{5'}$), 51.7 (C^{20}), 45.3 (C^{20}), 40.4 (C$^{2'}$), 31.0 (C^{21}), 29.3 (C^{21}), 28.2 (C^{10}), 20.2 (C^{22}), 19.8 (C^{22}), 13.9 (C^{23}), 13.7 (C^{23})

IR v$_{max}$/cm^{-1}: 2957 (w), 2931 (w), 2871 (w), 2360 (w), 2235 (w), 1771 (w), 1716 (s), 1656 (w), 1617 (w), 1571 (m), 1538 (s), 1425 (s), 1366 (m), 1337 (m), 1193 (m), 1093 (m), 1055 (m), 989 (w), 938 (m), 806.5 (w), 725 (m), 710 (m)

ES$^+$/MS: 573.3 (M+H)$^+$, 1145.4 (2M+H)$^+$

31. 7-[2'-Deoxy-β-D-*erythro*-pentofuranosyl]-*N*⁴-[(dimethylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[74]

30   **31**
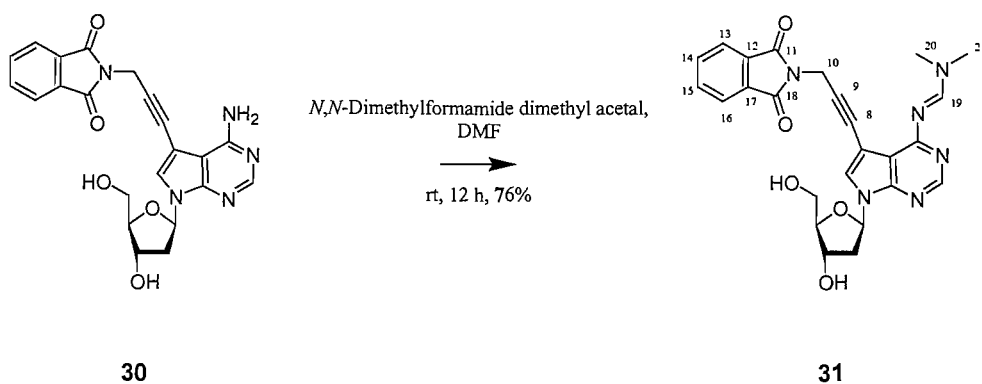
Compound **30** (0.5 g, 1.15 mmol) in MeOH (20 ml) was stirred with *N,N*-dimethylformamide dimethyl acetal (2.61 ml, 19.61 mmol) for 2 h at 40°C. After concentration *in vacuo* the product was purified using flash silica gel column chromatography (10% MeOH / DCM) resulting in **31** (0.43 g, 76%) a yellow amorphous solid.

R_f 0.29 (10% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 8.79 (1H, s, CH^2), 7.95-7.85 (6H, m, $CH^{6, \ 13\text{-}16, \ 19}$), 6.52 (1H, dd, J = 6.6, 7.4, $CH^{1'}$), 4.62 (2H, s, CH^{10}), 4.34 (1H, m, $CH^{3'}$), 3.82 (1H, m, $CH^{4'}$), 3.61-3.46 (2H, m, $CH^{5'}$), 3.13 (3H, s, CH^{20}), 3.03 (3H, s, CH^{21}), 2.48-2.43 (1H, m, $CH^{2'\beta}$), 2.21-2.14 (1H, m, $CH^{2'\alpha}$)

δ_C (75.5 MHz; DMSO-d6) 169.8 (C^4), 166.8 ($C^{11, \ 18}$), 156.4 (C^{19}), 150.7 (C^{7a}), 134.7 ($C^{14, \ 15}$), 131.4 ($C^{12, \ 17}$), 128.8 (C^6), 123.3 ($C^{13, \ 16}$), 87.4 ($C^{4'}$), 83.8 (C^9), 83.0 ($C^{1'}$), 76.8 (C^8), 70.9 ($C^{3'}$), 61.8 ($C^{5'}$), 40.3 (C^{20}), 39.7 ($C^{2'}$), 34.3 (C^{21}), 27.7 (C^{10})

IR v_{max}/cm⁻¹: 2922 (w), 2362 (w), 1770 (w), 1712 (s), 1629 (m), 1550 (m), 1448 (m), 1417 (m), 1363 (m), 1337 (m), 1197 (m), 1090 (m), 1053 (m), 990 (m), 937 (m), 881 (w), 806 (w), 757 (w)

32. 7-[2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*⁴-[(di-*n*-butyl-amino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

30   **32**
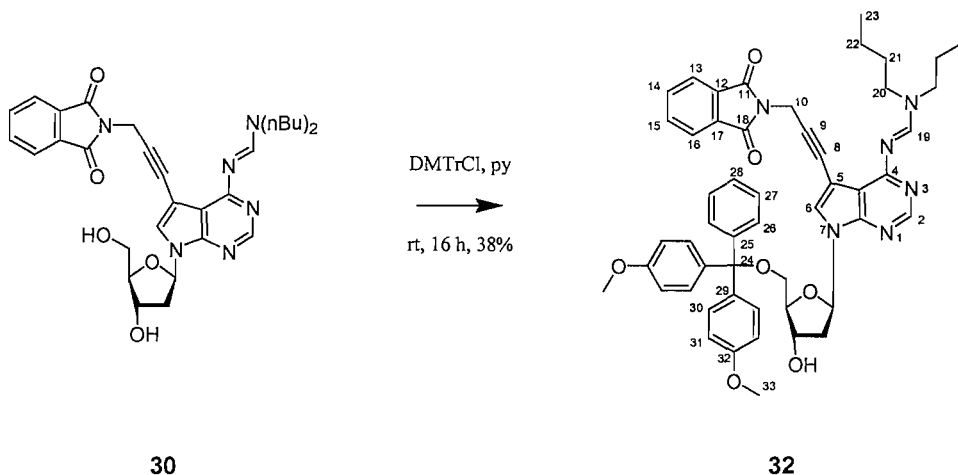
Compound **30** (0.43 g, 0.75 mmol) and 4,4'-dimethoxytrityl chloride (0.27 g, 0.79 mmol) were added to the same RB flask and were dissolved in pyridine (3 ml). The reaction mixture was allowed to stir at RT overnight. Upon completion as shown by TLC analysis, MeOH (5 ml) was added and the solvents removed *in vacuo*. The pyridine free product was purified by flash silica gel column chromatography (0-10% MeOH / DCM) to yield a yellow foam **32** (0.25 g, 38%).

R_f 0.22 (5% MeOH/DCM)

δ_H (300 MHz; CDCl₃) 8.76 (1H, s, CH²), 8.40 (1H, s, CH⁶), 7.86 (2H, dd, J = 3.0, 5.5 Hz, CH¹³,¹⁶), 7.73 (2H, dd, J = 3.0, 5.5 Hz, CH¹⁴,¹⁵), 7.43-7.17 (10H, m, CH¹⁹,²⁶⁻²⁸,³⁰), 6.81 (4H, dd, J = 2.4, 8.9 Hz, CH³¹), 6.65 (1H, ap.t., J = 6.5 Hz, CH¹'), 4.70 (2H, s, CH¹⁰), 4.53 (1H, m, CH³'), 4.07 (1H, m, CH⁴'), 3.78 (3H, s, CH³³), 3.77 (3H, s, CH³³), 3.75 (2H, m, CH²⁰), 3.43-3.34 (3H, m, CH⁵',²⁰), 3.27 (1H, dd, J = 5.5, 10.0 Hz, CH⁵'), 2.51-2.38 (2H, m, CH²'), 1.74-1.60 (4H, m, CH²¹), 1.46-1.32 (4H, m, CH²²), 0.96 (2H, t, J = 7.5 Hz, CH²³), 0.94 (2H, t, J = 7.5 Hz, CH²³)

δ_C (75.5 MHz; CDCl₃) 167.0 (C¹¹,¹⁸), 161.5 (C⁴), 158.4 (C³²), 156.3 (C²), 152.3 (C⁶), 151.1 (C⁷ᵃ), 144.6 (C²⁵), 135.7 (C²⁹), 135.6 (C¹²,¹⁷), 134.0 (C¹⁴,¹⁵), 132.1 (C³⁰), 130.0 (C²⁷), 129.9 (C²⁶), 128.0 (C²⁸), 127.9 (C¹⁹), 126.8 (C⁷), 123.3 (C¹³,¹⁶), 113.1 (C³¹), 111.0 (C⁵), 97.0 (C⁴ᵃ), 86.4 (C²⁴), 85.1 (C⁴'), 83.1 (C⁹), 83.0 (C¹'), 77.2 (C⁸), 72.5 (C³'), 64.1

($C^{5'}$), 55.1 (C^{33}), 51.6 (C^{20}), 45.2 (C^{20}), 40.2 ($C^{2'}$), 31.1 (C^{21}), 29.3 (C^{21}), 28.1 (C^{10}), 20.2 (C^{22}), 19.8 (C^{22}), 13.9 (C^{23}), 13.7 (C^{23})

IR ν_{max}/cm^{-1}: 2933 (w), 1771 (w), 1718 (s), 1612(m), 1571(m), 1544 (s), 1508 (m), 1425 (m), 1367 (m), 1338 (m), 1248 (m), 1177 (m), 1090 (m), 1033 (m), 939 (m), 828 (m), 725 (m), 708 (m)

ES$^+$/MS: 875.4 (M+H)$^+$

HRMS (ES$^+$) for $C_{52}H_{54}N_6O_7$ (M+H)$^+$: calculated 875.4127, found 875.4147

33. 7-[2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*⁴-[(dimethylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

31   **33**
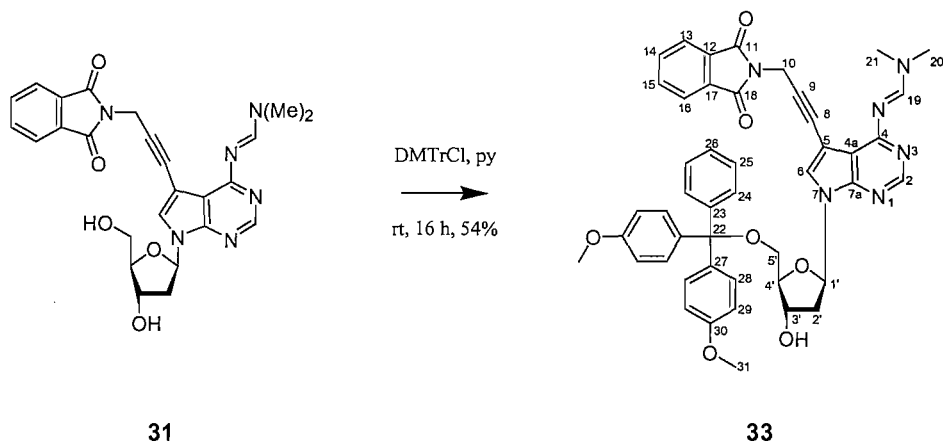
Compound **31** (0.38 g, 0.78 mmol) and 4,4'-dimethoxytrityl chloride (0.29 g, 0.86 mmol) were added to the same RB flask and were dissolved in pyridine (5 ml). The reaction mixture was allowed to stir at RT for 4 h. Upon completion of the reaction as shown by TLC, triethylsilane (1 ml) was added and the solvents removed *in vacuo*. The pyridine free product was purified by flash silica gel column chromatography (0-10% MeOH / DCM) to yield **33**, a yellow foam (0.33 g, 54%).

R$_f$ 0.20 (5% MeOH/DCM)

δ$_H$ (300 MHz; CDCl₃) 8.73 (1H, s, C**H**²), 8.40 (1H, s, C**H**⁶), 7.86 (2H, dd, *J* = 3.0, 5.3 Hz, C**H**¹³, ¹⁶), 7.73 (2H, dd, *J* = 3.0, 5.3 Hz, C**H**¹⁴, ¹⁵), 7.44-7.18 (10H, m, C**H**¹⁹, ²⁴⁻²⁶, ²⁸), 6.82 (4H, dd, *J* = 2.6, 9.1 Hz, C**H**²⁹), 6.67 (1H, app.t. *J* = 6.6 Hz, C**H**¹'), 4.72 (2H, s, C**H**¹⁰), 4.55 (1H, m, C**H**³'), 4.08 (1H, m, C**H**⁴'), 3.78 (6H, 2s., OC**H**₃), 3.41 (1H, m, C**H**⁵'), 3.27 (4H, m, C**H**⁵', NC**H**₃), 3.17 (3H, s, NC**H**₃), 2.97-2.89 (1H, m, C**H**²'), 2.53-2.38 (1H, m, C**H**²')

δ$_C$ (75.5 MHz; CDCl₃) 174.4 (**C**⁴), 167.1 (**C**¹¹, ¹⁸), 161.2 (**C**⁷ᵃ), 158.4 (**C**³⁰), 158.4 (**C**³⁰), 156.4 (**C**²), 152.3 (**C**⁶), 144.6 (**C**²³), 135.7(**C**²⁷), 135.6 (**C**²⁷), 134.0 (**C**¹⁴, ¹⁵), 132.0 (**C**¹², ¹⁷), 130.0 (**C**²⁸), 129.9 (**C**²⁸), 128.0 (**C**²⁵), 127.9 (**C**²⁴), 126.8 (**C**²⁶), 123.3 (**C**¹³, ¹⁶), 113.1 (**C**²⁹), 111.1 (**C**⁵), 96.8 (**C**⁴ᵃ), 86.4 (**C**²²), 85.1 (**C**⁴'), 83.4 (**C**⁹), 83.0 (**C**¹'), 76.9 (**C**⁸), 72.6 (**C**³'), 64.0 (**C**⁵'), 55.2 (**C**³¹), 40.9 (**C**²⁰), 40.2 (**C**²'), 35.1 (**C**²¹), 28.1 (**C**¹⁰)

IR ν_{max}/cm^{-1}: 2935 (w), 1772 (w), 1715 (s), 1628 (m), 1609 (m), 1556 (s), 1508 (m), 1446 (m), 1362 (m), 1338 (m), 1305 (m), 1249 (s), 1177 (m), 1116 (m), 1089 (m), 1032 (m), 939 (m), 882 (w), 828 (m), 807 (m), 757 (m), 725 (m), 709 (m)

ES$^+$/MS: 791.2 (M+H)$^+$

HRMS (ES$^+$) for $C_{46}H_{42}N_6O_7$ (M+H)$^+$: calculated 791.3188, found 791.3208

34. **7-[3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*⁴-[(di-*n*-butylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine**[77]

2-Cyanoethyl-*N,N*-diisopropyl chlorophosphine, DIPEA, DCM

rt, 1 h, 72%

32   **34**
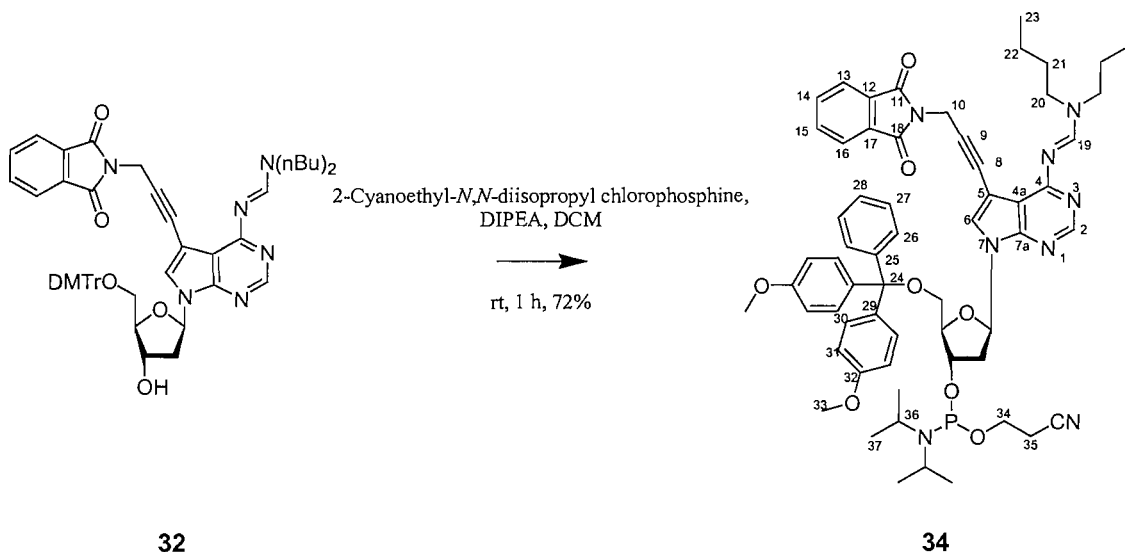
The nucleoside **32** (0.36 g, 0.41 mmol) was dissolved in anhydrous DCM (5 ml), diisopropylethylamine (0.18 ml, 1.03 mmol) and 2-cyanoethyl-*N,N*-diisopropyl chlorophosphine (0.11 ml, 0.49 mmol). Following completion (1 h), as shown by TLC, the solution was diluted with DCM (20 ml), washed (2 x 20 ml, Na_2CO_3 (aq), 20 ml, NaCl (aq)), dried over anhydrous Na_2SO_4 and then the solvents were removed *in vacuo*. The resultant oil was dissolved in EtOAc (5-10 ml) and precipitated into hexane (250 ml) twice. The solid precipitate was then dried overnight *in vacuo*. The precipitate was then purified by flash silica gel column chromatography (60/39/1 EtOAc/Hexane/Pyridine) under argon to yield **34**, a pale yellow foam (0.32 g, 72%).

R$_f$ 0.68 (90% EtOAc/Hexane)

δ_H (300 MHz; CDCl$_3$) 8.77 (1H, s, CH^2), 8.42 (1H, s, CH^6), 7.86 (2H, m, CH$^{13, 16}$), 7.72 (2H, dd, J = 3.1, 5.4 Hz, CH$^{14, 15}$), 7.44-7.16 (10H, m, CH$^{19, 26-28, 30}$), 6.80 (4H, m, CH^{31}), 6.63 (1H, m, CH$^{1'}$), 4.70 (2H, s, CH^{10}), 4.62 (1H, m, CH$^{3'}$), 4.22 (1H, m, CH$^{4'}$), 3.86-3.53 (12H, m, CH$^{20, 33, 34, 36}$), 3.38-3.24 (4H, m, CH$^{5', 20}$), 2.62-2.44 (4H, m, CH$^{2', 35}$), 1.74-1.60 (4H, m, CH^{21}), 1.45-1.31 (4H, m, CH^{22}), 1.18 (3H, d, J = 6.7 Hz, CH^{37}), 1.17 (3H, d, J = 6.7 Hz, CH^{37}), 1.16 (3H, d, J = 6.7 Hz, CH^{37}), 1.11 (3H, d, J = 6.7 Hz, CH^{37}), 0.96 (3H, t, J = 7.5 Hz, CH^{23}), 0.94 (3H, t, J = 7.5 Hz, CH^{23})

δ_P (90 MHz; CDCl$_3$) 149.3, 149.1

ES⁺/MS: 1075.4 (M+H)⁺

35. 7-[3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-*N*4-[(dimethylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine[77]

2-Cyanoethyl-*N*,*N*-diisopropyl chlorophosphine, DIPEA, DCM

rt, 1 h, 61%

33   **35**
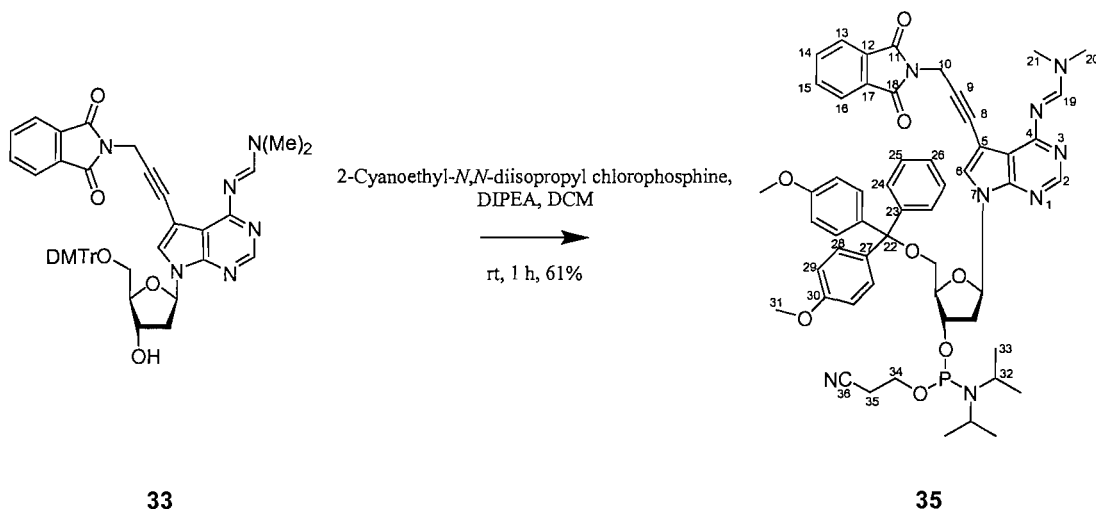
Nucleoside **33** (0.26 g, 0.33 mmol) was dissolved in anhydrous DCM (5 ml) and diisopropylethylamine (0.14 ml, 0.82 mmol) and 2-cyanoethyl-*N*,*N*-diisopropyl chlorophosphine (0.09 ml, 0.39 mmol). Following completion (1 h), as shown by TLC, the solution was diluted with DCM (20 ml), washed (2 x 10 ml, Na$_2$CO$_3$ (aq), 10 ml, NaCl (aq)), dried over anhydrous Na$_2$SO$_4$ and then the solvents were removed *in vacuo*. The resultant oil was dissolved in EtOAc (~2 ml) and precipitated into hexane (~100 ml) twice. The solid precipitate was then dried overnight *in vacuo* and purified by flash silica gel column chromatography to yield **35** as a pale yellow foam (0.2 g, 61%).

R$_f$ 0.34, 0.40 (90% EtOAc / Hexane)

δ_H (300 MHz; CDCl$_3$) 8.66 (1H, s, CH^2), 8.35 (1H, s, CH^6), 7.78 (2H, dd, J = 3.0, 5.2 Hz, CH$^{13,\,16}$), 7.65 (2H, dd, J = 3.0, 5.2 Hz, CH$^{14,\,15}$), 7.37-7.09 (10H, m, CH$^{19,\,24-26,\,28}$), 6.73 (4H, m, CH^{29}), 6.56 (1H, dd, J = 6.5, 11.5 Hz, CH$^{1'}$), 4.63 (2H, s, CH^{10}), 4.57 (1H, m, CH$^{3'}$), 4.15 (1H, m, CH$^{4'}$), 3.71 (7H, m, CH$^{31,\,34}$), 3.61 (1H, m, CH^{34}), 3.52 (2H, m, CH^{32}), 3.25 (2H, m, CH$^{5'}$), 3.19 (3H, s, CH^{21}), 3.09 (3H, s, CH^{20}), 2.53 (1H, m, CH^{35}), 2.46 (2H, m, CH$^{2'}$), 2.38 (1H, m, CH^{35}), 1.22-1.02 (12H, m, CH^{33})

δ_P (90 MHz; CDCl$_3$) 149.3, 149.1

ES$^+$/MS: 991.3 (M+H)$^+$

36. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-amine
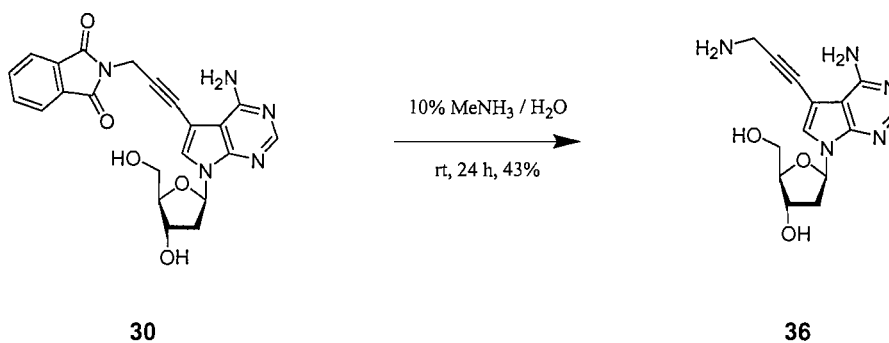
Compound **30** (0.1 g, 0.23 mmol) was dissolved in 10% MeNH$_2$ in water (2 ml) and left for 24 h. TLC analysis showed the reaction to be finished and the solvent was removed *in vacuo*. The residue was then purified by flash silica gel column chromatography (20-40% MeOH / DCM) to yield **36** (0.03 g, 43%) as an amorphous solid.

R$_f$ 0.10 (20% MeOH/DCM)

δ$_H$ (300 MHz; DMSO-d6) 8.09 (1H, s, C**H**^{2}), 7.64 (1H, s, C**H**^{6}), 6.47 (1H, dd, *J* = 6.2, 7.7 Hz, C**H**$^{1'}$), 4.34 (1H, m, C**H**$^{3'}$), 3.82 (1H, m, C**H**$^{4'}$), 3.57 (1H, dd, *J* = 4.3, 11.5 Hz, C**H**$^{5'}$), 3.52 (2H, s, C**H**^{10}), 3.50 (1H, dd, *J* = 4.3, 11.5 Hz, C**H**$^{5'}$), 2.46 (1H, ddd, *J* = 5.8, 7.7, 13.4 Hz, C**H**$^{2'}$), 2.17 (1H, ddd, *J* = 2.7, 6.2, 13.4 Hz, C**H**$^{2'}$)

δ$_C$ (75.5 MHz; DMSO-d6) 157.4 (**C**^{4}), 152.2 (**C**^{2}), 149.0 (**C**^{7a}), 125.3 (**C**^{6}), 102.1 (**C**^{4a}), 95.2 (**C**^{5}), 87.4 (**C**$^{4'}$), 83.1 (**C**$^{1'}$), 74.3 (**C**^{9}), 70.8 (**C**$^{3'}$), 63.0 (**C**$^{5'}$), 61.8 (**C**^{8}), 39.8 (**C**$^{2'}$), 31.5 (**C**^{10})

UV/vis (H$_2$O): 260 nm (1.7 x 10^3 cm^{-1}mol^{-1})

ES$^+$/MS: 304.1 (M+H)$^+$, 326.1 (M+Na)$^+$

HRMS (ES$^+$) for C$_{14}$H$_{17}$N$_5$O$_3$ (M+H)$^+$: calculated 304.1404, found 304.1409

37. 2-(Methylthio)pyrrolo[2,3-*d*]pyrimidin-4-one[113]
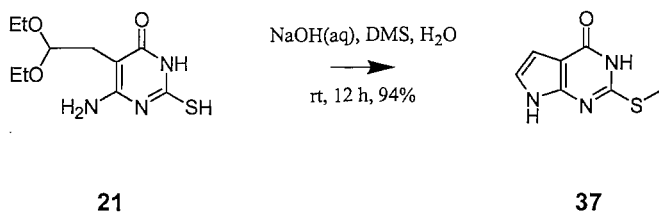
Compound **21** (100 g, 385.62 mmol) was dissolved in NaOH (1 M, 200 ml) and dimethylsulfate (38.31 ml, 404.90 mmol) was added at RT. The solution was stirred overnight, then filtered, washed and dried to give a yellow amorphous solid **37** (62.4 g, 94%).

R_f 0.37 (EtOAc:MeOH:c.NH$_{3(aq)}$)

δ_H (300 MHz; DMSO-d6) 12.06 (1H, s, **NH^7**), 11.77 (1H, s, **NH^3**), 6.90 (1H, dd, J = 2.2, 3.3 Hz, **CH^6**), 6.35 (1H, dd, J = 2.2, 3.3 Hz, **CH^5**), 2.50 (3H, s, **CH$_3$**)

δ_C (75.5 MHz; DMSO-d6) 158.9 (**C^4**), 154.2 (**C^2**), 148.3 (**C^{7a}**), 119.3 (**C^6**), 104.1 (**C^{4a}**), 101.9 (**C^5**), 12.8 (SCH$_3$)

ES$^+$/MS: 204.1 (M+Na)$^+$, 385.1 (2M+Na)$^+$

38. 4-Chloro-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine[123]
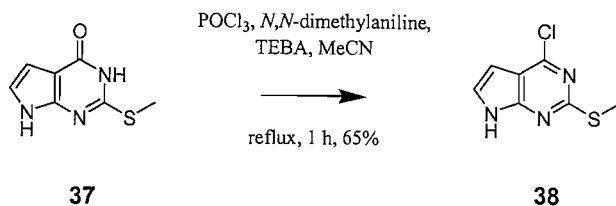
Compound **37** (1 g, 5.56 mmol) was dissolved in acetonitrile (10 ml) together with TEBA (0.63 g, 2.76 mmol) phosphorus oxychloride (5.14 ml, 55.18 mmol), *N,N*-dimethylaniline (3.50 ml, 27.59 mmol) and the solution was stirred at reflux for 1 h. Distilled water was added with caution to the mixture and the residue dissolved. The solution was stirred for another hour and the solution was cooled and adjusted to pH 4 with concentrated aqueous ammonia. The resulting precipitate which formed overnight at 4°C and was collected by filtration, washed with water and hot toluene to yield **38** (0.72 g, 65%) as a light green amorphous solid.

R_f 0.72 (5:1:1 EtOAc/MeOH/NH$_{3(aq)}$)

δ_H (300 MHz; DMSO-d6) 7.51 (1H, d, J = 3.7 Hz, C**H**^{6}), 6.51 (1H, d, J = 3.7 Hz, C**H**^{5}), 2.55 (3H, s, C**H**$_3$)

δ_C (75.5 MHz; DMSO-d6) 162.7 (**C**^{2}), 152.6 (**C**^{4}), 150.4 (**C**^{7a}), 126.8 (**C**^{6}), 113.2 (**C**^{4a}), 99.0 (**C**^{5}), 13.7 (S**C**H$_3$)

ES$^+$/**MS**: 200.1 (M+H)$^+$, 222.1 (M+Na)$^+$

39. 4-Chloro-5-iodo-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine[75]
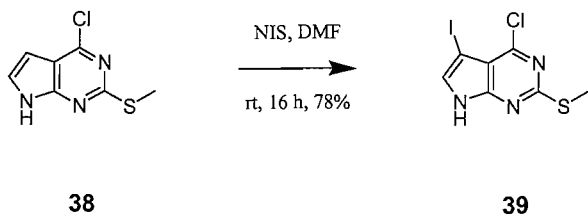
To a solution of **38** (3.6 g, 18.03 mmol) in DMF (72 ml), *N*-iodosuccinimide (4.26 g, 18.93 mmol) was added. After stirring (18 h) at RT, the solvent was evaporated. The residue was purified by crystallization from MeOH to yield **39** (4.6 g, 78%) a brown amorphous solid.

R_f 0.76 (5:1:1 EtOAc/MeOH/NH$_{3(aq)}$)

δ_H (300 MHz; DMSO-d6) 12.73 (1H, s, **NH**), 7.74 (1H, d, **CH^6**), 2.54 (3H, s, **CH$_3$**)

δ_C (75.5 MHz; DMSO-d6) 163.1 (**C^2**), 152.3 (**C^4**), 150.6 (**C^{7a}**), 132.2 (**C^6**), 112.4 (**C^{4a}**), 52.0 (**C^5**), 13.7 (SCH$_3$)

ES$^+$/MS: 326.1, 328.1 (M+H)$^+$

40. 4-Chloro-7-(2'-deoxy-3',5'-di-*O*-*p*-toluoyl-β-D-*erythro*-pentofuranosyl)-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine

38   **19**   **40**
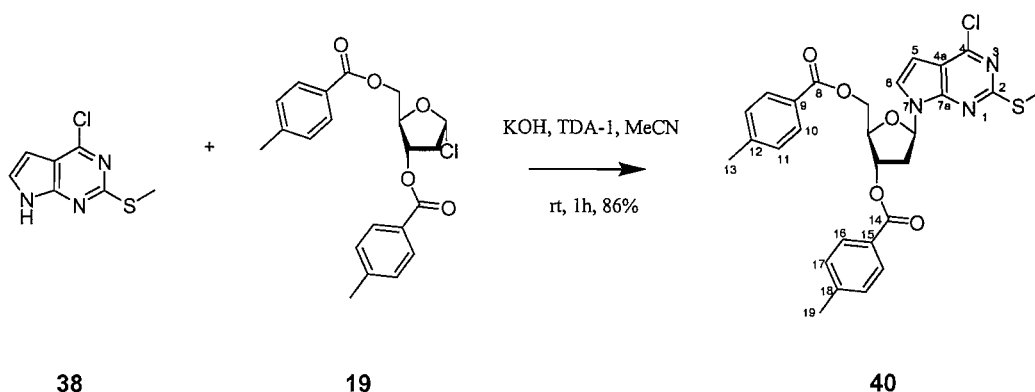
Powdered potassium hydroxide (3.09 g, 55.09 mmol), TDA-1 (0.59 ml, 1.84 mmol) and **38** (3.85 g, 19.28 mmol) were added to anhydrous acetonitrile (200 ml) within 5 min. After an additional 5 min, **19** (7.14 g, 18.36 mmol) was added and the mixture stirred for 10 min. Insoluble material was filtered off, the residue washed with hot acetone and the combined filtrates were evaporated *in vacuo*. The resultant solid was then purified using a flash master chromatography system in 20% EtOAc/Hexane to yield a white foam **40** (8.72 g, 86%).

R$_f$ 0.35 (1% EtOAc/DCM)

δ_H (300 MHz; CDCl$_3$) 8.98 (2H, d, J = 8.1 Hz, CH^{16}), 7.90 (2H, d, J = 8.1 Hz, CH^{10}), 7.28 (2H, d, J = 8.1 Hz, CH^{17}), 7.24 (1H, d, J = 4.0 Hz, CH^{6}), 7.23 (2H, d, J = 8.1 Hz, CH^{11}), 6.74 (1H, dd, J = 6.0, 8.0 Hz, CH$^{1'}$), 6.48 (1H, d, J = 4.0 Hz, CH^{5}), 5.75 (1H, m, CH$^{3'}$), 4.75-4.57 (3H, m, CH$^{4', 5'}$), 2.93-2.80 (1H, ddd, J = 6.7, 8.0, 14.3 Hz, CH$^{2'}$), 2.78-2.71 (1H, ddd, J = 2.4, 6.0, 14.3 Hz, CH$^{2'}$), 2.63 (3H, s, SCH$_3$), 2.44 (3H, s, CH^{19}), 2.41 (3H, s, CH^{13})

δ_C (75.5 MHz; CDCl$_3$) 166.1 (C$^{8, 14}$), 164.8 (C^{2}), 152.0 (C^{4}), 144.5 (C^{18}), 144.2 (C^{12}), 129.8 (C^{16}), 129.6 (C^{10}), 129.2 (C$^{11, 17}$), 126.7 (C^{15}), 126.4 (C^{9}), 124.3 (C^{7a}), 114.6 (C^{6}), 101.3 (C$^{4a, 5}$), 84.1 (C$^{1'}$), 82.2 (C$^{4'}$), 74.9 (C$^{3'}$), 64.1 (C$^{5'}$), 38.1 (C$^{2'}$), 21.8 (C^{19}), 21.7 (C^{13}), 14.5 (SCH$_3$)

41. **4-Chloro-7-(2'-deoxy-3',5'-di-*O*-*p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodo-2-(methylthio)pyrrolo[2,3-*d*]pyrimidine**[116]
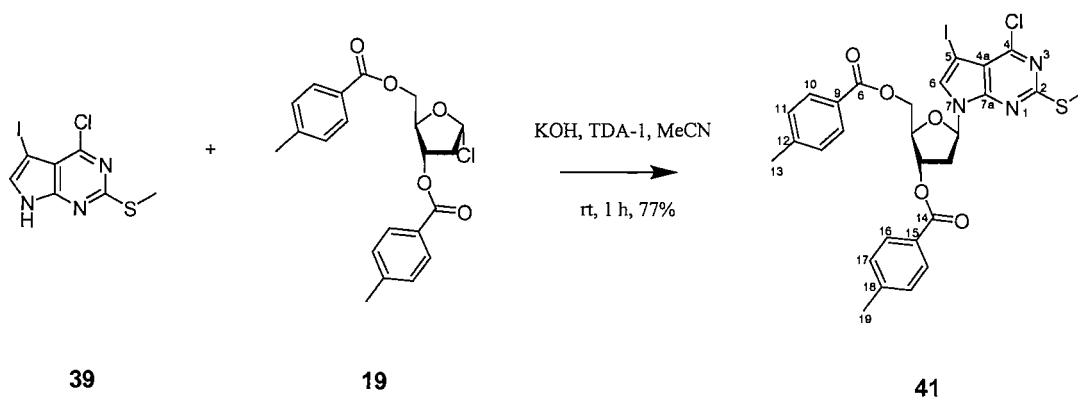
Powdered potassium hydroxide (0.49 g, 8.78 mmol), TDA-1 (0.11 ml, 0.29 mmol) and **39** (1 g, 3.07 mmol) were added to anhydrous acetonitrile (50 ml) within 5 min. After an additional 5 min, **19** (0.95 g, 2.93 mmol) was added and the mixture was stirred for 10 min. Insoluble material was filtered off, the residue washed with hot acetone and the combined filtrates were evaporated *in vacuo*. The resultant solid was then purified using a flash master chromatography system (20% EtOAc/Hexane) to yield **41** as a white foam (1.53 g, 77%).
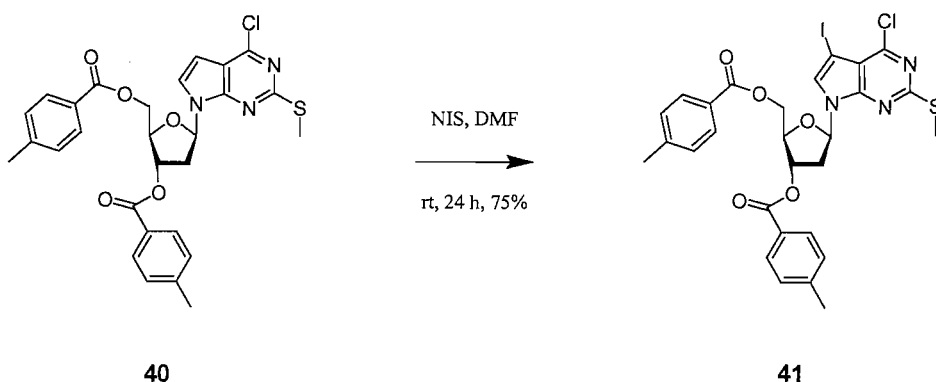
To a solution of **40** (1 g, 1.81 mmol) in DMF (50 ml) was added *N*-iodosuccinimide (2.85 g, 12.68 mmol), the mixture was heated at 96°C (6 h) and cooled to RT. After the addition of 0.2 ml 10% (aq) $NaHCO_3$ the mixture was concentrated to ~10 ml, diluted with 20 ml ethyl acetate, washed with H_2O, 5% $Na_2S_2O_4$ (aq), saturated NaCl, dried over Na_2SO_4 and evaporated. The residue was purified by flash silica gel column chromatography using DCM to yield **41** as a white foam (0.92 g, 75%).

R$_f$ 0.3 (DCM)

δ$_H$ (300 MHz; CDCl$_3$) 7.96 (2H, d, J = 8.3 Hz, C**H**^{16}), 7.91 (2H, d, J = 8.3 Hz, C**H**^{10}), 7.38 (1H, s, C**H**^{6}), 7.28 (2H, d, J = 8.3 Hz, C**H**^{17}), 7.26 (2H, d, J = 8.3 Hz, C**H**^{11}), 6.72 (1H, dd, J = 6.9, 7.0 Hz, C**H**$^{1'}$), 5.72 (1H, m, C**H**$^{3'}$), 4.77-4.57 (3H, m, C**H**$^{4', 5'}$), 2.76 (2H, dd, J = 4.5, 6.9 Hz, C**H**$^{2'}$), 2.62 (3H, s, SC**H**$_3$), 2.44 (3H, s, C**H**^{19}), 2.42 (3H, s, C**H**^{13})

δ$_C$ (75.5 MHz; CDCl$_3$) 166.2 (**C**^{14}), 166.0 (**C**^{8}), 165.3 (**C**^{2}), 151.4 (**C**^{4}), 144.6 (**C**^{18}), 144.3 (**C**^{12}), 129.8 (**C**^{6}), 129.7 (**C**^{16}), 129.6 (**C**^{10}), 129.5 (**C**^{17}), 129.3 (**C**^{11}), 126.5 (**C**^{15}), 126.3 (**C**^{9}), 113.8 (**C**^{7a}), 84.1 (**C**$^{1'}$), 82.7 (**C**$^{4'}$), 75.0 (**C**$^{3'}$), 64.0 (**C**$^{5'}$), 53.4 (**C**$^{4a, 5}$), 38.5 (**C**$^{2'}$), 21.8 (**C**$^{13, 19}$), 14.6 (SC**H**$_3$)

ES$^+$/MS: 700.2, 702.2 (M+Na)$^+$

42. 7-(2'-Deoxy-β-D-erythro-pentofuranosyl)-5-iodopyrrolo[2,3-d]pyrimidin-2,4-diamine

a) mCPBA, DCM, 0°C – rt, 2 h
b) NH₃, dioxane, 140°C, 9 h

c) NH₃, MeOH, 150°C, 24 h
43% over 3 steps

41   42
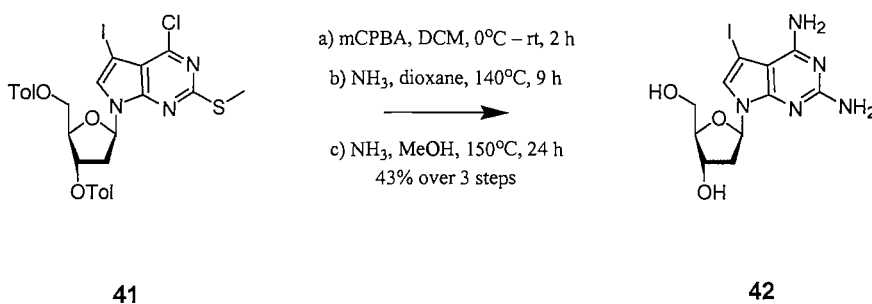
To compound **41** (5 g, 7.38 mmol) in DCM (200 ml) at 0°C was added 3-chloroperoxybenzoic acid (mCPBA, 77 % max, 2.48 g, 11.06 mmol). After 15 min the ice bath was removed and the mixture was stirred for a further 105 min. The reaction mixture was diluted with DCM, washed with sat. $NaHCO_3$(aq) and dried (Na_2SO_4). After removal of Na_2SO_4 by filtration, MeOH was added to the filtrate to make the solution 4% MeOH. This solution was directly chromatographed through a short column of silica gel, eluted with 4% MeOH/DCM and concentrated.

The sulphoxide (5.8 g, 8.36 mmol) was suspended in dioxane (50 ml) saturated with ammonia at 12°C in a Parr® general-purpose acid digestion bomb (stainless steel, 125 ml). The reaction vessel was sealed, heated at 140°C for 16 h, cooled and concentrated. The residue was taken up in DCM (100 ml), extracted with saturated $NaHCO_3$ (aq), dried (Na_2SO_4) and evaporated. Polar compounds were then removed by passing the reaction mixture through a silica gel plug (5% MeOH / DCM).

The protected nucleoside (5.5 g, 8.50 mmol) was dissolved in 2 M NH₃/MeOH (50 ml) in a Parr® general-purpose acid digestion bomb (stainless steel 125 ml). The sealed vessel was then heated (150°C, 24 h), cooled (3 h) and the contents then concentrated *in vacuo*. The residue was dissolved in H_2O, washed with Et_2O and then the combined Et_2O layers back extracted with H_2O. The combined aqueous phases were then concentrated *in vacuo*. After failing to crystallize the product from MeOH flash silica gel column chromatography (10-20% MeOH / DCM) was used yield **42** as a pale brown amorphous solid (1.2 g , 43%).

Rf 0.47 (20% MeOH/DCM)

δ_H (300 MHz; CD$_3$OD) 7.32 (1H, s, C**H**^{6}), 6.36 (1H, dd, J = 6.1, 8.0 Hz, C**H**$^{1'}$), 4.48 (1H, ddd, J = 2.8, 2.8, 5.8 Hz, C**H**$^{3'}$), 3.98 (1H, ddd, J = 2.8, 3.4, 4.0 Hz, C**H**$^{4'}$), 3.79 (1H, dd, J = 3.4, 12.0 Hz, C**H**$^{5'}$), 3.73 (1H, dd, J = 4.0, 12.0 Hz, C**H**$^{5'}$), 2.52 (1H, ddd, J = 5.8, 8.0, 13.2 Hz, C**H**$^{2'\beta}$), 2.28 (1H, ddd, J = 2.8, 6.1, 13.2 Hz, C**H**$^{2'\alpha}$)

δ_C (75.5 MHz; CD$_3$OD) 156.3 (**C**^{4}), 156.2 (**C**^{2}), 127.1 (**C**^{6}), 102.1 (**C**^{7a}), 89.0 (**C**$^{4'}$), 88.8 (**C**^{4a}), 85.9 (**C**$^{1'}$), 72.9 (**C**$^{3'}$), 63.5 (**C**$^{5'}$), 52.7 (**C**^{5}), 41.3 (**C**$^{2'}$)

ES$^+$/MS: 392.2 (M+H)$^+$

HRMS (ES$^+$) for C$_{11}$H$_{14}$IN$_5$O$_3$ (M+H)$^+$: calculated 392.0214, found 392.0222

Mp 122-128°C (from H$_2$O)

43. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine
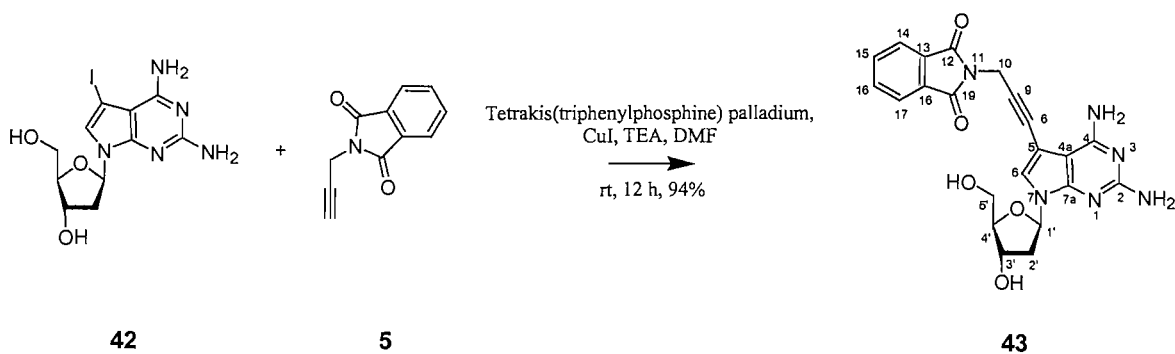
To a solution of compound **42** (0.1 g, 0.26 mmol) in anhydrous DMF (2 ml) under argon was added copper iodide(I) (10 mg, 0.05 mmol), anhydrous triethylamine (0.18 ml, 1.28 mmol) and **5** (100 mg, 0.56 mmol). The reaction was stirred in the dark (10 min), then tetrakis(triphenylphosphine)palladium (30 mg, 0.03 mmol) was added. The reaction was stirred overnight and then concentrated *in vacuo* to give a brown oil. The oil was then purified by flash silica gel column chromatography (0-20% MeOH / DCM) to yield a yellow amorphous powder **43**, (0.1 g, 94%).

R$_f$ 0.15 (10% MeOH/DCM)

δ$_H$ (300 MHz; DMSO-d6) 7.89 (4H, m, **CH$^{14\text{-}17}$**), 7.34 (1H, s, **CH^6**), 6.29 (1H, dd, *J* = 5.9, 8.1 Hz, **CH$^{1'}$**), 4.65 (2H, s, **CH^{10}**), 4.27 (1H, m, **CH$^{3'}$**), 3.75 (1H, m, **CH$^{4'}$**), 3.48 (2H, m, **CH$^{5'}$**), 2.32 (1H, m, **CH$^{2'}$**), 2.05 (1H, m, **CH$^{2'}$**)

δ$_C$ (75.5 MHz; DMSO-d6) 167.0 (**C$^{12,\,19}$**), 157.1 (**C^4**), 151.9 (**C$^{2,\,7a}$**), 134.7 (**C$^{15,\,16}$**), 131.5 (**C$^{13,\,18}$**), 123.3 (**C$^{14,\,17}$**), 123.2 (**C^6**), 95.2 (**C^{4a}**), 94.3 (**C^5**), 87.1 (**C$^{4'}$**), 85.6 (**C^8**), 82.4 (**C$^{1'}$**), 76.7 (**C^9**), 70.9 (**C$^{3'}$**), 61.9 (**C$^{5'}$**), 39.4 (**C$^{2'}$**), 27.9 (**C^{10}**)

IR v$_{max}$/cm^{-1}: 3344 (w), 2936 (w), 2357 (w), 1767 (w), 1815 (s), 1622 (m), 1594 (m), 1465 (m), 1420 (m), 1393 (m), 1338 (m), 1205 (m), 1087 (m), 1039 (m), 941 (m), 791 (m), 725 (s), 710 (s)

ES$^+$/MS: 449.3 (M+H)$^+$

HRMS (ES$^+$) for C$_{22}$H$_{20}$N$_6$O$_5$ (M+H)$^+$: calculated 449.1568, found 449.1573

44. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)- N^4,N^2-[di-(di-*n*-butylamino)methylidene]-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

43   **44**
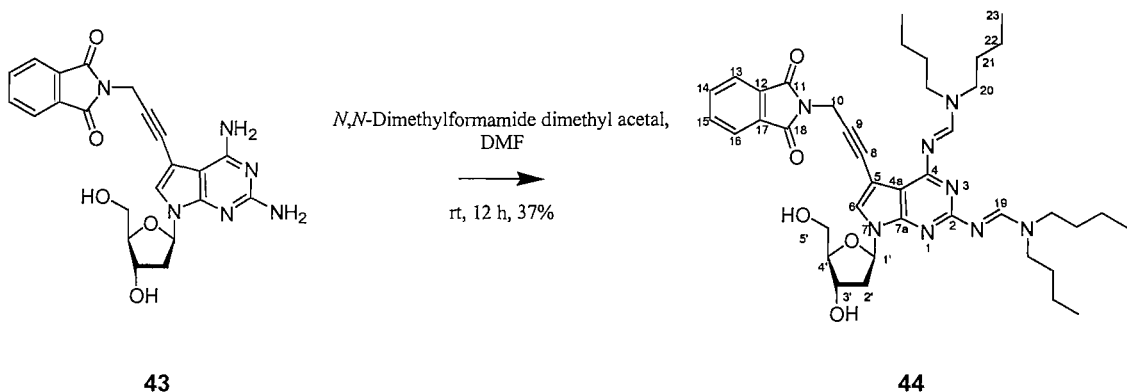
Compound **43** (0.5 g, 1.11 mmol) was dissolved in DMF (50 ml) and stirred at RT. *N,N*-Di-*n*-butylformamide dimethyl acetal (0.63 g, 3.12 mmol) freshly distilled was added drop-wise and the reaction was stirred at RT (48 h). TLC analysis after that time showed complete reaction to a single product. The reaction mixture was diluted with Et₂O (100 ml) and shaken with sat. KCl (3 x 30 ml). The organic phase was dried over Na₂SO₄ and concentrated *in vacuo*. The resulting oil/solid was purified by flash silica gel column chromatography (10% MeOH / DCM) to yield **44** (0.3 g, 37%) as a yellow amorphous solid.

R_f 0.27 (10% MeOH/DCM)

δ_H (300 MHz; CDCl₃) 8.78 (1H, s, CH^{19}), 8.56 (1H, s, CH^{19}), 7.84 (2H, m, CH^{13,16}), 7.70 (2H, m, CH^{14,15}), 7.24 (1H, s, CH^{6}), 6.35 (1H, m, CH$^{1'}$), 4.65 (3H, m, CH$^{3',10}$), 4.05 (1H, s, CH$^{4'}$), 3.81-3.27 (10H, 3m, CH$^{5',20}$), 2.70 (1H, m, CH$^{2'}$), 2.27 (1H, m, CH$^{2'}$), 1.68-1.55 (8H, m, CH^{21}), 1.39-1.21 (8H, m, CH^{22}), 0.96-0.86 (12H, m, CH^{23})

δ_C (75.5 MHz; CDCl₃) 167.3 (C^{11,18}), 157.4 (C^{19}), 134.3 (C^{13,16}), 132.2 (C^{6}), 123.6 (C^{14,15}), 88.2 (C$^{4'}$), 86.5 (C$^{1'}$), 72.8 (C$^{3'}$), 63.3 (C$^{5'}$), 52.3 (C^{20}), 45.9 (C^{20}), 40.8 (C$^{2'}$), 31.5 (C^{21}), 28.6 (C^{10}), 20.4 (C^{22}), 14.2 (C^{23})

ES⁺/MS: 727.4 (M+H)⁺

HRMS (ES⁺) for $C_{40}H_{54}N_8O_5$ (M+Na)⁺: calculated 727.4290, found 727.4315

45. 7-(2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)- *N*⁴-[(di-*n*-butylamino)methylidene]-*N*²-formyl-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

44   **45**
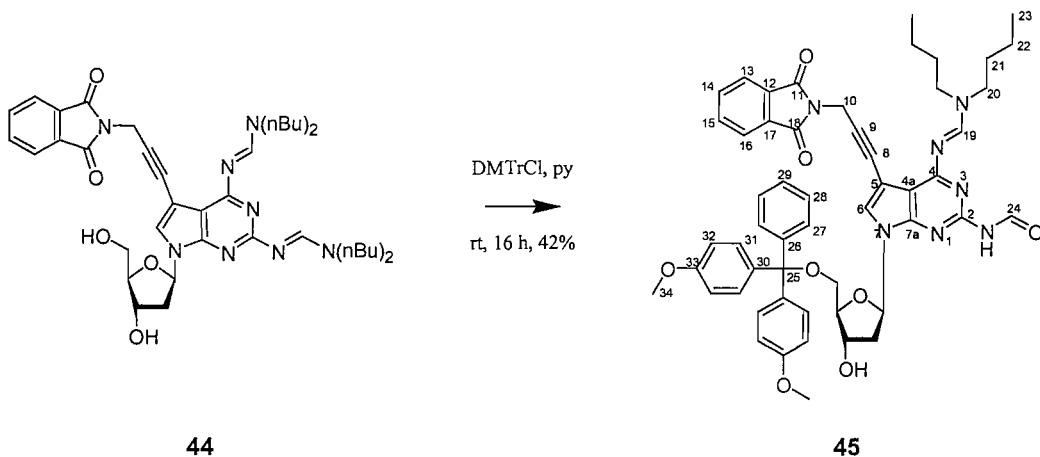
Compound **44** (0.30 g, 0.41 mmol) and 4,4'-dimethoxytrityl chloride (0.15 g, 0.43 mmol) were added to the same RB flask and dissolved in pyridine (2 ml). The reaction mixture was allowed to stir at RT overnight. Upon completion as shown by TLC analysis, MeOH (5 ml) was added and the solvents removed *in vacuo*. The pyridine free product was purified by flash silica gel column chromatography (7% MeOH/DCM) to yield a yellow foam **45** (0.16 g, 42%).

R$_f$ 0.23 (5% MeOH/DCM)

δ_H (300 MHz; CDCl₃) 9.47 (1 H, s, CH²⁴), 8.71 (1H, s, CH⁶), 7.86 (2H, dd, *J* = 3.1, 5.4 Hz, CH¹³˒¹⁶), 7.73 (2H, dd, *J* = 3.1 5.4 Hz, CH¹⁴˒¹⁵), 7.42-7.17 (10H, m, CH¹⁹˒²⁷⁻²⁹˒³¹), 6.81 (4H, dd, *J* = 8.3 Hz, CH³²), 6.49 (1H, ap.t., *J* = 6.6 Hz, CH¹'), 4.70 (2H, s, CH¹⁰), 4.51 (1H, m, CH³'), 4.03 (1H, m, CH⁴'), 3.78 (3H, s, CH³⁴), 3.77 (3H, s, CH³⁴), 3.75 (2H, m, CH²⁰), 3.43-3.35 (3H, m, CH⁵'˒²⁰), 3.23 (1H, dd, *J* = 5.6, 9.9 Hz, CH⁵'), 2.50-2.34 (2H, m, CH²'), 1.75-1.60 (4H, m, CH²¹), 1.46-1.33 (4H, m, CH²²), 0.98 (3H, t, *J* = 7.3 Hz, CH²³), 0.95 (3H, t, *J* = 7.3 Hz, CH²³)

δ_C (75.5 MHz; CDCl₃) 167.0 (C¹¹˒¹⁸), 163.2 (C⁴), 162.1 (C²), 158.5 (C²⁴), 156.8 (C⁶), 151.9 (C³³), 151.8 (C³³), 144.6 (C²⁶), 135.7 (C³⁰), 135.6 (C³⁰), 134.0 (C¹⁴˒¹⁵), 132.1 (C¹²˒¹⁷), 130.0 (C³¹), 129.9 (C²⁸), 128.0 (C¹⁹), 127.9 (C²⁷), 127.2 (C²⁹), 126.8 (C⁷ᵃ), 123.4 (C¹³˒¹⁶), 113.2 (C³²), 108.0 (C⁵), 97.6 (C⁴ᵃ), 86.5 (C²⁵), 85.0 (C⁴'), 83.3 (C⁹), 82.9 (C¹'),

77.2 (C^8), 72.8 ($C^{3'}$), 64.1 ($C^{5'}$), 55.2 (C^{34}), 51.8 (C^{20}), 45.4 (C^{20}), 39.8 ($C^{2'}$), 31.0 (C^{21}), 29.3 (C^{21}), 28.2 (C^{10}), 20.2 (C^{22}), 19.8 (C^{22}), 13.9 (C^{23}), 13.7 (C^{23})

IR ν_{max}/cm^{-1}: 2932 (w), 2871 (w), 2358 (w), 1772 (w), 1717 (s), 1608 (m), 1578 (m), 1508 (m), 1445 (m), 1415 (m), 1389 (m), 1345 (m), 1301 (m), 1247 (s), 1217 (m), 1217 (m), 1176 (m), 1114 (m), 1088 (m), 1033 (m), 940 (m), 828 (m), 801 (m), 755 (m), 725 (s), 709.5 (m)

ES$^+$/MS: 918.4 (M+H)$^+$, 940.4 (M+Na)$^+$

HRMS (ES$^+$) for $C_{53}H_{55}N_7O_8$ (M+H)$^+$: calculated 918.4185, found 918.4192

46. 7-(3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)- β-D -*erythro*-pentofuranosyl)-*N*⁴-[(di-*n*-butylamino)methylidene]-*N*²-formyl-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

45   **46**
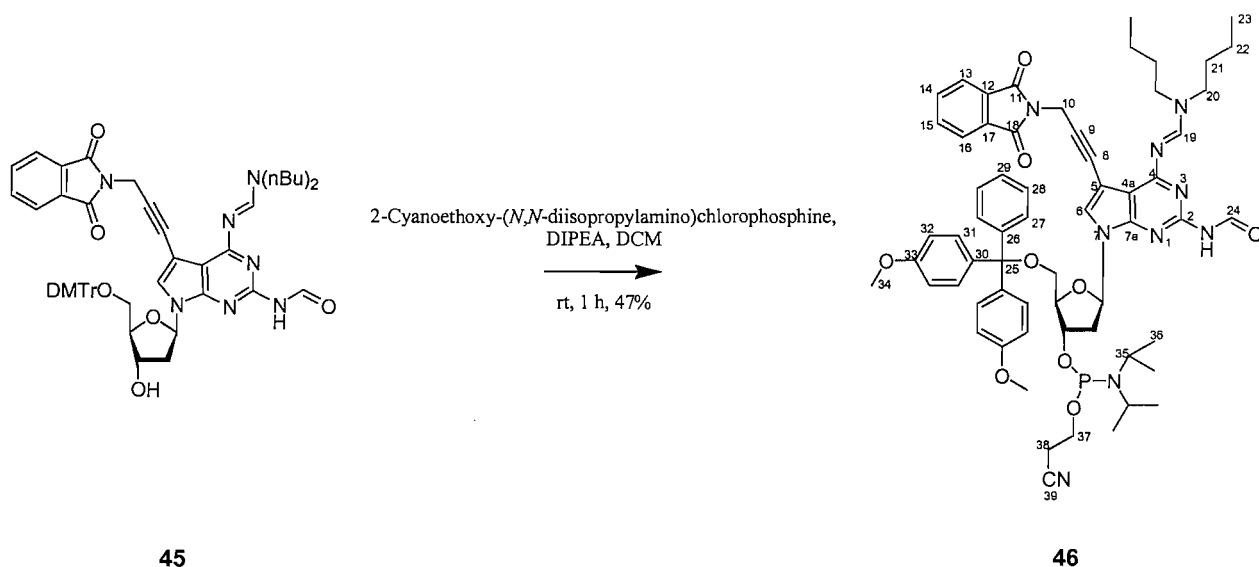
Compound **45** (0.16 g, 0.16 mmol) was dissolved in anhydrous DCM (3 ml) to this was added diisopropylethylamine (0.07 ml, 0.39 mmol) and 2-cyanoethyl-*N*, *N*-diisopropyl chlorophosphine (0.04 ml, 0.19 mmol). Following completion (1 h), as shown by TLC, the solution was diluted with DCM (10 ml), washed (2 x 5 ml, Na₂CO₃ (aq), 5 ml, NaCl (aq)), dried over anhydrous Na₂SO₄ and then the solvents were removed *in vacuo*. The resultant oil was purified by flash silica gel column chromatography (1:49.5:49.5 pyridine/EtOAc/Hexane) to yield the product **46** (0.08 g, 47%) as a pale yellow foam.

R_f 0.64 (90% EtOAc/Hexane)

δ_H (300 MHz; CDCl₃) 9.47 (1H, s, CH²⁴), 8.71 (1H, s, CH⁶), 7.85 (2H, m, CH¹³, ¹⁶), 7.72 (2H, dd, *J* = 3.1, 5.4 Hz, CH¹⁴, ¹⁵), 7.43-7.16 (10H, m, CH¹⁹, ²⁷⁻²⁹, ³¹), 6.80 (4H, m, CH³²), 6.48 (1H, ap.t., *J* = 6.7 Hz, CH¹'), 4.68 (2H, s, CH¹⁰), 4.60 (1H, m, CH³'), 4.21 (1H, m, CH⁴'), 3.88-3.54 (12H, m, CH²⁰, ³⁴, ³⁵, ³⁷), 3.39-3.20 (4H, m, CH⁵', ²⁰), 2.61 (1H, t, *J* = 6.4 Hz, CH³⁸), 2.50-2.42 (3H, m, CH²', ³⁸), 1.74-1.60 (4H, m, CH²¹), 1.45-1.33 (4H, m, CH²²), 1.20-1.10 (12H, m, CH³⁶), 0.97 (3H, t, *J* = 7.3 Hz, CH²³), 0.94 (3H, t, *J* = 7.3 Hz, CH²³)

δ_P (90 MHz; CDCl₃) 149.4, 149.1

ES⁺/MS: 1118.4 (M+H)⁺

47. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-2,4-diamine

43 → **47**

10% MeNH₃ / H₂O, rt, 24 h, 56%
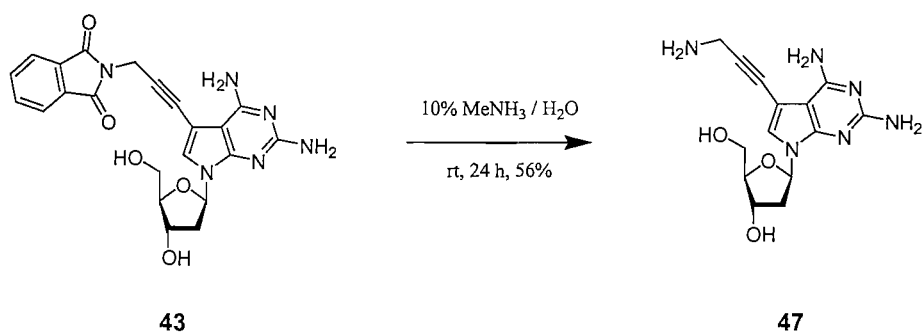
Compound **43** (0.05 g, 0.11 mmol) was dissolved in 10% MeNH₂ in water (2 ml) and left for 24 h. TLC analysis showed the reaction to be finished and the solvent was removed *in vacuo*. The residue was then purified by flash silica gel column chromatography (20-40% MeOH / DCM) to yield **47** (0.02 g, 56%) as an amorphous solid.

R_f 0.10 (20% MeOH / DCM)

δ_H (300 MHz; DMSO-d6) 7.22 (1H, s, CH⁶), 6.31 (1H, dd, J = 5.9, 8.1 Hz, CH¹'), 5.24 (1H, br.s, OH³'), 5.09 (1H, br.s, OH⁵'), 4.29 (1H, m, CH³'), 3.76 (1H, m, CH⁴'), 3.50 (4H, m, CH⁵', ¹⁰), 2.35 (1H, ddd, J = 5.7, 8.1, 13.1 Hz, CH²'), 2.05 (1H, ddd, J = 2.0, 5.9, 13.1 Hz, CH²')

δ_C (75.5 MHz; DMSO-d6) 157.8 (C⁴), 121.6 (C⁶), 95.5 (C⁹), 87.0 (C⁴'), 82.3 (C¹'), 71.0 (C³'), 63.0 (C⁸), 61.9 (C⁵'), 39.4 (C²'), 31.6 (C¹⁰)

UV/vis (H₂O): 260 nm (5.9 x 10³ cm⁻¹mol⁻¹)

ES⁺/MS: 319.1 (M+H)⁺

HRMS (ES⁺) for C₁₄H₁₈N₆O₃ (M+H)⁺: calculated 319.1513, found 319.1515

48. 7-(2'-Deoxy-3',5'-di-*O*-*p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodo-2-(methylthio)pyrrolo[2,3-*d*]pyrimidin-4-one[74]

41   **48**
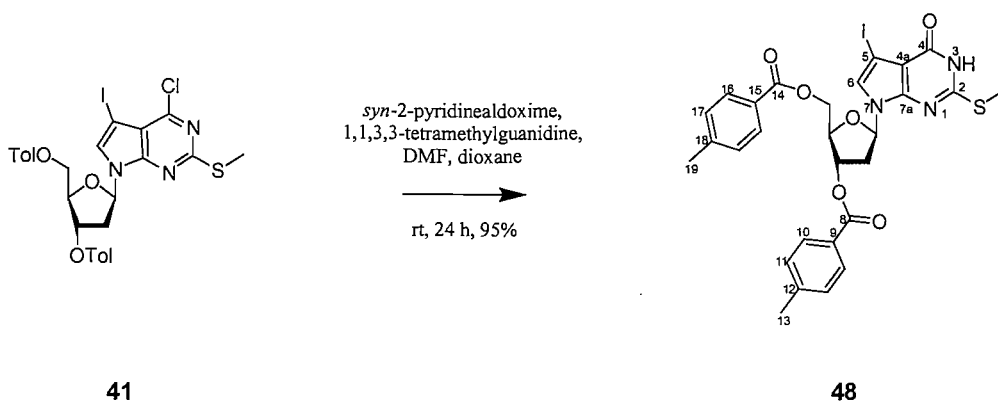
To compound **41** (10 g, 15.48 mmol) in DMF (150 ml) and dioxane (100 ml) was added *syn*-2-pyridinealdoxime (10.40 g, 85.16 mmol) and 1,1,3,3-tetramethylguanidine (11.66 ml, 92.90 mmol). The mixture was stirred at RT for 24 h and concentrated *in vacuo*. The residual solid was dissolved in DCM (40 ml), washed with 0.1 M aqueous citric acid (2 x 25 ml), H_2O (25 ml) and saturated aqueous $NaHCO_3$ (25 ml), dried (Na_2SO_4) and evaporated. The residue was dissolved in boiling iso-propanol and allowed to cool before collecting the white crystals. The solid **48** was dried *in vacuo* (9.2 g, 95%).

R$_f$ 0.53 (70% EtOAc / Hexane)

δ$_H$ (300 MHz; CDCl$_3$) 11.73 (1H, s, NH), 7.96 (2H, d, J = 8.0 Hz, CH^{10}), 7.94 (2H, d, J = 8.0 Hz, CH^{16}), 7.28 (4H, d, J = 8.0 Hz, CH$^{11, 17}$), 7.07 (1H, s, CH^6), 6.62 (1H, dd, J = 6.4, 7.4 Hz, CH$^{1'}$), 5.71 (1H, ddd, J = 2.9, 3.0, 6.1 Hz, CH$^{3'}$), 4.68 (1H, m, CH$^{4'}$), 4.57 (2H, dd, J = 3.5, 6.7 Hz, CH$^{5'}$), 2.82-2.67 (2H, m, CH$^{2'}$), 2.64 (3H, s, SCH$_3$), 2.44 (3H, s, CH^{13}), 2.42 (3H, s, CH^{19})

δ$_C$ (75.5 MHz; CDCl$_3$) 166.2 (C^8), 165.9 (C^{14}), 159.9 (C^4), 156.6 (C^2), 148.2 (C^{7a}), 144.4 (C^{12}), 144.1 (C^{18}), 129.7 (C^{10}), 129.6 (C^{16}), 129.4 (C^{11}), 129.2 (C^{17}), 126.7 (C^9), 126.5 (C^{15}), 123.9 (C^6), 105.7 (C^{4a}), 83.8 (C$^{1'}$), 82.3 (C$^{5'}$), 74.8 (C$^{3'}$), 63.9 (C$^{4'}$), 55.6 (C^5), 38.5 (C$^{2'}$), 21.7 (C$^{13, 19}$), 13.6 (SCH$_3$)

IR v_{max}/cm^{-1}: 2943 (w), 1720 (s), 1690 (s), 1610 (m), 1549 (m), 1521 (m), 1424 (w), 1409 (m), 1372 (m), 1317 (m), 1270 (s), 1096 (s), 1040 (m), 1017 (m), 959 (m), 937 (m), 842 (m), 752 (s)

ES$^+$/MS: 660.1 (M+H)$^+$, 682.0 (M+Na)$^+$, 1340.9 (2M+Na)$^+$

Mpt. 223-225°C (from i-PrOH)

49. 2-Amino-7-(2'-deoxy-3',5'-di-*O-p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-4-one[74]
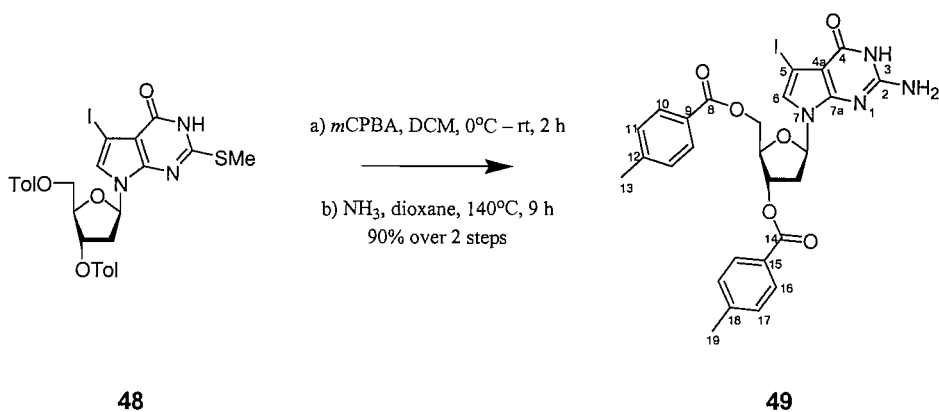
To compound **48** (4.5 g, 6.82 mmol) in DCM (190 ml) at 0°C was added MCPBA (~75 %), (2.2 g, 9.55 mmol). After 15 min the ice bath was removed and stirring continued for an additional 105 min. The reaction mixture was diluted with DCM, washed with saturated aqueous NaHCO₃ and dried (Na₂SO₄). After removal of Na₂SO₄ by filtration, MeOH was added to the filtrate to make the solution 4% MeOH. This solution was directly chromatographed through a short column of silica gel, eluted with 4% MeOH in DCM and concentrated.

The sulphoxide (4.5 g, 6.66 mmol) was suspended in dioxane (50 ml) saturated with ammonia at 12°C in a Parr® general-purpose acid digestion bomb (stainless steel, 125 ml). The reaction vessel was sealed, heated at 145°C for 16 h, cooled and concentrated. The residue was taken up in DCM (100 ml), extracted with saturated NaHCO₃ (aq), dried (Na₂SO₄) and evaporated. Polar compounds were then removed by passing the reaction mixture through a silica gel plug (5% MeOH / DCM) to yield **49** (3.75 g, 90% (over 2 steps)) as an amorphous solid.

R$_f$ 0.68 (10% MeOH/DCM)

δ_H (300 MHz; CDCl₃) 10.80 (1H, s, NH), 7.95 (4H, d, *J* = 7.1 Hz, CH¹⁰, ¹⁶), 7.26 (4H, m, CH¹¹, ¹⁷), 6.87 (1H, s, CH⁶), 7.48 (1H, dd, *J* = 6.2, 7.7 Hz, CH¹'), 6.09 (2H, s, NH₂), 5.69 (1H, m, CH³'), 4.71 (1H, dd, *J* = 3.8, 11.9 Hz, CH⁵'), 4.63 (1H, dd, *J* = 3.5, 11.9 Hz, CH⁵'), 4.54 (1H, m, CH⁴'), 2.73 (1H, m, CH²'), 2.61 (1H, m, CH²'), 2.43 (3H, s, CH¹⁹), 2.41 (3H, s, CH¹³)

δ_C (75.5 MHz; CDCl$_3$) 166.3 (**C**^{14}), 165.9 (**C**^{8}), 160.0 (**C**^{4}), 152.5 (**C**^{2}), 151.5 (**C**^{7a}), 144.4 (**C**^{18}), 144.1 (**C**^{12}), 129.8 (**C**^{16}), 129.7 (**C**^{10}), 129.4 (**C**^{17}), 129.2 (**C**^{11}), 126.8 (**C**^{15}), 126.5 (**C**^{9}), 122.6 (**C**^{6}), 101.1 (**C**^{4a}), 83.8 (**C**$^{1'}$), 82.2 (**C**$^{4'}$), 75.2 (**C**$^{3'}$), 64.1 (**C**$^{5'}$), 54.7 (**C**^{5}), 37.8 (**C**$^{2'}$), 21.7 (**C**$^{13, 19}$)

ES$^+$/MS: 629.1 (M+H)$^+$, 651.2 (M+Na)$^+$

173

50. 2-Amino-7-(2'-deoxy-β-ᴅ-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-4-one[74]

49   **50**
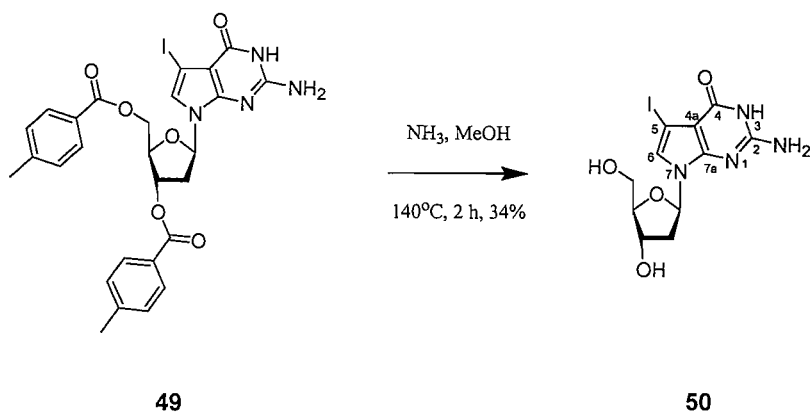
Compound **49** (3.75 g, 5.97 mmol) was dissolved in 2 M NH₃/MeOH (50 ml) in a Parr®
general-purpose acid digestion bomb (stainless steel, 125 ml). The sealed vessel was then
heated (140°C, 2 h), cooled (3 h) and the contents concentrated *in vacuo*. The residue was
dissolved in H₂O, washed twice with Et₂O, filtered and concentrated *in vacuo*. The solid
was then dissolved in a minimum of boiling water (~20 ml) and allowed to cool slowly
forming a precipitate. The precipitate was collected by filtration, washed with cold H₂O
and Et₂O and dried *in vacuo* to give **50** (0.8 g, 34%).

R$_f$ 0.12 (10% MeOH/DCM)

δ$_H$ (300 MHz; DMSO-d6) 10.45 (1H, s, N**H**), 7.11 (1H, s, C**H**^{6}), 6.31 (2H, s, N**H**$_2$), 6.26
(1H, dd, *J* = 5.8, 8.4 Hz, C**H**$^{1'}$), 4.27 (1H, m, C**H**$^{3'}$), 3.75 (1H, m, C**H**$^{4'}$), 3.49 (2H, m,
C**H**$^{5'}$), 2.31 (1H, ddd, *J* = 5.6, 8.4, 13.3 Hz, C**H**$^{2'}$), 2.05 (1H, ddd, *J* = 2.3, 5.8, 13.3 Hz,
C**H**$^{2'}$)

δ$_C$ (75.5 MHz; DMSO-d6) 158.0 (**C**^{4}), 152.6 (**C**^{2}), 150.5 (**C**^{7a}), 121.6 (**C**^{6}), 99.8 (**C**^{4a}),
87.0 (**C**$^{4'}$), 82.1 (**C**$^{1'}$), 70.8 (**C**$^{3'}$), 61.8 (**C**$^{5'}$), 54.9 (**C**^{5}), 39.5 (**C**$^{2'}$)

ES⁺/MS: 393.2 (M+H)⁺, 431.1 (M+K)⁺, 785.1 (2M+H)⁺, 822.9 (2M+K)⁺

51. **2-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[74]**
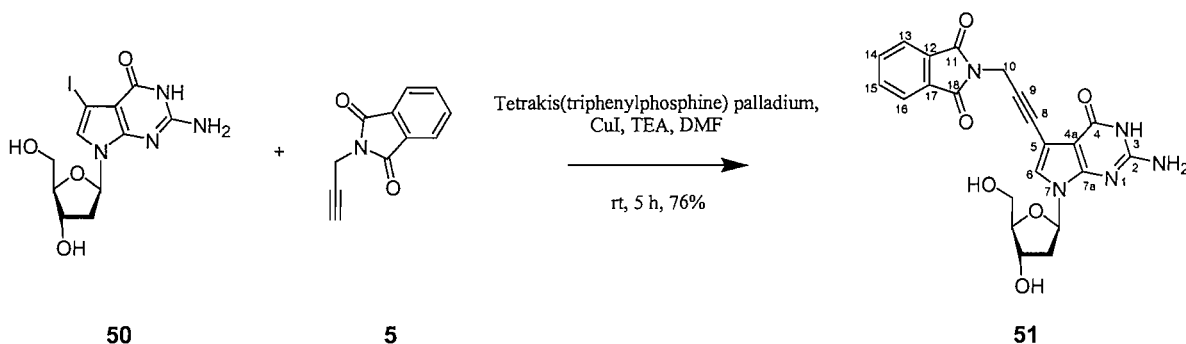
To a solution of **50** (1.29 g, 3.30 mmol) in anhydrous DMF (20 ml) under argon was added copper iodide(I) (0.13 g, 0.66 mmol), anhydrous triethylamine (2.30 ml, 16.50 mmol) and **5** (0.67 g, 3.63 mmol). The reaction was stirred in the dark (10 min), then tetrakis(triphenylphosphine)palladium (0.38 g, 0.33 mmol) was added. The reaction was stirred (5 h) after which TLC showed the reaction to be complete. The reaction was then concentrated *in vacuo* to give a brown oil. The oil was purified by flash silica gel column chromatography to yield **51** (1.12 g, 76%) as a yellow amorphous solid.

R$_f$ 0.50 (20% MeOH/DCM)

δ$_H$ (300 MHz; DMSO-d6) 10.45 (1H, s, NH), 7.92 (2H, dd, *J* = 3.1, 5.4 Hz, CH$^{13, 16}$), 7.87 (2H, dd, *J* = 3.1, 5.4 Hz, CH$^{14, 15}$), 7.25 (1H, s, CH^6), 6.31 (2H, s, NH$_2$), 6.25 (1H, dd, *J* = 5.8, 8.3 Hz, CH$^{1'}$), 5.17 (1H, d, *J* = 3.7, OH$^{3'}$), 4.86 (1H, t, *J* = 5.5 Hz, OH$^{5'}$), 4.57 (2H, s, CH^{10}), 4.25 (1H, m, CH$^{3'}$), 3.74 (1H, m, CH$^{4'}$), 3.47 (2H, m, CH$^{5'}$), 2.29 (1H, ddd, *J* = 5.7, 8.3, 13.3 Hz, CH$^{2'}$), 2.05 (1H, ddd, *J* = 2.3, 5.8, 13.3 Hz, CH$^{2'}$)

δ$_C$ (75.5 MHz; DMSO-d6) 166.7 (C$^{11, 18}$), 157.6 (C^4), 153.1 (C^2), 150.2 (C^{7a}), 134.6 (C$^{14, 15}$), 131.4 (C$^{12, 17}$), 123.2 (C$^{13, 16}$), 122.5 (C^6), 99.2 (C^{4a}), 97.6 (C^5), 87.1 (C$^{4'}$), 83.4 (C^9), 82.2 (C$^{1'}$), 76.6 (C^8), 70.8 (C$^{3'}$), 61.8 (C$^{5'}$), 39.5 (C$^{2'}$), 27.7 (C^{10})

IR v$_{max}$/cm^{-1}: 3328 (br. m), 2931 (w), 1771 (w), 1723 (m), 1679 (m), 1626 (s), 1590 (s), 1555 (m), 1506 (m), 1424 (m), 1385 (m), 1322 (m), 1177 (m), 1108 (m), 1044 (s), 996 (m), 943 (m), 870 (w), 782 (m), 723 (s)

ES$^+$/MS: 472.2 (M+Na)$^+$

52. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-2-{[(*N*,*N*-dimethylamino)methylidene]amino}-5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[74]

51 → **52**

N,*N*-Dimethylformamide diethylacetal, EtOH, DMF

rt, 24 h, 40%
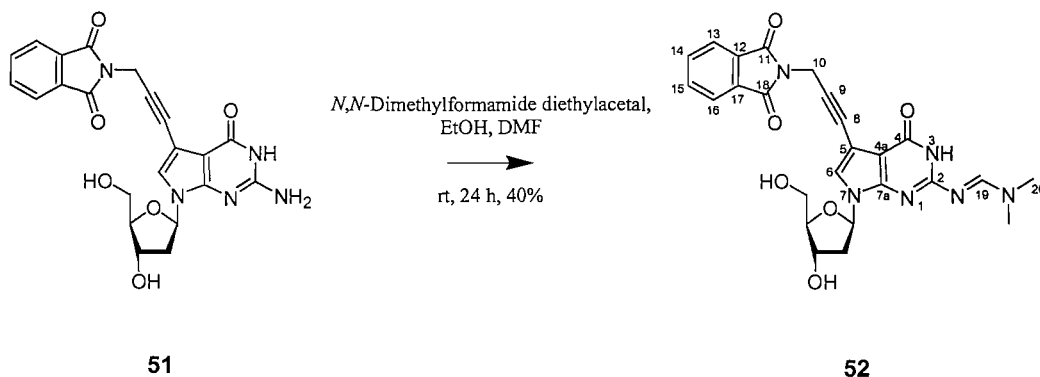
Compound **51** (1.12 g, 2.49 mmol) was dissolved in DMF (15 ml) and to this was added absolute EtOH (0.14 ml) and *N*,*N*-dimethylformamide diethyl acetal (0.51 ml, 2.99 mmol). After stirring for 24 h at RT TLC analysis showed that the reaction had reached completion and H_2O (0.5 ml) was added. The solvents were then removed *in vacuo* and the resulting oil was purified by flash silica gel column chromatography 10% MeOH in DCM to yield (0.5 g, 40%) of **52** as a yellow amorphous solid.

R$_f$ 0.52 (20% MeOH / DCM)

δ_H (300 MHz; DMSO-d6) 11.07 (1H, s, NH), 8.53 (1H, s, CH[19]), 7.92 (2H, dd, J = 3.1, 5.6 Hz, CH[13, 16]), 7.87 (2H, dd, J = 3.1, 5.6 Hz, CH[14, 15]), 7.38 (1H, s, CH[6]), 6.40 (1H, dd, J = 6.3, 7.7 Hz, CH[1']), 5.22 (1H, d, J = 3.9 Hz, OH[3']), 4.87 (1H, t, J = 5.4 Hz, OH[5']), 4.59 (2H, s, CH[10]), 4.29 (1H, m, CH[3']), 3.77 (1H, m, CH[4']), 3.55-3.45 (2H, m, CH[5']), 3.14 (3H, s, CH[20]), 3.01 (3H, s, CH[20]), 2.37 (1H, ddd, J = 6.0, 7.7, 13.3 Hz, CH[2']), 2.12 (1H, ddd, J = 2.5, 6.3, 13.3 Hz, CH[2'])

δ_C (75.5 MHz; DMSO-d6) 166.7 (C[11, 18]), 158.5 (C[19]), 157.4 (C[4]), 156.6 (C[2]), 134.6 (C[14, 15]), 131.4 (C[12, 17]), 123.9 (C[6]), 123.2 (C[13, 16]), 102.3 (C[4a]), 97.6 (C[5]), 87.2 (C[4']), 83.6 (C[9]), 82.3 (C[1']), 76.6 (C[8]), 70.9 (C[3']), 61.8 (C[5']), 40.4 (C[20]), 39.8 (C[2']), 34.5 (C[20]), 27.7 (C[10])

ES$^+$/MS: 505.5 (M+H)$^+$, 527.4 (M+Na)$^+$, 1009.6 (M+H)$^+$, 1031.5 (M+Na)$^+$

53. 7-(2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)-2-{[(*N*,*N*-dimethylamino)methylidene]amino}-5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[80]

52   **53**
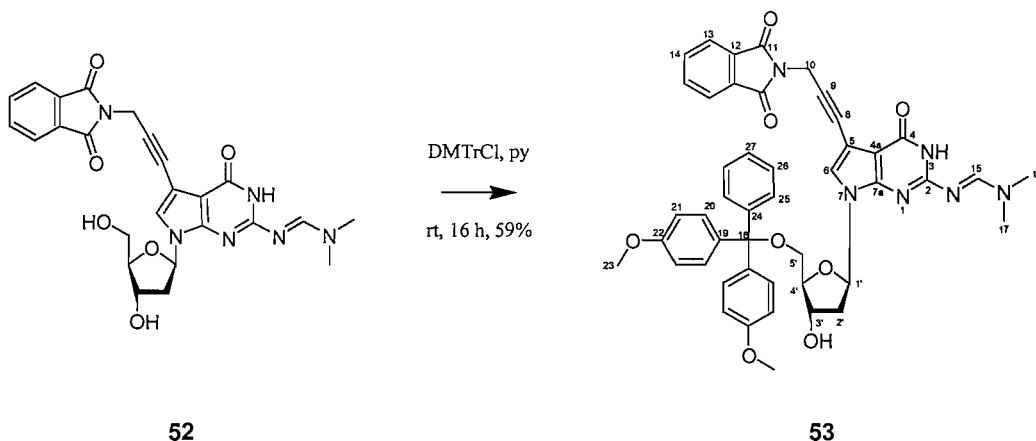
Compound **52** (0.33 g, 0.65 mmol) and 4,4'-dimethoxytrityl chloride (0.23 g, 0.69 mmol) were dissolved in anhydrous pyridine (20 ml) under argon. The solution was stirred for 4 h when TLC analysis showed the reaction to be complete. MeOH (3 ml) was added and stirred for 10 min then sat. aq NaHCO$_3$ solution (30ml) was added and the aqueous layer was extracted with DCM (3 x 60 ml). The organic fractions were combined, dried (Na$_2$SO$_4$) and the solvent was removed *in vacuo*. The resultant foam was purified by flash silica gel column chromatography (1:5:94 Py/MeOH/DCM) to yield **53** as a yellow foam (0.31 g, 59%).

R$_f$ 0.21 (5% MeOH/DCM)

δ_H (300 MHz; CDCl$_3$) 9.03 (1H, br. s, NH), 8.49 (1H, s, CH^{15}), 7.81-7.83 (2H, dd, *J* = 3.0, 5.5 Hz, CH^{13}) 7.67-7.69 (2H, dd, *J* = 3.0, 5.5 Hz, CH^{14}), 7.15-7.42 (9H, m, CH$^{20, 25-27}$), 6.94 (1H, s, CH^6), 6.79-6.82 (4H, dd, *J* = 3.1, 8.8 Hz, ArH^{21}), 6.54 (1H, app. t, *J* = 6.8 Hz, CH$^{1'}$), 4.69 (2H, s, NCH$_2$), 4.48 (1H, m, CH$^{3'}$), 4.05-4.08 (1H, dt, *J* = 5.3, 3.3 Hz, CH$^{4'}$), 3.76 & 3.77 (3H, s, OCH$_3$), 3.35-3.39 (1H, dd, *J* = 10.0, 5.3 Hz, CH$^{5'}$), 3.18-3.22 (1H, dd, *J* = 10.0, 5.3 Hz, CH$^{5'}$), 2.98 & 3.05 (6H, 2s, NCH$_3$), 2.28-2.41 (2H, m, CH$^{2'}$)

δ_C (75.5 MHz; CDCl$_3$) 167.03 (C^{11}), 159.17 (C^4), 158.50 (C^{22}) 158.46 (C^{22}), 157.60 (C^{15}), 155.9 (C^2), 149.6 (C^{7a}), 144.6 (C^{24}), 135.8 (C^{19}), 135.6 (C^{19}), 133.9 (C^{14}), 132.1 (C^{12}), 130.2 (C^{20}), 130.1 (C^{20}), 129.9 (C^{20}), 128.2 (C^{20}), 128.1(C^{26}), 127.8 (C^{26}), 126.8 (C^{25}), 125.24 (C^{27}), 123.33 (C^{13}), 123.07 (C^6), 113.15 (C^{21}), 103.46 (C^{4a}), 99.42 (C^5),

86.45 (C^{18}), 85.18 ($C^{4'}$), 83.67 (C^9), 82.43 ($C^{1'}$), 76.37 (C^8), 72.78 ($C^{3'}$), 64.30 ($C^{5'}$), 55.17 (C^{23}), 41.14 (C^{17}), 40.16 ($C^{2'}$), 34.93 (C^{16}), 28.31 (C^{10})

IR ν_{max}/cm^{-1}: 2932 (w), 1772 (w), 1718 (s), 1671 (m), 1627 (s), 1554 (s), 1526 (s), 1508 (s), 1421 (s), 1391 (m), 1339 (s), 1247 (s), 1175 (m), 1115 (s), 1031 (s), 939 (m), 828 (m), 787 (m), 724 (s)

ES$^+$/MS: 807 (M+H)$^+$, 829.3 (M+Na)$^+$

HRMS (ES$^+$) for $C_{46}H_{42}N_6O_8$ (M+Na)$^+$: calculated 829.2956 found 829.2978

54. 7-(3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)-2-{[(*N,N*-dimethylamino)methylidene]amino}-5-[4-phthalimidoprop-1-ynyl]-7*H*-pyrrolo[2,3-*d*]pyrimidin-4-one

2-Cyanoethoxy-(*N,N*-diisopropylamino)chlorophosphine, DIPEA, DCM
→
rt, 1 h, 75%

53   **54**
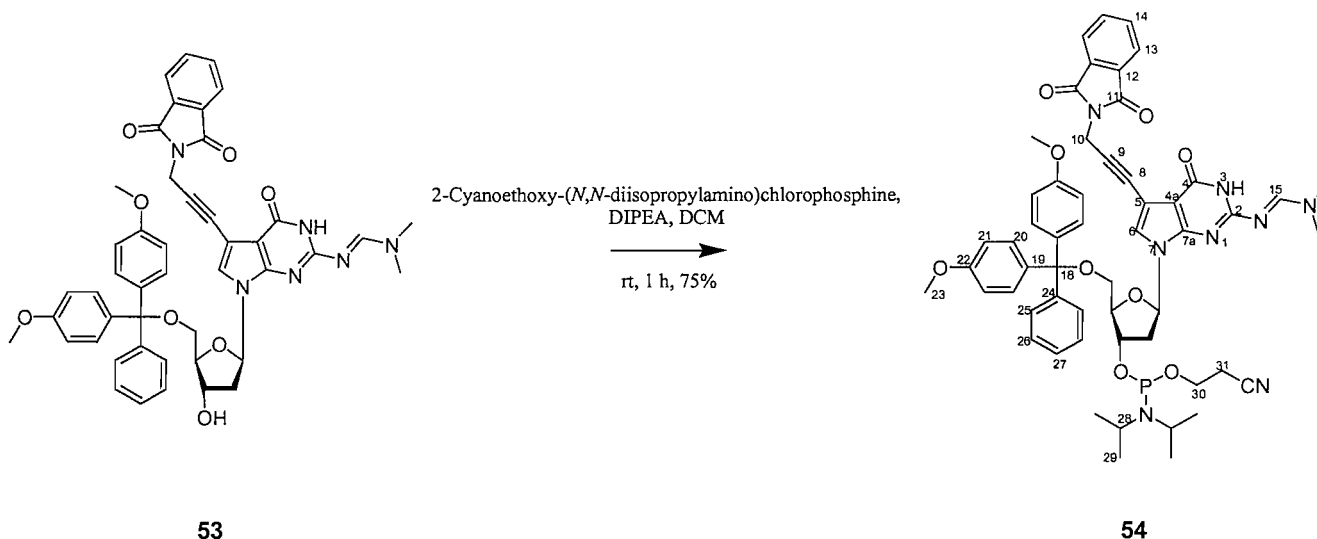
Compound **53** (0.32 g, 0.40 mmol) was dissolved in anhydrous DCM (10 ml) and diisopropylethylamine (0.17 ml, 0.99 mmol) and 2-cyanoethyl-*N,N*-diisopropyl chlorophosphine (0.11 ml, 0.48 mmol) were added. Following completion (1 h), as shown by TLC, the solution was diluted with DCM (30 ml), washed (2 x 10 ml, Na$_2$CO$_3$ (aq), 10 ml, NaCl (aq)), dried over anhydrous Na$_2$SO$_4$ and then the solvents were removed *in vacuo*. The resultant oil was purified by flash silica gel column chromatography (1:90:9 py/EtOAc/Hexane) to yield **54** (0.30 g, 75%) as a pale yellow foam.

R$_f$ 0.10 (EtOAc)

δ$_H$ (300 MHz; CDCl$_3$) 8.60 (1H, s, CH[15]), 7.84 (2H, dd, *J* = 3.1, 5.6 Hz, CH[13]), 7.69 (2H, dd, *J* = 3.1, 5.6 Hz, CH[14]), 7.32-7.16 (9H, m, CH[20, 25-27]), 6.97 (1H, s, CH[6]), 6.80 (4H, m, CH[21]), 6.53 (1H, m, CH[1']), 4.71 (2H, s, CH[10]), 4.57 (1H, m, CH[3']), 4.18 (1H, m, CH[4']), 3.87-3.46 (10H, m, CH[23, 28, 30]), 3.33-3.18 (2H, m, CH[5']), 3.14 (3H, s, CH[16]), 3.07 (3H, s, CH[17]), 2.61 (1H, t, *J* = 6.4 Hz, CH[31]), 2.49-2.33 (3H, m, CH[2', 31]), 1.18 (3H, d, *J* = 6.7 Hz, CH[29]), 1.17 (3H, d, *J* = 6.7 Hz, CH[29]), 1.15 (3H, d, *J* = 6.7 Hz, CH[29]), 1.09 (3H, d, *J* = 6.7 Hz, CH[29])

δ$_P$ (90 MHz; CDCl$_3$) 149.4, 149.1

ES$^+$/MS: 1007.4 (M+H)$^+$, 1029.4 (M+Na)$^+$

55. 2-Amino-7-(2'-deoxy-β-D-*erythro*-pentofuranosyl)- 5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimid-4-one
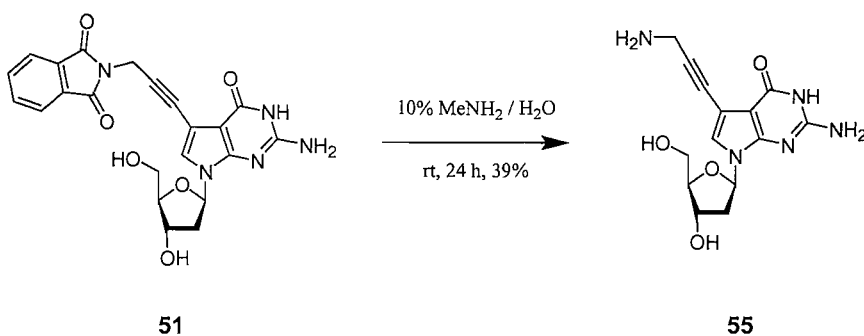
Compound **51** (0.05 g, 0.09 mmol) was dissolved in 10% $MeNH_2$ in water (2 ml) and left for 24 h. TLC analysis showed the reaction to be finished and the solvent was removed *in vacuo*. The residue was then purified by flash silica gel column chromatography (20-40% MeOH / DCM) to yield **55** (0.01 g, 39%) as an amorphous solid.

R_f 0.15 (40% MeOH / DCM)

δ_H (300 MHz; DMSO-d6) 7.16 (1H, s, CH^6), 6.49 (2H, br.s, NH_2), 6.27 (1H, dd, J = 5.9, 8.1 Hz, $CH^{1'}$), 5.23 (1H, br.s, $OH^{3'}$), 4.96 (1H, br.s, $OH^{5'}$), 4.27 (1H, m, $CH^{3'}$), 3.75 (1H, m, $CH^{4'}$), 3.49 (2H, m, $CH^{5'}$), 3.45 (2H, s, CH^{10}), 2.30 (1H, ddd, J = 5.7, 8.1, 13.1 Hz, $CH^{2'}$), 2.06 (1H, ddd, J = 2.2, 5.9, 13.1 Hz, $CH^{2'}$)

δ_C (75.5 MHz; DMSO-d6) 153.5 (C^2), 153.4 (C^4), 150.3 (C^{7a}), 121.2 (C^6), 99.2 (C^{4a}), 99.0 (C^5), 91.3 (C^9), 87.0 ($C^{4'}$), 82.2 ($C^{1'}$), 75.7 (C^8), 70.9 ($C^{3'}$), 61.8 ($C^{5'}$), 39.6 ($C^{2'}$), 31.6 (C^{10})

UV/vis (H_2O): 260 nm (4.6 x 10^3 $cm^{-1}mol^{-1}$)

ES$^+$/MS: 342.1 $(M+H)^+$

HRMS (ES^+) for $C_{13}H_{17}N_5O_4$ $(M+Na)^+$: calculated 342.1173, found 342.1177

56. 7-(2'-Deoxy-3',5'-di-*O*-*p*-toluoyl-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimidin-4-one[74]

syn-2-pyridinealdoxime,
1,1,3,3-tetramethylguanidine,
DMF, dioxane
⟶
rt, 24 h, 95%

27   **56**
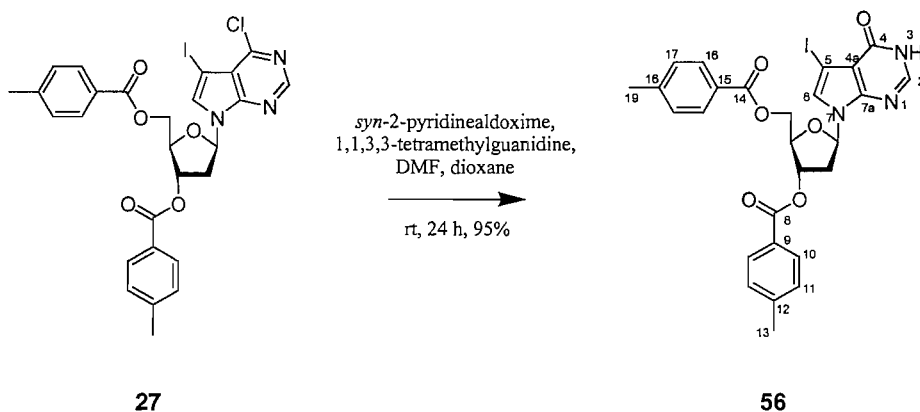
To compound **27** (0.5 g, 0.79 mmol) in DMF (8 ml) and dioxane (5 ml) was added *syn*-2-pyridinealdoxime (0.53 g, 4.35 mmol) and 1,1,3,3-tetramethylguanidine (0.6 ml, 4.75 mmol). The mixture was stirred at RT for 24 h and then concentrated *in vacuo*. The residual solid was dissolved in DCM (40 ml), washed with 0.1 M aqueous citric acid (2 x 25 ml), H_2O (25 ml) and saturated aqueous $NaHCO_3$ (25 ml), dried (Na_2SO_4) and evaporated. The residue was washed in boiling iso-propanol and allowed to cool to 4°C before collecting the white crystals. The solid, **56**, was dried *in vacuo* (0.46 g, 95%).

R$_f$ 0.23 (70% EtOAc/Hexane)

δ$_H$ (300 MHz; $CDCl_3$) 12.10 (1H, s, NH), 7.94 (1H, s, CH^2), 7.93 (2H, d, *J* = 8.1 Hz, CH^{10}), 7.88 (2H, d, *J* = 8.1 Hz, CH^{16}), 7.51 (1H, s, CH^6), 7.35 (2H, d, *J* = 8.1 Hz, CH^{11}), 7.33 (2H, d, *J* = 8.1 Hz, CH^{17}), 6.57 (2H, dd, *J* = 6.1, 7.2 Hz, CH$^{1'}$), 5.69 (1H, m, CH$^{3'}$), 4.64-4.48 (3H, m, CH$^{4',5'}$), 2.94 (1H, ddd, *J* = 7.2, 8.0, 14.5 Hz, CH$^{2'}$), 2.68 (1H, ddd, *J* = 2.3, 6.1, 14.5 Hz, CH$^{2'}$), 2.39 (3H, s, CH^{13}), 2.38 (3H, s, CH^{19})

δ$_C$ (75.5 MHz; $CDCl_3$) 165.4 (C^8), 165.1 (C^{14}), 157.4 (C^4), 147.3 (C^2), 144.7 (C^{7a}), 143.9 (C^{18}), 143.7 (C^{12}), 129.5 (C^{17}), 129.4 (C^{11}), 129.3 C^{16}), 129.2 (C^{10}), 126.5 (C^{15}), 126.4 (C^9), 125.2 (C^6), 108.2 (C^{4a}), 83.1 (C$^{1'}$), 81.3 (C$^{4'}$), 74.7 (C$^{3'}$), 64.0 (C$^{5'}$), 55.9 (C^5), 36.4 (C$^{2'}$), 21.1 (C$^{13, 19}$)

IR v_{max}/cm^{-1}: 3068 (br), 2361 (w), 1718 (s), 1679 (s), 1610 (m), 1592 (m), 1522 (m), 1426 (m), 1380 (m), 1272 (s), 1178 (s), 1101 (s), 1020 (m), 966 (m), 932 (m), 839 (m), 782 (m), 751 (s), 679 (m), 635 (w), 603 (m)

ES$^+$/MS: 614.1 (M+H)$^+$, 636.1 (M+Na)$^+$

HRMS (ES$^+$) for $C_{27}H_{24}IN_3O_6$ (M+H)$^+$: calculated 614.0783, found 614.0788

Mpt. 204-208°C (from iPrOH)

Elemental Analysis for $C_{27}H_{24}IN_3O_6$ calculated C 52.87, H 3.94, N 6.85; found, C 53.12, H 4.06, N 6.80

57.   7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)-5-iodopyrrolo[2,3-*d*]pyrimid-4-one

56   **57**
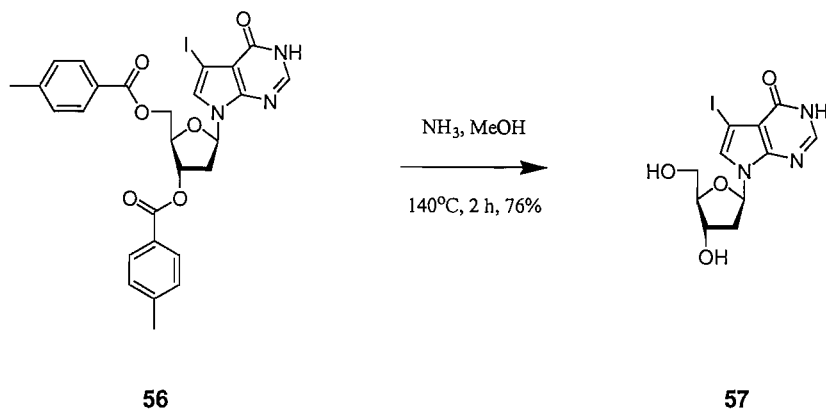
Compound **56** (2.2 g, 3.59 mmol) was dissolved in 2 M NH_3/MeOH (50 ml) in a Parr®
general-purpose acid digestion bomb (stainless steel, 125 ml). The sealed vessel was then
heated (140°C, 2 h), cooled (3 h) and the contents concentrated *in vacuo*. The residue was
dissolved in H_2O, washed twice with Et_2O, filtered and concentrated *in vacuo*. The solid
was then dissolved in a minimum of boiling water (~20 ml) and allowed to cool slowly
forming off white crystals. The tan crystals of **57** were collected by filtration, washed
with cold H_2O and Et_2O and dried *in vacuo* (1.03 g, 76%).

R$_f$ 0.43 (20% MeOH/DCM)

δ$_H$ (300 MHz; DMSO-d6) 12.04 (1H, s, NH), 7.93 (1H, s, CH^2), 7.54 (1H, s, CH^6), 6.43
(1H, dd, J = 6.1, 7.8 Hz, CH$^{1'}$), 5.25 (1H, d, J = 3.9 Hz, OH$^{3'}$), 4.93 (1H, t, J = 5.2 Hz,
OH$^{5'}$), 4.31 (1H, m, CH$^{3'}$), 3.81 (1H, m, CH$^{4'}$), 3.53 (2H, m, CH$^{5'}$), 2.41 (1H, ddd, J =
5.8, 7.8, 13.3 Hz, CH$^{2'}$), 2.18 (1H, ddd, J = 2.8, 6.1, 13.3 Hz, CH$^{2'}$)

δ$_C$ (75.5 MHz; DMSO-d6) 157.5 (C^4), 147.0 (C^{7a}), 144.4 (C^2), 125.4 (C^6), 107.8 (C^{4a}),
87.4 (C$^{4'}$), 83.0 (C$^{1'}$), 70.7 (C$^{3'}$), 61.6 (C$^{5'}$), 55.1 (C^5), 40.0 (C$^{2'}$)

ES$^+$/MS: (M+H)$^+$ 378.2

HRMS (ES)$^+$ for $C_{11}H_{12}IN_3O_4$ (M+H)$^+$: calculated 377.9945, found 377.9947

Mpt. 212-214°C (from H_2O)

Elemental Analysis for $C_{11}H_{12}IN_3O_4$ calculated C 35.03, H 3.21, N 11.14; found, C
34.87, H 3.20, N 10.96

58. 7-(2'-Deoxy-β-D-*erythro*-pentofuranosyl)- 5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one[74]

57   **5**   **58**
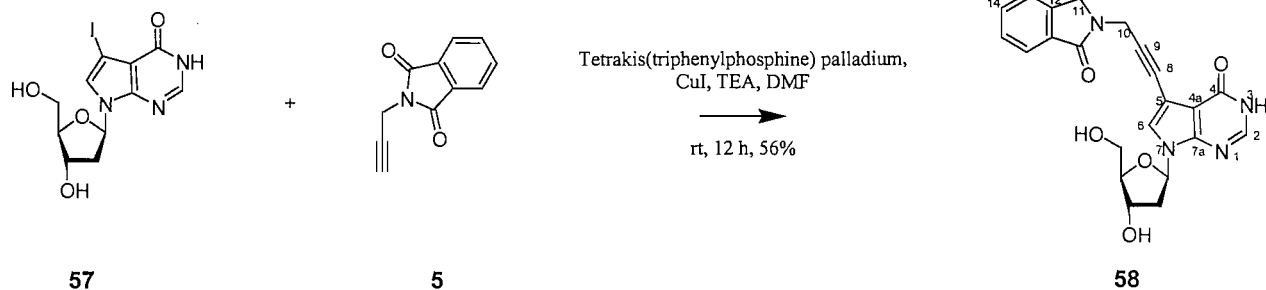
Compound **57** (0.2 g, 3.3 mmol) and CuI (0.13 g, 0.66 mmol) were dissolved in DMF (5 ml). To this was added **5** (0.11 g, 0.58 mmol) and triethylamine (0.37 ml, 2.65 mmol). After 10 min in the dark tetrakis(triphenylphosphine)palladium(0), (0.06 g, 0.05 mmol), was added. The reaction was stirred under argon and in the dark for 5 h, before TLC analysis showed the reaction to be complete. The solvents were removed *in vacuo*, the residue was suspended in DCM, filtered, then suspended in 10% MeOH / DCM and filtered to yield **58** (0.13 g, 56%) as an amorphous solid.

R_f 0.53 (20% MeOH/DCM)

δ_H (300 MHz; DMSO-d6) 12.07 (1H, s, NH), 7.93 (2H, dd, J = 3.1, 5.5 Hz, CH^{13}), 7.92 (1H, s, CH^2), 7.87 (2H, dd, J = 3.1, 5.5 Hz, CH^{14}), 7.53 (1H, s, CH^6), 6.42 (1H, dd, J = 6.1, 7.7 Hz, $CH^{1'}$), 5.24 (1H, d, J = 4.1 Hz, $OH^{3'}$), 4.91 (1H, t, J = 5.4 Hz, $OH^{5'}$), 4.61 (2H, s, CH^{10}), 4.30 (1H, m, $CH^{3'}$), 3.80 (1H, m, $CH^{4'}$), 3.51 (2H, m, $CH^{5'}$), 2.39 (1H, ddd, J = 5.8, 7.7, 13.3 Hz, $CH^{2'}$), 2.17 (1H, ddd, J = 2.8, 6.1, 13.3 Hz, $CH^{2'}$)

δ_C (75.5 MHz; DMSO-d6) 166.7 (C^{11}), 157.2 (C^4), 146.8 (C^{7a}), 145.0 (C^2), 134.6 (C^{14}), 131.4 (C^{12}), 125.7 (C^6), 123.2 (C^{13}), 107.5 (C^{4a}), 97.9 (C^5), 87.5 ($C^{4'}$), 84.1 (C^9), 83.0 ($C^{1'}$), 75.8 (C^8), 70.7 ($C^{3'}$), 61.6 ($C^{5'}$), 40.0 ($C^{2'}$), 27.6 (C^{10})

IR v_{max}/cm^{-1}: 3253 (w), 3085 (w), 2926 (w), 1769 (w), 1712 (s), 1680 (s), 1593 (m), 1557 (w), 1525 (w), 1455 (m), 1423 (m), 1388 (m), 1333 (m), 1188 (m), 1141 (w), 1089 (m), 1059 (m), 999 (m), 931 (m), 888 (m), 789 (m), 726 (s), 708 (m).

ES$^+$/MS: 457.3 (M+Na)$^+$, 891.3 (2M+Na)$^+$

HRMS (ES$^+$) for $C_{22}H_{18}N_4O_6$ (M+H)$^+$: calculated 457.1118, found 457.1120

59. 7-(2'-Deoxy-5'-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)- 5-(3-phthalimidoprop-1-ynyl)-4*H*-pyrrolo[2,3-*d*]pyrimidin-4-one

58   **59**
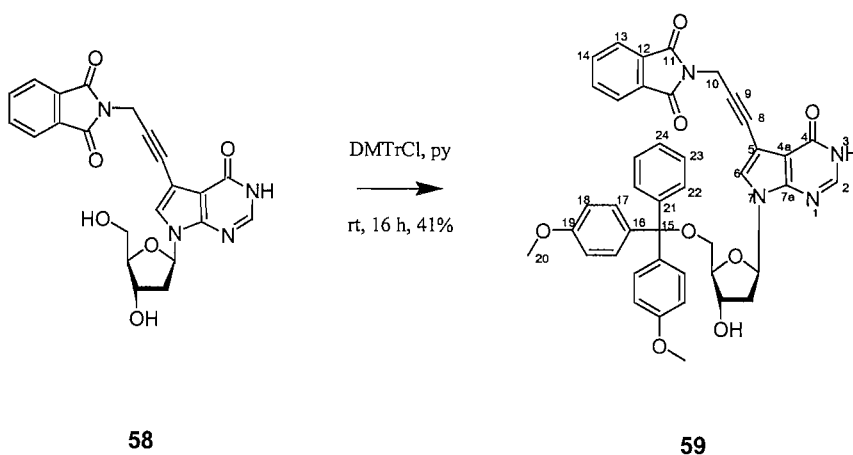
Compound **58** (0.2 g, 0.46 mmol) and 4,4'-dimethoxytrityl chloride (0.16 g, 0.48 mmol) were dissolved in anhydrous pyridine (10 ml) under argon and stirred for 16 h. After this time TLC analysis showed that the reaction was complete, MeOH (2 ml) was added and the reaction was stirred for 10 min. After this time sat. aq $NaHCO_3$ solution (30 ml) was added and the aqueous layer was extracted with DCM (3 x 40 ml). The solvent was then remove *in vacuo* and the residue was purified by flash column chromatography (5% MeOH / DCM) to yield **59** (0.14 g, 41%) a yellow foam.

R_f 0.18 (5% MeOH / DCM)

δ_H (300 MHz; CDCl$_3$) 12.59 (1H, s, NH), 8.00 (1H, s, CH^2), 7.85 (2H, dd, J = 3.1, 5.4 Hz, CH^{13}), 7.69 (2H, dd, J = 3.1, 5.4 Hz, CH^{14}), 7.42-7.18 (10H, m, CH$^{6, 17, 22-24}$), 6.82 (4H, dd, J = 2.3, 9.0 Hz, CH^{18}), 6.55 (1H, ap.t., J = 6.5 Hz, CH$^{1'}$), 4.73 (2H, s, CH^{10}), 4.53 (1H, m, CH$^{3'}$), 4.07 (1H, m, CH$^{4'}$), 3.78 (6H, s, CH^{20}), 3.41 (1H, dd, J = 4.7, 10.1 Hz, CH$^{5'}$), 3.28 (1H, dd, J = 5.2, 10.1 Hz, CH$^{5'}$), 2.44 (2H, ap.t., J = 5.8 Hz, CH$^{2'}$)

δ_C (75.5 MHz; CDCl$_3$) 167.2 (C^{11}), 160.1 (C^4), 158.5 (C^{19}), 147.5 (C^{7a}), 144.5 (C^{21}), 144.1 (C^2), 135.7 (C^{16}), 135.6 (C^{16}), 134.0 (C^{14}), 132.1 (C^{12}), 130.1 (C^{17}), 130.0 (C^{17}), 129.0 (C^{23}), 128.2 (C^{23}), 128.0 (C^{22}), 127.9 (C^{22}), 126.9 (C^{24}), 125.3 (C^6), 123.4 (C^{13}), 113.2 (C^{18}), 108.6 (C^{4a}), 99.4 (C^5), 86.6 (C^{15}), 85.5 (C$^{4'}$), 84.5 (C^9), 83.5 (C$^{1'}$), 75.6 (C^8), 72.6 (C$^{3'}$), 64.0 (C$^{5'}$), 55.2 (C^{20}), 40.5 (C$^{2'}$), 28.3 (C^{10})

IR v_{max}/cm^{-1}: 2935 (w), 1771 (w), 1715 (s), 1669 (s), 1607 (m), 1588 (m), 1508 (m), 1444 (m), 1421 (m), 1391 (m), 1339 (m), 1304 (m), 1248 (m), 1175 (m), 1115 (m), 1086 (m), 1031 (m), 940 (m), 828 (m), 790 (m), 725 (s), 710 (m)

ES$^+$/MS: 759.6 (M+Na)$^+$

HRMS (ES)$^+$ for $C_{43}H_{36}N_4O_8$ (M+Na)$^+$: calculated 759.2425, found 759.2420

60. 7-(3'-(2-Cyanoethyl-diisopropylphosphoramidyl)-2'-deoxy-5'-*O*-(4,4'-
dimethoxytrityl)-β-D-*erythro*-pentofuranosyl)-5-[4-phthalimidoprop-1-ynyl]-
7*H*-pyrrolo[2,3-*d*]pyrimid-4-one

2-Cyanoethoxy-(*N*,*N*-diisopropylamino)chlorophosphine,
DIPEA, DCM

rt, 1 h, 45 %

59   **60**
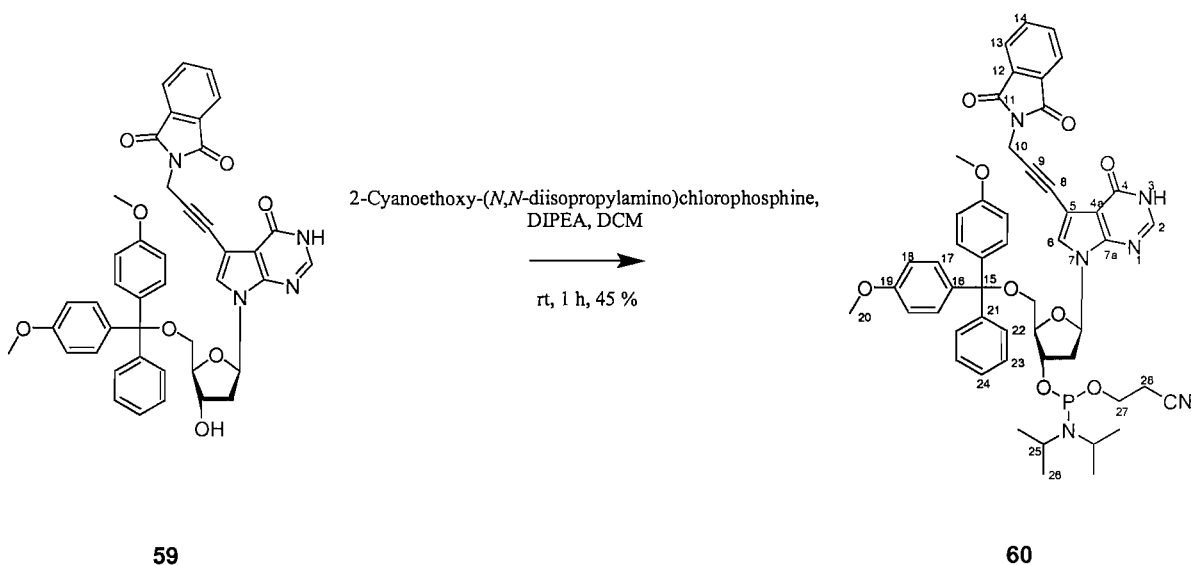
Compound **59** (0.14 g, 0.19 mmol) was dissolved in anhydrous DCM (2 ml),
diisopropylethylamine (0.08 ml, 0.48 mmol) and 2-cyanoethyl-*N*,*N*-diisopropyl
chlorophosphine (0.05 ml, 0.23 mmol) was added. Following completion (1 h), as shown
by TLC, the solution was diluted with DCM (10 ml), washed (2 x 10 ml, Na_2CO_3 (aq), 10
ml, NaCl (aq)), dried over anhydrous Na_2SO_4 and then the solvents were removed *in vacuo*.
The resultant oil was dissolved in EtOAc (5 ml) and precipitated into hexane (100 ml)
twice. The solid precipitate was then dried overnight *in vacuo*. The precipitate was then
purified by flash silica gel column chromatography (1:70:29 Py/EtOAc/Hexane) under
argon to yield **60** (0.08 g, 45%) a white foam.

R$_f$ 0.28 (90% EtOAc/Hexane)

δ_H (300 MHz; CDCl$_3$) 8.05 (1H, s, CH^2), 7.85 (2H, dd, J = 3.1, 5.5 Hz, CH^{13}), 7.70 (2H,
dd, J = 3.1, 5.5 Hz, CH^{14}), 7.43-7.18 (10H, m, CH$^{6, 17, 22-24}$), 6.81 (4H, m, CH^{18}), 6.56
(1H, ap.t., J = 6.6 Hz, CH$^{1'}$), 4.74 (2H, s, CH^{10}), 4.63 (1H, m, CH$^{3'}$), 4.24 (1H, m, CH$^{4'}$),
3.88-3.49 (10H, m, CH$^{20, 25, 27}$), 3.31 (2H, m, CH$^{5'}$), 2.64-2.45 (4H, m, CH$^{2', 28}$), 1.19 (3H,
d, J = 6.6 Hz, CH^{26}), 1.18 (3H, d, J = 6.6 Hz, CH^{26}), 1.17 (3H, d, J = 6.6 Hz, CH^{26}), 1.11
(3H, d, J = 6.6 Hz, CH^{26})

δ_P (90 MHz; CDCl$_3$) 149.4, 149.2

ES$^+$/MS: 959.6 (M+Na)$^+$

61. 7-(2-Deoxy-β-D-*erythro*-pentofuranosyl)-5-(3-aminoprop-1-ynyl)-7*H*-pyrrolo[2,3-*d*]pyrimid-4-one

58   **61**
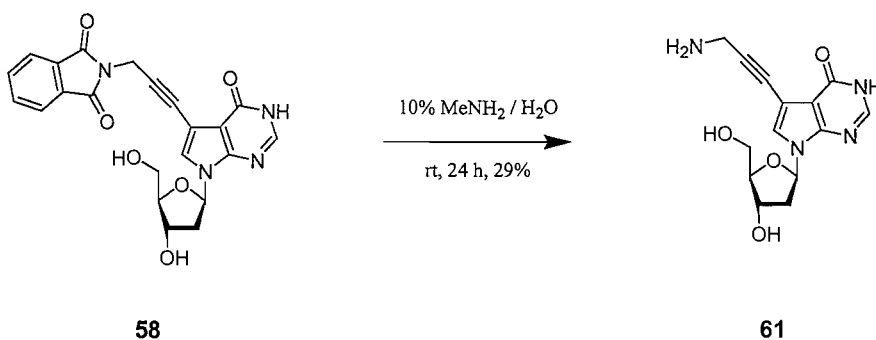
Compound **58** (0.05 g, 0.12 mmol) was dissolved in 10% MeNH$_2$ in water (2 ml) and left for 24 h. TLC analysis showed the reaction to be finished and the solvent was removed *in vacuo*. The residue was then purified by flash silica gel column chromatography (20-40% MeOH / DCM) to yield **61** (0.01 g, 29%) as an amorphous solid.

R$_f$ 0.15 (40% MeOH/DCM)

δ$_H$ (300 MHz; DMSO-d6) 7.92 (1H, s, **CH^2**), 7.59 (1H, s, **CH^6**), 6.44 (1H, dd, *J* = 6.4, 7.4 Hz, **CH$^{1'}$**), 5.29 (1H, br.s, **OH$^{3'}$**), 4.98 (1H, br.s, **OH$^{5'}$**), 4.32 (1H, m, **CH$^{3'}$**), 3.81 (1H, m, **CH$^{4'}$**), 3.55 (1H, dd, *J* = 4.2, 11.7 Hz, **CH$^{5'}$**), 3.50 (1H, dd, *J* = 4.0, 11.7 Hz, **CH$^{5'}$**), 3.48 (2H, s, **CH^{10}**), 2.40 (1H, ddd, *J* = 5.8, 6.7, 13.2 Hz, **CH$^{2'}$**), 2.18 (1H, ddd, *J* = 2.7, 5.9, 13.2 Hz, **CH$^{2'}$**)

δ$_C$ (75.5 MHz; DMSO-d6) 157.4 (**C^4**), 146.7 (**C^{7a}**), 144.9 (**C^2**), 124.6 (**C^6**), 107.4 (**C^{4a}**), 99.3 (**C^5**), 92.1 (**C^9**), 87.5 (**C$^{4'}$**), 83.0 (**C$^{1'}$**), 74.7 (**C^8**), 70.8 (**C$^{3'}$**), 61.7 (**C$^{5'}$**), 40.1 (**C$^{2'}$**), 31.5 (**C^{10}**)

UV/vis (H$_2$O): 260 nm (6.9 x 10^3 cm^{-1}mol^{-1})

ES$^+$/MS: 327.1 (M+Na)$^+$

HRMS (ES$^+$) for C$_{14}$H$_{16}$N$_4$O$_4$ (M+H)$^+$:calculated 305.1244, found 305.1246

4.5 SYNTHETIC PEPTIDES

4.5.1 *N*-Terminal Fmoc Removal

The resin was swollen in a minimum volume of DCM (30 min). The resin was then swollen in a minimum volume of piperidine:DMF (1:4 v:v) in sequential treatments (5 then 15 min). The resin was then filtered and washed DMF (2 ml, x 3), DCM (2 ml, x 3), MeOH (2 ml, x 3), Et_2O (2 ml, x 3) and dried *in vacuo*.

4.5.2 Solid Phase Peptide Coupling

The resin was swollen in a minimum of DCM (30 min). *N*-fmoc amino acid (2 eq) and HOBt (2 eq) were dissolved in DCM with a minimum number of drops of DMF and stirred (RT, 10 min).[*] DIC (2.2 eq) was added and the mixture was stirred (10 min) before the addition of the resin. The resin was agitated (RT, 2 h) then washed with DMF (2 ml, x 3), DCM (2 ml, x 3), MeOH (2 ml, x 3), Et_2O (2 ml, x 3) and dried *in vacuo*.

4.5.3 Qualitative Ninhydrin Test

A small amount of the resin was treated with three drops of reagent A, 2 drops of reagent B and heated (100°C, 10 min). If the resin turned blue/black this indicated the presence of primary amines. In such cases the coupling of the amino acid was repeated or if free amines were still present after a second attempt the resin was capped.

4.5.4 Peptide Cleavage

The resin was swollen in a minimum of DCM (30 min). A TFA solution (TFA 5.4 ml, DCM 0.6 ml and a drop of water) was then added to the swollen resin and agitated (1 h). The resin was then removed by filtration and the filtrate was concentrated to ~1 ml at 80°C under a stream of nitrogen gas. The concentrate was then added to Et_2O (25 ml) at 0°C. The resulting precipitate was collected by centriguation and the Et_2O was decanted off. The precipitate was washed with Et_2O (25 ml, 0°C) and collected by centrifugation.

---

[*] To dissolve *N*-fmoc glycine use DMF only and heat gently.

The bulk of the solvent was removed by decanting and the remainder was removed *in vacuo*.

4.5.5   Quantitative Trityl Analysis to Calculate Resin Loading

A sample of resin (5~mg) was accurately weighed, suspended in conc. HCl:EtOH (3:2 v:v, 25 ml) and shaken (5 min). Sonication (15 min) was required in certain cases to remove the trityl cation from the resin and into solution. The absorbance of the orange solution (1 ml) was measured on a Perkin Elmer UV/vis spectrometer (495 nm). The mass of the resin (M g) and the absorbance (A_{495}) was used to calculate the loading on the resin (L $\mu molg^{-1}$).

$$0.35A_{495} / M = L$$

A_{495} = absorbance for 1 ml (495 nm), M = mass of resin (g), 0.35 = 25 / 71.7 = volume of conc.HCl:EtOH (3:2) / the absorbance of 1 μmol of DMT^+ (495 nm), L = resin loading ($\mu molg^{-1}$).

4.5.6   Quantitative Fmoc Analysis to Calculate Resin Loading

Piperidine:DMF (1:4 v:v, 1 ml) was added to a known mass of resin and shaken (15 min). The filtered solution was made up to 25 ml with piperidine:DMF (1:4 v:v). The absorbance of the solution (1 ml) was measured on a UV spectrometer (302 nm). The mass of the resin (M g) and the absorbance (A_{302}) was used to calculate the loading on the resin (L $mmolg^{-1}$).

$$3.21A_{302} / M = L$$

A_{302} = absorbance for 1 ml (302 nm), M = mass of resin (mg), 3.21 = 25 x 1000 / 7800 = volume of piperidine:DMF (1:4) x 1000 / the extinction coefficient of the adduct at 302 nm ($M^{-1}cm^{-1}$), L = resin loading ($mmolg^{-1}$).

4.5.7   Quantitative Ninhydrin Analysis to Calculate Resin Loading[129]

Reagent A: Potassium cyanide (65 mg) was dissolved in water (100 ml). An aliquot (2 ml) of this solution was diluted with pyridine (freshly distilled from ninhydrin) and stirred over Amberlite mixed-bed resin MB-3 (4 g) before filtering. This was mixed with the

following solution: Phenol (40 g) was dissolved in absolute ethanol (10 ml) with warming which was then stirred over Amberlite mixed-bed resin MB-3 (4 g) for 45 min before filtering and adding to the previous solution.

Reagent B: Ninhydrin (2.5 g) was dissolved in absolute ethanol (50 ml).

Three drops of reagent A and 2 drops of reagent B were added to a known mass of resin (~mg) and then heated (100°C, 10 min). Once cool 60% aqueous ethanol (2 ml) was added, the mixture was then filtered and placed in a volumetric flask. Et_4NCl (0.5 M in DCM, 2 x 0.5 ml) was then added to the resin. This was then filtered and the solution placed in a volumetric flask (25 ml) and made up with 60% aqueous ethanol. The absorbance (570 nm) of this solution was then measured. The mass of the resin (M g) and the absorbance (A_{570}) were used to calculate the loading on the resin (L mmolg^{-1}).

$$1667A_{570} / M = L$$

A_{570} = absorbance for 1 ml (570 nm), M = mass of resin (mg), $1667 = 25$ x $(1x10^6)$ / 15000 = volume of piperidine:DMF (1:4) x $(1x10^6)$ / the extinction coefficient of the adduct at 570 nm ($M^{-1}cm^{-1}$), L = resin loading (μmolg^{-1}).

4.6   RESIN EXPERIMENTS

4.6.1   Fluorescence and UV Melting of Oligonucleotides on Resins

Resin (3 mg) was weighted into an Eppendorf tube (1.5 ml) and water (1 ml) was added and the suspension agitated (10 min). The Eppendorf tube was then spun to 10,000 rpm and the water was removed. This was repeated x 3 before the resin was dried *in vacuo*. A fluorescent oligonucleotide (10 μl, 0.0328 nmolμl^{-1}) was added to the resin together with a buffer solution (1 μl, (0.1 M NaCl, 1mM EDTA, 10 mM Na-phosphate (pH 7.0))) and the water was then removed *in vacuo*. Water (1 μl) was then added and the Eppendorf tube was spun to 10,000 rpm. The resin was then left overnight then washed with water (1 ml, x 3).

4.6.1.1 UV-Melting

The resin sample was heated in the above buffer (1 ml) at $1^\circ Cmin^{-1}$ from $25^\circ C$ to $75^\circ C$ measuring the absorbance of the sample every 10 s at 260 nm. The data was recorded using the PECSS 2, UV melting curve controller software.

4.6.1.2 Fluorescence Melting

The resin sample was heated in the above buffer (1 ml) at $1^\circ Cmin^{-1}$ from $25^\circ C$ to $85^\circ C$ measuring the emission of the sample at 520 nm with excitation at 495 nm. The experiments were performed in a Perkin Elmer Luminescence Spectrometer LS50B using a Jencons (Scientific) Ltd Julabo F25 heating bath and associated software, Autopole or FL Winlab.

4.6.2   MALDI-TOF MS Analysis

The experiments were carried out on a ThermoBioAnalysis Dynamo, which is a linear MALDI-TOF MS.

4.6.2.1 Sample preparation

The matrices α-cyano-4-hydroxycinnamic acid (alpha) and 2,5-dihydroxybenzoic acid (DHB) were prepared as saturated solutions of acetone and acetonitrile:water 1:1 respectively. The samples were prepared by placing the matrix (0.5 μl) on the metal plate and then mixing with the anaylate (0.5 μl, 10pmoleμl^{-1}). This order is reversed if using alpha. Initial studies used a serial dilution of anaylates starting at 1 mgml^{-1}. The plate was then allowed to evaporate to dryness.

4.6.2.2 MALDI-TOF calibration

Initial calibration was carried out using α-cyclodextrin. Peptide markers hexaglycine and Ala-Ala-Phe-7-amido-4-methyl-coumarin were dissolved in 1 M HCl and ethanol respectively. These markers were chosen as their masses fell above and below that expected for the anaylate.

4.6.2.3 MS of the resin bound peptide

The resin beads (many to five) were exposed to a TFA solution (1 µl, of TFA 5.4 ml, DCM 0.6 ml and a drop of water) for 10 min. Once the solution has evaporated the matrix (alpha, 1 µl) was added.

4.7   OLIGONUCLEOTIDES

4.7.1   Synthesis

Oligonucleotides were synthesised on either an Applied Biosystems 394 automated DNA synthesiser or an Applied Biosystems Expedite™ Nucleic Acid Synthesis System using the standard 0.2 µmole or the 1 µmole phosphoramidite cycle. Stepwise coupling yields and overall yields were calculated by the automated trityl conductivity monitoring facility. Standard DNA phosphoramidities, solid supports and additional reagents were purchased from Applied Biosystems, Glen Research Inc and Link Technologies Ltd.

Standard β-Cyanoethyl phosphoramidite monomers were used throughout. The phosphoramidites were dissolved in MeCN, passed through a non-sterile 25 mm Millex® syringe filter unit, and the MeCN was then removed *in vacuo*. The resultant oil was dissolved in DCM (10 mlg^{-1} + 0.2 ml) and 1 ml aliquots were syringed into individual monomer bottles containing rubber septa under argon. The solvent was then removed *in vacuo* and the bottles stored at -20°C. Immediately prior to use, the β-cyanoethyl phosphoramidites were dissolved in anhydrous acetonitrile to a concentration of 0.1 M.

4.7.2   Deprotection

Standard cleavage and deprotection of oligonucleotides was carried out by exposing the CPG to concentrated ammonia solution at 55°C for 5 h. Cleavage and deprotection of the modified oligonucleotides was achieved by exposure to 10 % methylamine in water (2 ml) containing phenol (5.0 mg) for 36 h at RT. Oligonucleotides that were permanently

attached to the resin were deprotected using ethanol:concentrated ammonia solution (1:4, v:v) at 55°C for 5 h, then washed MeOH(x5), DCM(x5) and Et$_2$O(x5) and dried *in vacuo*.

4.7.3 Purification

Purification was carried out by reversed phase HPLC on a Gilson system (column: ABI C8 (octyl) 8 mm x 250 mm, pore size = 300 Å), controlled by Gilson 7.12 software. (Buffer A: 0.1 M ammonium acetate, pH 7.0. Buffer B: 0.1 M ammonium acetate with 20% acetonitrile, pH 7.0). A 30 minute program with a flow rate of 3 ml/min was used with the following gradient: Time in mins (%Buffer B); 0 (0); 3 (0); 5 (20); 21 (100); 25 (100); 27 (0); 30 (0). Elution of oligonucleotides was monitored by UV detection at 290 nm and peaks collected manually.

4.7.4 Oligonucleotide sequences

Reference Code	Oligonucleotide Number	Sequence	Molecular Ion		Concentration / nmolesμl⁻¹
			Expect	Found	
-	O1	5'd-FAMGGGCTGAAAAGCTCCCGAT-3'	-	-	-
-	O2	5'd-CTACGTCTGC<u>CGGGAGC</u>TTTTCAGCTTGAGCT**R2**	-	-	-
00230R	O3	5'-dCAGATAGATAGC-3'	-	-	0.0316
00233R	O4	5'-dGCTATCTATCTG-3'	-	-	0.0463
00960R	O5	5'-dGCTATC<u>U</u>ATCTG-3'	3650	3645	0.0665
00961R	O6	5'-dGCTATC<u>U</u>A<u>U</u>CTG-3'	3689	3681	0.1056
00959R	O7	5'-dGCTA<u>U</u>C<u>U</u>A<u>U</u>CTG-3'	3728	3727	0.0875
00271R	O8	5'-dGCTATC**<u>dmU</u>**ATCTG-3'	3678	3677	0.0318
00272R	O9	5'-dGCTATC**<u>dmU</u>**A**<u>dmU</u>**CTG-3'	3745	3732	0.0603
00270R	O10	5'-dGCTA**<u>dmU</u>**C**<u>dmU</u>**A**<u>dmU</u>**CTG-3'	3812	3812	0.0515
00326R	O11	5'-dCAGATTGATAGC-3'	-	-	0.0976
00325R	O12	5'-dCAGATGGATAGC-3'	-	-	0.0885
00324R	O13	5'-dCAGATCGATAGC-3'	-	-	0.0954
00885R	O14	5'-dGCTATC**<u>C</u>**TATCTG-3'	3665	3662	0.1122
00887R	O15	5'-dGCTATC**<u>C</u>**TATC**<u>C</u>**TG-3'	3719	3715	0.0798
00886R	O16	5'-dG**<u>C</u>**TATC**<u>C</u>**TATC**<u>C</u>**TG-3'	3773	3769	0.0304
00271R	O17	5'-dGCTAT**<u>dmC</u>**TATCTG-3'	-	-	0.0440
00272R	O18	5'-dGCTAT**<u>dmC</u>**TAT**<u>dmC</u>**TG-3'	-	-	0.0502
00270R	O19	5'-dG**<u>dmC</u>**TAT**<u>dmC</u>**TAT**<u>dmC</u>**TG-3'	-	-	0.0336
00410R	O20	5'-dCAGATAGATAGC-3'	-	-	0.1868
00409R	O21	5'-dGCTATCTATCTG-3'	-	-	0.2766
00894R	O22	5'-dGCTATCT**A**TCTG-3'	3663	3662	0.0533
00895R	O23	5'-dGCT**A**TCT**A**TCTG-3'	3715	3715	0.0647
00926R	O24	5'-dGCTATCT**D**TCTG-3'	3678	3678	0.0813
00925R	O25	5'-dGCT**D**TCT**D**TCTG-3'	3745	3744	0.1015
01038R	O26	5'-dGCTATCTGTCTG-3'	-	-	0.1556
01039R	O27	5'-dGCTGTCTGTCTG-3'	-	-	0.1548
01043R	O28	5'-dCAGACAGATAGC-3'	-	-	0.1393
01044R	O29	5'-dCAGACAGACAGC-3'	-	-	0.1172
00997R	O30	5'-dGTCTATCTATCG-3'	-	-	0.2764
00970R	O31	5'-dCGATA**<u>G</u>**ATAGAC-3'	3730	3730	0.0672
00973R	O32	5'-dCGATA**<u>G</u>**ATA**<u>G</u>**AC-3'	3782	3781	0.0905
00971R	O33	5'-dC**<u>G</u>**ATA**<u>G</u>**ATA**<u>G</u>**AC-3'	3834	3835	0.0768
00968R	O34	5'-dCGATA**<u>I</u>**ATAGAC-3'	3715	3715	0.0820
00972R	O35	5'-dCGATA**<u>I</u>**ATA**<u>I</u>**AC-3'	3752	3751	0.0800
00969R	O36	5'-dC**<u>I</u>**ATA**<u>I</u>**ATA**<u>I</u>**AC-3'	3789	3789	0.0590
01040R	O37	5'-dCGATAAATAGAC-3'	-	-	0.1186
01041R	O38	5'-dCGATAAATAAAC-3'	-	-	0.1238
01042R	O39	5'-dCAATAAATAAAC-3'	-	-	0.1301
01045R	O40	5'-dGTCTATTTATCG-3'	-	-	0.2175
01046R	O41	5'-dGTTTATTTATCG-3'	-	-	0.1814
01047R	O42	5'-dGTTTATTTATTG-3'	-	-	0.2327

4.7.5   Job Plot

The oligonucleotide(s) were added to a buffer (100 mM NaCl, 1 mM EDTA, 10 mM Na-phosphate (pH 7.0) to make a final volume of 1.5 ml. The total single stranded oligonucleotide concentration was 2 μM. The buffer was prepared as a 10 times solution.

Ratio			Water	10 x Buffer
0:100	0	73.6	1276.4	0.15
20:80	3.6	58.9	1287.5	0.15
40:60	7.1	44.2	1298.7	0.15
60:40	10.7	29.4	1309.9	0.15
80:20	14.2	14.7	1321.1	0.15
100:0	17.8	0	1332.2	0.15

The solutions were mixed and then heated to 95°C and allowed to cool before the readings were taken. The values were then corrected for the different absorption coefficients of the oligonucleotides using the following formula,

$$A \,/\, E_c^{O2}\,(P_{O2}) + E_c^{O3}\,(P_{O3})$$

A = Absorbance, E_c = Extinction coefficient of **O2**, P = Proportion of **O2**

4.8   UV-MELTING STUDIES

Table 14. In order to make a buffer two solutions have to be prepared, one containing the acidic component ($NaH_2PO_4.2H_2O$ or NaH_2PO_4) and one the basic component (Na_2HPO_4). These solutions both require EDTA, and NaCl if the $[Na^+]$ is constant. The buffer is made up at 10x the required concentration as other components need to be added to the samples and in this case 100 ml is to be made. The table details the masses of each compound that are required to make up these solutions.

Volume of buffer required		100 mL	
Concentration of buffer		10 Times	
Compound	FW gmol^{-1}	Final Concentration Required mM (mmoleL^{-1})	Mass Required g
EDTA	372.24	1	0.37224
NaCl	58.44	100	5.844
$NaH_2PO_4.2H_2O$	156.01	10	1.5601
Na_2HPO_4	141.96	10	1.4196
NaH_2PO_4	120	10	1.2

Thermodynamic parameters for duplex formation were determined by UV melting in a
Cary 400 UV/visible spectrometer with a Cary Temperature Controller in 100 mM NaCl,
1 mM EDTA and 10 mM Na-phosphate buffer at pH 7.0. Thermodynamic parameters for
oligonucleotides containing the modified pyrimidines were also determined in 10 mM
Na-phosphate buffers at pH 7.0 with 1 mM EDTA and also with 0, 0.025, 0.05, 0.1, 0.2, 1
M NaCl. The proportions of buffer components were determined and are shown below in
the Table 14.

In order to reach the correct pH the two solutions must be mixed in the correct
proportions as shown in Table 15. The salt dependent studies required a constant pH of
7.0 so the proportions of the solutions making up the buffers had to be altered according
to the values shown in Table 16.

Table 15. The pH range available with a 10 mM Na-phosphate buffer containing 0.1 M NaCl and 1 mM
EDTA at 293 K.

For the preparation of buffers with 0.1M NaCl		
Proportion of solution containing 0.1 M NaH$_2$PO$_4$	Proportion of solution containing 0.1 M Na$_2$HPO$_4$	Resultant pH
0.1	0.9	7.26
0.2	0.8	7.01
0.3	0.7	6.88
0.4	0.6	6.64
0.5	0.5	6.51
0.6	0.4	6.31
0.7	0.3	6.12
0.8	0.2	5.93
0.9	0.1	5.46

Table 16. The proportions of each solution (that also contain EDTA 10 mM) required to produce a buffer
at pH 7.0 at different NaCl concentrations.

For the preparation of buffers at pH 7.0		
Proportion of solution containing 0.1 M NaH$_2$PO$_4$	Proportion of solution containing 0.1 M Na$_2$HPO$_4$	Final [NaCl] (M)
0.4	0.6	0
0.3	0.7	0.025
0.25	0.75	0.05
0.2	0.8	0.1
0.15	0.85	0.2
0.04	0.96	1

Single-strand DNA concentrations were determined from 90% of the absorbance at 260 nm at RT on the basis of monomer extinction coefficients. In order to estimate the concentration of modified and unmodified chimeric oligomers the monomer extinction coefficients of the modified bases were determined and are given in the compound data.

Complementary single strands were mixed at 1:1 concentration ratios to give single strand concentrations of 0.05, 0.1, 0.2, 0.4, 0.8, 1.2, 1.6, 2.0, 3.0, 4.0, 8.0, 16.0 μM. The samples were filtered, Kinesis syringe filters regenerated cellulose 13 mm, 0.45 μM, into Hellma® synthetic quartz SUPRASIL cuvettes (pathlength 10 mm). The samples were then denatured for 4 min at 80°C (the temperature was monitored using temperature probe 1 in a cuvette of buffer (1 ml) in position 3 in the multiple cell holder), and re-annealed at 1°Cmin^{-1} to 20°C and held at 20°C for 2 min. The samples were then heated to 80°C at 1°Cmin^{-1} collecting an absorbance reading every 0.2°C. The sample was then held at 80°C 2 min before cooling at 1°Cmin^{-1} again collecting a datum point every 0.2°C. The melt curves were carried out in triplicate.

The thermodynamics were calculated using the van't Hoff plot and by averaging the results from single melt analysis as determined by Meltwin3.5.

5 APPENDICES

5.1 APPENDIX 1 - A DEFINITION OF ABSORBANCE

Absorbance is the dimensionless product of the molar absorption coefficient ε, concentration of the absorbing species [J] and the sample thickness l. This is referred to as the Beer-Lambert Law and is derived as follows.

The reduction of intensity, I for the intensity and dI for the change, that occurs when light passes through a sample of thickness dx containing absorbing species J at a concentration [J] is proportional to the thickness, the concentration and the incident intensity,

$$dI = -\alpha[J]Idx, \text{ or } d\ln I = -\alpha[J]dx$$

where α is the proportionality coefficient. This equation applies to each successive layer into which the sample can be regarded as being divided to obtain the final intensity I_f that emerges from a sample of thickness l when the incident intensity is I_i we sum all the successive changes;

$$\int d\ln I = -\int \alpha[J]\,dx$$

If the concentration is uniform, [J] is independent of x and the expression integrates to;

$$\text{Beer-Lambert Law: } I_f = I_i e^{-\alpha[J]l}$$

If $\alpha = \varepsilon \ln 10$ then,

$$I_f = I_i 10^{-\varepsilon[J]l}, \text{ or } \log(I_f/I_i) = -\varepsilon[J]l$$

And if $\log(I_f/I_i) = A$ (absorbance) then,

$$A = -\varepsilon[J]l$$

The molar absorption coefficient of the absorbing species is dependent on the wavelength of the incoming light and the molecule.

5.2   APPENDIX 2 - THERMODYNAMIC DEFINITIONS

5.2.1   van't Hoff plot

Thermodynamic parameters for duplex formation can be derived by different van't Hoff analysis methods. The first analysis method requires at least six melting temperatures over a ten fold range of concentration for the single stranded oligonucleotides[88]. The equation that is used is shown below,

$$T_m^{-1} = R / \Delta H° \ln(C_T/a) + \Delta S°/ \Delta H°$$

where R is the gas constant, 1.987 cal K^{-1} mol^{-1} and a is 1 for self-complementary duplexes and 4 for non-self-complementary duplexes, C_T is the total concentration of single strands, T_m is the melting temperature of the duplex.

In order to determine $\Delta H°$ and $\Delta S°$, T_m^{-1} is plotted on the y-axis and $\ln(C_T/a)$ on the x-axis. A least squares linear regression line analysis was then carried out on the data, this gave the gradient $R / \Delta H°$ so giving you a value for $\Delta H°$ and the intercept is $\Delta S°/ \Delta H°$ so giving you $\Delta S°$. Then since you now know both $\Delta S°$ and $\Delta H°$ you can calculate $\Delta G°$ as;

$$\Delta G° = \Delta H° - T \, \Delta S°$$

T is temperature measured in Kelvin, where 0°C = 273.15 K, $\Delta G°$ is usually quoted at either 298.15 (standard room temperature) or 310.15 (body temperature) K.

5.2.2   The two-state model – Meltwin 3.5

The second analysis method fits the shape of a melting curve to a two-state model with linear sloping baselines using a non-linear least-squares program[86,108,130]. The calculation is carried out using a program that is available on the internet called Meltwin 3.5[131]. The equation that is used is shown below,

$$\alpha = \in_{ss}(T) - \in(T) / \in_{ss}(T) - \in_{ds}(T) =$$

$$(m_{ss}T + b_{ss}) - \in(T) \, / \, (m_{ss}T + b_{ss}) - (m_{ds}T + b_{ds})$$

where α is the fraction of total single strand in duplex form, $\in(T)$ is the extincition coefficient at any temperature, $\in_{ss}$ is the average extinction coefficient for the single-stranded state, $\in_{ds}$ is the extinction coefficient per strand for the double-stranded state, m_{ss}, m_{ds}, b_{ss}, b_{ds}, T are the slope of the single stranded state, double stranded state, intercept of the single stranded state, double stranded state and the temperature respectively.

α is related to the changes in enthalpy and entropy by the following equation,

$$K = \alpha \, / \, 2(1-\alpha)^2 C_T = \exp(\Delta S°/R - \Delta H°/RT)$$

The iterative nature of the algorithm requires an initial estimate of the entropy and of the enthalpy are and these are estimated using the following equations.

$$\Delta H° = C/(1/T_{\frac{1}{2}} - 1/T_{\frac{3}{4}})$$

$$\Delta S° = \Delta H°/ T_{\frac{1}{2}} + X$$

C is 4.4, 7.0 or 3.2 cal/(molK) and X is $\ln(C_t)$, $\ln(C_t/4)$, or 0 for self-complementary, non-complementary or unimolecular equilibria respectively. $T_{\frac{1}{2}}$ and $T_{\frac{3}{4}}$ are the absolute temperatures at the maximum and upper half maximum of the melt curve derivative respectively.

The program fits the experimental data treating $\Delta H°$, $\Delta S°$, m_{ds}, b_{ds}, m_{ss} and b_{ss} as variable parameters. In general, the root-mean-square difference between the data and calculated curve is less than 0.5%.[130]

5.3 APPENDIX 3 - DEFINITIONS FOR ERROR ANALYSIS

In order that the values for $\Delta H°$, $\Delta S°$, $\Delta G°$ and T_m are meaningful they need to be shown with errors. Both van't Hoff analysis methods provide ways of calculating the sample variances, σ^2. The equations for the $1/T_m$ vs $\ln(C_T/a)$ plots are shown below,

$$\sigma_{\Delta H°}{}^2 = (\Delta H°)^2(\sigma_m{}^2/m^2)$$

$$\sigma_{\Delta S°}{}^2 = (\Delta S°)^2(\sigma_b{}^2/b^2 + \sigma_m{}^2/m^2 - 2\mathrm{Cov}(m,b)/mb)$$

$$\sigma_{\Delta G°(T)}{}^2 = \sigma_m{}^2 \,(Tb - 1/m^2)^2 + \sigma_b{}^2\,(T^2/m^2) - 2T\mathrm{Cov}(m,b)\,((Tb - 1)/m^3)$$

$$\sigma_{Tm}{}^2 = (\sigma_{\Delta G°(Tm)}{}^2 \,/\, (\Delta H°)^2)\, T_m{}^2$$

where T, m, b, $\sigma_m{}^2$, $\sigma_b{}^2$ and $\mathrm{Cov}(m,b)$ are the temperature in Kelvin, slope, intercept, variance of slope, variance of intercept and covariance of slope and intercept respectively.[132]

Given the sample variances the standard deviation can be calculated and used to produce the confidence limits shown below,

$$\mu = \chi \pm t(s/\sqrt{n})$$

where μ is the true mean of a population, χ is the mean of the sample which is an estimate of the true mean, t is a value that depends on the number of degrees of freedom (n-1) where n is the number of measurements and the degree of confidence required (knowing both of these values allows you to look up the relevant value of t from a table), s is the standard deviation and n is the number of measurements[133].

As many of the values calculated are a mean of many values and each value has an error associated with it then all the individual errors need to be taken account of. This is done through the propagation of random errors and calculated as shown below,

$$\sigma_x = \sqrt{\sum(k_{(1-n)}\sigma_{(1-n)})^2}$$

where σ_x is the error in the final value x, $k_{(1-n)}$ are the constants associated with the values 1-n, $\sigma_{(1-n)}$ are the random errors associated with values 1-n.

6 REFERENCES

1. Chargaff, E., Vischer, E., Doniger, R., Green, C. & Misani, F. *J. Biol. Chem.* **177**, 405-416 (1951).
2. Chargaff, E. *Fed. Proc.* **10**, 654 (1951).
3. Goodsell, D. S., Grzeskowiak, K. & Dickerson, R. E. Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the Holliday junction. *Biochemistry* **34**, 1022-10029 (1995).
4. Franklin, R. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740-741 (1953).
5. Franklin, R. & Gosling, R. G. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* **172**, 156-157 (1953).
6. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature (London)* **171**, 737 (1953).
7. Watson, J. D. & Crick, F. H. C. Genetic implications of the structure of deoxyribonucleic acid. *Nature (London)* **172**, 156 (1953).
8. Beaucage, S. L. & Iyer, R. P. Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach. *Tetrahedron* **48**, 2223-2311 (1992).
9. Smith, M., Rammler, D. H., Goldberg, I. H. & Khorana, H. G. Studies on Polynucleotides. XIV. Specific Synthesis of the C3'-C5' Inter-ribonucleotide Linkage. Synthese of Uridylyl-(3'-5')-Uridine and Uridylyl-(3'-5')-Adenosine. *J. Am. Chem. Soc.* **84**, 430-440 (1962).
10. Sinha, N. D., Biernat, J. & Koster, H. Beta-Cyanoethyl N,N-Dialkylamino/N-Morpholinomonochloro Phosphoamidites, New Phosphitylating Agents Facilitating Ease of Deprotection and Work-up of Synthesized Oligonucleotides. *Tetrahedron Lett.* **24**, 5843-5846 (1983).
11. Adams, S. P., Kavka, K. S., Wykes, E. J., Holder, S. B. & Galluppi, G. R. Hindered Dialkylamino Nucleoside Phosphite Reagents in the Synthesis of 2 DNA 51-Mers. *J. Am. Chem. Soc.* **105**, 661-663 (1983).
12. McBride, L. J. & Caruthers, M. H. Nucleotide Chemistry .10. An Investigation of Several Deoxynucleoside Phosphoramidites Useful for Synthesizing Deoxyoligonucleotides. *Tetrahedron Lett.* **24**, 245-248 (1983).
13. Topal, M. D. & Fresco, J. R. Complementary base pairing and the origin of substitution mutations. *Nature* **263**, 285-289 (1976).
14. Sinha, N. K. & Haimes, M. D. Molecular Mechanisms of Substitution Mutagenesis - an Experimental Test of the Watson-Crick and Topal-Fresco Models of Base Mispairings. *J. Biol. Chem.* **256**, 671-683 (1981).
15. Blackburn, G. M. & Gait, M. J. *Nucleic Acids in Chemistry and Biology* (Oxford University Press, 1996).
16. Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z. M., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N. S., Zody, M. C., Linton, L., Lander, E. S. & Altshuler, D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-933 (2001).

17.   Tishkoff, S. A., Dietzsch, E., Speed, W., Pakstis, A. J., Kidd, J. R., Cheung, K.,
       Bonne-Tamir, B., Santachiara-Benerecetti, A. S., Moral, P., Krings, M., Paabo, S.,
       Watson, E., Risch, N., Jenkins, T. & Kidd, K. K. Global Patterns of Linkage
       Disequilibrium at the CD4 Locus and Modern Human Origins. *Science* **271**, 1380-
       1387 (1996).
18.   Lovmar, L., Fredriksson, M., Liljedahl, U., Sigurdsson, S. & Syvanen, A. C.
       Quantitative evaluation by minisequencing and microarrays reveals accurate
       multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids
       Res.* **31**, art. no.-e129 (2003).
19.   Kwok, P. Y. Methods for genotyping single nucleotide polymorphisms. *Annu.
       Rev. Genomics Hum. Genet.* **2**, 235-258 (2001).
20.   Mir, K. U. & Southern, E. M. SEQUENCE VARIATION IN GENES AND
       GENOMIC DNA: Methods for Large-Scale Analysis. *Annu. Rev. Genomics Hum.
       Genet.* **1**, 329-360 (2000).
21.   Lee, L. G., Connell, C. R. & Bloch, W. Allelic discrimination by nick-translation
       PCR with fluorogenic probes. *Nucleic Acids Res.* **21**, 3761-3766 (1993).
22.   Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of
       Specific Polymerase Chain-Reaction Product by Utilizing the 5'- 3' Exonuclease
       Activity of Thermus-Aquaticus DNA-Polymerase. *Proc. Natl. Acad. Sci. U. S. A.*
       88, 7276-7280 (1991).
23.   Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes that fluoresce upon
       hybridization. *Nature Biotech.* **14**, 303-308 (1996).
24.   Nazarenko, I. A., Bhatnager, S. K. & Hohman, R. J. A closed tube format for
       amplification and detection of DNA based on energy transfer (SUNRISE primers).
       Nucleic Acids Res. **25**, 2516 (1997).
25.   Wittwer, C. T., Herrmann, M. G., Moss, A. A. & Rasmussen, R. P. Continuous
       fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22**,
       130-& (1997).
26.   Whitcombe, D., Theaker, J., Guy, S. P., Brown, T. & Little, S. Scorpion primers:
       efficient and homogeneous detection of PCR amplification products using self-
       probing amplicons and fluorescence. *Nature Biotech.* **17**, 804-807 (1999).
27.   Thelwell, N., Millington, S., Solinas, A., Booth, J. A. & Brown, T. Mode of action
       and application of Scorpion primers in mutation detection. *Nucleic Acids Res.* **28**,
       3752-3761 (2000).
28.   Solinas, A., Brown, L. J., McKeen, C., Mellor, J. M., Nicol, J. T. G., Thelwell, N.
       & Brown, T. Duplex Scorpion primers in SNP analysis and FRET applications.
       Nucleic Acids Res. **29**, art. no.-e96 (2001).
29.   Todd, A. V., Fuery, C. J., Impey, H. L., Applegate, T. L. & Haughton, M. A.
       DzyNA-PCR: Use of DNAzymes to detect and quantify nucleic acid sequences in
       a real-time fluorescent format. *Clin. Chem.* **46**, 625-630 (2000).
30.   Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. & Syvanen, A. C.
       A system for specific, high-throughput genotyping by allele- specific primer
       extension on microarrays. *Genome Res.* **10**, 1031-1042 (2000).
31.   Dubiley, S., Kirillov, E. & Mirzabekov, A. Polymorphism analysis and gene
       detection by minisequencing on an array of gel-immobilized primers. *Nucleic
       Acids Res.* **27**, e19 (1999).
32.   Chen, X. N., Zehnbauer, B., Gnirke, A. & Kwok, P. Y. Fluorescence energy
       transfer detection as a homogeneous DNA diagnostic method. *Proc. Natl. Acad.
       Sci. U. S. A.* **94**, 10756-10761 (1997).
33.   Tost, J. & Gut, V. G. Genotyping single nucleotide polymorphisms by mass
       spectrometry. *Mass Spectrom. Rev.* **21**, 388-418 (2002).

34. Tang, K., Opalsky, D., Abel, K., van den Boom, D., Yip, P., Del Mistro, G., Braun, A. & Cantor, C. R. Single nucleotide polymorphism analyses by MALDI-TOF MS. *Int. J. Mass Spectrom.* **226**, 37-54 (2003).

35. Walters, J. J., Fox, K. F. & Fox, A. Mass spectrometry and tandem mass spectrometry, alone or after liquid chromatography, for analysis of polymerase chain reaction products in the detection of genomic variation. *J. Chromatogr. B* **782**, 57-66 (2002).

36. Sauer, S. & Gut, I. G. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *J. Chromatogr. B* **782**, 73-87 (2002).

37. Jackson, P. E., Scholl, P. F. & Groopman, J. D. Mass spectrometry for genotyping: an emerging tool for molecular medicine. *Mol. Med. Today* **6**, 271-276 (2000).

38. Ahmadian, A., Gharizadeh, B., Gustafsson, A. C., Sterky, F., Nyren, P., Uhlen, M. & Lundeberg, J. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**, 103-110 (2000).

39. Chen, J. W., Iannone, M. A., Li, M. S., Taylor, J. D., Rivers, P., Nelsen, A. J., Slentz-Kesler, K. A., Roses, A. & Weiner, M. P. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res.* **10**, 549-557 (2000).

40. Kellar, K. L. & Iannone, M. A. Multiplexed microsphere-based flow cytometric assays. *Exp. Hematol.* **30**, 1227-1237 (2002).

41. Armstrong, B., Stewart, M. & Mazumder, A. Suspension arrays for high throughput, multiplexed single nucleotide polymorphism genotyping. *Cytometry* **40**, 102-108 (2000).

42. Cai, H., White, P. S., Torney, D., Deshpande, A., Wang, Z. L., Marrone, B. & Nolan, J. P. Flow cytometry-based minisequencing: A new platform for high-throughput single-nucleotide polymorphism scoring. *Genomics* **66**, 135-143 (2000).

43. Lizardi, P. M., Huang, X. H., Zhu, Z. R., Bray-Ward, P., Thomas, D. C. & Ward, D. C. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nature Genet.* **19**, 225-232 (1998).

44. Chen, X. N., Livak, K. J. & Kwok, P. Y. A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res.* **8**, 549-556 (1998).

45. Iannone, M. A., Taylor, J. D., Chen, J. W., Li, M. S., Rivers, P., Slentz-Kesler, K. A. & Weiner, M. P. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* **39**, 131-140 (2000).

46. Kaiser, M. W., Lyamicheva, N., Ma, W. P., Miller, C., Neri, B., Fors, L. & Lyamichev, V. I. A comparison of eubacterial and archaeal structure-specific 5'-exonucleases. *J. Biol. Chem.* **274**, 21387-21394 (1999).

47. Lyamichev, V., Mast, A. L., Hall, J. G., Prudent, J. R., Kaiser, M. W., Takova, T., Kwiatkowski, R. W., Sander, T. J., de Arruda, M., Arco, D. A., Neri, B. P. & Brow, M. A. D. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* **17**, 292-296 (1999).

48. Olivier, M., Chuang, L. M., Chang, M. S., Chen, Y. T., Pei, D., Ranade, K., de Witte, A., Allen, J., Tran, N., Curb, D., Pratt, R., Neefs, H., Indig, M. D., Law, S., Neri, B., Wang, L. & Cox, D. R. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res.* **30**, art. no.-e53 (2002).

49. Rao, K. V. N., Stevens, P. W., Hall, J. G., Lyamichev, V., Neri, B. P. & Kelso, D. M. Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage reaction on microspheres. *Nucleic Acids Res.* **31**, art. no.-e66 (2003).

50. Singh-Gasson, S., Green, R. D., Yue, Y. J., Nelson, C., Blattner, F., Sussman, M. R. & Cerrina, F. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17**, 974-978 (1999).

51. Heller, M. J. DNA microarray technology: Devices, systems, and applications. *Annu. Rev. Biomed. Eng.* **4**, 129-153 (2002).

52. Freier, S. M. & Altmann, K. H. The ups and downs of nucleic acid duplex stability: structure- stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.* **25**, 4429-4443 (1997).

53. Sonogashira, K., Tohda, Y. & Hagihara, N. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Lett.* **16**, 4467-4470 (1975).

54. Hobbs, F. W. Palladium-Catalyzed Synthesis of Alkynylamino Nucleosides - a Universal Linker for Nucleic-Acids. *J. Org. Chem.* **54**, 3420-3422 (1989).

55. Froehler, B. C., Wadwani, S., Terhorst, T. J. & Gerrard, S. R. Oligodeoxynucleotides Containing C-5 Propyne Analogs of 2'- Deoxyuridine and 2'-Deoxycytidine. *Tetrahedron Lett.* **33**, 5307-5310 (1992).

56. Barnes, T. W. & Turner, D. L. H. Long-range cooperativity due to C5-propynylation of oligopyrimidines enhances specific recognition by uridine of ribo-adenosine over ribo-guanosine. *J. Am. Chem. Soc.* **123**, 9186-9187 (2001).

57. Barnes, T. W. & Turner, D. H. Long-range cooperativity in molecular recognition of RNA by oligodeoxynucleotides, with multiple C5-(1-propynyl) pyrimidines. *J. Am. Chem. Soc.* **123**, 4107-4118 (2001).

58. Barnes, T. W. & Turner, D. H. C5-(1-Propynyl)-2 '-deoxy-pyrimidines enhance mismatch penalties of DNA : RNA duplex formation. *Biochemistry* **40**, 12738-12745 (2001).

59. Wagner, R. W., Matteucci, M. D., Lewis, J. G., Gutierrez, A. J., Moulds, C. & Froehler, B. C. Antisense Gene Inhibition by Oligonucleotides Containing C-5 Propyne Pyrimidines. *Science* **260**, 1510-1513 (1993).

60. Gutierrez, A. J., Matteucci, M. D., Grant, D., Matsumura, S., Wagner, R. W. & Froehler, B. C. Antisense gene inhibition by C-5-substituted deoxyuridine-containing oligodeoxynucleotides. *Biochemistry* **36**, 743-748 (1997).

61. Cruickshank, K. A. & Stockwell, D. L. Oligonucleotide Labeling - a Concise Synthesis of a Modified Thymidine Phosphoramidite. *Tetrahedron Lett.* **29**, 5221-5224 (1988).

62. Heystek, L. E., Zhou, H. Q., Dande, P. & Gold, B. Control over the localization of positive charge in DNA: The effect on duplex DNA and RNA stability. *J. Am. Chem. Soc.* **120**, 12165-12166 (1998).

63. Gibson, K. J. & Benkovic, S. J. Synthesis and Application of Derivatizable Oligonucleotides. *Nucleic Acids Res.* **15**, 6455-6467 (1987).

64. Giller, G., Tasara, T., Angerer, B., Muhlegger, K., Amacker, M. & Winter, H. Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. I. Chemical synthesis of various reporter group- labeled 2 '-deoxyribonucleoside-5 '-triphosphates. *Nucleic Acids Res.* **31**, 2630-2635 (2003).

65. Bijapur, J., Keppler, M. D., Bergqvist, S., Brown, T. & Fox, K. R. 5-(1-propargylamino)-2 '-deoxyuridine (U-P): a novel thymidine analogue for

generating DNA triplexes with increased stability. *Nucleic Acids Res.* **27**, 1802-1809 (1999).

66.   Sollogoub, M., Dominguez, B., Fox, K. R. & Brown, T. Synthesis of a novel bis-amino-modified thymidine monomer for use in DNA triplex stabilisation. *Chem. Commun.*, 2315-2316 (2000).

67.   Ahmadian, M., Zhang, P. M. & Bergstrom, D. E. A comparative study of the thermal-stability of oligodeoxyribonucleotides containing 5-substituted 2 ' deoxyuridines. *Nucleic Acids Res.* **26**, 3127-3135 (1998).

68.   Seela, F., Ramzaeva, N., Leonard, P., Chen, Y., Debelak, H., Feiling, E., Kroschel, R., Zulauf, M., Wenzel, T., Frohlich, T. & Kostrzewa, M. Phosphoramidites and oligonucleotides containing 7-deazapurines and pyrimidines carrying aminopropargyl side chains. *Nucleosides Nucleotides Nucleic Acids* **20**, 1421-1424 (2001).

69.   Battersby, T. R., Ang, D. N., Burgstaller, P., Jurczyk, S. C., Bowser, M. T., Buchanan, D. D., Kennedy, R. T. & Benner, S. A. Quantitative analysis of receptors for adenosine nucleotides obtained via in vitro selection from a library incorporating a cationic nucleotide analog. *J. Am. Chem. Soc.* **121**, 9781-9789 (1999).

70.   Lee, S. E., Sidorov, A., Gourlain, T., Mignet, N., Thorpe, S. J., Brazier, J. A., Dickman, M. J., Hornby, D. P., Grasby, J. A. & Williams, D. M. Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Res.* **29**, 1565-1573 (2001).

71.   Hardwidge, P. R., Lee, D. K., Prakash, T. P., Iglesias, B., Den, R. B., Switzer, C. & Maher, L. J. DNA bending by asymmetrically tethered cations: influence of tether flexibility. *Chem. Biol.* **8**, 967-980 (2001).

72.   Graham, D., Mallinder, B. J. & Smith, W. E. Detection and identification of labeled DNA by surface enhanced resonance Raman scattering. *Biopolymers* **57**, 85-91 (2000).

73.   Sauer, S., Lechner, D., Berlin, K., Plancon, C., Heuermann, A., Lehrach, H. & Gut, I. G. Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res.* **28**, e100 (2000).

74.   Buhr, C. A., Wagner, R. W., Grant, D. & Froehler, B. C. Oligodeoxynucleotides containing C-7 propyne analogs of 7- deaza-2'-deoxyguanosine and 7-deaza-2'-deoxyadenosine. *Nucleic Acids Res.* **24**, 2974-2980 (1996).

75.   Gourlain, T., Sidorov, A., Mignet, N., Thorpe, S. J., Lee, S. E., Grasby, J. A. & Williams, D. M. Enhancing the catalytic repertoire of nucleic acids. II. Simultaneous incorporation of amino and imidazolyl functionalities by two modified triphosphates during PCR. *Nucleic Acids Res.* **29**, 1898-1905 (2001).

76.   Balow, G., Mohan, V., Lesnik, E. A., Johnston, J. F., Monia, B. P. & Acevedo, O. L. Biophysical and antisense properties of oligodeoxynucleotides containing 7-propynyl-, 7-iodo- and 7-cyano-7-deaza-2-amino-2 '-deoxyadenosines. *Nucleic Acids Res.* **26**, 3350-3357 (1998).

77.   Seela, F. & Zulauf, M. Oligonucleotides containing 7-deazaadenines: The influence of the 7-substituent chain length and charge on the duplex stability. *Helv. Chim. Acta* **82**, 1878-1898 (1999).

78.   Seela, F., Ramzaeva, N. & Zulauf, M. Duplex stability of oligonucleotides containing 7-substituted 7-deaza- and 8-aza-7-deazapurine nucleosides. *Nucleosides Nucleotides* **16**, 963-966 (1997).

79.   McKeen, C. M., Brown, L. J., Nicol, J. T. G., Mellor, J. M. & Brown, T. Synthesis of fluorophore and quencher monomers for use in Scorpion primers and nucleic acid structural probes. *Org. Biomol. Chem.* **1**, 2267-2275 (2003).

80. Rosemeyer, H., Ramzaeva, N., Becker, E. M., Feiling, E. & Seela, F. Oligonucleotides incorporating 7-(aminoalkynyl)-7-deaza-2 '- deoxyguanosines: Duplex stability and phosphodiester hydrolysis by exonucleases. *Bioconjugate Chem.* **13**, 1274-1285 (2002).

81. Ramzaeva, N., Mittelbach, C. & Seela, F. 7-deaza-2 '-deoxyguanosines functionalized with 7-(omega- aminoalk-1-ynyl) residues. *Nucleosides Nucleotides* **18**, 1439-1440 (1999).

82. He, J. L. & Seela, F. Propynyl groups in duplex DNA: stability of base pairs incorporating 7-substituted 8-aza-7-deazapurines or 5- substituted pyrimidines. *Nucleic Acids Res.* **30**, 5485-5496 (2002).

83. Becher, G., He, J. L. & Seela, F. Major-groove-halogenated DNA: The effects of bromo and iodo substituents replacing H-C(7) of 8-aza-7-deazapurine-2,6-diamine or H-C(5) of uracil residues. *Helv. Chim. Acta* **84**, 1048-1065 (2001).

84. Herdewijn, P. Heterocyclic modifications of oligonucleotides and antisense technology. *Antisense Nucleic Acid Drug Dev.* **10**, 297-310 (2000).

85. Savitzky, A. & Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **36**, 1627-1639 (1964).

86. Gralla, J. & Crothers, D. M. Free Energy of Imperfect Nucleic Acid Helices. *J Mol Biol* **78**, 301-319 (1973).

87. Martin, F. H., Uhlenbeck, O. C. & Doty, P. Self-complmentary Oligoribonucleotides: Adenylic Acid-Uridylic Acid Block Copolymers. *J Mol Biol* **57**, 201-215 (1971).

88. Borer, P. N., Dengler, B. & Tinoco, I. Stability of Ribonucleic acid Double-stranded Helices. *J. Biol. Chem.* **86**, 843-853 (1974).

89. Holbrook, J. A., Capp, M. W., Saecker, R. M. & Record, M. T. Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: Interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry* **38**, 8409-8422 (1999).

90. Kool, E. T. Preorganization of DNA: Design principles for improving nucleic acid recognition by synthetic oligonucleotides. *Chem. Rev.* **97**, 1473-1487 (1997).

91. Sponer, J., Leszczynski, J. & Hobza, P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **61**, 3-31 (2001).

92. Spink, C. H. & Chaires, J. B. Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA. *Biochemistry* **38**, 496-508 (1999).

93. Battersby, B. J., Lawrie, G. A. & Trau, M. Optical encoding of microbeads for gene screening: alternatives to microarrays. *Drug Discov. Today* **6**, S19-S26 (2001).

94. Wagner, T. & Pfleiderer, W. Synthesis of 2 '-deoxyribonucleoside 5 '-phosphoramidites: New building blocks for the inverse (5 '-3 ')-oligonucleotide approach. *Helv. Chim. Acta* **83**, 2023-2035 (2000).

95. Matyas, G., Giunta, C., Steinmann, B., Hossle, J. P. & Hellwig, R. Quantification of single nucleotide polymorphisms: A novel method that combines primer extension assay and capillary electrophoresis. *Hum. Mutat.* **19**, 58-68 (2002).

96. Syvanen, A. C. From gels to chips: "Minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum. Mutat.* **13**, 1-10 (1999).

97. Syvanen, A. C., Aaltosetala, K., Harju, L., Kontula, K. & Soderlund, H. A Primer-Guided Nucleotide Incorporation Assay in the Genotyping of Apolipoprotein-E. *Genomics* **8**, 684-692 (1990).

98.   Battersby, B. J., Bryant, D., Meutermans, W., Matthews, D., Smythe, M. L. & Trau, M. Toward larger chemical libraries: Encoding with fluorescent colloids in combinatorial chemistry. *J. Am. Chem. Soc.* **122**, 2138-2139 (2000).

99.   Matthews, D. C., Grondahl, L., Battersby, B. J. & Trau, M. Multi-fluorescent silica colloids for encoding large combinatorial libraries. *Aust. J. Chem.* **54**, 649-656 (2001).

100.   Nielsen, J., Brenner, S. & Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **115**, 9812-9813 (1993).

101.   Birchhirschfeld, E., Foldespapp, Z., Guhrs, K. H. & Seliger, H. A Versatile Support for the Synthesis of Oligonucleotides of Extended Length and Scale. *Nucleic Acids Res.* **22**, 1760-1761 (1994).

102.   BirchHirschfeld, E., FoldesPapp, Z., Guhrs, K. H. & Seliger, H. Oligonucleotide synthesis on polystyrene-grafted poly(tetrafluoroethylene) support. *Helv. Chim. Acta* **79**, 137-150 (1996).

103.   Rink, H. Solid-Phase Synthesis of Protected Peptide-Fragments Using a Trialkoxy-Diphenyl-Methylester Resin. *Tetrahedron Lett.* **28**, 3787-3790 (1987).

104.   Chang, C. D. & Meienhofer, J. Solid-phase peptide synthesis using mild base cleavage of N alpha-fluorenylmethyloxycarbonylamino acids, exemplified by a synthesis of dihydrosomatostatin. *Int. J. Pept. Protein Res.* **11**, 246-249 (1978).

105.   Lane, S. J. & Pipe, A. Single bead and hard tag decoding using accurate isotopic difference target analysis-encoded combinatorial libraries. *Rapid Commun. Mass Spectrom.* **14**, 782-793 (2000).

106.   Brown, B. B., Wagner, D. S. & Geysen, H. M. A single-bead decode strategy using electrospray ionization mass spectrometry and a new photolabile linker: 3-Amino-3-(2- nitrophenyl)propionic acid. *Mol. Divers.* **1**, 4-12 (1995).

107.   Ferrer, E., Neubauer, G., Mann, M. & Eritja, R. Synthesis of oligodeoxynucleotides containing 5-aminouracil and its N-acetyl derivative. *J. Chem. Soc.-Perkin Trans. 1*, 2051-2057 (1997).

108.   McDowell, J. A. & Turner, D. H. Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAGGUCUC)(2) by two-dimensional NMR and simulated annealing. *Biochemistry* **35**, 14077-14089 (1996).

109.   Hobbs, F. W. & Cocuzza, A. J. 40 (DU PONT (US), European Patent Application, 1988).

110.   Froehler, B. C. & Matteucci, M. D. Dialkylformamidines - Depurination Resistant N-6-Protecting Group for Deoxyadenosine. *Nucleic Acids Res.* **11**, 8031-8036 (1983).

111.   Hoffer, M. a-Thymidin. *Chem. Ber.* **93**, 2777-2781 (1960).

112.   Rolland, V., Kotera, M. & Lhomme, J. Convenient preparation of 2-deoxy-3,5-di-O-p-toluoyl-alpha-D- erythro-pentofuranosyl chloride. *Synth. Commun.* **27**, 3505-3511 (1997).

113.   Davoll, J. Pyrrolo[2,3-*d*]pyrimidines. *J. Chem. Soc.*, 131-138 (1960).

114.   Seela, F. & Lupke, U. Mannich Reaction at 2-Amino-3,7-dihydropyrrolo[2,3-*d*]pyrimidin-4-one, the Chromophore of the Ribonucleoside "Q". *Chem. Ber.-Recl.*, 1462-1469 (1977).

115.   Pudlo, J. S., Nassiri, M. R., Kern, E. R., Wotring, L. L., Drach, J. C. & Townsend, L. B. Synthesis, Antiproliferative, and Antiviral Activity of Certain 4-Substituted and 4,5-Disubstituted 7-[(1,3-Dihydroxy-2- Propoxy)Methyl]Pyrrolo[2,3-D]Pyrimidines. *J. Med. Chem.* **33**, 1984-1992 (1990).

116.   Seela, F., Westermann, B. & Bindig, U. Liquid Liquid and Solid Liquid Phase-Transfer Glycosylation of Pyrrolo[2,3-D]-Pyrimidines - Stereospecific Synthesis

of 2- Deoxy-Beta-D-Ribofuranosides Related to 2'-Deoxy-7- Carbaguanosine. *J. Chem. Soc.-Perkin Trans. 1*, 697-702 (1988).

117. Seela, F. & Zulauf, M. Palladium-catalyzed cross coupling of 7-iodo-2'-deoxytubercidin with terminal alkynes. *Synthesis*, 726-& (1996).

118. Ramasamy, K., Imamura, N., Robins, R. K. & Revankar, G. R. A Facile Synthesis of Tubercidin and Related 7-Deazapurine Nucleosides Via the Stereospecific Sodium-Salt Glycosylation Procedure. *Tetrahedron Lett.* **28**, 5107-5110 (1987).

119. Kazimierczuk, Z., Cottam, H. B., Revankar, G. R. & Robins, R. K. Synthesis of 2'-Deoxytubercidin, 2'-Deoxyadenosine, and Related 2'-Deoxynucleosides Via a Novel Direct Stereospecific Sodium- Salt Glycosylation Procedure. *J. Am. Chem. Soc.* **106**, 6379-6382 (1984).

120. McBride, L. J., Kierzek, R., Beaucage, S. L. & Caruthers, M. H. Amidine Protecting Groups for Oligonucleotide Synthesis. *J. Am. Chem. Soc.* **108**, 2040-2048 (1986).

121. Seela, F. & Becher, G. Pyrazolo 3,4-d pyrimidine nucleic acids: adjustment of dA-dT to dG-dC base pair stability. *Nucleic Acids Res.* **29**, 2069-2078 (2001).

122. Seela, F. & Richter, R. Aminomethylation 0f 3,7-Dihydropyrrolo[2,3-*d*]pyrimidines at C-5 - a Method for the Synthesis of Aglycone Analogues of the Nucleoside "Q". *Chem. Ber.-Recl.*, 2925-2930 (1978).

123. Seela, F., Kehne, A. & Winkeler, H. Synthesis of Acyclo-7-Deazaguanosine by Regiospecific Phase- Transfer Alkylation of 2-Amino-4-Methoxy-7h-Pyrrolo[2,3-D]Pyrimidine. *Liebigs Annalen Der Chemie*, 137-146 (1983).

124. Seela, F., Chen, Y. M., Bindig, U. & Kazimierczuk, Z. Synthesis of 2'-Deoxyisoinosine and Related 2'- Deoxyribonucleosides. *Helv. Chim. Acta* **77**, 194-202 (1994).

125. Ramzaeva, N., Mittelbach, C. & Seela, F. 7-halogenated 7-deaza-2 '-deoxyinosines. *Helv. Chim. Acta* **82**, 12-18 (1999).

126. Perrin, D. D. & Armarego, W. L. F. *Purification of Laboratory Chemicals* (Pergamon Press, 1989).

127. Still, W. C., Kahn, M. & Mitra, A. Rapid Chromatographic Technique for Preparative Separations with Moderate Resolution. *J. Org. Chem.* **43**, 2923-2925 (1978).

128. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.* **62**, 7512-7515 (1997).

129. Sarin, V. K., Kent, S. B. H., Tam, J. P. & Merrifield, R. B. Quantitative Monitoring of Solid-Phase Peptide-Synthesis by the Ninhydrin Reaction. *Anal. Biochem.* **117**, 147-157 (1981).

130. Petersheim, M. & Turner, D. H. Base-Stacking and Base-Pairing Contributions to Helix Stability - Thermodynamics of Double-Helix Formation with Ccgg, Ccggp, Ccgap, Accggp, Ccggup, and Accggup. *Biochemistry* **22**, 256-263 (1983).

131. http://www.meltwin.com.

132. Xia, T. B., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X. Q., Cox, C. & Turner, D. H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**, 14719-14735 (1998).

133. Miller, J. C. & Miller, J. N. in *Statistics for Analytical Chemistry* (eds. Masson, M., Tyson, J. & Stockwell, P.) 31-52 (Ellis Horwood PTR Prentice Hall, 1994).