University of Southampton

Novel techniques for genetic analysis

Rohan T. Ranasinghe

A thesis submitted for the degree of Doctor of Philosophy

School of Chemistry

January 2005

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

SCHOOL OF CHEMISTRY

<u>Doctor of Philosophy</u>

NOVEL TECHNIQUES FOR GENETIC ANALYSIS

by Rohan Tissa Ranasinghe

Genetic analysis involves detection of nucleic acids in a sequence-specific manner. Typically, oligonucleotide probes labelled with fluorescent dyes are used to facilitate detection of their complementary sequences. The development of two fluorescent probe formats, each employing DNA-intercalators as fluorescence quenchers is described. The first, in which the intercalator 9-amino-6-chloro-2-methoxyacridine is covalently linked to the probe, adjacent to a 5'-fluorophore, increases its fluorescence upon hybridisation to the target sequence, due to interaction of the quencher with the probe/target duplex. The hybridisation specificity of these probes has been demonstrated, culminating in their use in real time PCR. The second format, which involves binding of intercalators to the probe/target duplex from free solution, leads to a decrease in fluorescence upon hybridisation. A range of DNA-binding ligands have been screened for use in this context, leading to the use of ethidium bromide as an intercalating quencher in real time PCR.

Efficient hybridisation of labelled oligonucleotide probes to their complementary sequences in PCR products is important for sensitive PCR-based genetic analysis. This can be hindered by competition between the probe and one amplicon strand for the target sequence. Several solutions to this problem are evaluated. Among these, the use of 2'-deoxyinosine-5'-triphosphate (dITP) to produce amplicons with reduced T_m is outlined.

Synthetic oligonucleotide probes used for genetic analysis must be obtained in high purity. Two hydrophobic tagging monomers for synthesis of the probes used in the multiplex ligation dependent probe amplification (MLPA) assay have been developed. Introduction of the tags at the 5'-end of synthetic oligonucleotides >100 nt in length facilitates their purification by RP HPLC.

i

Contents

Chapter 2 - DNA Intercalators as fluorescence quenchers in genetic analysis – covalent approach

Chapter 3 - DNA Intercalators as fluorescence quenchers in genetic analysis – noncovalent approach

Chapter 4 - Problems with detecting PCR products with oligonucleotide probes and possible solutions

Abbreviations

δ	chemical shift (parts per million)
ε	extinction coefficient
λ_{em}	emission wavelength
λ_{ex}	excitation wavelength
λ_{max}	UV-absorption maximum
λ_{obs}	observed wavelength
τ	fluorescence lifetime
ν_{max}	Infrared absorption maximum
-ve	negative control
^{1}H	proton (NMR)
7-AACTD	7-aminoactinomycin D
^{13}C	carbon (NMR)
^{31}P	phosphorus (NMR)
A	adenine
Å	Ångstrom (10^{-10} m)
ABI	Applied Biosystems
ACD	9-amino-6-chloro-2-methoxyacridine
AcOH	acetic acid
ACTD	actinomycin D
AEGIS	an expanded genetic information system
ARMS	amplification refractory mutation system
BHQ	Black hole quencher
B_{max}	total binding sites
bp	base pairs
BSA	bovine serum albumin
C	cytosine
cDNA	complementary DNA
C_T	threshold cycle
CFTR	cystic fibrosis transmembrane conductance regulator
Cy	cyanine (dye)
dA	2'-deoxyadenine

DABCYL	4- (4'-*N*,*N*,-dimethylaminophenylazo) benzoic acid
dATP	2'-deoxyadenosine 5'-triphosphate
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
dC	2'-deoxycytosine
DCM	dichloromethane
ddNTP	2',3'-dideoxynucleoside 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
DCM	dichloromethane
$d^{4Et}C$	*N*4-ethyl-2'-deoxycytidine
$d^{4Et}CTP$	*N*4-ethyl-2'-deoxycytidine 5'-triphosphate
DFG	double-flap gap
dG	2'-deoxyguanine
dGTP	2'-deoxyguanosine 5'-triphosphate
dI	2'-deoxyinosine
DIPEA	*N*,*N*-diisopropylethylamine
dITP	2'-deoxyinosine 5'-triphosphate
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DMT	4,4'-dimethoxytrityl
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
dR	2'-deoxyribose
ds	double stranded
DTT	dithiothreitol
DU	2'-deoxyuridine
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
eq.	equivalents
ERα	estrogen receptor α
ES	electrospray (mass spectrometry)
EtBr	ethidium bromide
EtOAc	ethyl acetate
F	fluorophore/fluorescence intensity

FAM	5(6)-carboxyfluorescein
FITC	5(6)-carboxyfluorescein isothiocyanate
FRET	fluorescence resonance energy transfer
G	guanine
H-CRCA	hyperbranched cascade rolling circle amplification
HEG	hexa(ethylene glycol)
het.	heterozygote
HIV	human immunodeficiency virus
HPA	3-hydroxypicolinic acid
HPLC	high-pressure liquid chromatography
iC	*iso*-cytosine
iG	*iso*-guanine
K_d	equilibrium dissociation constant
LATE-PCR	linear-after-the-exponential polymerase chain reaction
LCR	ligase chain reaction
LDR	ligation detection reaction
LED	light-emitting diode
LNA	locked nucleic acid
LRET	luminescence resonance energy transfer
LRMS	low resolution mass spectroscopy
MALDI-TOF	matrix-assisted laser desorption/ionisation time of flight
^{MD}A	methoxybenzodeazaadenine
^{MD}I	methoxybenzodeazainosine
MeCN	acetonitrile
MeOH	methanol
MLPA	multiplex ligation-dependent probe amplification
mRNA	messenger RNA
MS	mass spectrometry
MTHFR	methylene tetrahydrofolate reductase
NAP	nucleic acid purification
NH_4OAc	ammonium acetate
NMR	nuclear magnetic resonance
nt	nucleotides

OCT	octanediol
OLA	oligonucleotide ligation assay
p	phosphate
PA	picolinic acid
PCR	polymerase chain reaction
PNA	peptide nucleic acid
Pu	purine
Py	pyrimidine
RCA	rolling circle amplification
RNA	ribonucleic acid
ROX	6-carboxy-X-rhodamine
RP HPLC	reverse phase high-pressure liquid chromatography
RT	room temperature (25 °C)
RT-PCR	reverse transcription polymerase chain reaction
sat.	saturated
SNP	single nucleotide polymorphism
ss	single stranded
T	thymine/thymidine
TAMRA	6-carboxytetramethylrhodamine
Taq	*Thermus aquaticus*
TBE	tris-borate-ethylenediamine tetraacetic acid
TEAA	triethylamonium acetate
TLC	thin layer chromatography
T_m	melting temperature (temperature at which duplex is half dissociated)
Tris	tris(hydroxymethyl)aminomethane
Trt	trityl
TTP	thymidine 5'-triphosphate
U	uracil/units of enzyme
UV	ultra violet
wt.	wild type
YO-PRO-1	4-[(3-methyl-2(3*H*)-benzoxazolylidene) methyl]-1-[3-propyltrimethyl ammonium]-

Declaration of authorship

I, Rohan Ranasinghe, declare that the thesis entitled "Novel Techniques for Genetic Analysis" and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- no part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always clearly given. With the exceptions of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- parts of this work have been published as:

1. R. T. Ranasinghe, L. J. Brown, and T. Brown, *Chem. Commun.*, 2001, 1480.
2. R. T. Ranasinghe, L. J. Brown and T. Brown. In 'Innovation & Perspectives in Solid Phase Synthesis & Combinatorial Libraries 2002: Peptides, Proteins and Nucleic Acids, Small Molecule Organic Chemical Diversity (Proceedings of the 7th International Symposium, Southampton, UK)', ed. R. Epton., Mayflower Worldwide., 2002, 253.

September 2004

Acknowledgements

1. Introduction

1.1 Chemical and physical properties of nucleic acids

1.1.1 Composition and structure of nucleic acids

DNA and RNA are condensed polymers of monomeric units called nucleotides, the phosphate ester of a nucleoside, which is the condensation product of a pentose sugar with a nitrogenous base. The pentose sugar is either ribose (in RNA) or deoxyribose (in DNA), and the nitrogenous bases found in nucleic acids are either monocyclic pyrimidines (cytosine, thymine – found in DNA, and uracil – found in RNA) or bicyclic purines (guanine and adenine) (Figures 1.1 and 1.2).

The secondary structure of DNA, as well as some of its properties, is derived from the nature of these bases. Chargaff, using a very early UV spectrophotometer, discovered from comparison of a wide variety of DNA sources that the base composition could vary within the constraint that the number of adenines was always equal to the number of thymines, and that the number of guanines was always equal to the number of cytosines.[1]
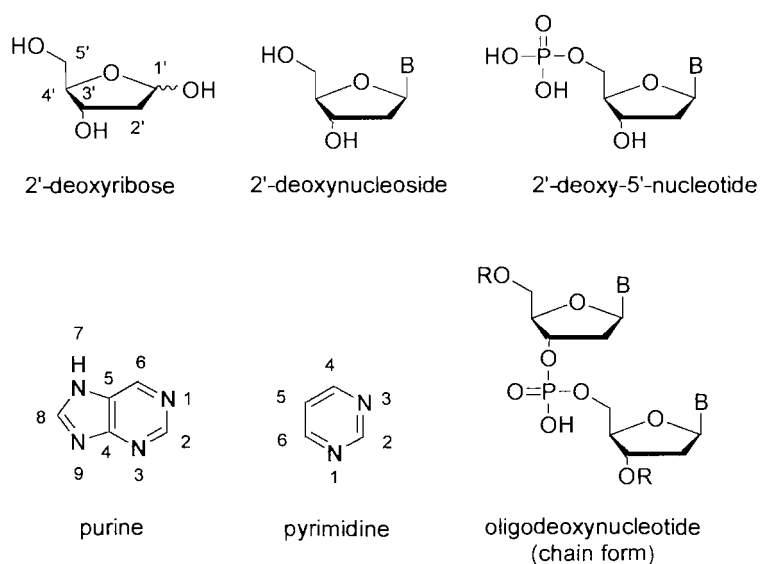
Figure 1.1 Structural units found in nucleic acids. B = a heterocyclic nucleobase (A, G, C or T), and R = the rest of the chain.

1
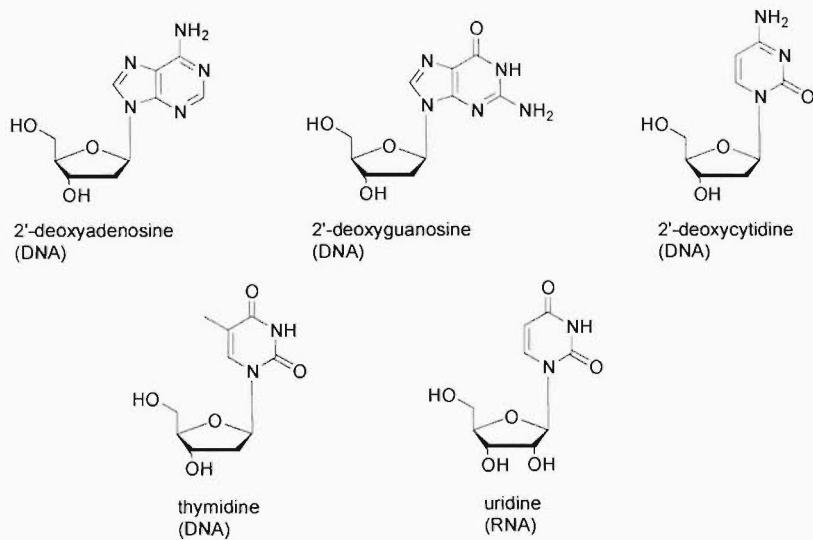
Figure 1.2 Structures of nucleosides found in DNA and RNA.

The relevance of Chargaff's rules was fully uncovered when Watson and Crick elucidated the secondary structure of DNA in 1953 (Figure 1.3).[2] Coupling these rules with, amongst other data, X-ray diffraction patterns,[3] they used models to postulate structures, and eventually arrived at a structure which satisfied all the physical observations which had been made about DNA.
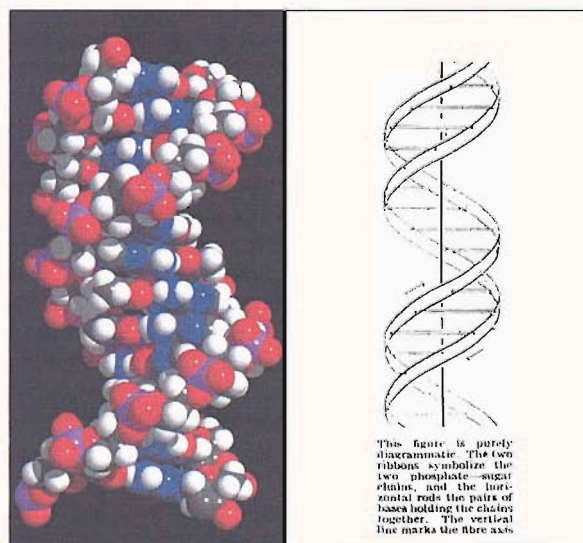
Figure 1.3 Two views of the double helix – one derived from molecular modelling technology (left)[2] and that published by Watson and Crick in 1953 (right).[1, 4]

The structure is an antiparallel double helix with the hydrophilic, anionic phosphate groups at the edge (allowing them to be solvated by water), and the bases packed in the central core separated by 3.4 Å (forming the base pairs implied by Chargaff's work), stabilising the structure by both hydrophobic and π-stacking interactions. The specificity of these base pairs is brought about by the orientation of the hydrogen bonds that form between bases. All the Watson-Crick base pairs (Figure 1.4) have the same overall shape (i.e. they are pseudosymmetric) and can form with no disruption to the double helix.
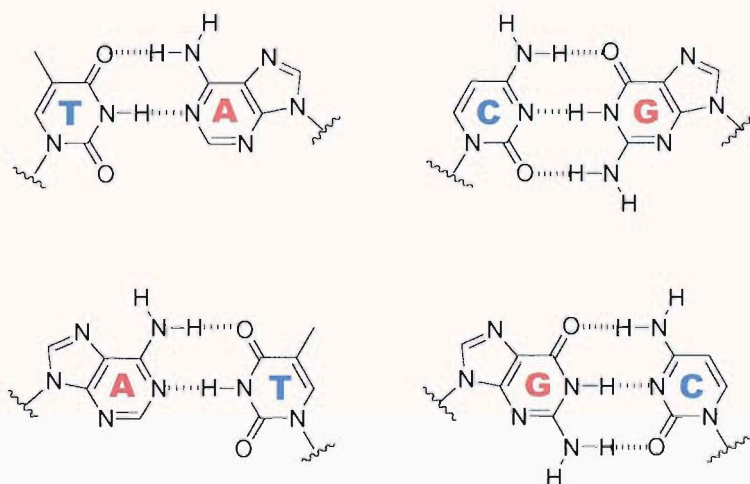
Figure 1.4 DNA base pairs.

The Watson-Crick model for the secondary structure of DNA was attractive since it immediately suggested a mechanism for its replication. Since the two intertwined strands are complementary, it was speculated that one single strand may act as a template for some cellular apparatus to assemble its complement.[5]
Confirmation of this theory has been provided by the work of Kornberg,[6] who showed that replication (or polymerisation) of DNA is catalysed in cells by a group of enzymes called polymerases, by the mechanism first postulated by Watson and Crick.[7]

1.1.2 Intercalation

In 1961, Lerman described the interaction of planar, protonated acridine compounds with DNA.[8] Reduction of the sedimentation constant and enhanced viscosity of the

duplex indicated that these compounds caused a lengthening and stiffening of the DNA helix. This behaviour was accounted for in a model with the acridines sandwiched between adjacent base pairs, which he described as intercalation (Figure 1.5). The intercalation process draws adjacent base pairs apart from their usual 3.4 Å spacing to 6.8 Å,[9] and unwinds the helix.[10] When a singly charged cationic drug is intercalated, distortion of the helix is offset by the 20-35 kJ.mol[-1] gain in enthalpy from increased π-π stacking, and the release of bound cations from the duplex,[11] which results in an increase in T_m (the temperature at which the duplex is half-dissociated).[12] In general, planar intercalators show a binding preference for G-C base pairs, which is explained by the greater polarisibility of the G-C base pair compared to A-T.[13]

B-DNA, native form   B-DNA with bound intercalator
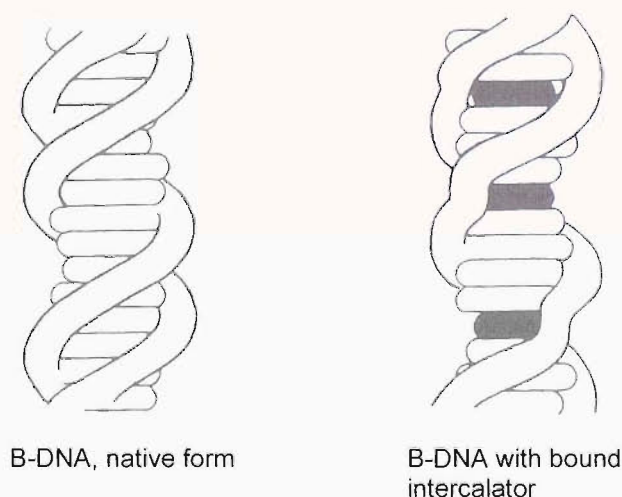
Figure 1.5 Model for DNA duplex containing intercalated drug molecules.[14]

Many families of intercalators are known (Figure 1.6). Classical tricyclic acridines and phenanthridiniums, such as ethidium and propidium, intercalate according to Lerman's model, but other classes are known, where intercalation is accompanied by groove binding.

One such class of compounds is the anthracyclines, including daunomycin, adriamycin and nogalamycin, which contain planar aromatic rings and sugars bearing hydrogen bonding residues (Figure 1.6). Intercalation of daunomycin involves intercalation of the B and C rings and protrusion of the D ring into the major groove. The A ring and the amino sugar bind to the minor groove, with the A

ring hydroxyl group hydrogen bonding to the N3 and 2-amino groups of G. The groove binding provides extra stabilisation of the complex by hydrogen bonding and exclusion of water and imparts sequence-specificity to the interaction – daunomycin preferentially binds CpG sites.[15] Nogalamycin has two sugars - the amino sugar mounted on ring D binds to the major groove, hydrogen bonding to G-C base pairs, and the nogalose sugar attached to ring A binds to the minor groove. Nogalamycin is termed a threading intercalator because the two sugars interact with grooves on opposite sides of the intercalated aromatic system.[16]

R=H; Actinomycin D
R=NH₂; 7-Aminoactinomycin D

R₁=R₂=R₃=H; Acridine
R₁=NH₂, R₂=R₃=H; 9-Amnoacridine
R₁=H, R₂=R₃=NH₂; Proflavine
R₁=H, R₂=R₃=NMe₂; Acridine Orange

Nogalamycin

R=H; Daunomycin
R=OH; Adriamycin

R=Et; Ethidium bromide
R=(CH₂)₃N⁺(Me)Et₂; Propidium bromide

Thiazole orange
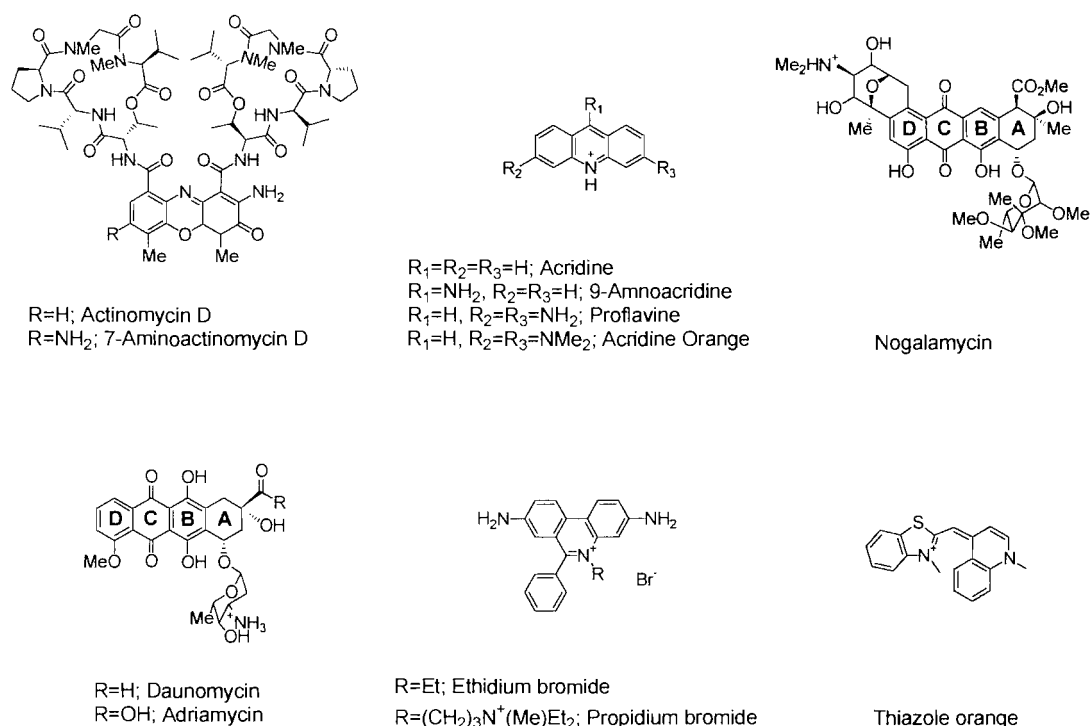
Figure 1.6 Common DNA intercalators.

Actinomycin D and 7-aminoactinomycin D, intercalators containing a phenoxazine heterocycle and two cyclic peptide arms (Figure 1.6), are unusual in that they are uncharged and bind GpC, rather than the usual Py-Pu steps.[17] The thermodynamics of intercalation involve an enthalpy change of between 0 and +8 kJ.mol⁻¹, which is accompanied by a negative change in entropy, in contrast to the usual gain in enthalpy and loss of entropy.[18] The two peptide arms bind to the minor groove, with the threonine residues hydrogen bonding to the N3 and 2-amino groups of Gs in opposite strands, accounting for the observed selectivity for GpC sites.[19]

Intercalators have found applications in medicine and biotechnology due to their

interactions with dsDNA. As inhibitors of DNA synthesis, they are frequently used as antibiotics and as anticancer drugs. Fluorescent intercalators are used extensively in nucleic acid staining,[20] and their stabilising effect on nucleic acid structures has led to the use of intercalator-oligonucleotide conjugates in duplex,[21] and triplex formation.[22]

1.2 Factors to be considered in genetic analysis

1.2.1 Importance of detecting nucleic acids

An organism's genome (the sum of its genetic information) encodes its inherited characteristics. This is because it is translated by the cellular apparatus to express the proteins, (e.g. enzymes), which mediate all cellular processes. The mechanism by which this occurs can be described crudely as: "DNA makes RNA makes proteins." The sequence of the protein is determined by the sequence of the DNA. This is what Crick referred to as "the central dogma of molecular biology". The detailed processes by which this occurs can be found in many textbooks.[23]
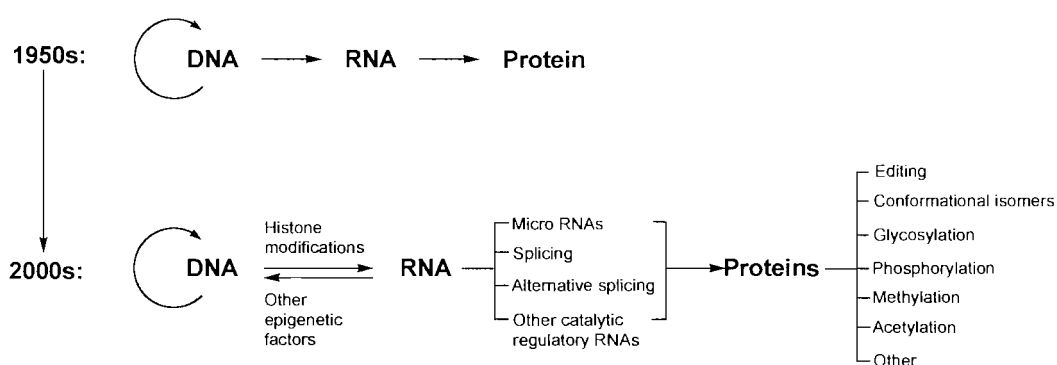
Figure 1.7 Representations of Crick's central dogma (top) and current models for protein synthesis *in vivo* (bottom).[24]

The view of central dogma is now more complex and no longer unidirectional (Figure 1.7).[24] First, the discovery of reverse transcription made it clear that RNA and DNA are interconvertible in the genetic code. Then post-translational modifications (glycosylations, methylations etc.) showed that single genes were not wholly responsible for the chemical structure of functional proteins. The results of

the human genome project and other work have shown that single genes (alternatively termed "transcription units") often encode more than one protein by alternative splicing, showing the "one gene-one protein" principle of the central dogma to not be true.[25] This explains, in part, how perhaps as few as 30,000 genes could encode the estimated >1,000,000 proteins.[26]

Nevertheless, genetic analysis, the sequence-specific detection of nucleic acid sequences is an important tool in modern science. The detection of species-specific genes in host tissues can show the presence of pathogens or viral genes embedded in the host's genome. Infectious diseases such as HIV,[27] *Chlamydia trachomatis*[28] and hepatitis C[29] can be diagnosed in this way, and bioterrorism agents can be detected in sterile body fluids.[30]

Genes are first transcribed to messenger RNA (mRNA) then translated into proteins. Semi-quantitative detection of mRNAs can therefore show which genes are being expressed, and at what levels. This can provide information about key processes that are occurring in an organism at any one time.[31, 32] For example, increased expression of estrogen receptor α (ERα) is found in human breast carcinomas.[33]

Mutations to an organism's genome can lead to genetic diseases. Large genomic duplications and deletions (in the kilobase order) are the recognised cause of α-thalassaemia,[34] Duchenne and Becker muscular dystrophies[35] and in familial breast cancer.[36] In contrast, point mutations in an organism's genome can lead to Huntingdon's disease,[37] cystic fibrosis,[38] sickle cell anaemia[39] and β-thalassemia[40] and can be diagnosed by DNA-based technologies.

Single nucleotide polymorphisms (SNPs) that have not been linked to disease phenotypes are currently the subject of significant research efforts.[41] SNPs represent the largest source of genetic variation between individuals. The preliminary results of the human genome project indicate there are approximately 2.1 million SNPs.[25] While most SNPs may not directly confer disease, it is thought that they may be important in the complex genotype-phenotype interactions that are likely to determine whether an individual suffers from a polygenic disease,[42] their susceptibility to various diseases,[43] and the efficacy of their response to drug treatments.[44] Useful information can be extracted from the non-coding regions of the genome, too. Analysis of short tandem repeats (sequences 2-5 nt in length, which are repeated many times in a head to tail fashion) can be used to uniquely identify

individuals.[45] This technique is widely used in paternity testing[46] and crime scene investigations.[47]

1.2.2 Labelling and amplification of nucleic acids

1.2.2.1 Fluorescent labelling

The heterocyclic bases in nucleic acids absorb UV-light, with the absorption maximum of mixed sequence nucleic acids lying at ~260 nm. However, the extinction coefficients of the nucleobases are not large, and many other compounds absorb at this wavelength, so the UV-absorption of the nucleobases is insufficiently sensitive for detection of cellular DNA. As a result, an easily detectable label is required for nucleic acid detection. Although radiolabelling of nucleic acids has been extensively used, fluorescent labels are now preferred, since they are easier to handle, have longer shelf lives and, with suitable instrumentation, are more sensitive. Fluorescence is the name given to the property of some atoms or molecules, which absorb a photon of light of one wavelength (excitation wavelength, λ_{ex}) and emit after some period (the fluorescent lifetime, τ) at another, higher, wavelength (emission wavelength, λ_{em}), a photon of lower energy (Figure 1.8).[48]
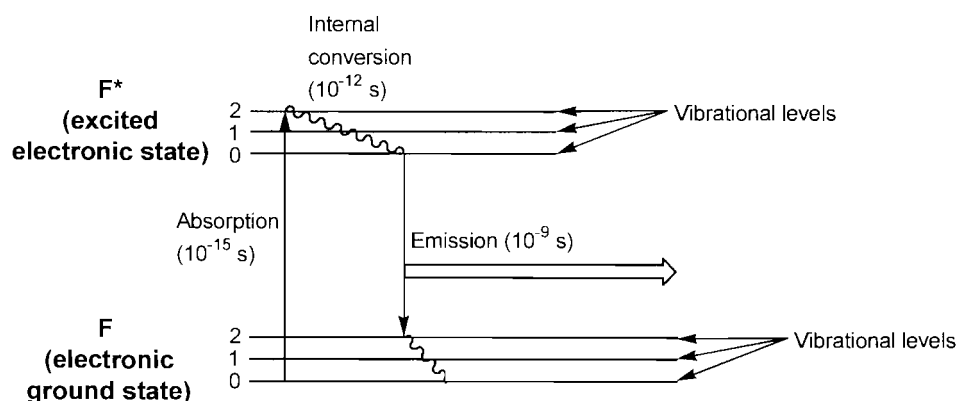
Figure 1.8 A Jablonski Energy Level diagram, depicting Fluorescence.

The molecule accepts a photon at its excitation wavelength, promoting it from the ground state (S_0) to the first excited singlet state (S_1). After intersystem crossing

within the excited state (dissipating energy), relaxation to the ground state occurs, with emission of a photon of a specific wavelength (λ_{em}).
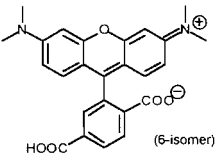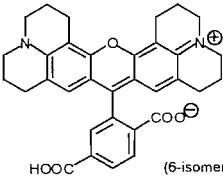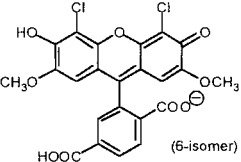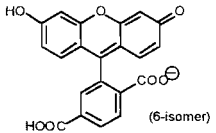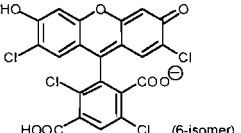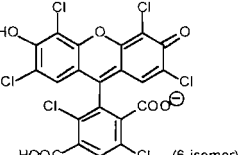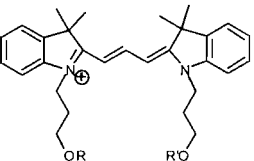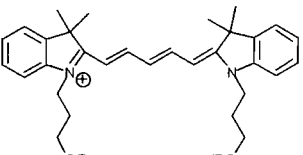
Name	Chemical structure	λ_{ex}	λ_{em}	ε
TAMRA		540	564	103,000
ROX		570	590	113,000
JOE		520	548	71,000
FAM		492	515	81,000
TET		521	536	---
HEX		535	556	---
Cy3		550	570	150,000
Cy5		650	670	250,000

Table 1.1 Fluorescent dyes commonly used in oligonucleotide labelling. Excitation and emission maxima (λ_{ex}, λ_{em}) are reported in nm, while extinction coefficients (ε) are reported in $mol^{-1}.dm^3.cm^{-1}$. Values are for the un-conjugated (free) dyes.[49]

Since the electronic structures of molecules vary depending on their constituent

groups and elements, the specific wavelengths of excitation and emission can be tuned by synthetic derivatisation of dyes. The difference between the excitation and emission wavelengths of a fluorophore is called the Stokes shift, and this varies between dyes. Many fluorescent dyes are commonly used to label oligonucleotides (Table 1.1).

1.2.2.2 Fluorescence quenching

Although fluorescent intercalators are often used in free solution to stain nucleic acids,[20] fluorescently-labelled oligonucleotide probes usually require the use of a quencher in addition to the label. Hybridisation of the probe causes a change in quenching of the fluorophore, indicating the presence of its complementary nucleic acid sequence. Typically, quenching of the label occurs in one of three ways:

i) Collisional (or Dexter) fluorescence quenching (Figure 1.9A).[50] An extremely short-range interaction involving spatial overlap of excited fluorophore and quencher molecular orbitals. The efficiency of energy transfer decays as a function of e^{-R} (where R is the distance between fluorophore and quencher). This mechanism prevails in applications where a fluorophore and quencher are in close proximity e.g. Methyl Red quenching FAM in Molecular Beacons (Figure 1.23).[51]

ii) Fluorescence Resonance Energy Transfer (FRET, or Förster) quenching (Figure 1.9B).[52] Energy is transferred from an excited fluorescent donor to a quencher non-radiatively, *via* a dipole-dipole interaction. The acceptor dye can then emit a photon at its characteristic emission wavelength. The phenomenon is slightly longer-ranging than Dexter quenching, decaying with R^{-6}, and is the mechanism of quenching of the 5'-FAM by the 3'-TAMRA in TaqMan® probes (Figure 1.21).[53] The Förster and Dexter mechanisms are known as dynamic quenching because energy transfer occurs from an excited fluorophore.

iii) Static quenching. The fluorophore and quencher form a non-fluorescent ground state complex, usually a π-stacked dimer, which has unique electronic properties distinct from the individual dyes, leading to a shifted UV spectrum. This occurs in unstructured probes labelled with Cy5 and BHQ-1.[54]

Quenching by electron transfer is possible, but causes rapid photobleaching and is not commonly used in oligonucleotide probes.

10
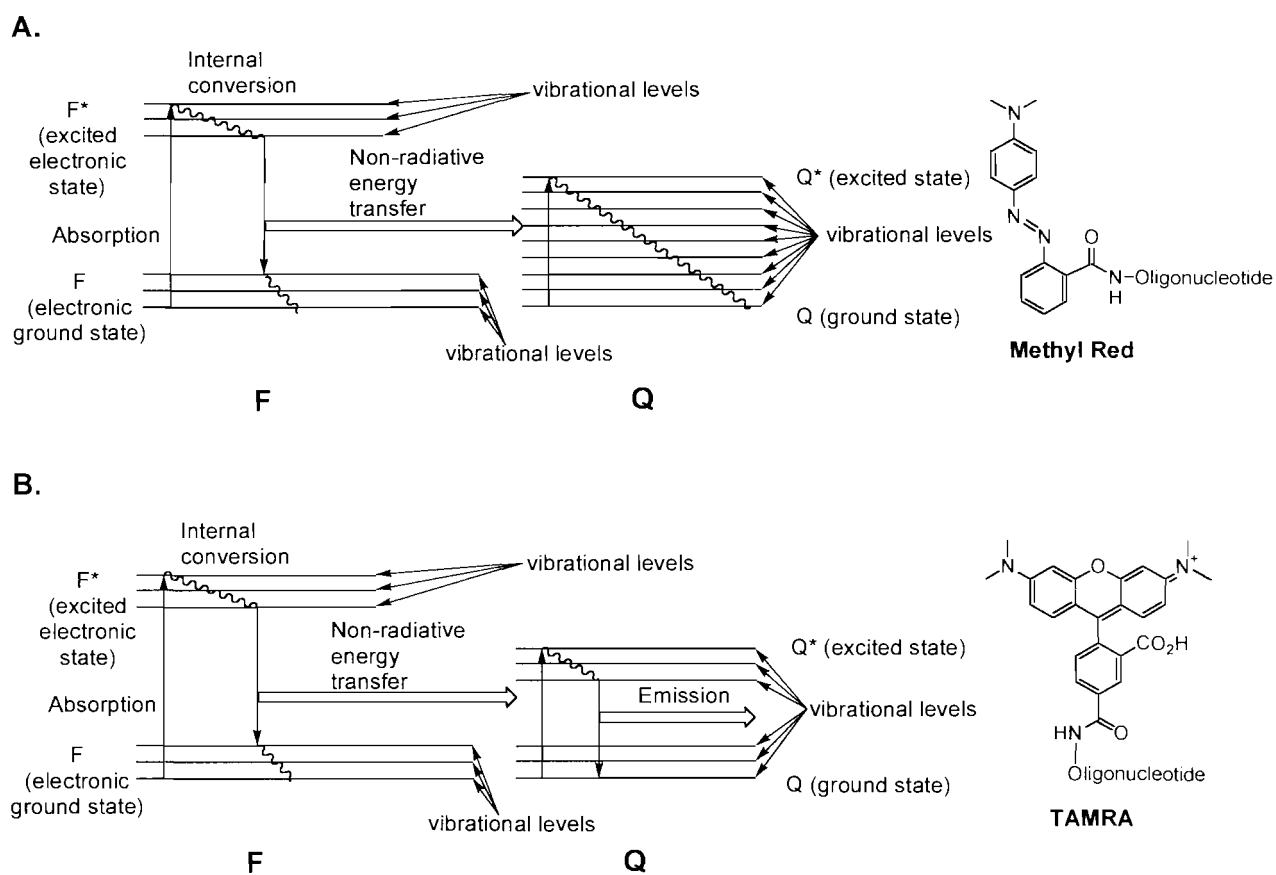
Figure 1.9 Dynamic quenching mechanisms. Collisional (Dexter) quenching (A) and FRET (Förster) quenching (B).
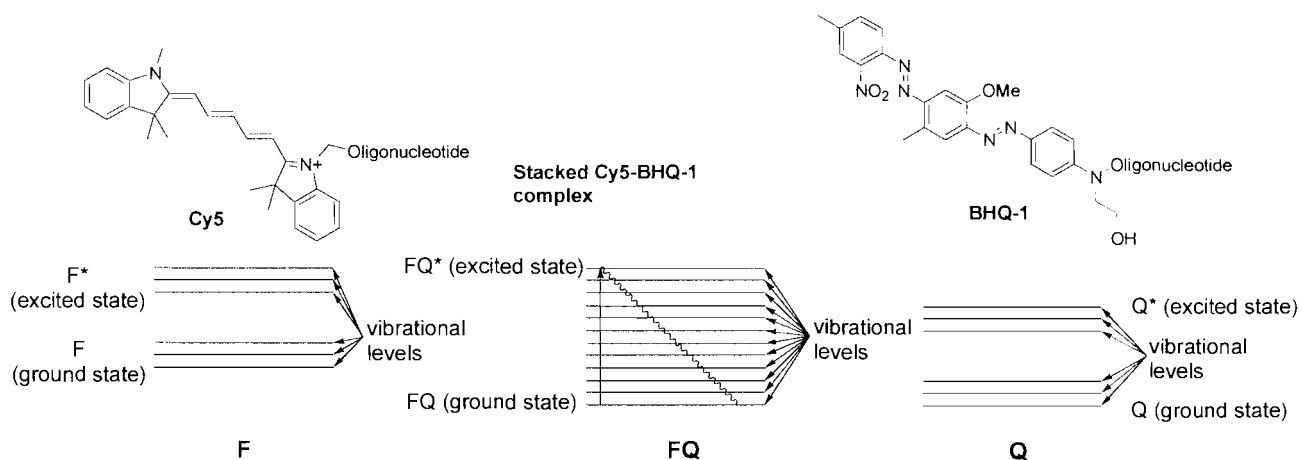
Figure 1.10 Static quenching of Cy5 by BHQ-1.[54]

1.2.2.3 The Polymerase Chain Reaction

In addition to fluorescent labelling, it is usually necessary to amplify nucleic acids extracted from biological samples in order to facilitate detection. The most commonly used method to achieve this is the polymerase chain reaction, or PCR (Figure 1.11).
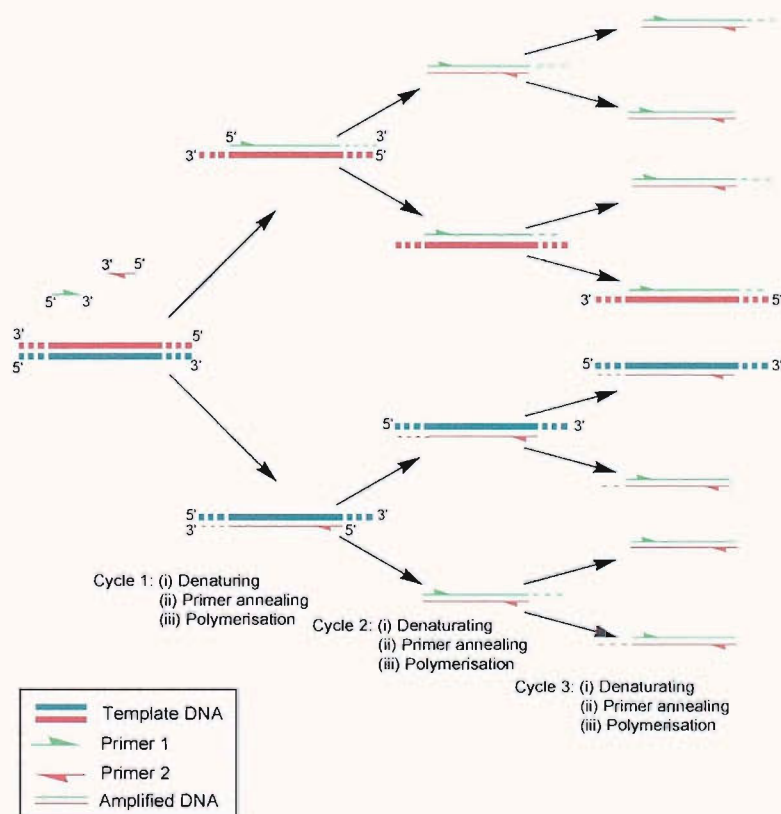
Figure 1.11 A schematic representation of the polymerase chain reaction.

PCR was first described by Saiki *et al.*[55, 56] It is a very simple method for replicating many copies of a specific section of DNA. The procedure employs two synthetic oligonucleotides called primers, which flank the section of DNA to be amplified, each primer being complementary to one strand of the target dsDNA. The first stage in the cycle involves heating to 95 °C, to denature all DNA. The reaction mixture is then cooled rapidly to a lower temperature in order to allow the primers to anneal to their targets. This temperature is determined by the T_ms of the primer-target hybrids. In addition to this, the primers are present in large excess ensuring that primer-target hybrids are favoured over re-formation of target duplexes. The temperature is then

elevated to circa 72 °C, near to the optimal for activity of a thermostable polymerase (often extracted from *Thermus Aquaticus*), which then extends both primers in the 5' to 3' direction, copying each strand once to give two-fold amplification (Figure 1.11).

Since each primer extended by this process contains a binding site for the opposite primer, when the cycle is repeated, four copies of the target can be produced, i.e. a chain reaction is occurring.

This cycle is repeated many times to produce 2^n copies, where n is the number of cycles. A rudimentary thermal cycler can perform 25 cycles in 1 hour, producing 2^{25} (~33.5 million) copies of the target. The product of a PCR is called an amplicon, and most genetic analysis technologies rely upon specific detection of one such target.

An elegant adaptation of the polymerase chain reaction, useful in genetic analysis, was described by Newton *et al.* in 1989.[57] The amplification refractory mutation system (ARMS), or allele-specific PCR, harnesses the selectivity of the polymerase enzyme to effect allelic discrimination (Figure 1.12). One of the primers is designed to be complementary (at its 3'- end) to one allele of the mutation site of interest (an allele- specific primer). If the nucleotide at the 3'-end of the primer is complementary to the template PCR proceeds. However, if the base at the 3'-end of the primer is not complementary to the template, it is not extended by the polymerase, and PCR does not proceed. A variety of mutation types (including point mutations) can be analysed with this technique, by detection of the presence or absence of a PCR product.
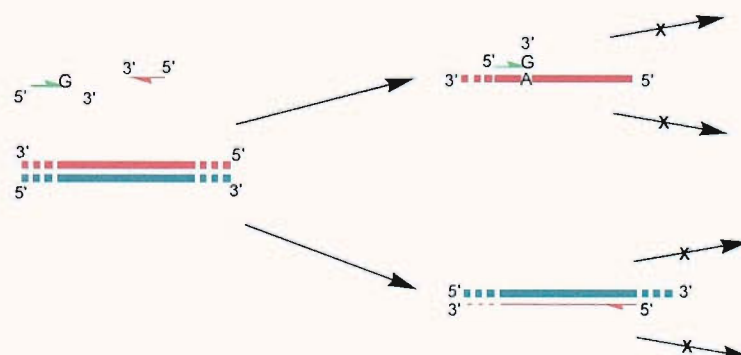
Figure 1.12 The ARMS PCR method.

1.3 Modern techniques for genetic analysis

1.3.1 Sequencing

Sequencing is used to determine the sequences of nucleic acids of unknown composition.
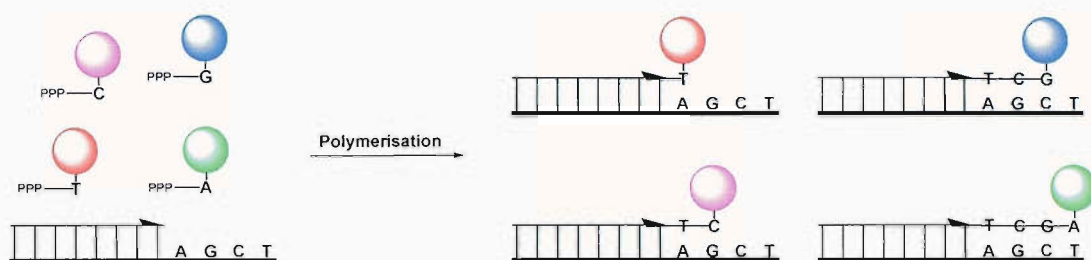
Figure 1.13 Dye-terminator Sanger dideoxy sequencing.

Sanger (or dideoxy) sequencing involves production of a number of labelled DNA fragments which are complementary to the sequence of interest, but which vary in length, i.e. each fragment is a fraction of the total sequence. These fragments each differ in length by one nucleotide, the smallest fragment being one nucleotide longer than the primer, and the longest being the length of the entire sequence.[58]

The fragments are resolved by gel or capillary electrophoresis, and from this the sequence can be deduced. A new strand is synthesised with a polymerase, using the original strand as a template. In addition to the dNTPs this requires, ddNTPs (dideoxynucleoside triphosphates) are introduced. Since ddNTPs contain no 3`-hydroxyl group, the chain can be extended no further after their incorporation - they are "chain terminators". There are two popular methods of Sanger sequencing. In dye-terminator sequencing, four ddNTPs are added, each bearing a different fluorophore, allowing the terminal nucleotide of each fragment to be identified (Figure 1.13).[59] Energy transfer dyes have been developed in order to maximise the signal obtained from different dyes with single excitation source equipment.[60] In dye-primer sequencing, the PCR-primer is fluorescently-labelled.[61] This method requires four separate sequencing reactions (one with each ddNTP) to be performed before electrophoresis.

However, resolution by electrophoresis is inherently undesirable, adding another step to the protocol, and also representing a possible source of cross-contamination.

With draft versions of many genomic sequences now available, the sequencing of DNA is becoming less important in the detection of known mutations. Ideally, the information obtained from sequencing can be used to design probes which can detect DNA in a sequence specific manner, but without the need for sequencing, thus creating a more efficient, higher throughput, and therefore more cost effective assay. The ideal system dispenses with the tedious and time-consuming steps of sample purification and separation after the allele discrimination reaction. This allows minimal sample handling, reducing sources of contamination, with potential for automation.

1.3.2 Real Time PCR

In 1996, a new technique for detecting nucleic acids was described.[62] Real time, or kinetic, PCR is a method in which the accumulation of amplicon is monitored throughout the PCR, rather than at the end. This allows detection in a 'closed-tube' format, where no post-amplification manipulation is required, removing the risk of cross contamination between amplified samples. A fluorescent signal whose strength is proportionate to the amount of PCR product present is measured in each cycle to give an amplification curve (Figure 1.14).
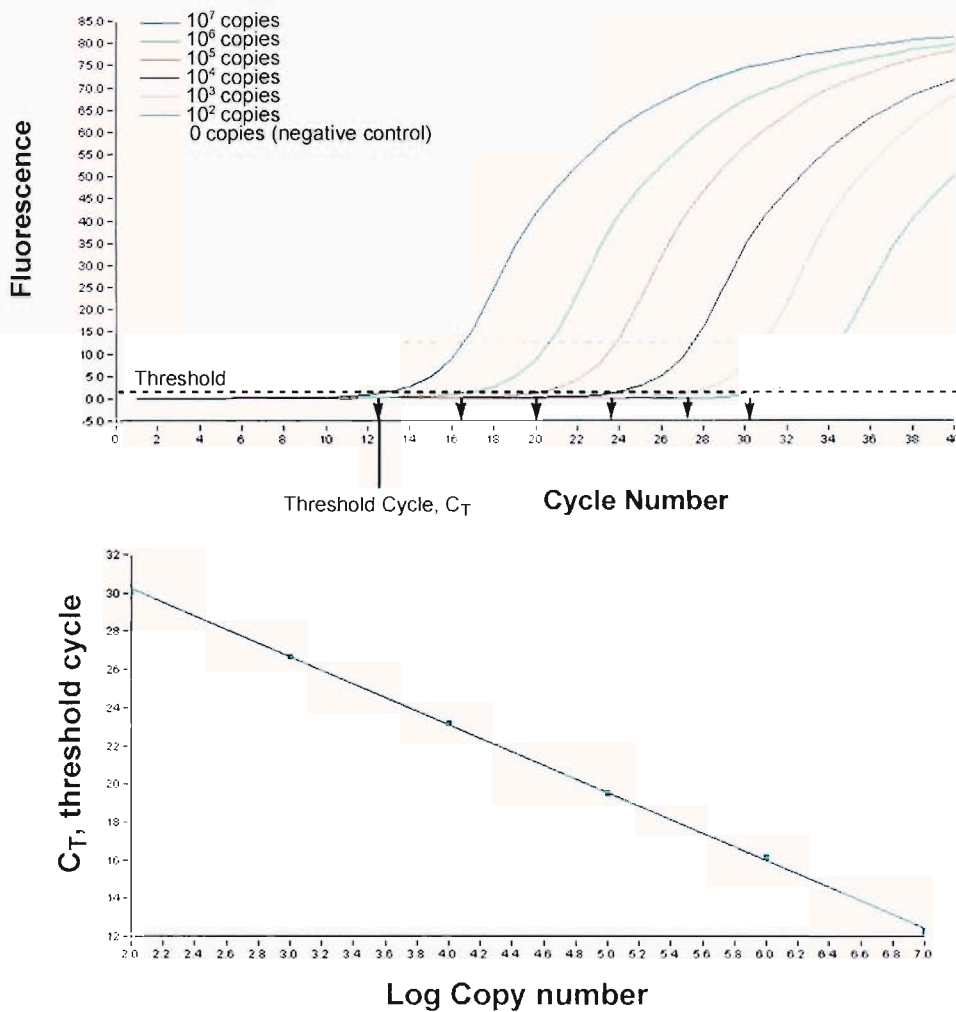
Figure 1.14 Real time PCR traces obtained by varying the number of copies of template, showing the effect on threshold cycle number (C_T), (top) and the relationship between threshold cycle number (C_T) and the log of template copy number (bottom).

Real time PCR represented a major advance in quantification of DNA, allowing a wide range (about eight orders of magnitude) of concentrations to be assayed with high accuracy. The quantity of template present is determined by the cycle, C_T, in which the amplification signal crosses a pre-determined threshold (Figure 1.14). The fewer copies of template present, the later the threshold is crossed. A plot of C_T versus log (copy number) gives a straight line from which the number of copies present in unknown samples can be determined. Quantification of RNA is also possible by reverse transcription-PCR (RT-PCR), where RNA is first transcribed to cDNA, which can be used as the template in PCR.[63] This method is particularly important as it allows quantification of mRNAs, allowing measurement of expression levels and viral loads by quantification of viral RNA.[31, 32]

Real time PCR can be used for genotyping – wild type, heterozygous and mutant genotypes can give different amplification curves if suitable allele-specific probes or primers are used (Figure 1.15). Real time PCR is useful for genotyping SNPs in this way.[64]
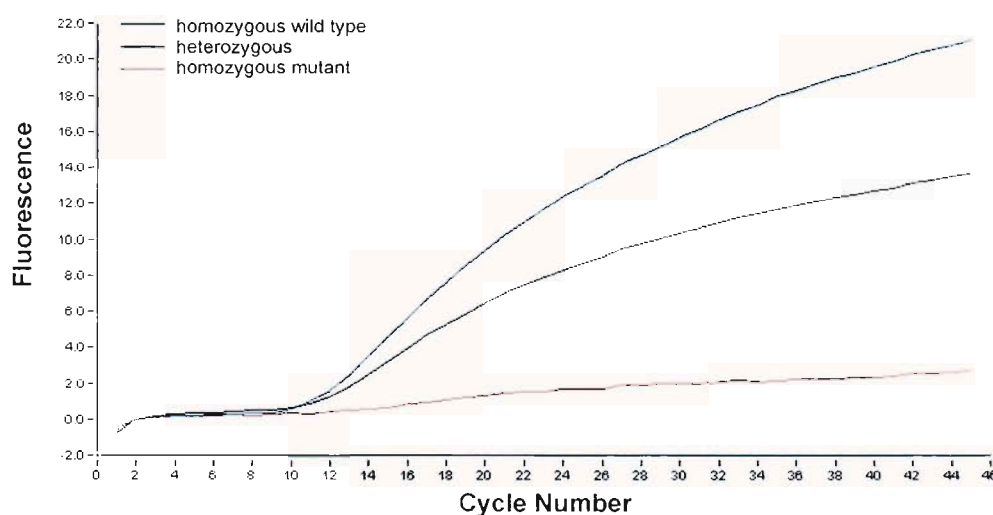
Figure 1.15 Amplification curves from genotypically different samples.

The ability of real time PCR to detect DNA makes it amenable to applications in parasitology,[65] oncology,[66] and food technology.[67, 68] Although these applications only require the presence or absence of specific sequences to be established, and

could be accomplished by endpoint detection, the speed and **convenience** of real time PCR makes it a commonly used technique.

In addition to amplification curves, fluorescence melting curves, acquired at the end of the PCR, can establish the identity of species that are generating the signal. This is particularly useful when the amplification curve is ambiguous. When labelled probes are used, fully complementary sequences can be discriminated from those containing single mismatches by T_m measurement, useful for SNP genotyping. If PCR is not optimal a signal can result from mispriming or primer-dimer extension products, and the presence of these can be revealed by amplicon melting when non strand-specific methods are used.
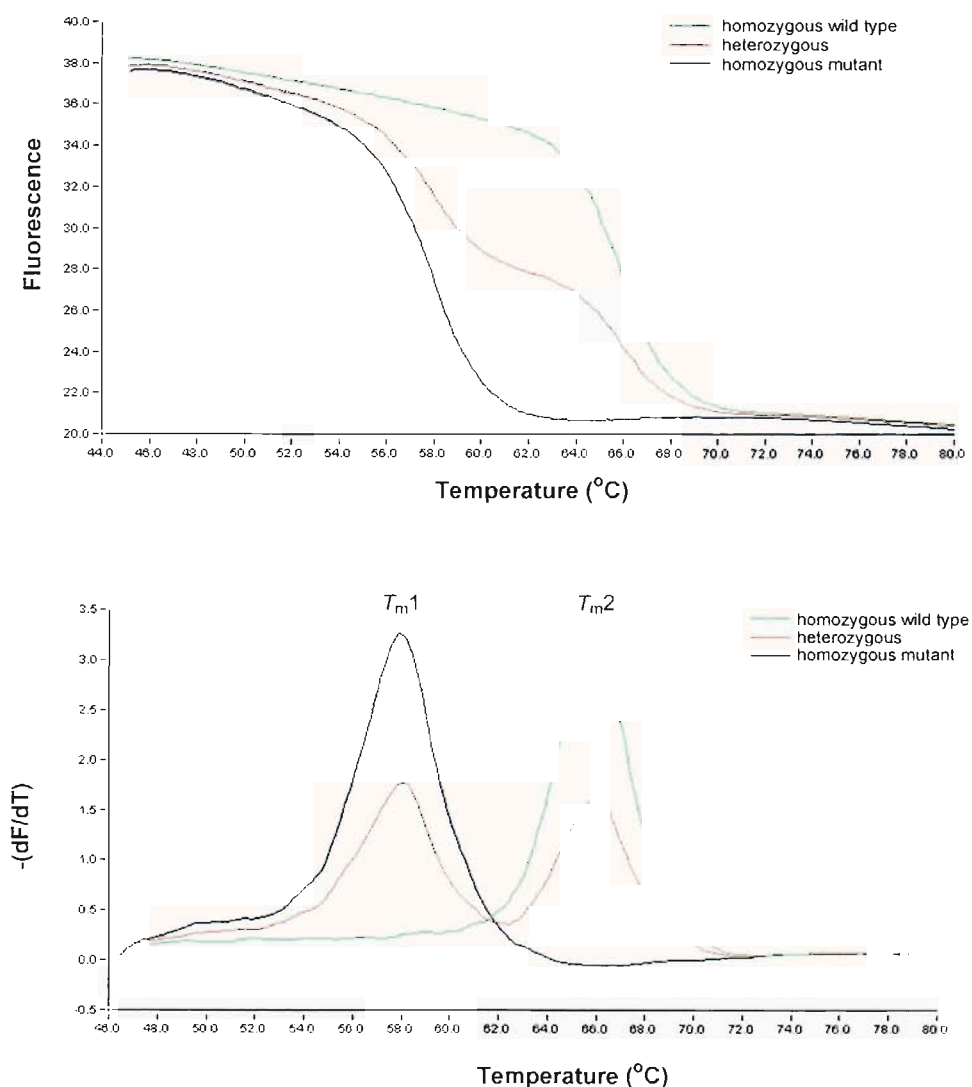
Figure 1.16 Melting curves and melting peaks determined from post-PCR melting curves.

18

The development of real time PCR has been made possible largely by the introduction of two real time thermal cyclers with on-line fluorescence monitoring: in 1997 the ABI 7700, a conventional blockcycler, and not long after, the Roche LightCycler, a carousel instrument that uses glass capillaries as reaction vessels. Although the sample capacity of the LightCycler (32 capillaries) is less than that of the 7700 (96 wells), the more efficient heat transfer through the thin capillary walls allows the use of rapid cycling conditions. A PCR of 40 cycles can be completed in 15-20 min, compensating for the lower sample capacity. The number of machines on the market has increased, with models from BioRad (iCycler), MJ Research (DNA Engine Opticon Continuous Fluorescence Detection System), Stratagene (Mx400), ThermoHybaid (Chimaera Quantitative Detection System), Corbett Research (Rotor-Gene 3000) and Cepheid (SmartCycler) broadening the price range at which real time thermal cyclers are available, and offering different features and interfaces. In 2002, sales of thermal cyclers in the United States and Europe were worth $430 million.[64]

Apart from suitable thermal cyclers, the other technology required for applications of real time PCR are suitable fluorescence-based detection methods. The problem of how to detect nucleic acids by accumulation of a fluorescent signal can conceptually be tackled in an immense variety of ways. This has led to a proliferation of DNA diagnostic methodologies, which are described below, concentrating on those most commonly used and those relevant to the work described in later chapters.

1.3.2.1 Fluorescent intercalator binding
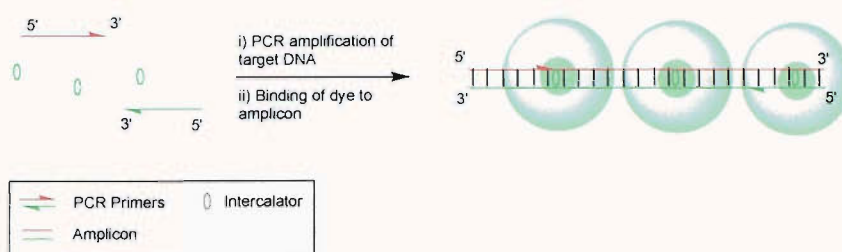
Figure 1.17 Signal generation by binding of fluorescent intercalator to amplicon.

Perhaps the simplest way to generate a signal upon production of PCR product is by the inclusion of a fluorescent intercalator in the reaction mixture.[69] Intercalators such as SYBR Green I,[70] ethidium bromide[71] and YO-PRO-1[72] bind to dsDNA with a dramatic increase in quantum yield, allowing them to be used as a detector for accumulation of amplicon during PCR (Figure 1.17). SYBR Green I is the most commonly used dye in real time PCR, because of its minimal binding to dsDNA, high quantum yield and the similarity of its spectral properties to FAM, convenient for real time thermocyclers which are set up to excite fluorescein. This approach has been used for quantification with high sensitivity.[73] While this method is the cheapest and simplest signalling chemistry available, it does have major disadvantages. As the intercalators bind indiscriminately to any dsDNA, the amplicon can be of any sequence, meaning that the assay is liable to score false positives for samples where primer-dimers or mispriming sites have been amplified. Also, there is no facility for allelic discrimination unless ARMS primers are used.

Fluorescence-based melting profile analysis (fluorescence melting) of the whole amplicon is a possibility though, which can demonstrate the presence of non-specific products, which generally have a lower T_m than the correct amplicon.[74] Others have suggested that the amplicon T_m can also be used to genotype variations such as small insertions, deletions and even SNPs.[75, 76] Short amplicons are required to enhance sensitivity to these small variations in sequence. This is likely to be of limited reliability and applicability, as amplicon T_m will depend not only on DNA sequence, length and G/C content, which are constant, but amplicon concentration and buffering conditions, which may vary from sample to sample.[77] A more general

approach has been described based on the use of GC-tailed allele-specific primers to exaggerate the difference in amplicon T_m between genotypes (Figure 1.18).[78]
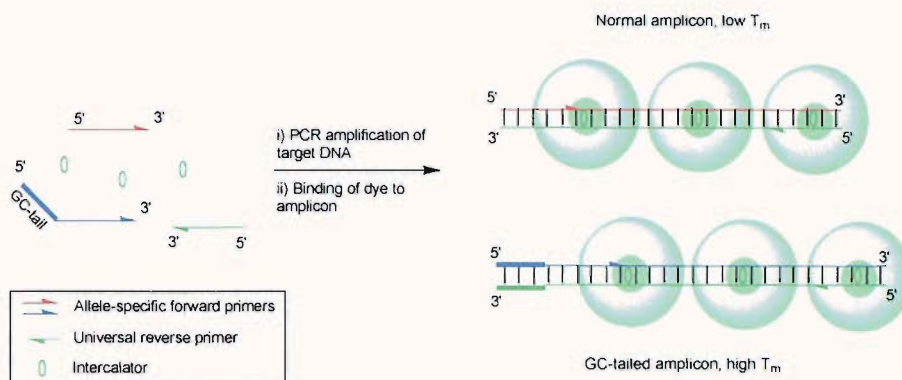
Figure 1.18 Tailed-primer approach to amplicon melting.

1.3.2.2 Primer-probes

Just as fluorescent intercalators can be used to generate a signal with accumulation of PCR product, so can labelled primers, called primer-probes. They are labelled so that extension of the primer leads to enhanced fluorescence.

An early example of this methodology was Sunrise primers, later called Amplifluor primers.[79] The Amplifluor primer is designed so it can form a stable hairpin-loop, at each end of which a fluorophore and quencher are attached. Extension of the Amplifluor primer leads to production of one strand of the amplicon. When this strand is copied, the polymerase 'reads-through' the hairpin loop, leading to the other strand of the amplicon being complementary to the whole of the labelled strand. When the two hybridise, the hairpin loop structure is abolished, leading to separation of fluorophore and quencher, and the release of a fluorescent signal. Later, the assay was refined by the use of tailed unlabelled primers, containing a universal sequence, the 'A sequence'.[80] If PCR occurs, this universal sequence is incorporated into the amplicon. The universal sequence is complementary to the primer portion of the Amplifluor primer, which will bind to the amplicon, be extended, copied and generate fluorescence (Figure 1.19). The primer-probe used is called the Amplifluor UniPrimer I, as the same labelled primer can be used in any PCR, as long as the tailed sequence is incorporated into the locus-specific primers.

This has obvious advantages in lowering the cost of the assay. Allele-specific (ARMS) primers can be used to distinguish point mutations.
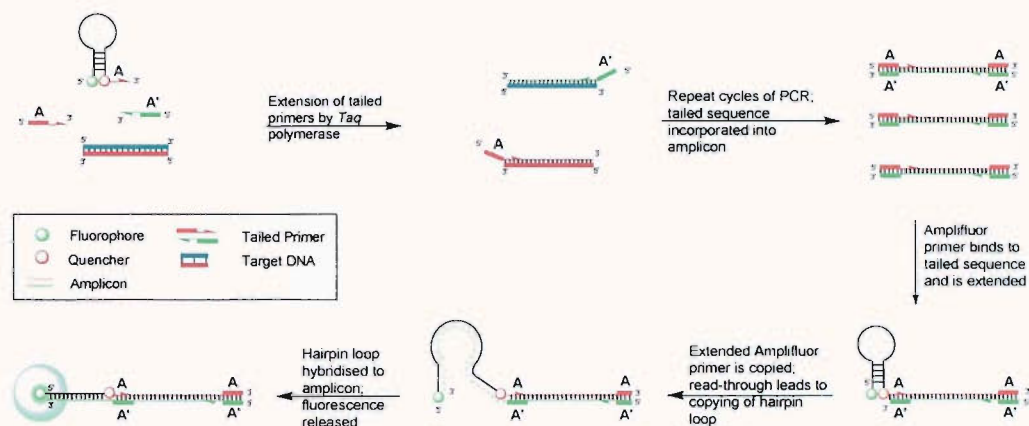
Figure 1.19 Mode of action of Amplifluor UniPrimers.[80]

Later still, it was found that primers with blunt-ended hairpin-loop structures and a dye attached at the C5 position of thymidine could enhance their fluorescence upon incorporation into PCR products (Figure 1.20A.). The design rules, established in a biophysical study, are that the terminal base pair of the hairpin-loop must be CG or GC and that the T bearing the fluorophore must be 2 or 3 nucleotides from the 3`-end.[81] The application of these hairpin primers to real-time quantitation and end-point SNP detection has been demonstrated.[82]
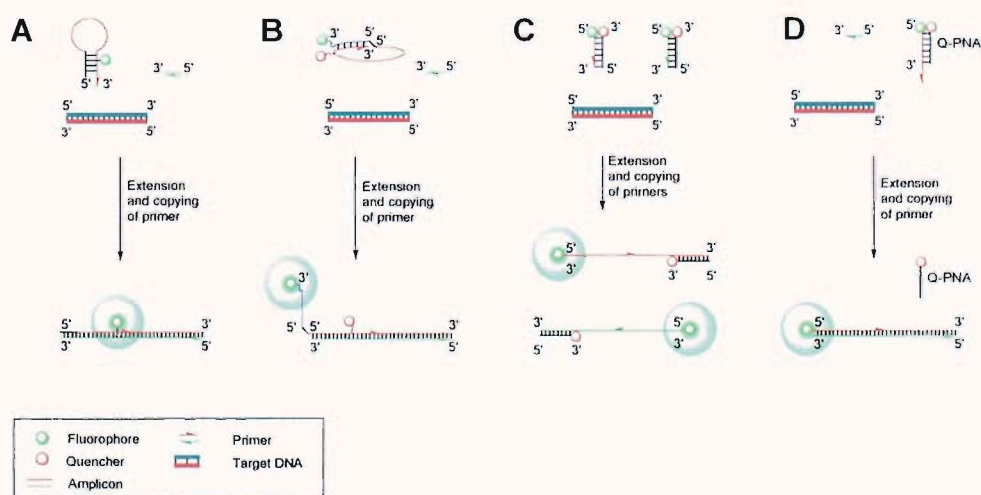
Figure 1.20 Hairpin primers labelled with a single fluorophore (A),[82] Cyclicons (B),[83] double-stranded primers (C)[84] and self-reporting PNA/DNA primers (D).[85]

22

Cyclicons,[83] 5'-5' linked pseudocyclic oligonucleotides bearing a fluorophore and quencher (Figure 1.20B), and double-stranded primers, with dye-labelled primers and quenchers attached to either a complementary oligonucleotide (Figure 1.20C)[84] or PNA oligomer (Figure 1.20D),[85] can be used in similar assays.

While all primer-probes can generate false positives as a result of primer-dimer formation and mispriming, the chance of this is reduced as all the primer-probes have secondary structure, making them natural hot-start primers. Unlabelled primers with hairpin-loop[86] and duplex[87] structures have already been shown to enhance the specificity of PCR. In any case, PCR must be optimised for all real time assays, because if reagents and primers are used up in synthesising spurious amplicons then loss of sensitivity and false negatives can be observed. The UniPrimer strategy loses this natural hot-start specificity, but is attractive, as only one labelled oligonucleotide has to be synthesised for any number of loci. This approach could be applied to any primer-probe format.

1.3.2.3 Signal generation by probe cleavage

Primer-probes can generate a signal upon incorporation into PCR products, but are vulnerable to the production of spurious amplicons resulting from primer-dimers or mispriming. Probes that become fluorescent during the PCR due to hybridisation with the polymorphic site are potentially more specific. Cleavage of the probe was one of the first methods presented to generate a fluorescent signal – the probe is quenched prior to hybridisation, but hybridisation leads to enzymatic cleavage of the probe, whereupon fluorescence is released.

The TaqMan® assay exploits the 5-nuclease activity of polymerase enzymes.[88] The TaqMan® probe is an oligonucleotide, labelled at either end with dyes,[53] which is complementary to some region of the section of DNA flanked by the primers. As the primer is extended by the polymerase, replication is obstructed by the TaqMan® probe. In this conformation the fluorophore is quenched, *via* FRET to an acceptor dye at the other of the oligonucleotide. The polymerase exploits its inherent 5' to 3' exonuclease activity at the point where replication is obstructed and hydrolytically cleaves the probe. Since the donor and acceptor dyes are no longer in the same molecule, the interfluorophore distance is now dramatically increased, and FRET is

essentially eliminated (Figure 1.21). This leads to an accumulation of the fluorescent signal observed from the donor over the cycles of the PCR, until digestion reaches a plateau.[89] TaqMan® has been used extensively as a tool in genetic analysis, detecting the Hepatitis C virus,[90] *Salmonella* in raw meat[91] and cancer susceptibility in humans.[92]
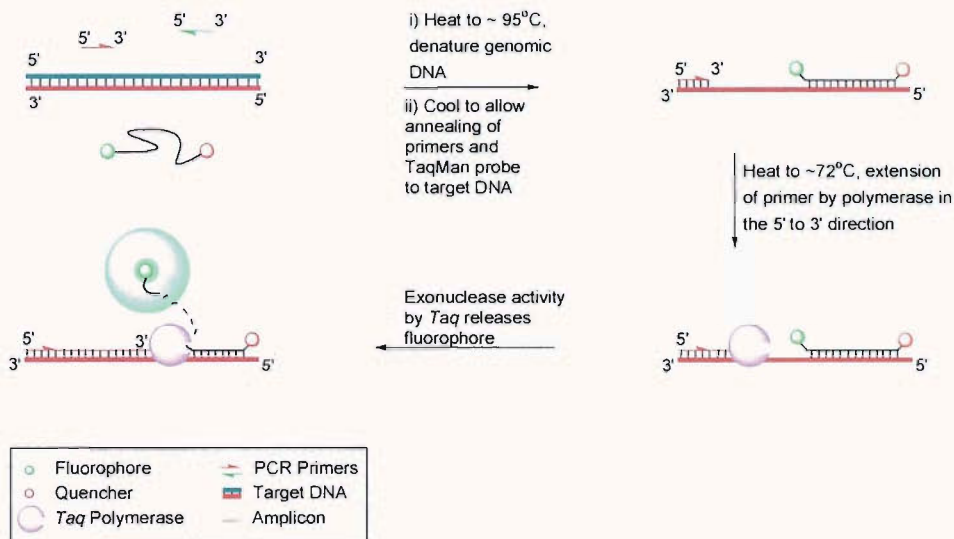
Figure 1.21 A schematic representation of the TaqMan® assay.[89]

While the TaqMan® technology is well established, several improvements to the technique have been presented. Terbium chelates can be used as the reporter, leading to an increase in sensitivity. Low background fluorescence from these luminescent labels can be achieved by internal quenching by the probe,[93] or by the use of a complementary oligonucleotide bearing a quencher.[94] An intramolecular version of the TaqMan® has been described (Figure 1.21A.),[95] as has the use of a minor groove binder (MGB) attached to the 3'-end of the TaqMan probe®.[96] The incorporation of the groove binder significantly increases the T_m of the TaqMan®-MGB probe/target duplex so that shorter probes can be used. TaqMan®-MGB probes are so commercially important that Applied Biosystems recently introduced 'Assays-On-Demand' – pre-formulated kits containing probes and primers for >146,000 SNPs.[97]

24
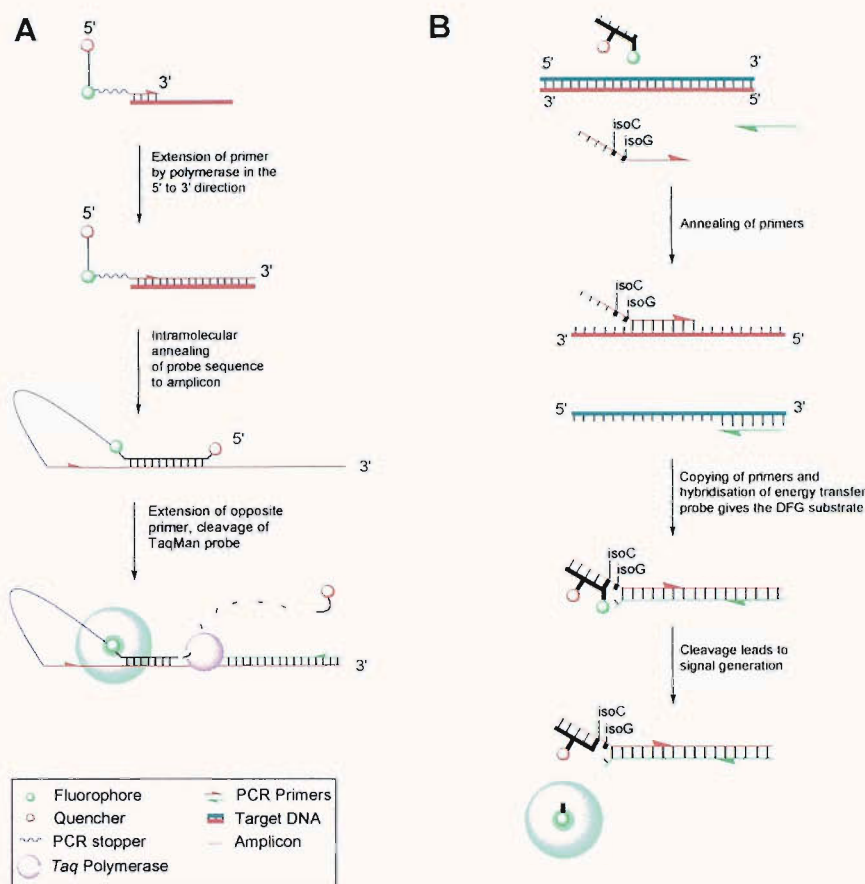
Figure 1.22 Intramolecular TaqMan® (A),[95] and AEGIS probes (B).[98]

A universal probe based on an expanded genetic information system (AEGIS) has been described recently. This uses the unnatural nucleosides isoC (iC) and isoG (iG),[99] and the double-flap gap (DFG) nucleolytic activity of *Taq* polymerase. One normal primer, and one tailed primer, containing consecutive iC and iG residues between the allele-specific primer and the universal tail, are used. During extension of the standard primer, one nucleotide is inserted opposite iG of the tailed primer then replication stops at iC. This creates a duplex with a 5'-single stranded overhang and a 3'-flap (single unpaired base). Hybridisation of the universal fluorescently labelled probe to the 5'-overhang gives the double-flap gap structure. This is recognised by the polymerase and is cleaved, increasing fluorescence of the cleaved fluorophore (Figure 1.22B).[98] While the universal probe is not extended and is therefore not a primer-probe, the assay is still sensitive to mispriming and primer-dimer formation, as production of any amplicon will lead to the cleavage of the probe. As with Amplifluor UniPrimers, though, the requirement for only a small

25

number of fluorescently labelled probes to be synthesised for a large number of loci is appealing.

A major disadvantage with probes cleaved upon hybridisation is that the probe is necessarily destroyed to generate a signal, so no further information (e.g. T_m data) can be obtained post-PCR.

1.3.2.4 Competitive hybridisation probes

Competitive hybridisation requires two dyes, one a fluorophore and the other a fluorescence quencher, which are held in close proximity by the formation of a duplex. The duplex may be inter- or intramolecular. PCR is used to amplify a fragment of DNA to be interrogated. A region of the amplicon is complementary to the dye-labelled probe, which is designed to preferentially bind to the amplicon rather than reform the quenched species. Upon hybridisation to the amplicon, the fluorophore and quencher become distal, and a fluorescent signal is obtained. When the probe oligonucleotide is not cleaved, competitive hybridisation offers the advantage that fluorescence melting curves can be obtained at the end of the PCR. Molecular Beacons are one example of competitive hybridisation probes. They are designed so that at the temperature where fluorescence is measured (~60 °C), they form a stable hairpin loop. The probe is therefore designed so that at either end there is a small region of DNA (~5-12 base pairs), which is self-complementary, and thus forms a double stranded stem. The fluorophore and quencher are covalently attached to either end of the stem. When the stem is formed, two dyes are held close together and fluorescence quenching occurs.
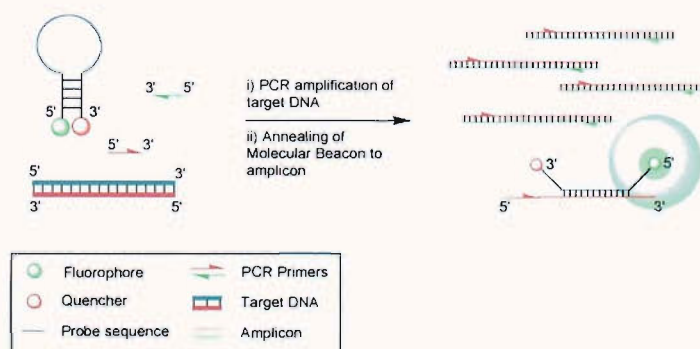
Figure 1.23 The mechanism of action of a Molecular Beacon.[100]

When a suitable amount of amplicon is present, the probe sequence may anneal to the target, in preference to formation of the hairpin loop. To aid this, the probe sequence is much longer than the stem (~25-30 nt), so that formation of the beacon/target hybrid is thermodynamically favoured over the reformation of the hairpin loop. Since in this open conformation the fluorophore and quencher do not lie in close proximity to one another, an increase in emission from the fluorophore occurs - indicating that the target DNA sequence is present in the amplicon (Figure 1.23). Molecular Beacons have found application in many nucleic acid detection assays – DNA in genotyping of human alleles[101] and drug resistance in *Mycobacterium Tuberculosis*,[102] and RNA in detection of splice variants,[103] quantitation of viral loads[104] and detection of *Potato leafroll virus* and *Potato virus Y*.[105]

This assay suffers from its bimolecular nature (due to difficulties in ensuring formation of probe-target hybrids), as reformation of the stem is an intramolecular process, and thus is kinetically and entropically favoured over the intermolecular formation of the probe/target hybrid.[106] Since the fluorophore and quencher remain in the same molecule, residual quenching in the open form is also observed. Various modifications have been made to Molecular Beacons to improve their sensitivity. Molecular Beacons constructed from PNA, a nucleic acid with an uncharged peptide backbone that has a greater affinity for DNA than DNA or RNA, have been used to enhance hybridisation.[107] Shared stem Molecular Beacons, where one half of the stem hybridises to the target, have been used to improve hybridisation and disfavour stem reformation (Figure 1.24A.),[108] and to make energy transfer Molecular Beacons, where a donor and acceptor beacon hybridise adjacent to each other (Figure 1.24B.).[109]
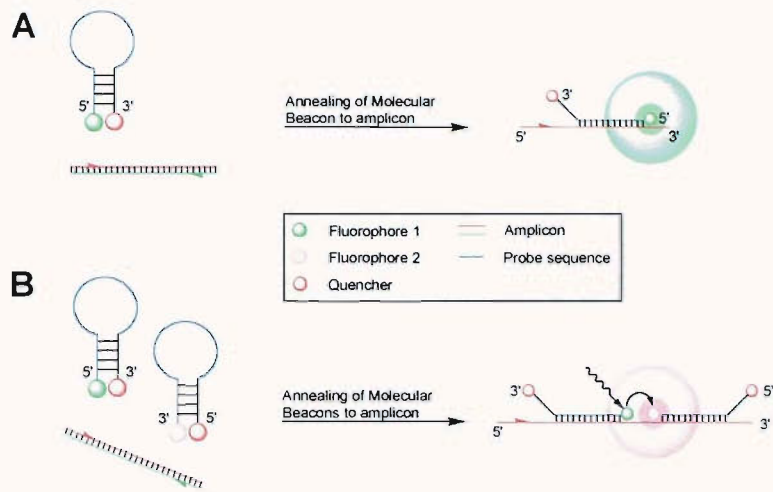
Figure 1.24 Shared stem Molecular Beacons (A),[108] and shared stem energy transfer Molecular Beacons (B).[109]

A Scorpion is the name given to a detection system that combines a probe sequence and a primer in a single oligonucleotide.[110] Scorpions are not considered to be primer-probes due to the inclusion of a PCR stopper, which avoids copying of the stem upon extension of the reverse primer. This results in specific probing of the amplicon, as distinct from primer-probes, where accumulation of signal is non strand-specific.
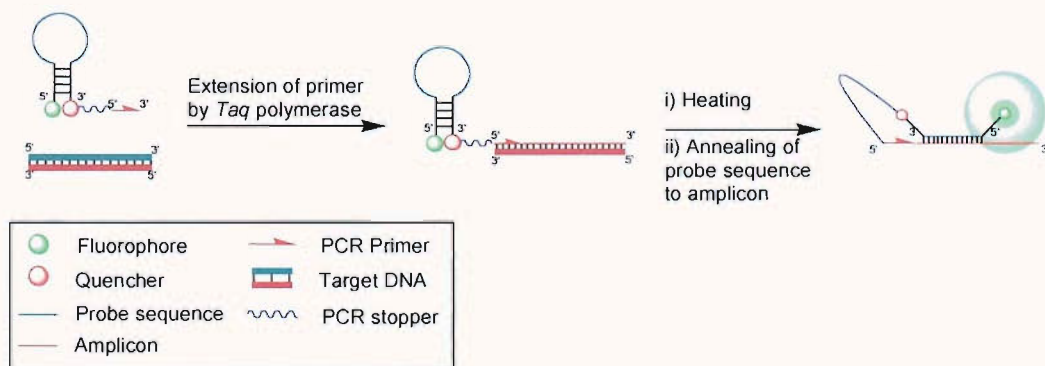
Figure 1.25 Generation of a fluorescent signal from one Scorpions Primer.[111]

The Scorpion incorporates a stem-loop section, as in the Molecular Beacon. When the Scorpion primer element is extended, the result is a strand of DNA, a section of which is complementary to the probe sequence. This sequence is longer than that of the stem and hence its binding is favoured over reformation of the stem, so the

28

Scorpion undergoes a conformational reorganisation, whereupon fluorescence is released (Figure 1.25).

Since the probe and target sequences are now in the same molecule, issues regarding unimolecular vs. bimolecular kinetics are relieved. This results in the Scorpions Primer having advantages over other assays. The speed of the unimolecular probing event enables rapid signal appearance and high signal to noise ratios, because probe-target binding is kinetically favoured over duplex reannealing and stem reformation, and the assay does not rely on an enzymatic cleavage for signal generation. ARMS can be used to effect allelic discrimination,[112] but the polymorphic sequence can also be located in the probe sequence, enabling thermodynamic discrimination, as in Molecular Beacons.[111] The disadvantages associated with Scorpions are that the probes are long and complex (~65 nt), and rely on secondary structure for their operation. Both of these factors introduce complications into the design and synthesis of Scorpions. Despite these, Scorpions have been used to effect quantitation of HIV-1 DNA/RNA,[113] and mRNAs linked to breast cancer[114] and muscular dystrophy.[103]
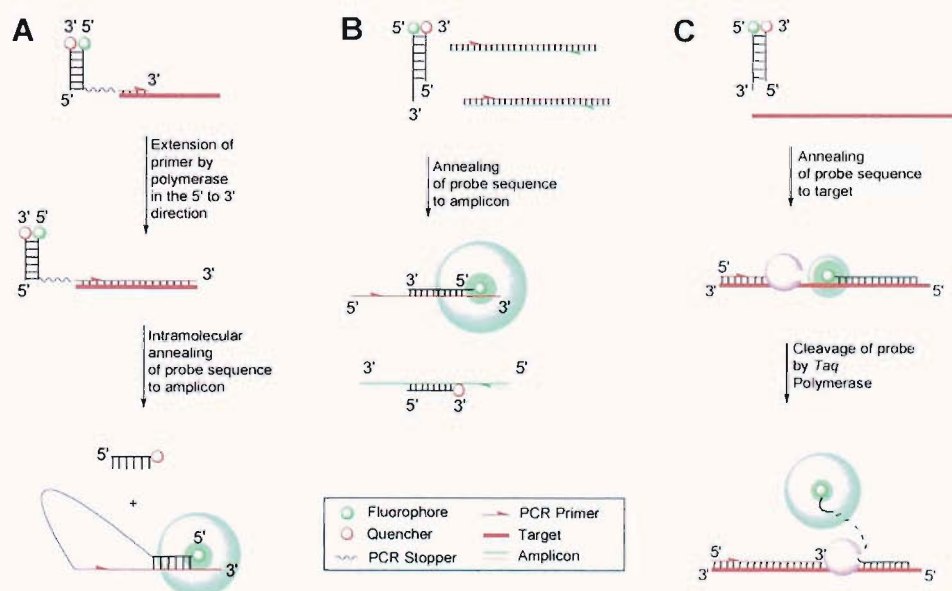
Figure 1.26 The Duplex Scorpion (A),[115] 'Yin-Yang' (B)[84] and Duplex Probe (C)[116] formats.

To overcome some of the difficulties associated with Scorpions, Duplex Scorpions have been developed, where the stem consists of the probe element and a separate

29

oligonucleotide, complementary to the probe, to which a quencher is attached. Upon extension, the probe binds to its complement in the amplicon, as before, but now this event has only to compete with the reformation of an intermolecular stem. Since the quencher oligonucleotide is expelled into solution, residual quenching is minimised, leading to a greater signal (Figure 1.26A). This advance greatly simplifies Scorpion synthesis – the probe requires the incorporation of just one dye, and the total length of the oligonucleotide to be synthesised is reduced to 35-50 nt.[115]

Subsequent to the development of Duplex Scorpions, related detection systems 'Yin-Yang'[84] and Duplex probes have been reported. 'Yin-Yang' probes differ from Duplex Scorpions only in that the tailed primer has been eliminated, so hybridisation is an intermolecular process. The quencher oligonucleotide is shorter than the dye-labelled probing strand, so hybridisation can occur spontaneously, facilitating its use in isothermal amplification protocols. Duplex probes are similar to 'Yin-Yang' probes, except that the two probing oligonucleotides are of similar length and the cycling conditions are designed to promote exonucleolytic cleavage of the probe.[116]

Competitive hybridisation is therefore a powerful technique, less susceptible to spurious amplicons than primer-probes and not requiring enzymatic cleavage for signal generation. However, purification of oligonucleotides with secondary structure (as in Molecular Beacons and Scorpions) is difficult, although 'two-oligo' versions of these have been described which alleviate this problem. Sensitivity can be limited by the nature of competitive hybridisation, as binding to the target is disfavoured by the presence of a complementary sequence elsewhere in the system. This has a hidden benefit though - specificity of hybridisation by hairpin oligonucleotides is enhanced over their linear counterparts,[117] and this is likely to also be the case for duplex structures.

1.3.2.5 Linear probes that fluoresce upon hybridisation

While primer-probes, cleaved oligonucleotides and competitive hybridisation probes have been used extensively in real time PCR, linear probes that fluoresce upon hybridisation have potential advantages over all of these techniques. Non-competitive hybridisation offers the potential for shorter and more specific oligonucleotides to be used, and non-destructive signal generation allows extra information to be obtained from post-PCR melting curves. A PCR stopper, such as phosphate or octanediol, is required at the 3'-end to prevent priming by the probe and possible accumulation of non-specific signal.

Labelling of a linear oligonucleotide with a single fluorophore is a conceptually attractive approach, since the probe can be synthesised and purified cheaply and easily. Fluorescence polarisation increases upon hybridisation due to the hybridised high molecular weight species having restricted molecular motion, and this phenomenon has been used for genotyping.[118] This technique requires more sophisticated equipment, which is not as widely available as thermocyclers that measure fluorescence intensity. However, the intensity change upon hybridisation is unpredictable, with varying degrees of quenching or fluorescence enhancement observed, depending on sequence and dye-incorporation chemistry. HyBeacons (Figure 1.27A.)[119] exploit modes of dye-incorporation that reliably lead to fluorescence enhancement on hybridisation. Fluorescein is positioned in the major or minor groove by conjugation to either the base[119] or 2'-position[120] of uridine analogues. HyBeacons have been used to detect and discriminate polymorphic targets amplified directly from saliva.[121]

Fluorescence quenching by nucleobases is a known phenomenon, with Guanine particularly effective due to its good electron-donating properties.[122] Two fluorescent probing technologies use this to achieve reliable quenching or dequenching upon hybridisation. Fluorescein quenching can be achieved by positioning two contiguous Cs next to the fluorophore in the probe. Hybridisation leads to fluorescein being in close proximity to two guanines, leading to a decrease in fluorescence (Figure 1.27B.). This methodology has been used in real time PCR to detect and discriminate polymorphic amplicons by both C_T and melting curve (T_m) analysis.[123] Target sequences with one G in the 5'-overhang position also allowed detectable quenching, enabling this method to be applicable to almost all

31

sequences. Inversion of the real time fluorescence trace is necessary for calculation of C_T, but this is a simple operation. Dequenching can also be accomplished, by placing two Gs next to the fluorophore in the probe. Internal quenching is relieved upon hybridisation, leading to signal generation (Figure 1.27C). Melting curve analysis was used to score a number of SNPs.[124] The requirement for two contiguous Gs in the probe is likely to limit the sequences for which this method can be used, though as with the fluorescein-quenching technique, the oligonucleotides are relatively cheap and easy to synthesise, requiring only widely available reagents.

FAM-dU monomer (major groove)   FAM-dU monomer (minor groove)   5'-FAM monomer
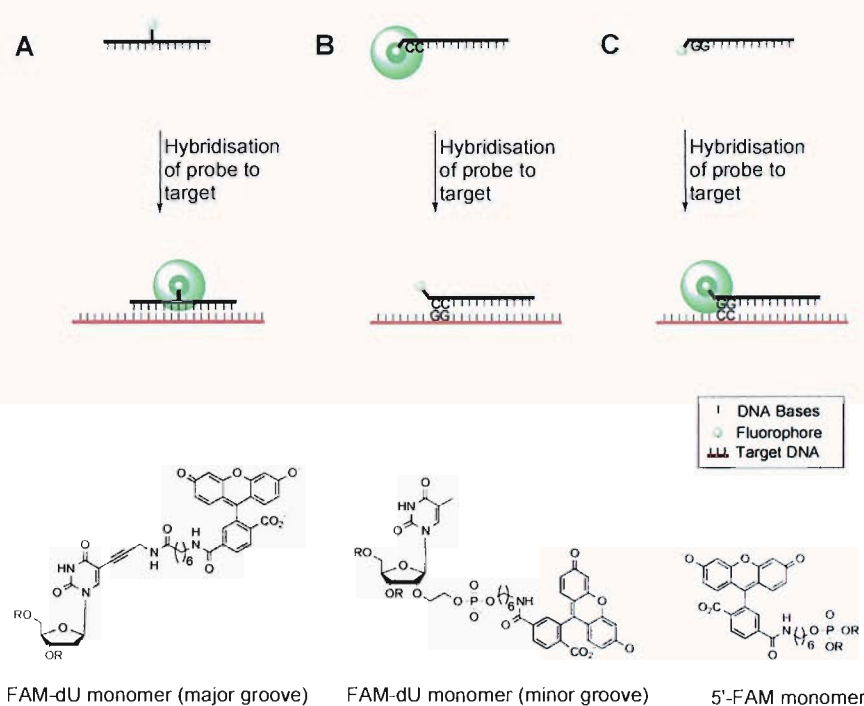
Figure 1.27 Mode of action of HyBeacons (A),[119, 120] fluorescein quenching probes (B)[123] and dequenching probes (C).[124]

FRET can be used to excite a linear probe labelled with a single dye. The donor dye must be brought close enough to the acceptor upon hybridisation to initiate energy transfer.[125]

An assay of this type using two labelled oligonucleotides called hybridisation probes has been developed. One probe is labelled with a fluorescent dye at its 3'-end (the donor dye), and the other is labelled with an acceptor dye at its 5'-end. The oligonucleotides are designed so that they hybridise to the amplified DNA fragment in a head to tail arrangement, and thus the dyes are conveniently placed for FRET to

occur (Figure 1.28). The fluorophores can be organic dyes such as fluorescein and LC Red 640,[126] but the use of luminescent Europium[127, 128] or Terbium[129] chelates as LRET (luminescence resonance energy transfer) donors has also been described.
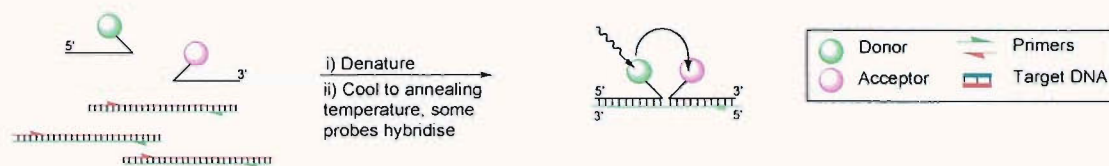
Figure 1.28 Mode of action of hybridisation probes.

The donor can be free in solution, but intercalate into the duplex formed by hybridisation of the probe to its target, causing FRET to occur, and an increase in fluorescence from the reporter, such as in ResonSense®[130] or iFRET[131] (Figure 1.29A). This has the advantage that just one labelled probe has to be synthesised. An intramolecular version of ResonSense®, Angler®, has also been reported (Figure 1.29C).[130]
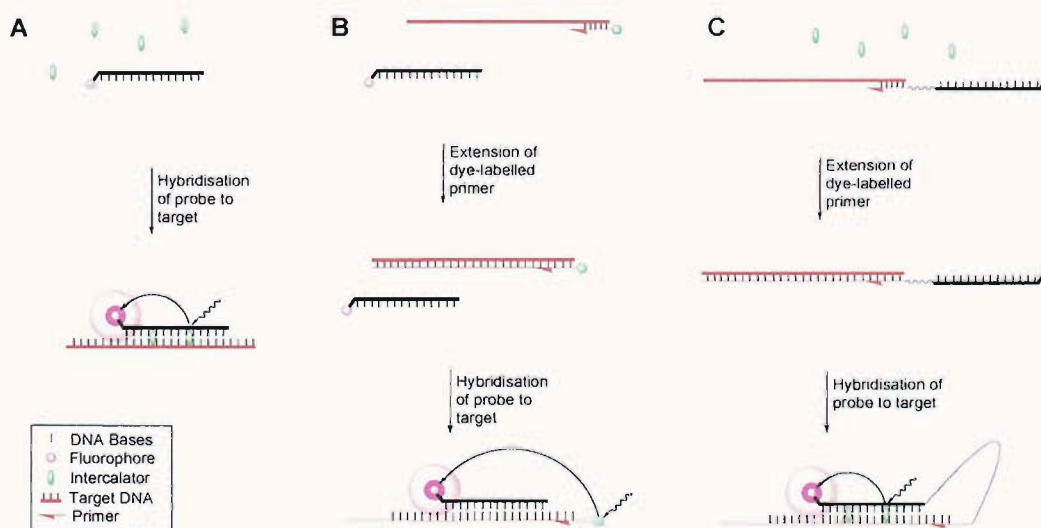
Figure 1.29 ResonSense®[130]/iFRET (A)[131] PriProEt (B)[132] and Angler® (C)[130] formats.

Similarly, if one primer is labelled with a donor fluorophore, an acceptor-labelled probe can be excited by hybridisation to the labelled strand of the amplicon. This

approach is used in PriProEt probes (Figure 1.29C), which have been used to detect the *Foot and Mouth disease virus*.[132]

More elaborate chemistry can also provide probes whose secondary structure does not change upon hybridisation, but whose fluorescence is enhanced. Intercalators often display enhanced fluorescence quantum yields upon binding to dsDNA, and many probes contain tethered intercalators in order to harness this effect.[133-136] This strategy is introduced more thoroughly in Section 2.1. Fluorescent nucleobase analogues methoxybenzodeazaadenine (^{MD}A) and methoxybenzodeazainosine (^{MD}I) become more brightly fluorescent when base paired in a duplex with T or C, but not when mispaired. These analogues have been incorporated into oligonucleotides and can be used to detect synthetic oligonucleotides.[137] Hairpin polyamides are a class of ligands that can bind to the minor groove of DNA duplexes in a sequence specific manner.[138] When labelled with TAMRA, the fluorescence of hairpin polyamides increase significantly when binding to a duplex containing their target sequence present in synthetic oligonucleotides.[139] Such probes are truly non-competitive, since disruption of the target duplex is not required for binding. Though the application of these assays requiring more complex probe synthesis are currently limited, they may prove more important as key reagents become more widely available.

1.3.3 Other methods for genetic analysis

1.3.3.1 DNA Microarrays

DNA microarrays are assemblies of oligonucleotides immobilised on a solid support. Each oligonucleotide is complementary to a DNA sequence of interest, e.g. a gene or splice variant. Hybridisation of a fluorescently labelled PCR product to the tethered probe indicates the presence of the sequence in the amplification mixture. Microarrays are used primarily in gene expression analysis. Since Schena *et al.*[140] described the use of a 45-member complementary DNA microarray to monitor gene expression in *Arabidopsis thaliana* in 1995, several companies have immobilised probes for each of the 30,000-40,000 genes in the human genome onto single chips.[141]

More recently, microarrays have been used for SNP genotyping applications. Hybridisation of allele-specific immobilised oligonucleotide probes to labelled PCR-

products has been used to this end, but stringent hybridisation conditions are required, and since not all matched and singly mismatched probe-target duplexes can be designed with the same T_ms, the scope for multiplexing is limited in this approach. Rather, extension of tethered allele-specific primers with fluorescent dNTPs is used to generate the signal.[142, 143]

1.3.3.2 Ligation

DNA ligases catalyse the formation of a phosphodiester bond between the 5'-phosphate and 3'-hydroxyl groups of two adjacently hybridised oligonucleotides. The requirement for the two base pairs either side of the reaction site to be fully matched makes ligation a good candidate as an allelic discrimination reaction. This led to the development of the oligonucleotide ligation assay (OLA).[144] The invention of the ligase chain reaction (LCR) facilitated its use, either as the sole means of amplification,[145] or in conjunction with PCR.[146]

Ligation of a labelled primer to an immobilised one[144] or ELISA-based methods[147] were initially used to assay the progress of the reaction, but as with real time PCR, homogenous methods where the reaction gives rise to a change in fluorescence signal are preferred. The most common method involves the use of two fluorescently labelled primers, one bearing a donor, the other a quencher. Ligation yields a product that can be excited at the donor's λ_{ex}, leading to emission by the acceptor via FRET (Figure 1.30A).[148-150]
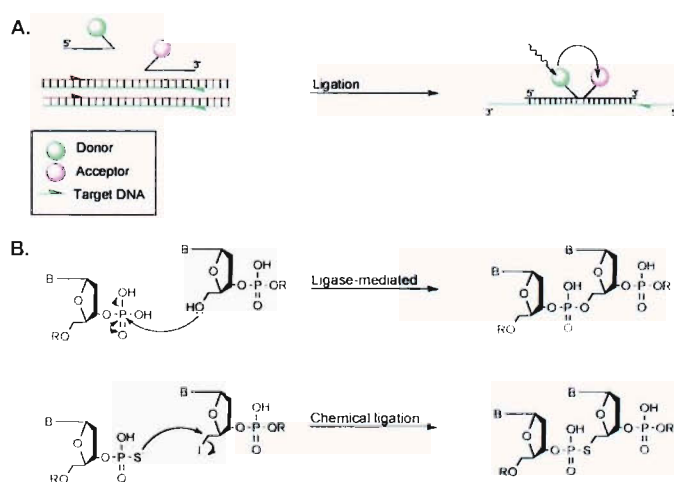
Figure 1.30 FRET-based detection of ligation reactions

The ligation primers can be a 5'-phosphorylated oligonucleotide and an unmodified oligodeoxynucleotide (in which case the reaction is catalysed by a ligase) or a nucleophilic 3'-thiophosphorylated oligonucleotide and a 5'-iodinated electrophilic oligonucleotide, which ligate spontaneously upon hybridisation. This is called 'chemical ligation' or 'autoligation' (Figure 1.30B). The autoligation has been found to have similar fidelity to the enzymatic reaction.[151] Sando and Kool have described an elegant adaptation of the autoligation, in which the quencher is expelled as the leaving group in the reaction, leading to a ~100-fold increase in fluorescence of the reporter (quenched autoligation, Figure 1.31).[152] Chemical ligations have the advantage that the template can be DNA or RNA, whereas enzymatic ligations only proceed with a DNA template. However, chemical ligations are slower than their enzymatic counterparts.
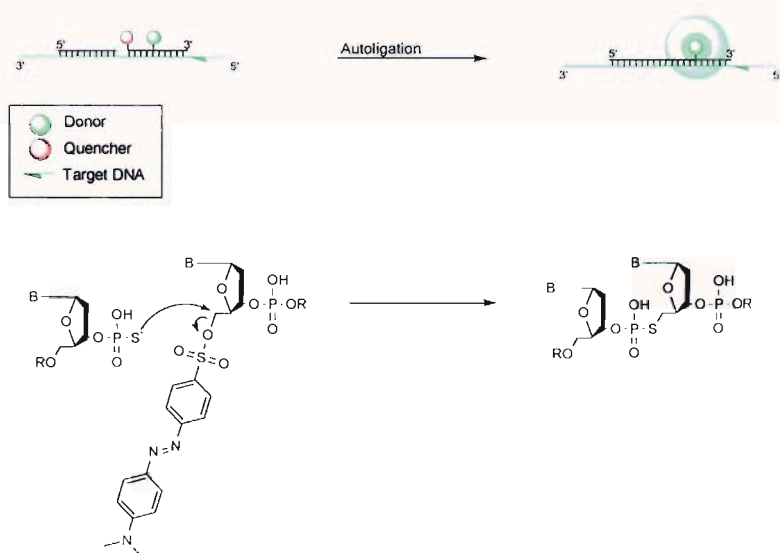
Figure 1.31 Quenched autoligation.[152]

Although in most genetic analysis methods post-discrimination electrophoretic resolution is deemed to be undesirable due to increased time and labour costs, these can be offset if the method is highly multiplexed, allowing for many alleles to be genotyped from one electropherogram. These factors were taken into account in the design of MLPA (multiplex ligation-dependent probe amplification).[153] Two MLPA probes are required for each allele: A 'short' synthetic oligonucleotide 40-50 nt in length comprising a universal primer sequence (~20 nt) and a hybridising sequence

(20-30 nt), and a 'long' phosphorylated cloned oligonucleotide (80-420 nt long) containing a universal primer sequence (~35 nt), a 'stuffer' sequence (20-370 nt) and a hybridising sequence (~25-45 nt). If the two hybridising sequences are ligated, two universal primers initiate PCR. The length of the stuffer sequence determines the electrophoretic mobility of the amplicons, which enables resolution by electrophoresis (Figure 1.32). Probe sets of up to 79 members have been used in this technique, allowing for highly multiplexed genotyping.[154] Trisomy mutations can be easily detected by the variation in peak height corresponding to their specific amplicon, and SNPs have been genotyped by the selectivity of the ligation reaction and the resulting varying amplicon length.[153] The ΔF508 deletion mutation in the ABCC7 gene was also detected by separating PCR products differing by just 3bp in length *via* high resolution capillary electrophoresis, suggesting scope for greater multiplexing. MLPA is most frequently used to determine gene dosage[155] or to detect large genomic deletions and duplications or rearrangements,[156-158] but has also been used for expression analysis.[159]
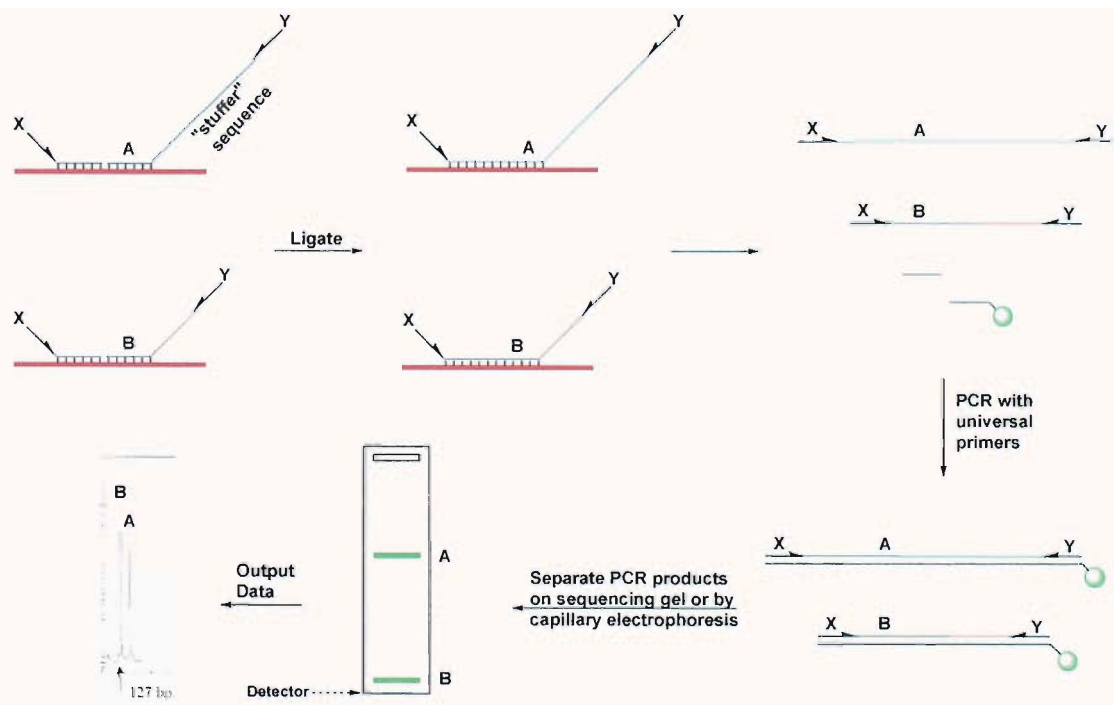
Figure 1.32 Principles of the MLPA technique.

Ligation can be made to circularise an oligonucleotide in which the two primers are separated by a linker, called a padlock probe (Figure 1.33).[160] The selectivity of the ligation reaction imparts the discrimination properties required for SNP genotyping. Hybridisation of padlock probes can be detected in a number of ways. If denaturing gel electrophoresis is used, the presence of the circularised probe is shown by its characteristic low mobility. The circularised padlock probes are insensitive to digestion by alkaline phosphatase or exonuclease VII, meaning that only ligated probes will retain their radiolabel. Under non-denaturing conditions, hybridisation to plasmid DNA containing the polymorphic site can be detected by migration of the labelled probe with the plasmid on an agarose gel, or retention of the label on a nylon membrane despite stringent washing. This has enabled padlock probes to be used in SNP genotyping.[161] In this form the method suffers from its requirement for radiolabelling and a post-ligation detection step. The ligation reaction only allows for modest amplification and lacks the sensitivity required for some applications. These issues have been addressed by coupling the ligation to an isothermal rolling circle amplification (RCA) step, which is discussed in 1.3.3.3.
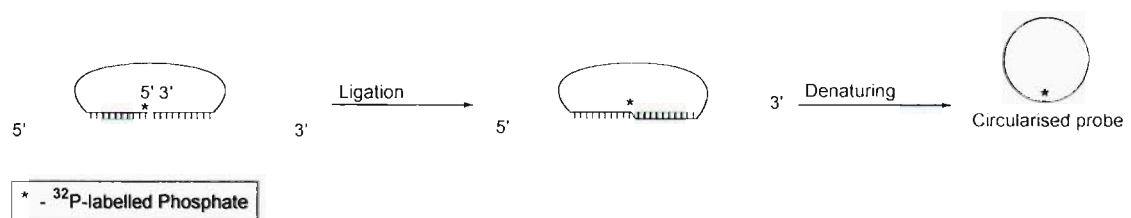
Figure 1.33 Circularisation of padlock probes by ligation.

1.3.3.3 Isothermal amplification

The signal obtained from padlock probes can be amplified by the use of rolling circle amplification, a type of DNA synthesis catalysed by strand displacing polymerases. The strand displacement obviates the need for thermal cycling. A circular DNA template is repeatedly copied to form a concatamer of its complement. The reaction therefore requires the padlock probe to be circularised before RCA can commence, so allelic discrimination results from the specificity of ligation. The use of two primers, the first complementary to the circularised padlock probe and the

second complementary to the RCA product are used. This gives rise to a hyperbranched cascade rolling circle amplification (H-CRCA) (Figure 1.34). The linear products can be detected as a characteristic ladder of bands separated by x nucleotides (where x is the length of the original padlock probe) by electrophoresis, or if the second primer is a primer-probe (e.g. an Amplifluor primer, Figure 1.19), endpoint fluorescence can be used.
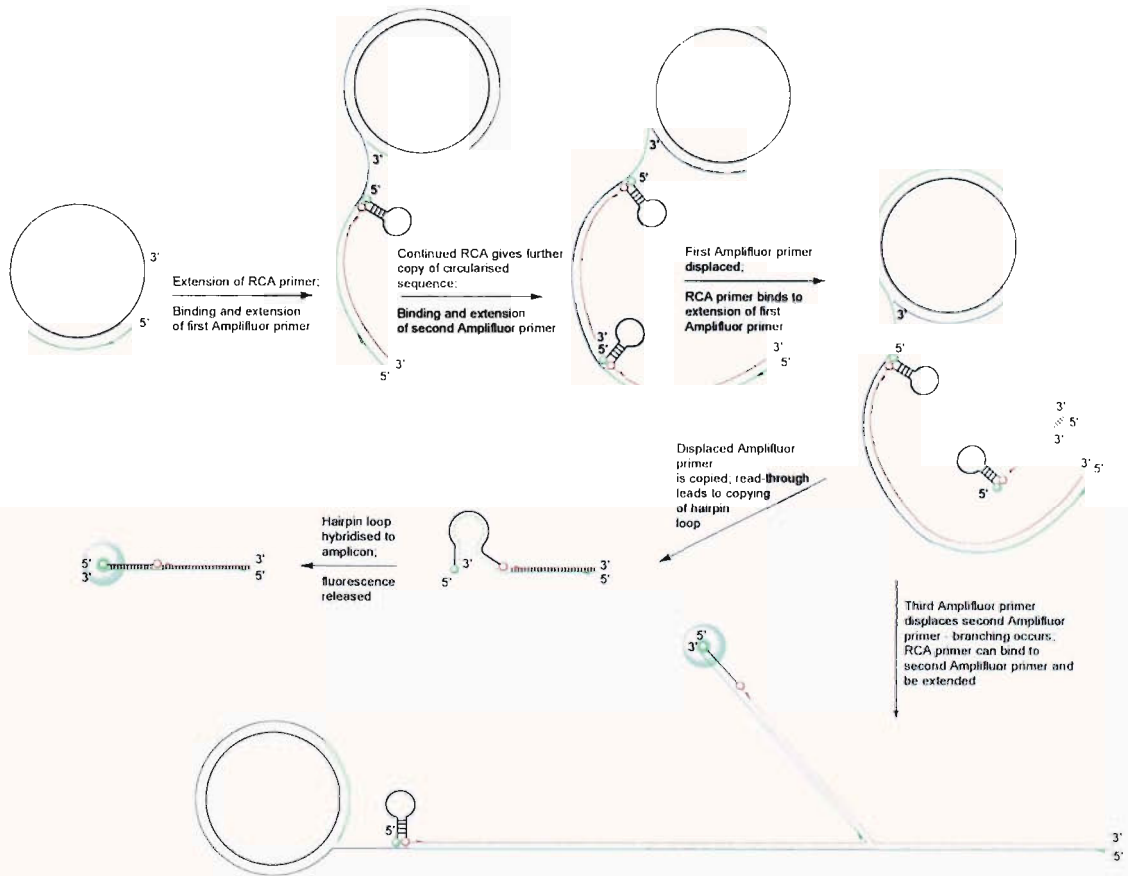
Figure 1.34 Early cycles of an H-CRCA with Amplifluor detection.

The Invader® assay uses the flap endonuclease activity of thermostable Cleavase® enzymes isolated from archaebacteria to perform allelic discrimination and generate a signal.[162] If the base pair at the flap junction is matched, the flap is cleaved. A fluorescent signal can be generated if the released flap hybridises to a FRET cassette forming a flap, which is then cleaved to release the fluorophore (Figure 1.35).[163] This method is most suitable for SNP detection.[164, 165]
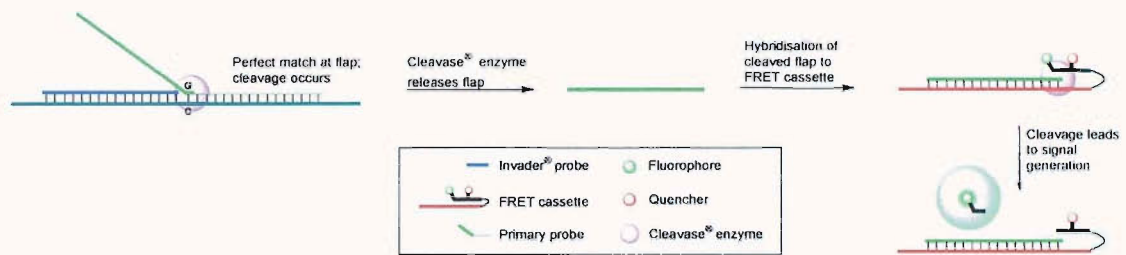
Figure 1.35 Invader® assay with fluorescence detection.[163]

Conclusions and project aims

It is clear that methods for sequence-specific detection of nucleic acid sequences (genetic analysis) are of great importance in medicine and biotechnology. Many such detection platforms using fluorescently-labelled oligonucleotides have been presented, falling into the two broad categories of homogeneous and heterogeneous assays. The work presented in the following chapters concentrates on homogeneous methods, which have the potential to provide the analyst with more information e.g. template copy number can be obtained from real-time PCR, and fluorescence melting curves can clarify genotyping if analysis of real-time traces is ambiguous.

Although many homogeneous techniques are available for genotyping, at the commencement of this project, a satisfactory approach based on the use of a single stranded probe, whose fluorescence is altered by hybridisation, but which is not enzymatically cleaved during the genetic analysis reaction (e.g. PCR) had not been described. This type of probe should present simple binding kinetics, and, since it is not destroyed, allow collection of fluorescence melting data after PCR. The initial aim of this project was to develop a novel fluorescence signalling methodology, which facilitated the design of probes meeting these criteria. Two such methods,

based on fluorescence quenching by DNA-intercalators, are presented in Chapters 2 and 3.

Homogeneous genetic analysis techniques typically rely on the formation of a duplex between a specific oligonucleotide probe and its complementary sequence in a PCR product. Despite this, little attention had been paid to this fundamental process. The efficiency of probe hybridisation to amplicons is studied in Chapter 4, and methods to enhance it are evaluated.

Synthetic oligonucleotides of defined sequence are used in all genetic analysis technologies. The recently developed MLPA assay requires the use of long (>100 nt) oligonucleotide probes. These are difficult to obtain from automated DNA synthesis, since failure sequences are difficult to remove by reversed-phase HPLC (RP HPLC). Two novel hydrophobic phosphoramidite monomers were synthesised in order to facilitate purification by RP HPLC. Their use in the synthesis of long oligonucleotides is evaluated in Chapter 5.

2. DNA Intercalators as fluorescence quenchers in genetic analysis – covalent approach

2.1 Intercalators as labels in genetic analysis

Intercalators possess extended aromatic heterocyclic or carbocyclic rings and can be attractive labels in fluorescence assays, often being fluorophores or possessing strong chromophores. In addition to this, the intercalation process enhances the quantum yield of certain fluorescent intercalators, by perturbation of their electronic structure. This property has been exploited in light-up probes, PNA oligomers functionalised with the asymmetric cyanine dye, thiazole orange (Figure 2.1).[133, 166, 167]
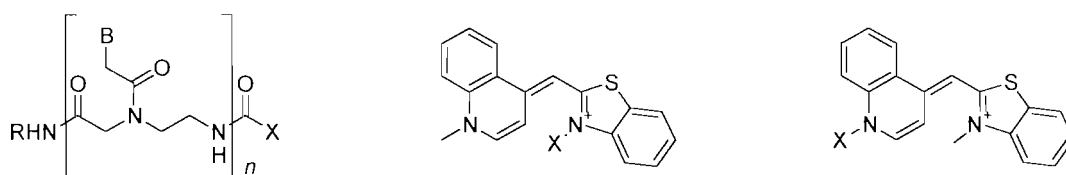
Figure 2.1 Chemical structure of light-up probes. B denotes nucleobases and R is H, lysinyl amide, or lys-lysinyl amide. X denotes the linker, which is either $(CH_2)_5NHCO(CH_2)_5$, or $(CH_2)_y$ where y is 5 or 10. The linker is attached either to the quinoline or benzothiazole nitrogen of thiazole orange.

Thiazole orange in free solution binds to calf thymus DNA with an enhancement of quantum yield from 2×10^{-4} to 0.11, a 550-fold increase.[168] To minimise electrostatic interaction of the cationic dye with the phosphates of oligonucleotides the probe is assembled from PNA, leading to an increase of quantum yield by 2.2-6.0-fold upon hybridisation.[166] Light-up probes have been used to detect PCR products from the human β-actin gene and the *invA* gene of *Salmonella*.[167]

Yamana and co-workers have described a series of intercalator-labelled oligonucleotides to detect DNA and RNA sequences.[134, 135, 169-173] The fluorophores pyrene and anthracene were attached the 2'- oxygen of RNA probes by a methylene bridge (Figures 2.2A and 2.2B). Quenching of pyrene fluorescence by the nucleobases is relieved upon formation of an RNA/RNA duplex, leading to an increase in quantum yield of 14-33.9-fold.[171] Importantly, the phenomenon does not

occur on hybridisation of an RNA probe to DNA sequences, or the hybridisation of a DNA probe to DNA or RNA sequences. This means that probes labelled in this way cannot be used in PCR-based genetic analysis techniques. Anthracene-labelled oligodeoxynucleotides also showed an increase in fluorescence upon duplex formation with RNA but not DNA.[173]
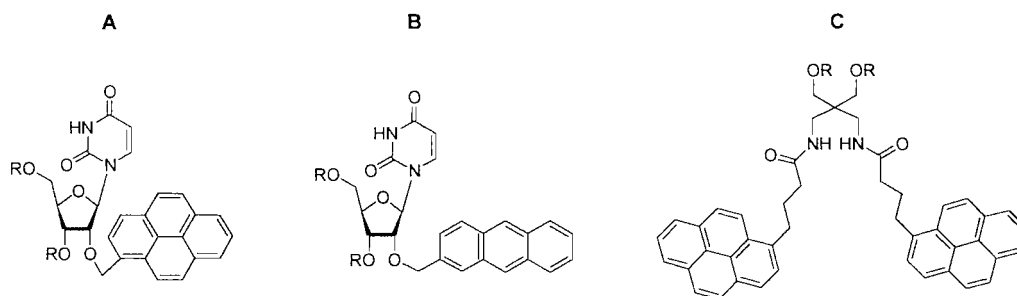
Figure 2.2 Intercalator-labelled probes described by Yamana *et al.*[134, 169, 171-174]

Bis-pyrene modified oligonucleotides labelled at the 5'-terminus (Figure 2.2C) do show enhanced fluorescence upon hybridisation to the complementary DNA sequence. The authors attributed this to excimer formation upon duplex formation, which is interrupted by one or both pyrene residues interacting with nucleobases in the single strand.[135]

Intercalators can also be used in combination with conventional fluorophores. Shinozuka *et al.* post-synthetically modified DNA probes with an acridine-fluorescein pair, as shown in Scheme 2.1.[175]
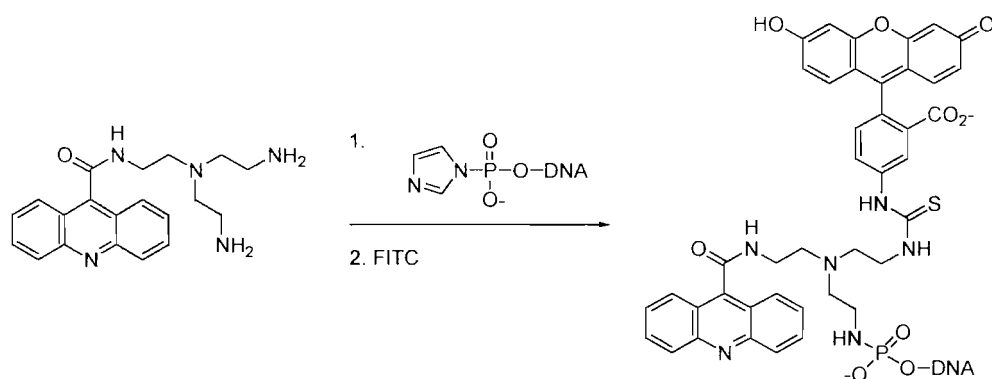
Scheme 2.1 Labelling strategy employed by Shinozuka *et al.*[175]

The acridine is a donor in a FRET system, exciting the reporter, fluorescein. Upon hybridisation, the acridine becomes less efficient in energy transfer and the fluorescence of fluorescein is reduced (Figure 2.3).
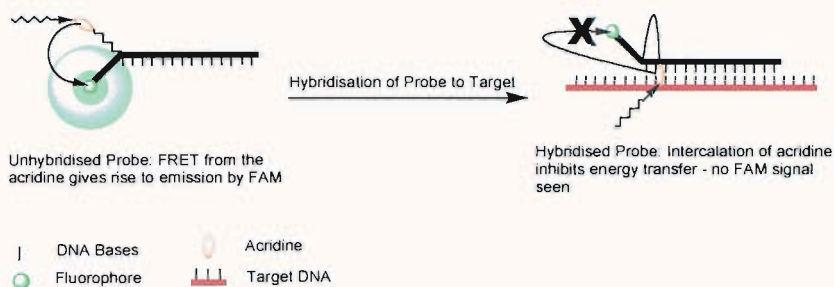
Figure 2.3 Mode of action of the acridine-containing probe described by Shinozuka et al.[175]

Recently, Yamane has described MagiProbe,[176] a probe similar to that described in this work, employing pyrene rather than an acridine as a quencher. Interaction between fluorescein and pyrene gives rise to quenching of fluorescein in the single strand, which is removed upon hybridisation by intercalation of the quencher (Figure 2.4).
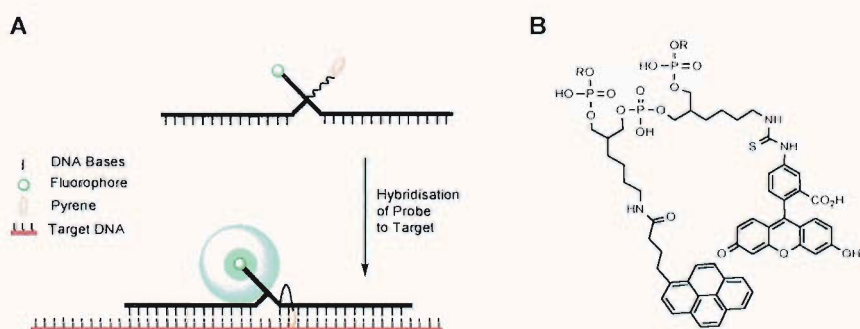
Figure 2.4 Mode of action (A), and chemical structure (B) of MagiProbe.[176]

2.2 Physical properties of the fluorophore-intercalator probes

2.2.1 Design of the novel probe

The system chosen for study employs an oligonucleotide probe labelled with a fluorophore and an intercalator. The unhybridised linear probe is expected to be non-fluorescent due to energy transfer between the fluorophore and the 9-amino-6-chloro-2-methoxyacridine quencher. However, formation of a probe/target hybrid should allow the acridine moiety to interact with the double stranded DNA, and no longer act as an efficient quencher. A fluorescent signal should therefore be generated (Figure 2.5).

Probe in "Dark" State. Stacking interactions in the single strand allow quenching of fluorophore by acridine

Hybridisation of Probe to Target

Hybridised Probe: Intercalation of acridine between base pairs releases fluorescence

| DNA Bases | Acridine |
| Fluorophore | Target DNA |
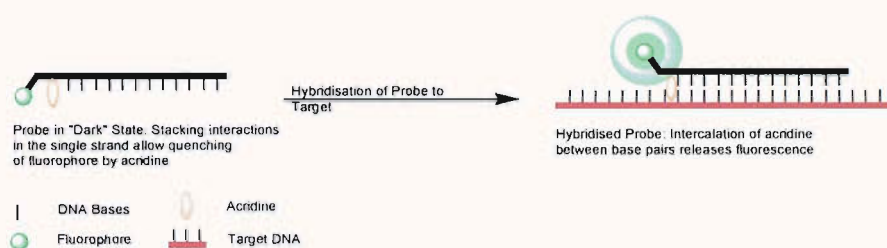
Figure 2.5 Proposed mode of action of dual-labelled probes.

The probe is synthesised with a fluorophore (fluorescein) and an intercalator (9-amino-6-chloro-2-methoxyacridine[177]) at the 5'-end (Figure 2.6). The advantages of the method lie in its simplicity: the probes are relatively short and inexpensive to synthesise (commercially available phosphoramidite monomers are used), and the intercalator inherently stabilises probe/target hybrids. The probes were 2'-O-methyl oligonucleotides (RNA). RNA-like oligonucleotides are not substrates for hydrolysis by *Taq* polymerase, and the probes were composed of RNA to avoid a TaqMan-like cleavage in the PCR.[178]
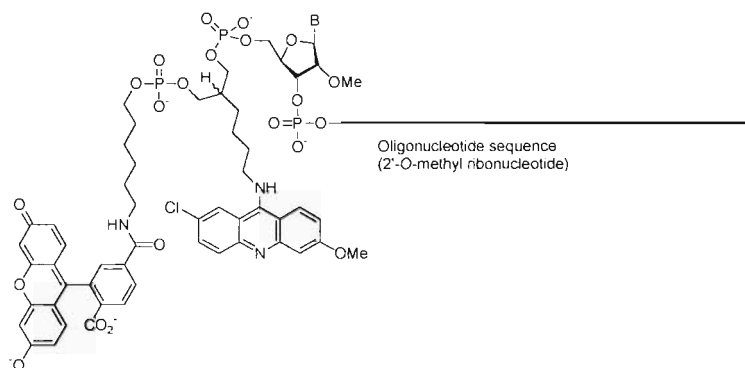
Figure 2.6 Chemical structure of dual-labelled probes bearing a fluorophore and an intercalator. B = A, G, C or T.

2.2.2 Probe fluorescence enhancement upon hybridisation

In order to investigate the properties of the dual-labelled oligonucleotides, four probes (FA12-18) and a synthetic target oligonucleotide (ST1303) were prepared (Table 2.1). When the synthetic target was added to a solution containing the probe in a suitable hybridisation buffer, and the probes excited at 495 nm, (near to the absorption maximum for the fluorescein moiety), the fluorescence emission at 520 nm (due to fluorescein) increased, by up to 5.8-fold relative to the unhybridised (quenched) fluorescence (Figure 2.7).
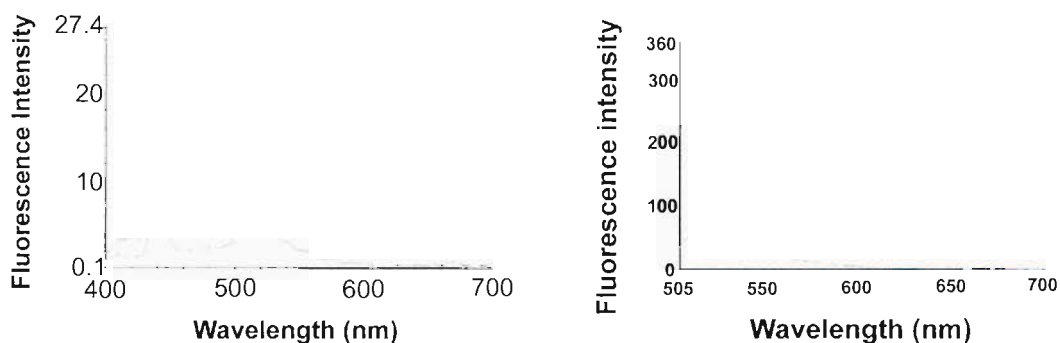
Figure 2.7 Simple hybridisation of probe FA12 to its single stranded synthetic complement, ST1303. **A.** Emission spectra obtained from excitation at 384nm in the absence of synthetic target (blue line), and after addition of 50 equivalents of the synthetic target (red line). **B.** Emission spectra obtained from excitation at 495nm in the absence of target (blue line), and after addition of 50 equivalents of the synthetic target (red line).

46

Labelled Probe	Probe Sequence (5' to 3')
FA12	**FAM-ACD-AAACTTGGATCC-OCT**
FA14	**FAM-ACD-AAAAACTTGGATCC-OCT**
FA16	**FAM-ACD-GAAAAAACTTGGATCC-OCT**
FA18	**FAM-ACD-TAGAAAAAACTTGGATCC-OCT**
ST1303	GGGATCCAAGTTTTTTCTAAATGTTCC

Table 2.1 Oligonucleotide sequences of probes with varying lengths. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine, **OCT** = octanediol.

Labelled Probe used	Probe length (nucleotides)	Excitation wavelength (nm)	Absorption wavelength monitored (nm)	Maximum Fluorescence enhancement upon hybridisation ($F_{hybridised}/F_{unhybridised}$)
FA12	12	384	520	6.4
FA12	12	495	520	5.8
FA14	14	384	520	4.9
FA14	14	495	520	4.5
FA16	16	384	520	3.1
FA16	16	495	520	2.8
FA18	18	384	520	2.8
FA18	18	495	520	3.9

Table 2.2 Fluorescence enhancements of probes upon hybridisation.

This is likely due to interaction of the acridine moiety with the probe/target duplex, reducing quenching (Figure 2.5). The effect was the same when the probes were excited by a light source of 384 nm (to excite the acridine moiety), analogous to the experiment for the probe described by Shinozuka *et al.*,[175] where a decrease in fluorescein emission was seen upon hybridisation. This suggests that the change in fluorescence seen here is not due to FRET from the acridine moiety, rather weak excitation of fluorescein itself. The probes present different stacking environments at the 5'-end, demonstrating that the observed phenomenon is not sequence dependent.

2.2.3 Optimisation of the fluorescence enhancement upon hybridisation

The ideal probe would emit a high fluorescence signal when hybridised, and no background signal arising from unhybridised probe. The 22mer oligonucleotide probes (P0-P11) were designed on the N1303K ABCC7 locus.[179] to hybridise to the

27mer N1303K ABCC7 wild type gene sequence complement (ST1303). Mutations to the ABCC7 gene can result in the Cystic Fibrosis phenotype, and the N1303K mutation was chosen because it is a point mutation,[180] presenting a good test of discriminating properties of the novel probes. The probes were designed with varying distances between the fluorophore/acridine pair, as this distance may affect the magnitude of fluorescence signal:noise ratio (the ratio of fluorescence intensity of the hybridised probe/target duplex to the unhybridised single-stranded acridine probe).

Oligonucleotide	Sequence (5' to 3')	Description
ST1303	GGGATCCAAGTTTTTTCTAAATGTTCC	Target
P0	**FAM-ACD**-ATTTAGAAAAAACTTGGATCCC	0 base separation probe
P1	**FAM-A-ACD**-TTTAGAAAAAACTTGGATCCC	1 base separation probe
P2	**FAM**-AT-**ACD**-TTAGAAAAAACTTGGATCCC	2 base separation probe
P4	**FAM**-ATTT-**ACD**-AGAAAAAACTTGGATCCC	4 base separation probe
P6	**FAM**-ATTTAG-**ACD**-AAAAAACTTGGATCCC	6 base separation probe
P9	**FAM**-ATTTAGAAA-**ACD**-AAACTTGGATCCC	9 base separation probe
P11	**FAM**-ATTTAGAAAAA-**ACD**-ACTTGGATCCC	11 base separation probe

Table 2.3 Sequences of oligonucleotides used in the optimisation study. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine.

Figure 2.8 shows the fluorescence intensity of each of the single-stranded acridine probes at identical concentrations in the absence of target. There is a direct correlation between the distance of the acridine quencher from the fluorophore and the amount of fluorescence background observed in the single strand (Figure 2.8). The efficiency of quenching is greatest when the acridine is placed adjacent to the fluorophore. It is important that the fluorescence signal in the unhybridised state is minimised to allow the maximum ratio of probe fluorescence in the 'light' state to the probe in its 'dark' state.
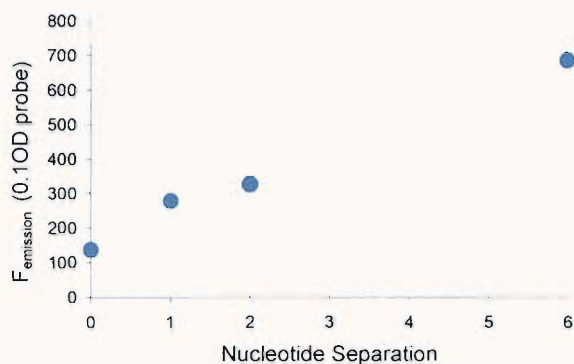
Figure 2.8 Fluorescence intensity of the single-stranded acridine probes at identical concentrations in the absence of target.

To each probe was added a five-fold excess of the complement and the fluorescence spectra recorded. All of the probes showed enhanced fluorescence in the presence of target (Figure 2.9). The most significant increase in fluorescence (over three-fold) was observed when the intercalator was placed directly adjacent to the 5'-fluorophore (P0).
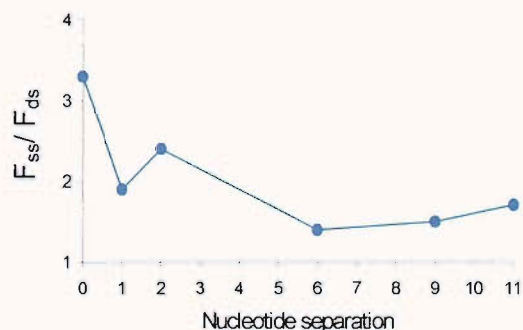
Figure 2.9 Fluorescence increase upon hybridisation for oligonucleotides with varying fluorophore-quencher separation.

2.2.4 Mode of action studies by UV/visible spectroscopy

UV spectroscopy allows study of the behaviour of aromatic compounds upon hybridisation. The nucleobases show a 20-30% decrease in UV absorptivity at 260 nm, or hypochromicity, upon formation of a duplex from single strands.[181]

Interaction of the conjugated dyes with the π-systems of nucleobases leads to shifts in absorption maxima and changes in the extinction coefficient of the dyes interacting with DNA. The oligonucleotides in Table 2.4 were used in the study.

Labelled probe	Description	Sequence (5' to 3')
A1	Acridine only probe	ATT-**ACD**-TAGAAAAAAC
F1	FAM only probe	**FAM**-ATTTAGAAAAAACTTGGATCCC
FA14	FAM-Acridine probe 1	**FAM-ACD**-AAAAACTTGGATCC-**OCT**
FA16	FAM-Acridine probe 2	**FAM-ACD**-GAAAAAACTTGGATCC-**OCT**
ST1303	Synthetic target	TAGGGATCCAAGTTTTTTCTAAAT
UP12	Unlabelled probe (12mer)	AAACTTGGATCC
UP14	Unlabelled probe (14mer)	AAAAACTTGGATCC
UP16	Unlabelled probe (16mer)	GAAAAAACTTGGATCC
UP18	Unlabelled probe (18mer)	TAGAAAAAACTTGGATCC

Table 2.4 Oligonucleotides used in UV/visible study of dye behaviour. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine, **OCT** = octanediol.

The results of the experiments are shown in Table 2.5. Intercalation of the acridine in A1 causes a decrease in its absorption at both λ_{max} of 11-13 %, consistent with increased π-stacking with the nucleobases. The FAM-only labelled oligonucleotide (F1) shows a similar drop in UV/visible absorbance upon hybridisation. The dual-labelled probes (FA14 and FA16) show a strikingly decreased extinction coefficient compared with the F1, 33-41 % lower in the single stranded form, indicating stacking with the acridine moiety.

Labelled probe	Dye studied	$\lambda_{max, \text{No target}}$	$\lambda_{max, \text{Target}}$	$\varepsilon_{\text{No target}}$	$\varepsilon_{\text{target}}$	$\varepsilon_{\text{target}}/\varepsilon_{\text{No target}}$
A1	Acridine	428	428	8,450	7,510	0.89
		452	452	8,590	7,510	0.87
F1	FAM	495	495	73,200	65,300	0.89
FA14	FAM	506	502	49,700	38,300	0.77
FA16	FAM	506	500	66,600	44,000	0.66

Table 2.5 UV/visible spectroscopy results. Absorption maxima (λ_{max}) are reported in nm, while extinction coefficients (ε) are reported in $mol^{-1}.dm^3.cm^{-1}$

Despite the increase in fluorescence (Table 2.2), the UV/visible absorbance for FAM in the dual-labelled probes (FA14 and FA16) falls on hybridisation. There is therefore still a strong interaction, presumably with the acridine moiety, which lowers ε. This indicates that the mode of action is not simply due to an increase in separation between fluorophore and quencher. It could be that there is a change in relative orientation between the acridine and FAM, which still involves π-stacking, but allows less efficient quenching of the excited fluorophore.

UV melting allows comparison of the T_m dual-labelled probe/target hybrid to that of an analogous unlabelled probe/target hybrid. The results are shown in Table 2.6. The labelled probe/target hybrids exhibited higher melting temperatures than the analogous unlabelled probe/target hybrids, supporting the proposal that the acridine moiety interacts with the duplex upon hybridisation.

Labelled probe	Analogous unlabelled probe	Complement	T_m of labelled probe with target (°C)	T_m of unlabelled probe with target (°C)	ΔT_m (°C)
FA12	UP12	ST1303	51.7	49.4	2.3
FA14	UP14	ST1303	54.0	52.3	1.7
FA16	UP16	ST1303	58.7	56.2	2.5
FA18	UP18	ST1303	58.4	57.6	0.8

Table 2.6 T_m results of UV melting experiments. Data were collected from monitoring UV-absorbance at 260 nm from each probe (1 μM) and ST1303 (1.3 μM) in a hybridisation buffer (100 mM Sodium Phosphate, 1 mM EDTA, 0.1 M NaCl, pH 7.0).

2.3 Application of fluorophore-intercalator probes to mutation detection

2.3.1 Fluorescence melting to demonstrate mismatch discrimination

To demonstrate the ability of the probes to discriminate between matched targets, and targets containing a single mismatch, each probe was melted with one synthetic target (ST1303, or MM1-3), whilst monitoring fluorescent emission in Channel 1 of the LightCycler (λ_{obs} = 520 nm). In samples containing the synthetic targets, the fluorescence intensity decreased upon melting of the probe/target hybrid (Figure 2.10). The first derivatives of the melting curves were taken, allowing calculation of the T_m of probe/target hybrids present in each tube (Table 2.7). Analysis of this data

51

shows considerable discrimination between matched and mismatched probe/target hybrids, with the T_ms differing by up to 22 °C.
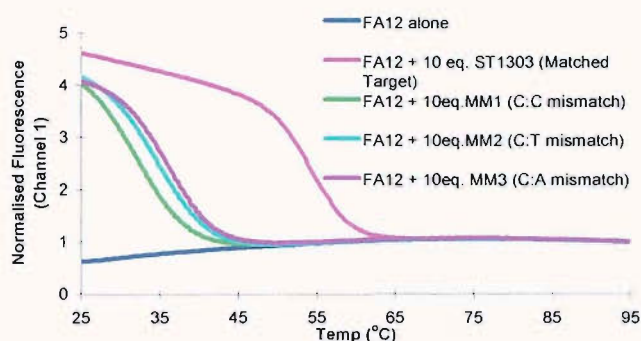
Figure 2.10 Melting curves obtained from melting the probe FA12 with synthetic targets. Data were collected in Channel 1 (λ_{obs} = 520 nm) of the LightCycler, from a solution of probe (0.5 μM) and the complement (0.5 μM) in a PCR buffer (20 mM $(NH_4)_2SO_4$, 75 mM tris-HCl, 0.1 % tween, 4 mM $MgCl_2$, 250 ng.μL^{-1} BSA, pH 8.0).

Probe	Probe length (nucleotides)	Complement	Description of complement	T_m (°C)	ΔT_m (°C)
FA12	12	ST1303	Matched	55.0*	-
FA12	12	MM1	C:C mismatch	32.9	-22.1
FA12	12	MM2	C:T mismatch	35.6	-19.4
FA12	12	MM3	C:A mismatch	36.1	-18.9
FA14	14	ST1303	Matched	56.8*	-
FA14	14	MM1	C:C mismatch	38.5	-18.3
FA14	14	MM2	C:T mismatch	40.6	-16.2
FA14	14	MM3	C:A mismatch	40.9	-15.9
FA16	16	ST1303	Matched	60.6*	-
FA16	16	MM1	C:C mismatch	46.0	-14.5
FA16	16	MM2	C:T mismatch	42.6	-18.0
FA16	16	MM3	C:A mismatch	43.0	-17.6
FA18	18	ST1303	Matched	59.4*	-
FA18	18	MM1	C:C mismatch	43.1	-16.3
FA18	18	MM2	C:T mismatch	44.7	-14.7
FA18	18	MM3	C:A mismatch	45.5	-13.9

Table 2.7 T_m results of fluorescent melting experiments. Data were collected in Channel 1 (λ_{obs} = 520 nm) of the LightCycler, from a solution of probe (0.5 μM) and the complement (0.5 μM) in a PCR buffer (20 mM $(NH_4)_2SO_4$, 75 mM tris-HCl, 0.1 % tween, 4 mM $MgCl_2$, 250 ng.μL^{-1} BSA, pH 8.0). *T_ms differ from those obtained by UV-melting **(Table 2.6)** as buffer composition and oligonucleotide concentrations were changed.

Oligonucleotide	Sequence (5' to 3')
MM1	GGGATCCAACTTTTTTCTAAATGTTCC
MM2	GGGATCCAATTTTTTTCTAAATGTTCC
MM3	GGGATCCAAATTTTTTCTAAATGTTCC

Table 2.8 Sequences of mismatched complements used in fluorescence melting experiments.

An advantage of this type of probe is that its linear nature renders interpretation of melting curves simple. A single transition is seen, which due to the discrimination of mismatches can be used to call genotypes if the real time trace is ambiguous. The potential for genotyping from melting curves also means that cheaper equipment can be used, e.g. a rudimentary thermal cycler with no fluorescence capabilities and a fluorimeter. Probes with secondary structure, such as Molecular Beacons, give complicated melting profiles with multiple transitions, which can be difficult to interpret (Figure 2.11).

Oligonucleotide	Description	Sequence (5' to 3')
MB1303	Molecular	FAM-CCCGCGCGG-AACATTTAGAAAAAACTTGGA
	Beacon	TCCCGCGCGGG-MR
ST1303	Target	GGGATCCAAGTTTTTTCTAAATGTTCC

Table 2.9 Oligonucleotides used to obtain Molecular Beacon melting profiles. **FAM** = fluorescein, **MR** = Methyl Red dR.
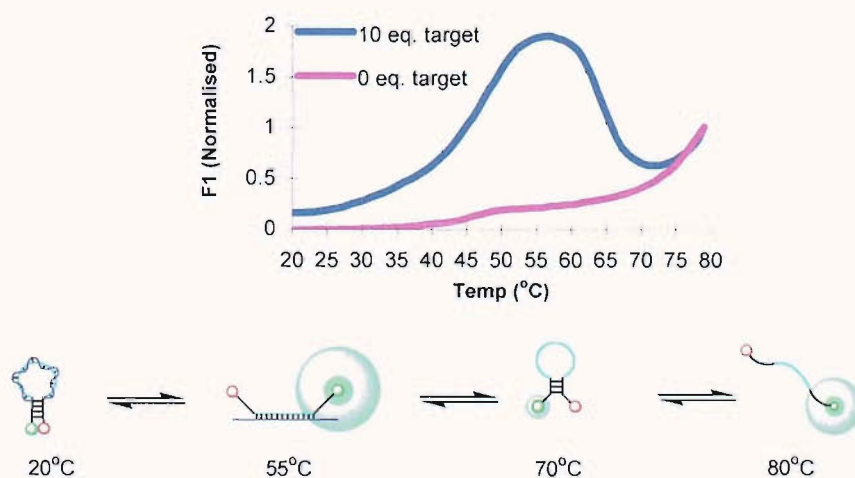
Figure 2.11 Fluorescence melting curves obtained for a Molecular Beacon in the presence and absence of target.

53

2.3.2 Real time PCR detection with the dual-labelled probes

The initial *Taq* polymerase PCR assay with the probe FA12 and primers FP1303 and RP1303 produced a single amplicon with $T_m \sim 80$ °C (deduced from reactions monitored with SYBR Green - data not shown), presumably corresponding to the 103 bp amplicon expected. Samples without any template DNA (negative controls) gave a PCR product with $T_m \sim 75$ °C, due to formation of primer-dimer species. Attempts to detect the correct amplicon with the 12mer probe (FA12) were successful, but signal:noise ratio and heterozygote/homozygote sample discrimination were poor (Figure 2.12). To test whether this was due to a TaqMan-like cleavage of the probe, a 5'-3' exonuclease deficient enzyme was used, and similar results obtained (data not shown), eliminating this possibility.

Oligonucleotide	Description	Sequence (5' to 3')
FP1303	Forward primer	TTTCTTGATCACTCCACTGTTC
RP1303	Reverse primer	CATACTTTCTTCTTCTTTTCTTT

Table 2.10 Sequences of primers used to amplify N1303K locus.
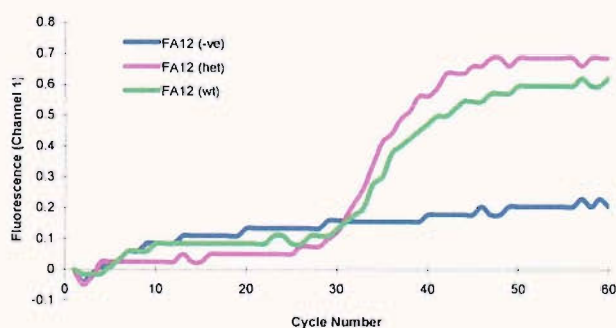
Figure 2.12 Real Time PCR with FA12. wt = wild type, het. = heterozygous, -ve = negative control.

Addition of the single stranded synthetic target (ST1303) to the reaction mixture at the end of the PCR yielded melting curves with a strong change in fluorescence, whereas addition of excess probe at the end of a PCR reaction performed without the probe gave rise to melting curves with weak changes in fluorescence (Figure 2.13). This demonstrates that the probe is intact at the end of PCR, and suggests that it does not compete well for the target strand of the double stranded PCR product, only producing a strong signal in the presence of excess single stranded target. This was

also supported by results from experiments where the time for which samples were held at denaturing (95 °C) and annealing (44 °C) temperatures prior to the post-PCR melting experiment were varied (Figure 2.13). Increasing the time at which the samples were held at these temperatures increased the change in fluorescence upon melting, indicating a greater population of bound probes. This suggests that the hybridisation problem is due to kinetic as well as thermodynamic factors.
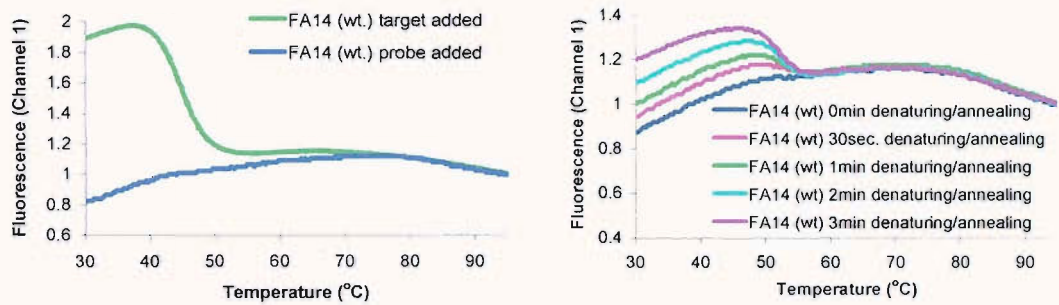
Figure 2.13 Melting curves obtained upon addition of excess probe or single stranded target to reaction mixture after PCR (left) and post-PCR melting with varying times for annealing (30 °C) and denaturing (95 °C) (right). wt = wild type.

Attempts to overcome these problems by simply increasing the time samples were held at annealing, extension and denaturing temperature in the PCR cycle, and changing to the 14mer probe (FA14) failed, due to production of mispriming products (Figure 2.14).
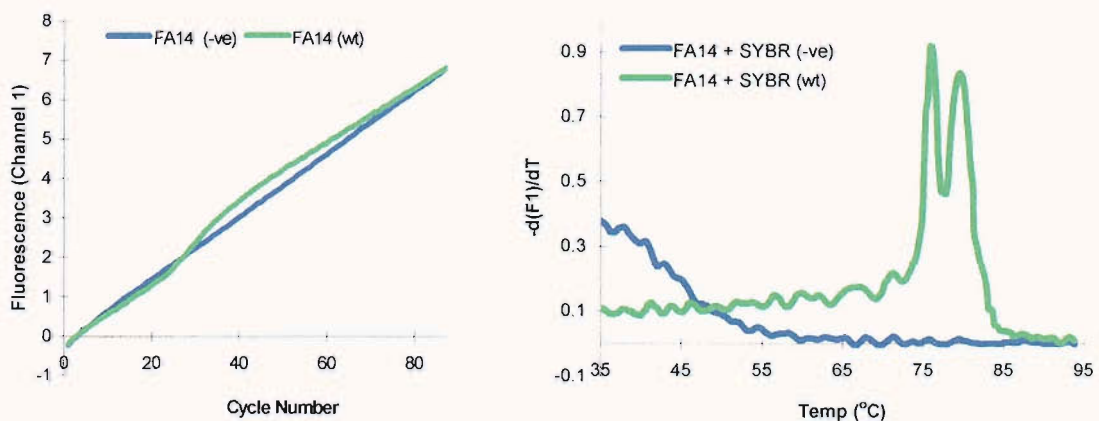
Figure 2.14 Mispriming caused by long cycle PCR. wt = wild type, -ve = negative control.

This problem was eliminated by use of a 'step-down' PCR experiment with the same long annealing time, where in the first five cycles the annealing temperature was 55 °C, (compared to 44 °C in 'standard' PCR reactions), and then reduced in the next 10 cycles to the standard annealing temperature. The effect of this is to create more stringent annealing conditions in the earlier cycles, preventing annealing of primers to alternative positions in the template, giving mispriming, or to each other, forming primer-dimers. An excess of the desired amplicon is formed in these early cycles, so that when the annealing temperature is reduced later, the primers bind preferentially in their 'correct' positions. This assay produced a single PCR product, which the 14mer probe (FA14) was able to detect in real time (Figure 2.15). Negative controls in this assay also gave no primer-dimer extension product.
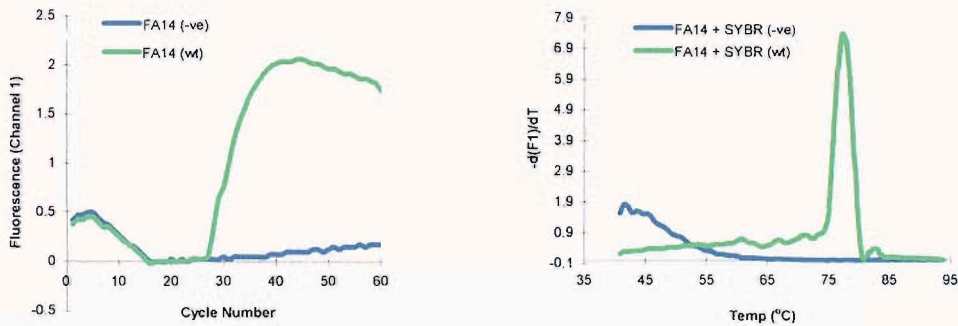
Figure 2.15 Real time PCR (left) and post-PCR SYBR Green melting curves (right) obtained with a 'step-down' cycle. wt = wild type, -ve = negative control.

Despite the successful detection of PCR product with the 'step-down' PCR cycle, the primers were redesigned in order to minimise primer-dimer formation and mispriming, producing a 72 bp amplicon. With these primers, detection of amplicon was possible under fast cycling conditions, allowing genotypic discrimination of human samples. Importantly, the intact probe could be used to obtain a melting curve at the end of PCR (Figure 2.16). This is not possible with probes that are cleaved upon hybridisation, such as TaqMan® probes.

Oligonucleotide	Description	Sequence (5' to 3')
RFP1303	Forward primer	ATTTCTTGATCACTCCACTGTT
RRP1303	Reverse primer	CTTTTTTGCTATAGAAAGTATTTA

Table 2.11 Redesigned primers used to amplify N1303K locus.
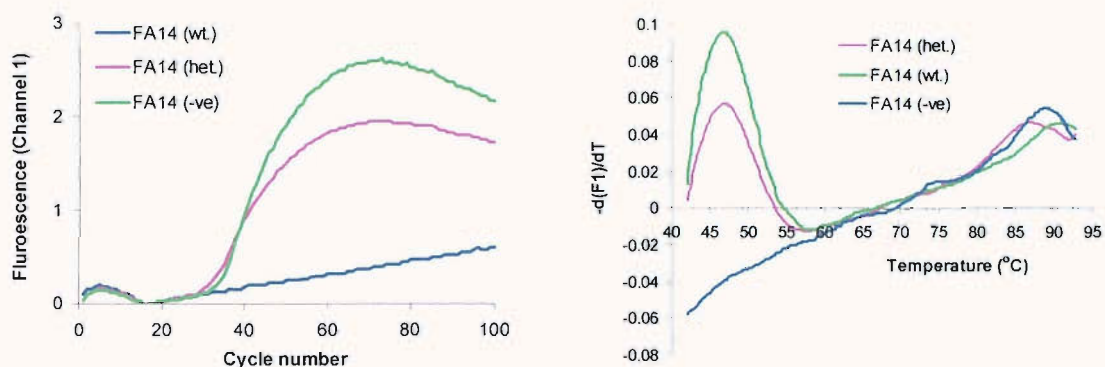
Figure 2.16 Real time PCR (left) and post-PCR melting curves (right) obtained with redesigned PCR primers RFP1303 and RRP1303. wt = wild type, het. = heterozygous, -ve = negative control.

2.4 Modification of probe design

2.4.1 Alternative quenchers for use in this system

While the probes assembled from commercially available monomers could be used satisfactorily in real time PCR, the signal:noise ratio obtained was less than that from other formats previously studied within the group e.g. TaqMan[®] and Scorpions.[111] Modification of the chemical nature of the fluorophore/quencher pair could produce a greater signal change upon hybridisation. Given that the LightCycler LED can only excite at 495 nm, limiting the choice of fluorophore for this particular application, the method may best be improved by utilising a more suitable quencher for FAM. The absorption maximum for the commercially available acridine moiety lies at 452 nm, making it a less than ideal quencher for FAM, whose emission maximum is 520 nm. It has previously been found that use of quenchers with appropriate absorption spectra can improve the signal:noise ratio obtained from other genetic analysis platforms.[182, 183]

A chromophore with the optimal properties required for this application therefore has its absorption maximum at 520 nm, and is a known DNA intercalator. The most suitable available compound found was acridine orange (Figure 2.17), whose absorption maximum is 490 nm.

Derivatisation from the 9- position of the acridine ring was the strategy chosen for attachment of the acridine orange moiety to oligonucleotides. However, fluorescence spectra obtained from the acridine orange derivatised oligonucleotide revealed that

57

the fluorescence properties of the acridine orange moiety had been altered – λ_{abs} was shifted to 429 nm, and λ_{em} to 515 nm.

Comparison of the fluorescence spectra of acridine orange with its N9- and S9-derivatives (Figure 2.17) shows the shifting absorption maxima are the result of the electronic influence of the 9-substituent. The S9-substituted 9-(methylsulfanyl)-acridine orange has an ideal absorption spectrum for quenching of fluorescein. However, the aromatic thioether is unlikely to be stable to oligonucleotide deprotection conditions (e.g. ammonia). This approach was therefore not explored further.

A	B	C
acridine	N9-(5-carboxypentyl)	9-(methylsulfanyl)-
orange	-acridine orange	acridine orange
λ_{abs} ~497nm	λ_{abs} ~429nm	λ_{abs} ~490-520nm
λ_{em} ~525nm	λ_{em} ~515nm	λ_{em} ~555nm
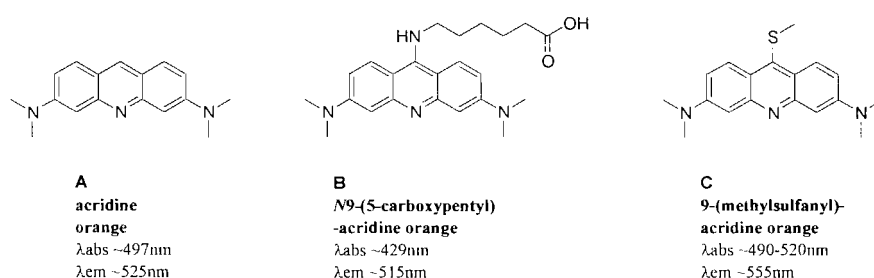
Figure 2.17 Fluorescence properties of acridine orange and two of its derivatives.

Whilst modification of the chromophore of the quenching moiety proved an unsuccessful method for improving the signal:noise obtained from the probes, the alternative strategy of attaching the same heterocycle to a different backbone could improve performance. The commercially available acridine-labelling monomer has a chiral centre, but is synthesised as a racemate. Others have found that incorporation of functional molecules into DNA by means of different enantiomers of chiral backbones can have significant effects on the properties of the resultant oligonucleotides. Asanuma *et al.* have found that a D-threoninol-mounted *cis*-azobenzene stabilised duplex formation more than did one attached to L-threoninol.[184] It has also been found that acridine-oligodeoxynucleotide conjugates where the acridine was attached to L-threoninol *via* a spacer were 2.5 times more effective at activating the phosphodiester backbone of a complementary RNA towards metal catalysed strand scission than the D-threoninol-tethered acridine.[185] This occurred despite there being little difference between the T_ms of the heteroduplexes, and was attributed to a subtle change in the orientation of the

acridine. Since, as suggested by UV-spectroscopy, the dequenching effect observed upon hybridisation for the dual-labelled probes described here is due to a change in relative orientation between the fluorophore/acridine pair upon hybridisation, it is reasonable to believe that two enantiomers of the acridine-bearing moiety might behave differently.

The length of linkage between the acridine and the backbone was also varied. Although it has previously been found that a linker of 8 atoms between the heterocyclic ring and the chiral carbon of an L-threoninol was necessary to allow optimal interaction of an acridine with an artificial abasic site,[186] a linker of 1 atom was selected for study. The 'three quarter intercalation' model proposed to accompany hybridisation is stereochemically different than intercalation at an artificial abasic site so there is no reason to assume the same backbone is appropriate for the two applications. Also, since the acridine is at the 5'-end of the probe, it has more translational/rotational freedom. Use of a shorter linker could limit translations/rotations of the acridine when interacting with duplex DNA, minimising quenching in the 'open' form. It may also accentuate any difference in properties between the two enantiomers, making it clear which to use if further optimisation were necessary. For these reasons, the backbone shown in Figure 2.18 was selected for incorporation of the acridine.

Original backbone for incorporation of acridine   **Proposed backbone for incorporation of acridine**
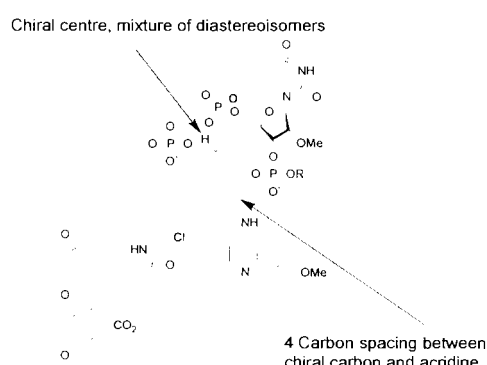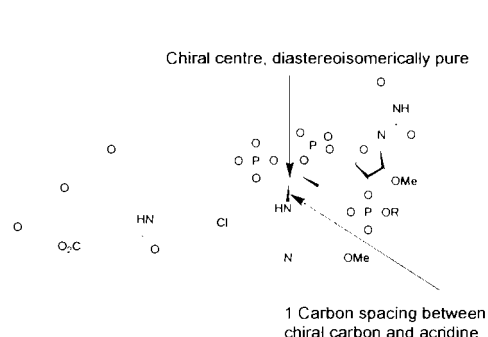
Figure 2.18 Backbones for incorporation of acridine chromophore into oligonucleotides.

Initially, the L-enantiomer of threoninol was used. Attachment of the acridine moiety to L-threoninol by S_NAr substitution, followed by tritylation and

phosphitylation of the resultant diol **2** provided the phosphoramidite monomer **4**. For unknown reasons, compound **4** could not be used in automated DNA synthesis. Coupling efficiency, even with extended times of up to 30 minutes, was less than 5%. While the proximity of the bulky acridine tricycle to the phosphoramidite might reduce coupling efficiency, complete failure would not be expected. Attempts to couple the phosphoramidite **4** to alcohols in solution yielded only its H-phosphonate derivative, suggesting the presence of water. Rigorous drying could not overcome this problem, so this strategy was also abandoned.
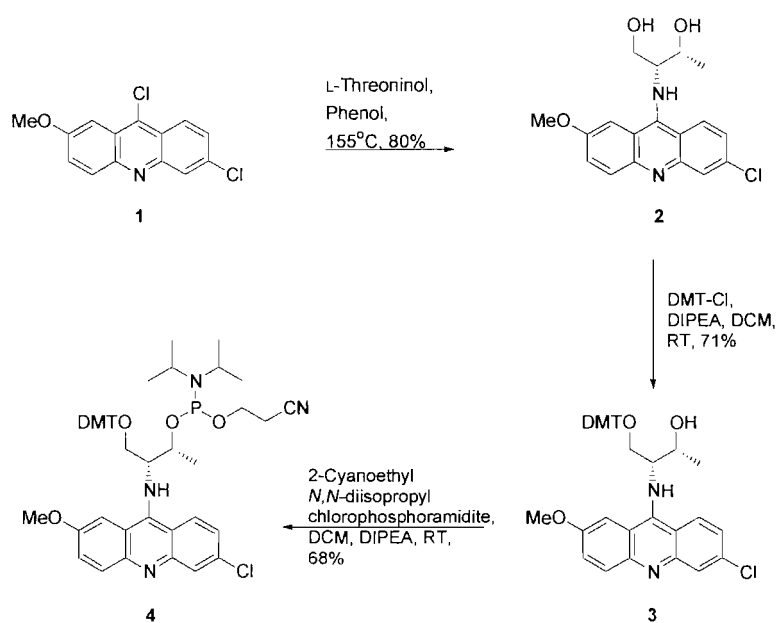
Scheme 2.2 Synthesis of alternative acridine-labelling monomer.

2.5 Conclusions

A novel dual-labelled probe has been designed and shown to increase its fluorescence upon hybridisation.[187] UV-spectroscopy suggests the effect is due to a 'three quarter intercalation' interaction between the quencher and the probe/target duplex, disrupting the fluorophore/quencher interaction. The format has been validated in real time PCR, and can discriminate between homozygote and heterozygote genotypes (Figure 2.16). The linear nature of the probe leads to it having advantages over existing genetic analysis techniques. Firstly, the lack of secondary structure allows the use of shorter probes with enhanced mismatch

discrimination. Secondly, since the probe is inert to PCR conditions it can be used to record a melting curve at the end of PCR (Figure 2.16), which is not possible with probes that are cleaved throughout the course of PCR, such as TaqMan® probes. The melting profiles obtained display a single transition and are easily interpreted (Figure 2.10). This makes the use of melting curves to discriminate mismatches more effective than with Molecular Beacons, whose melting are complicated and contain multiple transitions (Figure 2.11).

However, the signal change obtained from hybridisation of these probes has been found to be smaller than that obtained for TaqMan® or Scorpions. This is perhaps not surprising. In TaqMan®, the fluorophore and quencher are separated into different molecules by the probe being cleaved, and in Scorpions, they are separated by >30 nucleotides when the hairpin loop does not form. The solution to this problem may be found by chemically modifying the probe, specifically the quenching moiety. Initial attempts to do this described here have been unsuccessful for specific chemical reasons. Modification of the chromophore to reduce background signal gave rise to a new heterocycle with unsuitable spectral properties (Figure 2.17), and modification of the backbone to which the quencher was attached in order to produce a greater signal upon hybridisation gave a phosphoramidite which could not be used in automated DNA synthesis. In all likelihood, a combination of both these complementary strategies would be necessary to produce the desired effect. Due to the time that would be required to undertake a thorough study of potential modifications to both chromophore and backbone, this approach was not pursued, but could be the subject of future work. A modification of the assay, where the quencher is not conjugated to the probe, was preferred. This is described in Chapter 3.

In addition to the reduced signal generated by this particular probe, the results so far indicate that a severe general limitation of hybridisation as a method for detecting nucleic acid sequences is the tendency of short probe sequences to be displaced by the longer competing strands generated by the PCR. This phenomenon, and potential solutions to it, is the subject of Chapter 4.

3. DNA Intercalators as fluorescence quenchers in genetic analysis – noncovalent approach

3.1 Use of intercalators free in solution to quench singly-labelled probes upon hybridisation

Given that the approach of incorporating of the quenching and signalling moieties covalently into a single probe molecule led to a relatively small signal change upon hybridisation, it was postulated that a method whereby a greater change in the separation of the fluorophore and quencher is caused by probe binding may improve signal:noise ratios. The proposed method of detection of probe hybridisation involved quenching of a single fluorescent label by an intercalator free in solution (Figure 3.1). The converse approach, use of the intercalator SYBR Green I as a FRET donor for a labelled probes is used in ResonSense®, Angler® and iFRET platforms (Section 1.3.2.5).[130, 131] Although the proposed method would lead to a decrease in fluorescence on hybridisation and therefore require inversion of the real time curve, this is a simple operation that is used in other methods.[123]
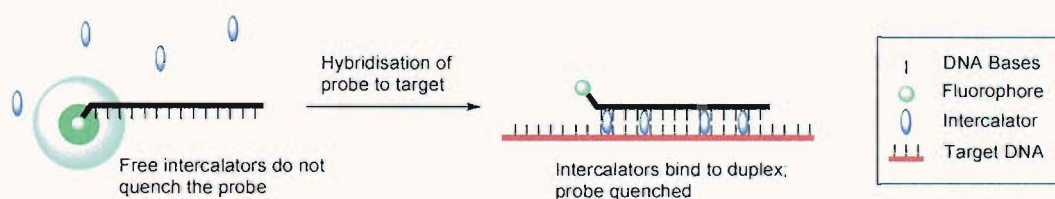
Figure 3.1 The use of an intercalator free in solution to effect quenching of a singly labelled probe upon hybridisation.

The signal change that can be obtained depends upon the efficiency of the intercalator as a fluorescence quencher and the selectivity of the intercalator for dsDNA over ssDNA, so that fluorescence of the single stranded (unquenched) probe is maximal compared to the hybridised (quenched) probe. If there is significant binding to unhybridised probe, there will be a loss of sensitivity. These properties will depend upon the nature of the intercalator as well as the sequence of interest. The idealised intercalating quencher binds dsDNA selectively and in sequence-

independent manner at low ligand concentration, effectively extinguishing fluorescence as it does so.

3.1.1 Assay for screening of intercalators as quenchers

In order to identify suitable DNA-binding agents for use in this context, eleven intercalators and one minor groove binder were screened against four short oligonucleotides of varying sequence. To make use of the high throughput possible on the LightCycler, the fluorescence melting technique was chosen to study drug-DNA interactions. Melting curves were obtained for both the duplex and the single stranded oligonucleotide at each drug concentration (Figure 3.2).
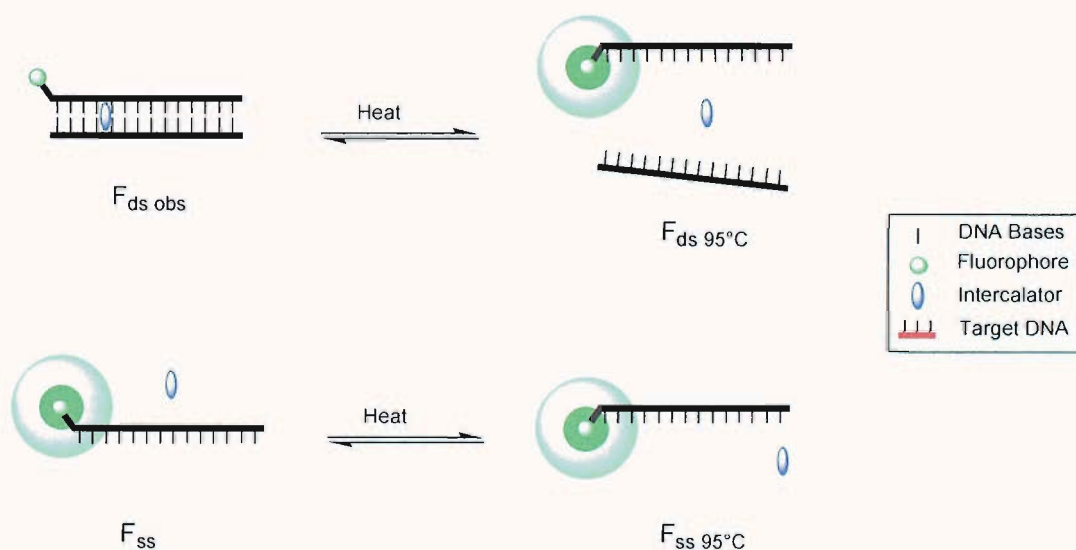
Figure 3.2 Fluorescence melting screening assay for intercalators.

To correct for small variations in labelled oligonucleotide concentration resulting from pipetting errors, the observed intensities ($F_{ds\ obs}$) in the double stranded dataset were first normalised to the fluorescence of the probe when complete denaturation had taken place, at 95°C ($F_{ss\ 95°C}$). Normalised values (F_{ds}) were calculated from observed values ($F_{ds\ obs}$) by the following formula:

$$F_{ds} = (F_{ds\ obs}/F_{ds\ 95°C})\ F_{ss\ 95°C}$$

This results in melting curves, with and without complement, which meet at 95 °C (Figure 3.3).
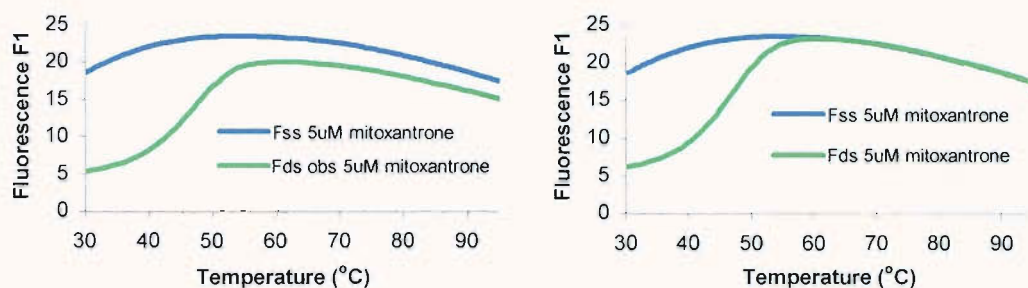
Figure 3.3 Uncorrected (left) and normalised (right) fluorescence melting curves. Data were collected in Channel 1 (λ_{obs} = 520 nm) of the LightCycler, from a solution of probe (0.5 µM) in the presence and absence complement (0.5 µM) in a buffer (10 mM tris-HCl, 50 mM NaCl, 1.5 mM $MgCl_2$, 250 ng.µL^{-1} BSA, pH 8.0) containing 5µM mitoxantrone.

Corrected fluorescence intensities were used to calculate the signal:noise ratio, which is the ratio of F_{ss} to F_{ds} at a given temperature. The efficacy of a drug in the quenching assay is proportional to this value. Melting curves were collected at six drug concentrations, allowing the "signal:noise curves" of the type in Figure 3.4 to be plotted for each sequence.
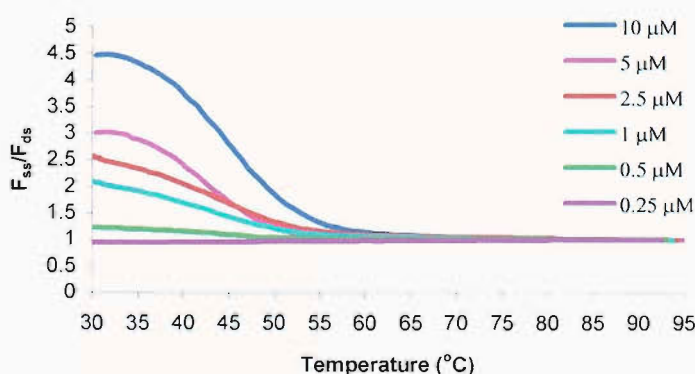
Figure 3.4 "Signal:noise curve" for a single drug (mitoxantrone) at six concentrations against a single sequence (165R/160R).

DNA-binding agents often display sequence selectivity. The sequences were designed to contain known binding sites for DNA-binding agents: CpG steps for the anthracyclines, GpC for the actinomycins and AATT for minor groove binders.

64

Inclusion of the GpC/CpG steps in the model sequence is reasonable, since the probability of either site appearing in the 20 nucleotides surrounding a polymorphic site is 71% assuming an equal mixture of nucleotides. The chance of the AATT motif appearing is very small (7% assuming an equal distribution of nucleotides), but was included in order to study the properties of minor groove binders. Sequences were 'scrambled' to allow quenching as a function of distance between drug and fluorophore to be studied. This is essential because if the presence of a particular dinucleotide is required in the probe for quenching to occur, it is important that the site does not have to occur close to the fluorophore, or the number of sequences that can be probed will be diminished. If, for example, a CpG site must occur within 6 nucleotides of the fluorophore, the chance of it occurring is only 28 %.

Oligonucleotide	Sequence (5'-3')	Description
163R	FAM-ACGCAATTGGT	FAM-labelled probe
158R	ACCAATTGCGT	Complement to 163R
164R	FAM-AGCGAATTGGT	FAM-labelled probe
159R	ACCAATTCGCT	Complement to 163R
165R	FAM-AATTGCGTCAG	FAM-labelled probe
160R	CTGACGCAATT	Complement to 163R
172R	FAM-CGCATCGCAATT	FAM-labelled probe
173R	AATTGCGATGCG	Complement to 163R

Table. Sequences of oligonucleotides used in the intercalator screening assay.

3.1.2 Intercalators studied

3.1.2.1 Mitoxantrone and the anthracyclines

Mitoxantrone   Daunomycin   Nogalamycin
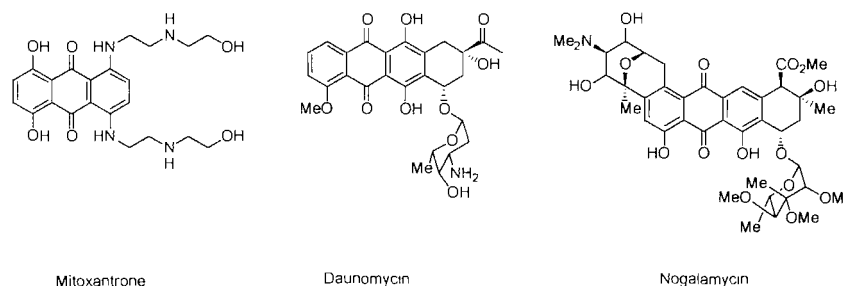
Figure 3.5 Chemical structures of mitoxantrone and the anthracycline antibiotics daunomycin and nogalamycin.

65

The intercalators mitoxantrone, daunomycin and nogalamycin, whose chromophores are all based upon substituted anthraquinones were all found to be suitable compounds, giving higher fluorescence of the fluorescently-labelled oligonucleotides when single stranded. Nogalamycin, which has an amino-sugar that binds to the major groove of dsDNA,[188] and daunomycin, which has an amino-sugar that protrudes into the minor groove,[15] gave better results than mitoxantrone, giving higher signal:noise ratios (Figures 3.6-3.8). Daunomycin was most effective at 5 μM for all sequences, mitoxantrone showed best signal:noise ratios at 5 or 10 μM depending on sequence, and the optimal concentration for nogalamycin was always 10 μM. These results do not indicate tighter binding of DNA by daunomycin, but a combination of binding, quenching and selectivity for duplexes over single strands.
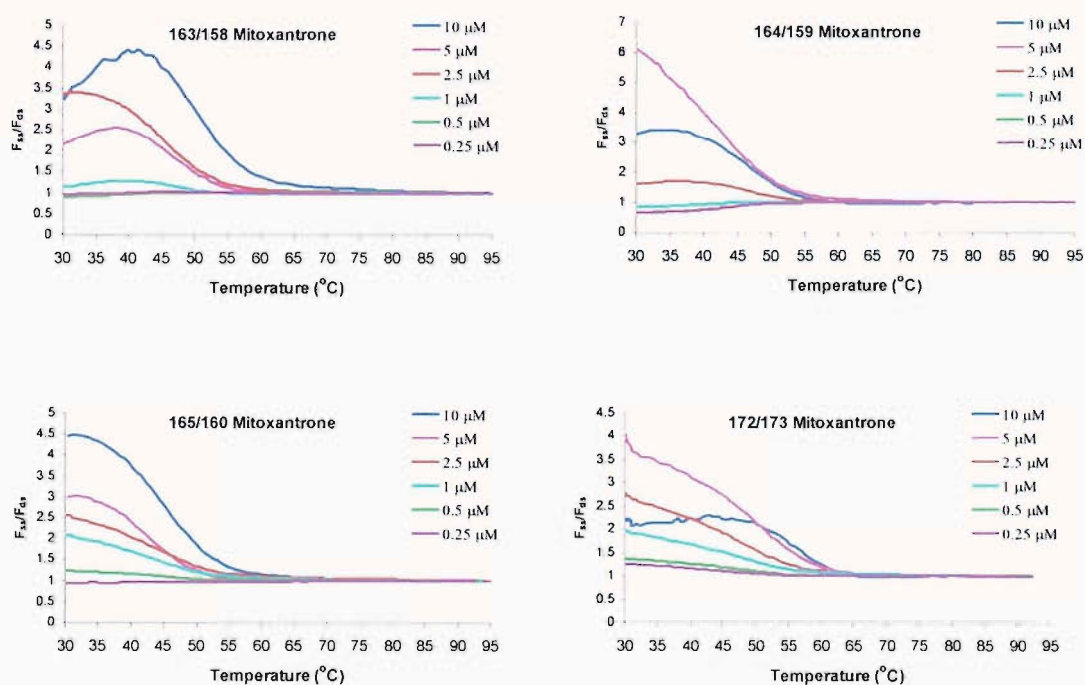
Figure 3.6 Signal:noise curves for mitoxantrone.
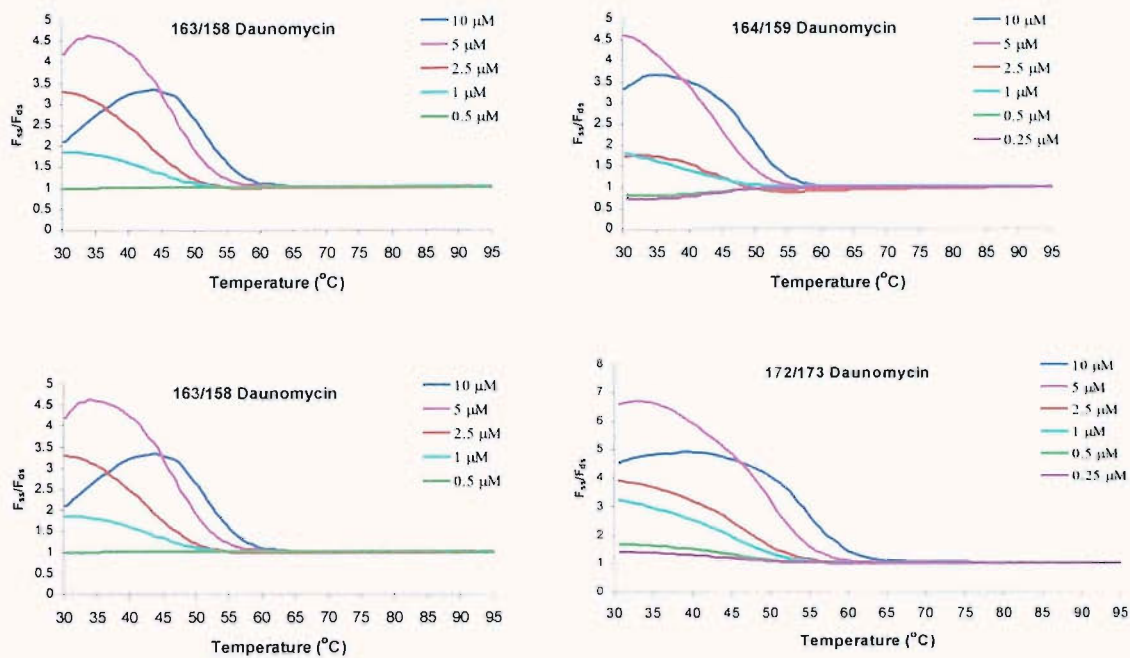
Figure 3.7 Signal:noise curves for daunomycin.
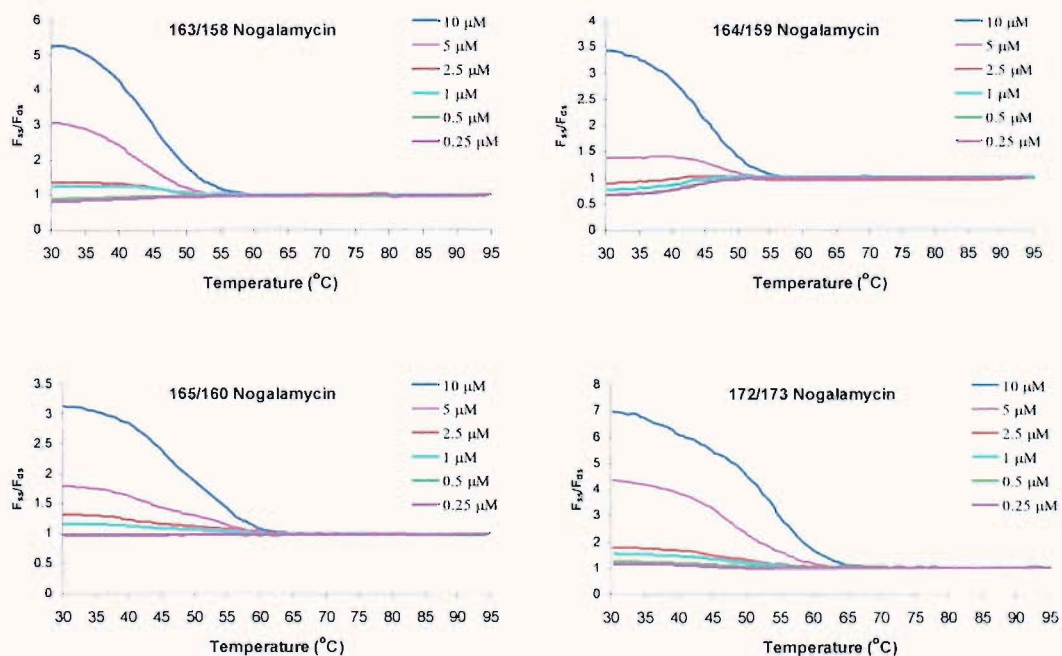
Figure 3.8 Signal:noise curves for nogalamycin.

The anthracyclines daunomycin and nogalamycin, are known to intercalate specifically at CpG sites in double stranded DNA, while mitoxantrone shares this preference but exhibits less stringent selectivity.[189] A plot of maximal fluorescence

67

quenching at 10 µM drug concentration against number of nucleotides to the nearest CpG site reveals that while fluorescence quenching by nogalamycin falls off as the distance between the fluorophore and CpG site increases, daunomycin and mitoxantrone quenching is maintained over a longer distance. This is probably due to the spectroscopic properties of daunomycin (λ_{max} = 495, 530 nm),[190] where FRET (a long range interaction compared to contact quenching) is likely to be the quenching mechanism, due to its higher absorption maximum compared with nogalamycin (λ_{max} = 480 nm).[191] Mitoxantrone, whose binding site is not clearly defined, might be expected to quench each sequence equally well, which is the case. Despite the drop in quenching with fluorophore-binding site separation observed for nogalamycin, the signal:noise ratio obtained remains high, indicating a decrease in quenching in the single stranded form.
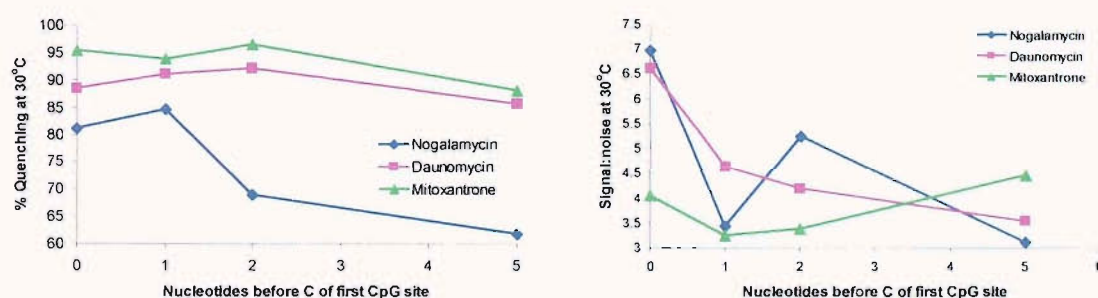
Figure 3.9 Variation of fluorescence quenching with distance between fluorophore and the nearest CpG site (left), and relationship between signal:noise ratio and distance of CpG site from the fluorophore (right).

3.1.2.2 Ethidium Bromide

The fluorescent phenanthridinium dye ethidium bromide (Figure 3.10) is a sequence-indiscriminate intercalator frequently used in the staining of nucleic acids ($\lambda_{ex} = 518$ nm, $\lambda_{em} = 605$ nm).[20] As expected from its properties, ethidium bromide quenched all the fluorescently-labelled sequences when hybridised, and showed little binding to single stranded probes, giving good signal:noise ratios (Figure 3.11).
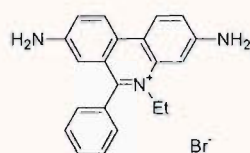
Figure 3.10 Chemical structure of ethidium bromide.
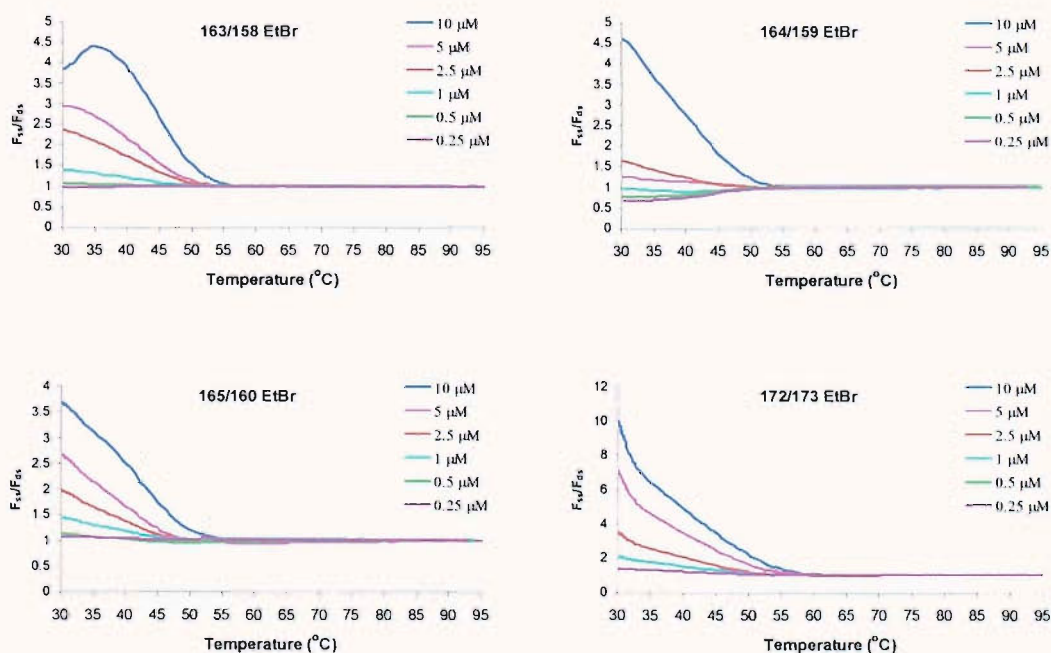
Figure 3.11 Signal:noise curves obtained for ethidium bromide.

Melting in the presence of ethidium bromide also gave a profile in Channel 2 of the LightCycler ($\lambda_{obs} = 650$nm) due to fluorescence enhancement of ethidium when intercalated into DNA (Figure 3.12). This property could be used for monitoring PCR in Channel 2, when Channel 1 is being used to perform mutation analysis. This

is analogous to the use of SYBR Gold to monitor progress of PCR in Channel 1 as well as a FRET donor to acceptor-labelled probes in ResonSense® and iFRET.[130, 131] Experiments with unlabelled duplexes (data not shown) indicated that ethidium bromide was excited by the LightCycler LED and not just FRET from the labelled probe.
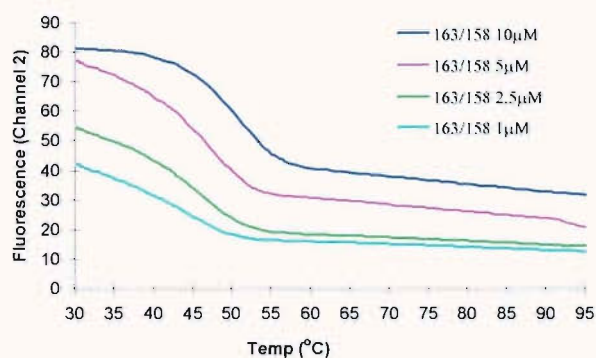
Figure 3.12 Melting curves obtained for duplex 163/158 in Channel 2 due to ethidium bromide fluorescence at four concentrations of intercalator.

3.1.2.3 The actinomycins

Actinomycin D (ACTD) and 7-aminoactinomycin (7-AACTD) are intercalators containing a phenoxazine heterocycle and two cyclic peptide arms (Figure 3.13) bind GpC, rather than the usual Py-Pu steps.[17] The two peptide arms bind to the minor groove, with the threonine residues hydrogen bonding to the N3 and 2-amino groups of Gs in opposite strands.

Actinomycin D

7-Aminoactinomycin D
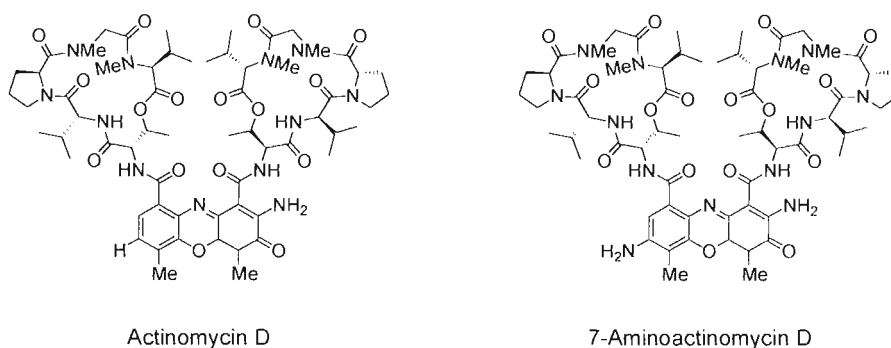
Figure 3.13 Chemical structures of actinomycin D and 7-aminoactinomycin D.

Both compounds showed significant signal:noise ratios with only one sequence, 164/159 (Figures 3.14 and 3.15). The reason for the poor performance with the other sequences was binding to the single stranded probes, which was clearly apparent from melting curves obtained for the probe in the absence of complement (Figure 3.16).
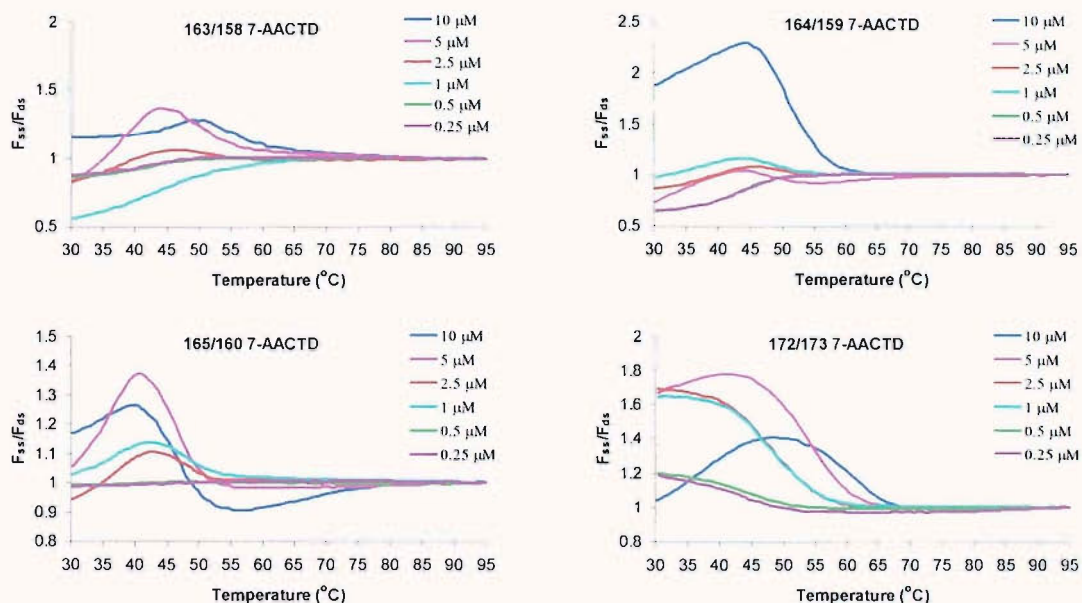
Figure 3.14 Signal:noise curves obtained for actinomycin D.
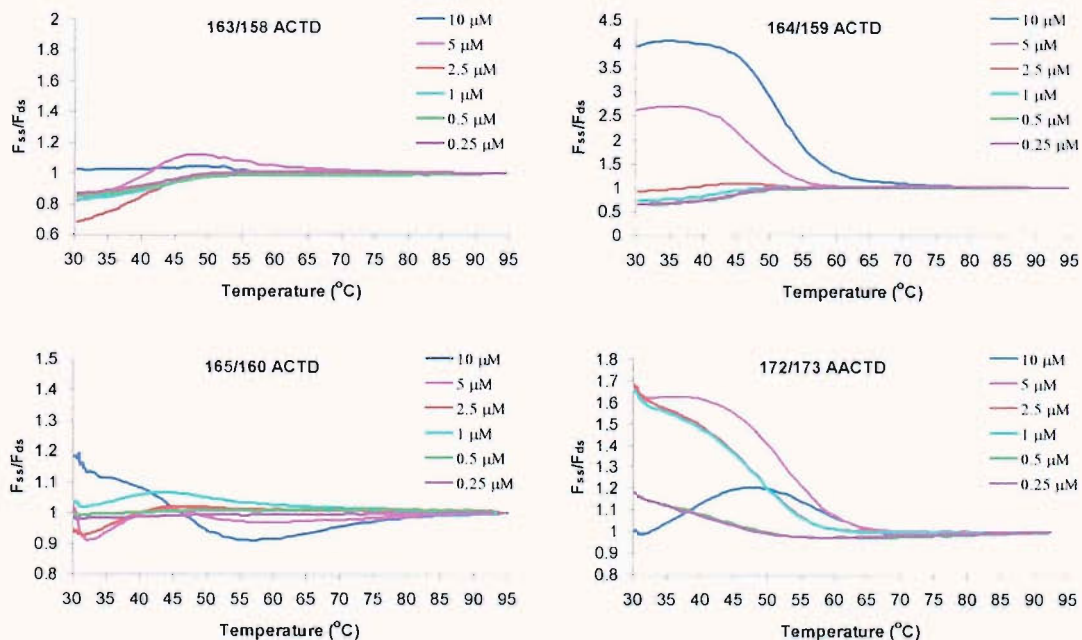
Figure 3.15 Signal:noise curves obtained for 7-aminoactinomycin D.
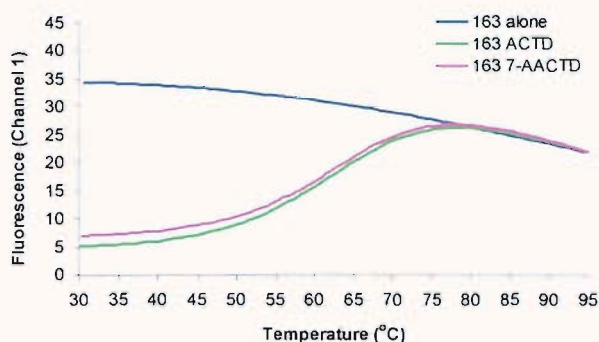
Figure 3.16 The effect of ACTD and 7-AACTD addition on the melting behaviour of single stranded probe 163R.

Binding of actinomycins to single stranded nucleic acids has been previously reported, and is likely to contribute to their inhibition of HIV-1 reverse transcriptase.[192] While some interactions of these drugs with apparently single stranded structures may result from the formation of metastable intramolecular hairpin loops,[193, 194] or short intermolecular duplexes, there is evidence for a hemi-intercalative model involving hydrogen bonding of the threonine residues to the 2-amino group of G and stacking of the phenoxazine ring with the H-bonded G and an adjacent base.[195] The formation of this structure not only requires the presence of G, but is also governed by complex and not yet fully understood sequence dependence.[196, 197] Whilst establishing these rules is well beyond the scope of this study, it is interesting to note that 164R (the sequence with minimal single stranded binding) differs from 163R (the sequence which displayed the most single stranded binding) only in that the nucleotides 2-4 are GCG rather than CGC.

The quenching efficiencies observed in duplexes for the non-fluorescent ACTD (λ_{max} = 442 nm)[20] and the fluorescent 7-AACTD (λ_{ex} = 546 nm, λ_{em} = 647 nm)[20] are largely as expected from their spectroscopic and binding properties. In general 7-AACTD is the more effective quencher as a result of its greater spectral overlap with FAM (λ_{em} = 520 nm), and the quenching efficiency generally falls off with the distance between the GpC binding site and the fluorophore (Figure 3.17). Sequence 172R, which has two GpC sites (1 bp and 6bp from the 5'-dye), shows anomalous behaviour, with ACTD being the more effective quencher. This is likely due to the two actinomycins (usually considered to have the same sequence specificity)

72

occupying different binding sites: ACTD 1 bp from FAM, and 7-AACTD 6 bp away.
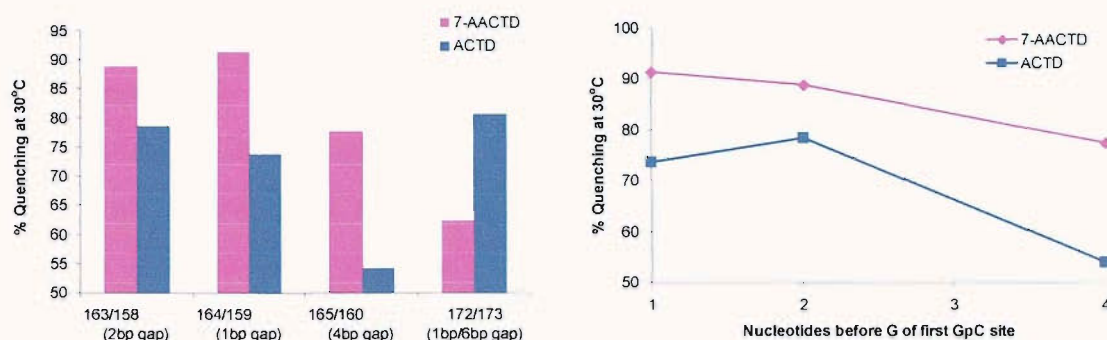
Figure 3.17 Quenching of the FAM-labelled duplexes by ACTD and 7-AACTD.

3.1.2.4 Hoechst 33258

Hoechst 33258 is a fluorescent (λ_{ex} = 352 nm, λ_{em} = 467 nm)[20] benzimidazole minor groove binder used in nucleic acid staining, which binds to A/T rich sequences, and specifically the AATTN motif.[198] Although this sequence is unlikely to occur in a 20mer probe sequence, tailing the sequence to an allele-specific primer-probe would allow it to be incorporated into the amplicon, as with the universal primer sequence used in the Amplifluor UniPrimer format,[80] or the GC-tailed primers used to exaggerate T_m differences between PCR products.[78] The possible advantages of using a minor groove binder with high sequence-specificity are that selectivity for dsDNA over ssDNA should be high, and that few drug molecules should be bound to each amplicon, minimising the interference to the PCR.
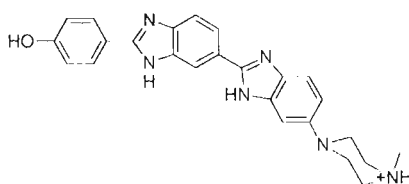
Figure 3.18 Chemical Structure of Hoechst 33258.

The lack of spectral overlap between Hoechst 33258 and FAM means there is no possibility of FRET quenching. Only Dexter quenching or static quenching are plausible. Since these are the shortest range quenching effects, it is likely that the fluorophore-binding site distance will have a large bearing on the quenching efficiency. This is clearly the case – only when the AATTN motif is directly adjacent to the fluorophore (in duplex 165R/160R) is significant quenching (Figure 3.19), and hence good signal:noise ratios (Figure 3.20) observed. Given these results, it seems likely that static quenching, requiring formation of a π-stacked fluorophore-quencher complex is the likely mode of action, since even Dexter quenching would not be expected to drop off so dramatically with fluorophore-binding site separation.
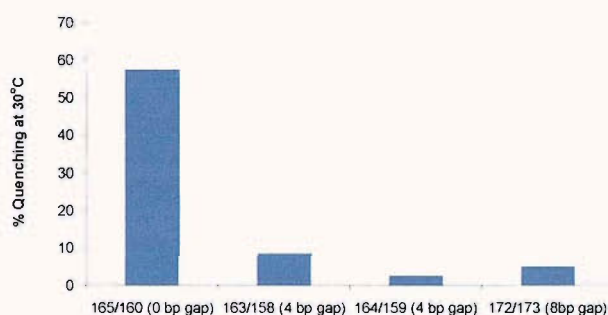
Figure 3.19 Fluorescence quenching of labelled duplexes by Hoechst 33258.
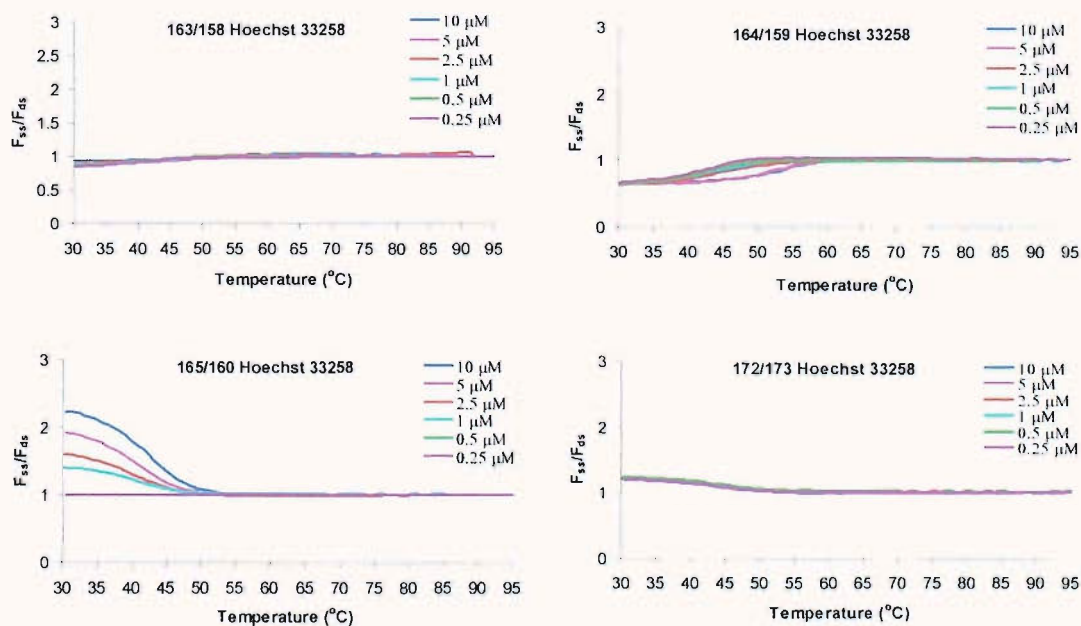
Figure 3.20 Signal:noise curves obtained for Hoechst 33258.

3.1.2.5 Coralyne Chloride

Coralyne chloride (Figure 3.21) is a member of the berberine class of alkaloids known to have antitumour activity[199] as a result of its topoisomerase inhibitory properties.[200] It can intercalate into double[201] and triple helices.[202] Binding to ssDNA has been observed, but only to poly (dA).[203]
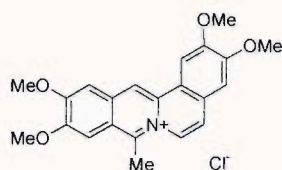
Figure 3.21 Chemical structure of coralyne chloride.
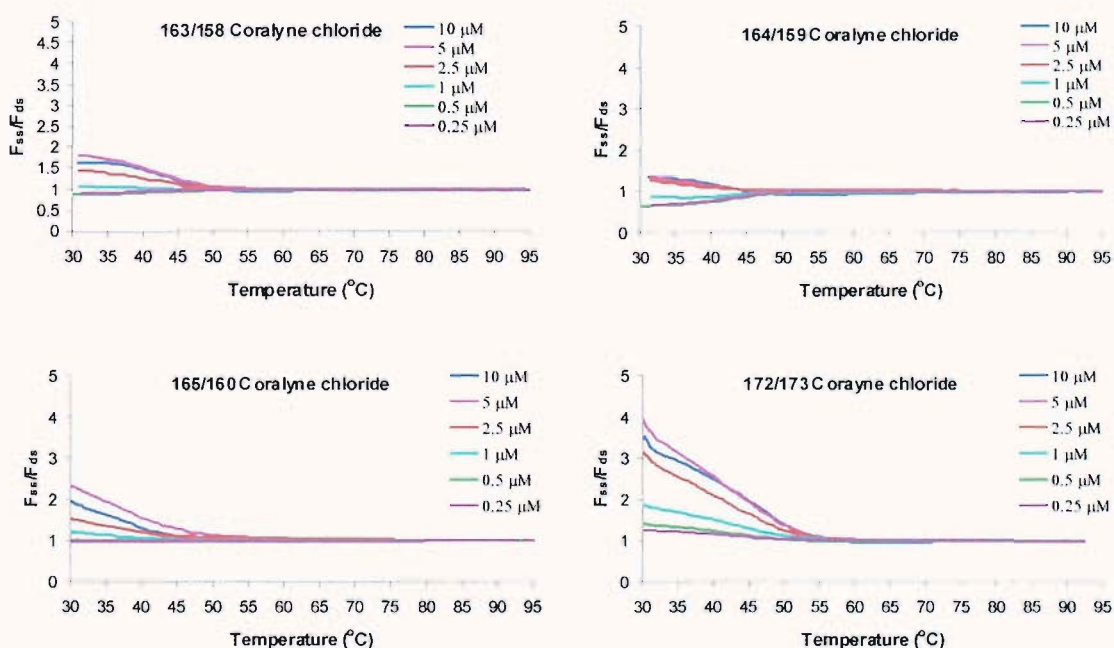
Figure 3.22 Signal:noise curves obtained for coralyne chloride.

The signal:noise curves for each sequence have a maximum >1, indicating greater quenching of the duplex than of the single stranded oligonucleotide, but the signal:noise ratios are smaller than those seen with other compounds (Figure 3.22). This is unsurprising since its absorption spectrum (λ_{max} = 334 nm, 440 nm) is not optimal for quenching of FAM.

3.1.2.6 Other compounds tested

In addition to the established DNA-binding agents, some less well-known chromophores were screened for selective quenching of dsDNA (Figure 3.23). Acridine orange (λ_{ex} = 497 nm, λ_{em} = 525 nm), despite its use for *in vivo* staining of nucleic acids did not quench sufficiently, and as a result of its spectral similarity to FAM, increased the background fluorescence. For this reason, its 9-(methylsulfanyl)- derivative (λ_{ex} = 490-520 nm, λ_{em} = 555 nm), described in Section 2.4.1, was screened. Although the compound did bind to dsDNA, quenching was weak (probably due to weak binding), leading to low signal:noise ratios (Figure 3.24).
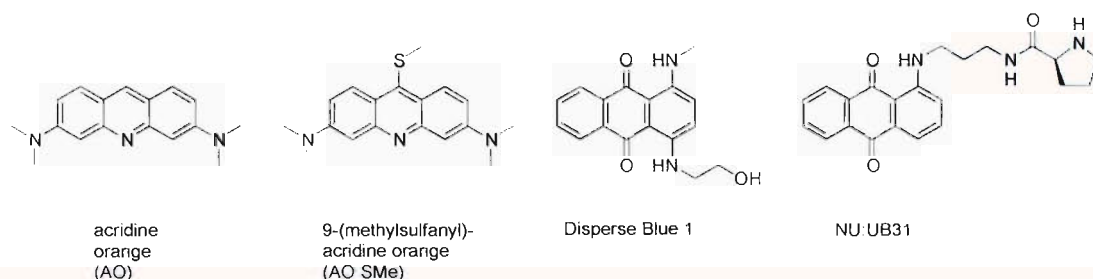
Figure 3.23 Other compounds screened.
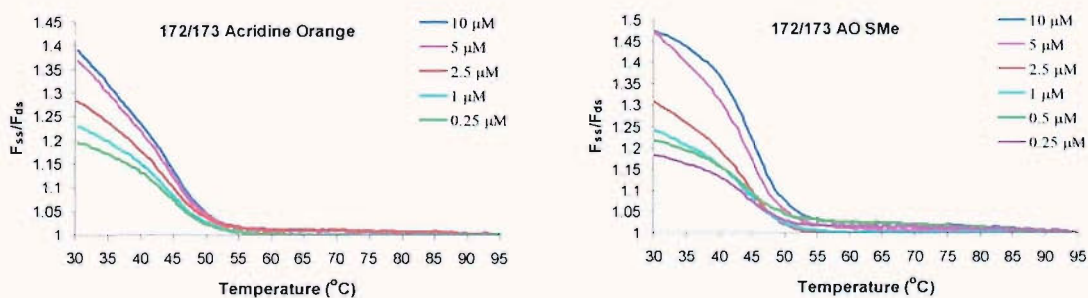
Figure 3.24 Signal:noise curves obtained for Acridine orange and 9-(methylsulfanyl)-acridine orange (AO-SMe).

Disperse blue 1 (λ_{max} = 595 nm, 660 nm), a fluorescence quencher previously conjugated to labelled oligonucleotide probes,[182] did not bind dsDNA to any significant degree, so quenching was not observed (Figure 3.25). The antitumour compound NU:UB 31,[204] whose binding to nucleic acids has not previously been studied, showed significant quenching of single stranded DNA, which led to poor

observed signal:noise ratios (Figure 3.25). As with the actinomycins, the apparent binding to single stranded DNA may be significant to its mode of action.
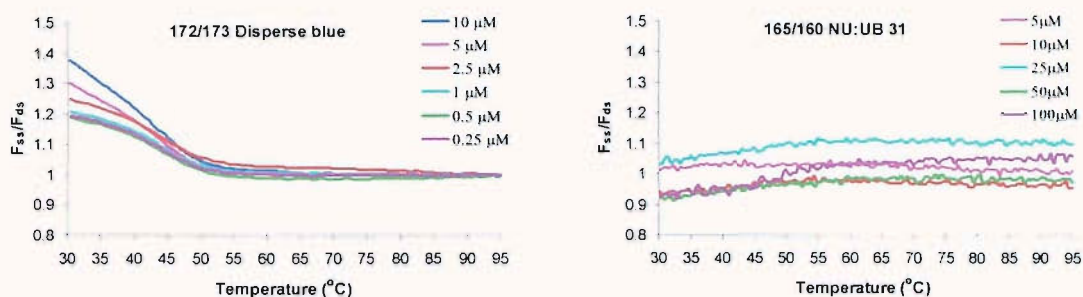
Figure 3.25 Signal:noise curves for disperse blue 1 (left) and NU:UB 31 (right).

3.1.3 Identification of compounds for further study

As outlined in 3.1, the compounds selected from the screening assay are those that exhibit the highest signal:noise ratios across the range of sequences. The four suitable compounds identified are therefore mitoxantrone, daunomycin, nogalamycin and ethidium bromide (Figure 3.26). Since the substituted anthraquinones have very similar properties, the most promising of these is daunomycin, which displays highest signal:noise ratios at 5 μM, compared to 10 μM for the other two. Intercalators that can be used at lower concentration are preferred because they are less likely to interfere with DNA polymerisation. Ethidium bromide has another useful property in addition to its quenching – it can be used to monitor the accumulation of PCR product in Channel 2 of the LightCycler. Also, the FRET mode of quenching, which is the longest ranging of quenching interactions, is likely to be most effective when the target is not a short duplex but a long PCR product. Furthermore, its sequence-independent binding means that no complications are introduced into primer and probe design. Ethidium bromide is therefore the best candidate for an intercalating quencher identified from the screening assay.
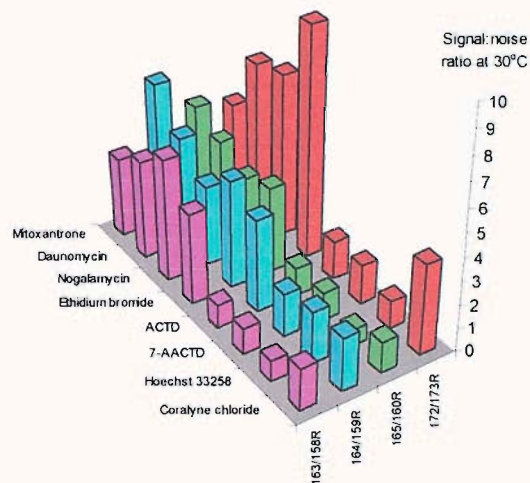
Figure 3.26 Maximum signal:noise ratios for the DNA-binding compounds studied observed at 30 °C. Values are reported at the optimal drug concentration.

3.2 Initial real time PCR with labelled probes and primer-probes

Primer-probe and hybridisation probe approaches (Figure 3.27) were both used to evaluate the efficacy of the intercalating quencher ethidium bromide. An exonuclease deficient polymerase was used to avoid cleavage of the probe. The W1282X locus of the ABCC7 gene was chosen for study. Primer and TaqMan®️ probe sequences for this locus have been previously published.[111]
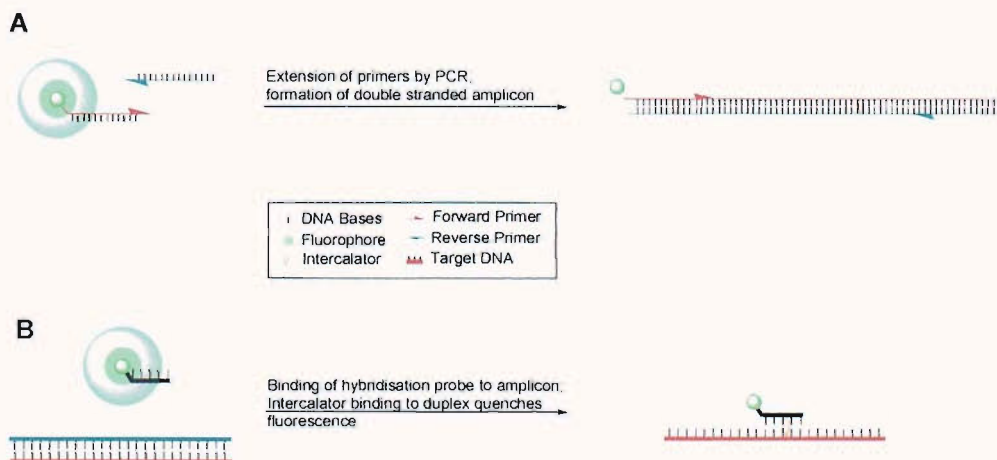
Figure 3.27 Primer-probe (A) and bimolecular hybridisation probe (B) approaches to real time PCR detection.

Oligonucleotide	Sequence (5'-3')	Description
FP1282	GATGGTGTGTCTTGGGATTCA	Forward primer
RP1282	TGGCTAAGTCCTTTTGCTCAC	Reverse primer
PRP1282	**FAM**-CTTTCCTCCACTGTTGC	Primer-probe
HP1282	**FAM**-CTTTCCTCCACTGTTGC-**p**	Hybridisation probe

Table 3.2 Primer and probe sequences. **FAM** = fluorescein; **p** = Phosphate (PCR stopper).

The progress of PCR in the presence of ethidium bromide was confirmed by SYBR Green monitored PCRs (data not shown). The bimolecular probe could not be used to generate a fluorescent signal. This is probably due to poor probe hybridisation, intercalator binding elsewhere in the amplicon, or both.

However, the primer-probe approach was successful (Figure 3.28). A large drop in fluorescence occurs as PCR product accumulates, due to ethidium binding to the labelled amplicon. The inverted trace resembles those usually obtained from real time PCR experiments. Unfortunately, since these experiments were performed on a Rotorgene instrument rather than the LightCycler, ethidium bromide fluorescence could not be monitored.
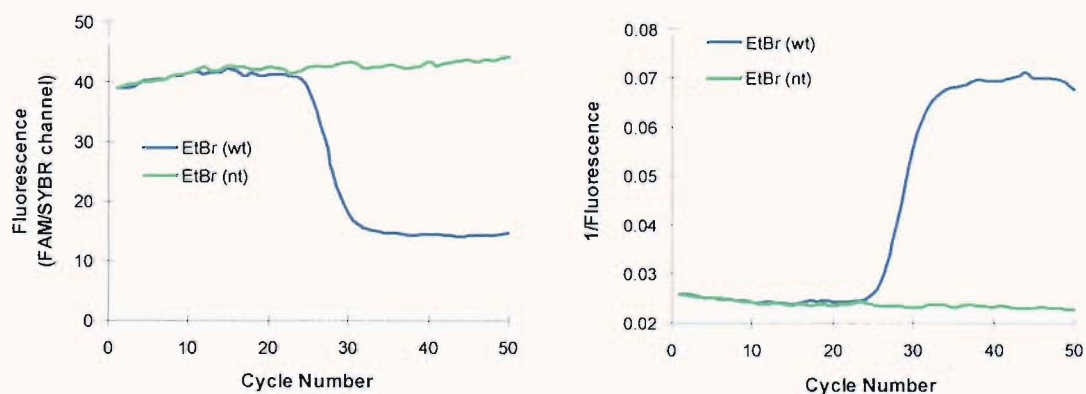
Figure 3.28 Uncorrected (left) and inverted (right) real time PCR traces obtained using primer-probe PRP1282 and ethidium bromide.

3.3 Extension of the screening assay to study drug-DNA binding –establishing stoichiometry and dissociation constant from fluorescence quenching titration

Studying drug-DNA interactions provides insight into their biological function. Two important parameters are the binding stoichiometry and equilibrium dissociation constant (K_d) of a drug to particular sequences. Both values can be determined from an appropriate saturation binding curve. Typically, fluorescence or absorbance changes of the ligand are used to monitor binding.[205] Fluorescence has the advantage of greater sensitivity, but can only be used if the ligand is a fluorophore. However, such measurements are not ideal, as the target nucleic acid concentration is usually varied as high concentrations of ligand can lead to high background fluorescence. A ligand-ethidium bromide competition method is also frequently used.[198] Drug binding displaces ethidium bromide, causing attenuation of ethidium fluorescence. A non-competitive fluorescence method applicable to non-fluorescent ligands has not yet been described. Fluorescence quenching of labelled nucleic acids, as used in Section 3.1.2 to screen DNA-binding compounds as fluorescence quenchers, has the potential to be adapted to provide such a method. The degree of quenching of the FAM-labelled duplex can be used to evaluate how much ligand is bound. The ligand chosen for study was the minor groove binder Hoechst 33258, whose binding properties have been well-established.[198, 206] The ligand was titrated into a 0.1 μM solution of duplex 165R/160R in a suitable hybridisation buffer. The curve obtained (Figure 3.29) could be used to evaluate binding stoichiometry. Solving the simultaneous equations of the two linear parts of the curve yields the expected value for B_{max} (total number of binding sites) of 0.091 μM, or 0.91 binding sites per duplex. With the binding stoichiometry in hand, the concentration of bound ligand at a given concentration could be calculated from:

$$[D]_b = ((F_0\text{-}F)/(F_0\text{-}F_b))\ B_{max}$$

Where $[D]_b$ = concentration of drug-DNA complex, F_0 = fluorescence observed in the absence of ligand, F = observed fluorescence, F_b = fluorescence of ligand-satu rated duplex and B_{max} = total number of binding sites.
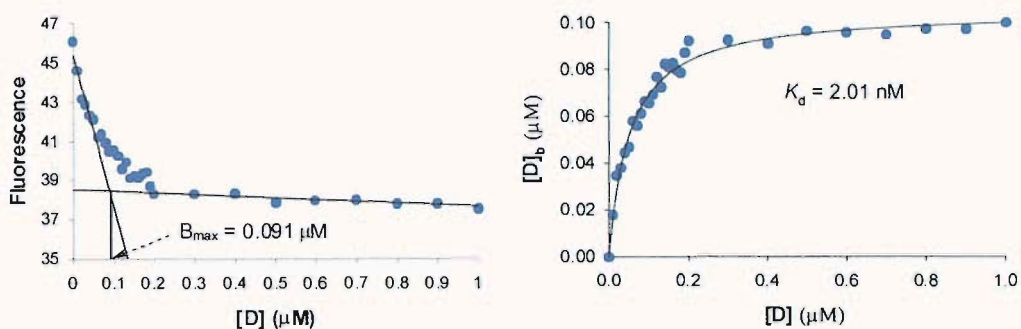
Figure 3.29 Fluorescence quenching titration (left) and saturation binding curve (right) obtained from fluorescein quenching by Hoechst 33258.

Plotting $[D]_b$ vs. total concentration of added drug $[D]$ gives a saturation binding curve (Figure 3.29), which can be used to determine the equilibrium dissociation constant, K_d:

$$K_d = [D][NA]/[D]_b$$

Where $[D]$ = total concentration of added drug, $[NA]$ = total concentration of nucleic acid and $[D]_b$ = concentration of drug-DNA complex.

Non-linear curve fitting, rather than Scatchard analysis, was found to be the most reliable method of calculating K_d. A value of 2.01 nM was obtained, close to those obtained for short duplexes containing similar A/T tracts from Hoechst 33258 fluorescence titration (3.12 nM)[206] and ethidium bromide competition (11.4 nM).[198]

3.4 Conclusions

A versatile screening methodology based on fluorescence melting has been developed and used to evaluate twelve DNA-binding compounds for use in a new genetic analysis technique. The screening method allowed study of selectivity of binding agents for dsDNA over ssDNA, efficiency of quenching and sequence dependence of quenching. The screening program identified four good candidates for use in the proposed genetic analysis technique: mitoxantrone, daunomycin, nogalamycin and ethidium bromide.

The use of ethidium bromide in the primer-probe variant of the intercalating quencher assay was demonstrated, giving good signal:noise ratios in real time PCR. This primer-probe method has advantages over others that have been previously described (Section 1.3.2.2). The method requires no secondary structure to be incorporated into the primer-probe and uses only one fluorescent label. These two factors simplify oligonucleotide synthesis and purification. Amplifluor primers,[79] hairpin primers[82] and Cyclicons[83] all have secondary structure and both Cyclicons and Amplifluor primers are labelled with a fluorophore and a quencher. Cyclicons introduce a further complication to oligonucleotide synthesis, which must be reversed partway through probe construction. Although double-stranded primers[84] and self-reporting PNA/DNA dimers[85] are linear probes modified with a single fluorophore, a separate quencher oligonucleotide or PNA oligomer must be used, raising the cost. In contrast, ethidium bromide can be obtained cheaply and in large quantities - 1 g costs ~£12. Based on 20 μL reaction volumes and 10 μM ethidium concentration, the cost of ethidium is ~£0.001 per assay. Given that the primer-probe is not modified at or near its 3'-end, ARMS primer-probes should be able to effect allelic discrimination, but this should be demonstrated in future work.

Of course, not all DNA-binding agents have been tested, and more could be screened in the future. The screening assay provides some basic principles about which ligands are likely to be suitable quenchers. Firstly, quenchers with λ_{max}>500 nm, as expected, appear to be the compounds that most effectively quench fluorescein. So, more ligands containing substituted anthraquinone chromophores (which generally have λ_{max}>500 nm), such as adriamycin, could be tested. Secondly, the results obtained suggest that tailing any primer-probe sequence with AATT at its 5'-end should provide a PCR product that would be quenched by the minor groove binder Hoechst 33258. This approach was not pursued because Hoechst 33258 was not a particularly good quencher (<60 %, even with the optimal sequence, Figure 3.19). Although this may be improved by ensuring the dye is nearer to the groove binder (perhaps by appending it to to the nucleobase of T,[207] or a shorter linker), there are many other A/T specific minor groove binders, such as berenil,[208] DAPI and SN 6999,[209] that could be tested.

The hybridisation probe (Figure 3.27B) approach failed, possibly as a result of poor binding of the probe. If this is the case, methods to circumvent this problem are

presented in Chapter 4. However, the possibility that the lack of signal is a result of ethidium binding to the amplicon rather than the probe-target duplex cannot be ruled out. If this is the cause of the problem, it may be overcome by optimising the ethidium bromide concentration in real time PCRs.

The use of the fluorescence quenching titration to study dissociation constants and binding stoichiometry appears to provide reliable data and could be used to gain information about less well-studied drug-DNA interactions, such as the binding of the actinomycins to ssDNA, or the binding properties of NU:UB 31.

In principle, since T_ms are obtained, fluorescence melting could also be used to provide thermodynamic information about drug-DNA interactions. UV-melting with van't Hoff analysis is often used to do this, but fluorescence melting can be more sensitive and provide higher throughput. This has been demonstrated in thermodynamic studies of triplexes.[210] Another possible advantage of fluorescence melting in studying the thermodynamics of drug-DNA interactions is that binding to 1-2 bases in ssDNA (as seen with the binding of actinomycins to ssDNA) gives a clear melting transition (Figure 3.16), which is unlikely to be seen in UV-melting experiments.

4. Problems with detecting PCR products with oligonucleotide probes and possible solutions

4.1 Competition for the target sequence in PCR products between its complementary amplicon strand and an oligonucleotide probe

As alluded to in Section 2.3.2, one factor in the relatively low signal obtained from the dual-labelled probes in real time PCR is the tendency of the short probe sequences (12 nt) to be displaced by the longer competing strand generated by the PCR (103 nt) (Figure 4.1). Although the short oligonucleotide probe may have a kinetic advantage over the competing strand of the amplicon, the competing strand forms 103 base pairs with the target compared with 12 base pairs for the probe, so is thermodynamically favoured. The equilibrium position in Figure 4.1 is therefore well to the right.
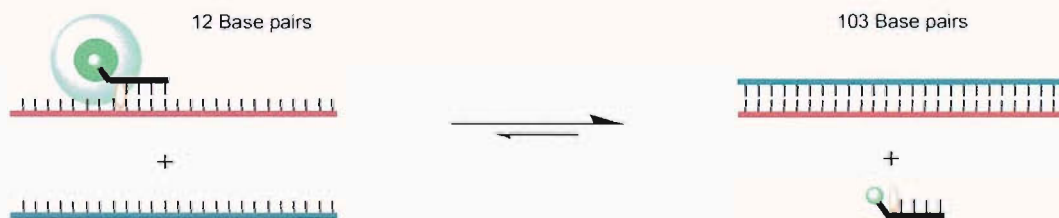
Figure 4.1 Displacement of fluorescent probe by the competing strand of the amplicon.

Fluorescence melting was used to demonstrate this phenomenon. The separate strands of the N1303K PCR product were produced by solid-phase oligonucleotide synthesis. It is clear that as the concentration of the competing strand is increased, the signal obtained from both the dual-labelled probe (FA12) and the Molecular Beacon (MB1303) is reduced (Figure 4.2). Since the signal is a result of probe hybridisation, the cause must be reduced binding of probe due to competition with the added strand.

Oligonucleotide	Description	Sequence (5'-3')
MB1303	Molecular Beacon	**FAM**-<u>CCCGCGCGG</u>-AACATTTAGAAAAAACTTGGA <u>TCCCGCGCGGG</u>-**MR**
FA12	Fluorophore-Intercalator probe	**FAM-ACD**-AAACTTGGATCC-**OCT**
T1303	Target strand of amplicon	TTTCTTGATCACTCCACTGTTCATAGGGATCCAAG TTTTTTCTAAATGTTCCAGAAAAAATAAATACTTT CTATAGCAAAAAAGAAAAGAAGAAGAAAGTATG
C1303	Competing strand of amplicon	CATACTTTCTTCTTCTTTTCTTTTTTGCTATAGAAA GTATTTATTTTTTCTGGAACATTTAGAAAAAACTT GGATCCCTATGAACAGTGGAGTGATCAAG AAA

Table 4.1 Sequences of N1303K oligonucleotides used to demonstrate hybridisation competition by the other strand of amplicon. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine, **MR** = Methyl Red dR, **OCT** = octanediol. FA12 is composed of 2'-*O*-methyl ribonucleotides. Molecular Beacon stem sequence is underlined.
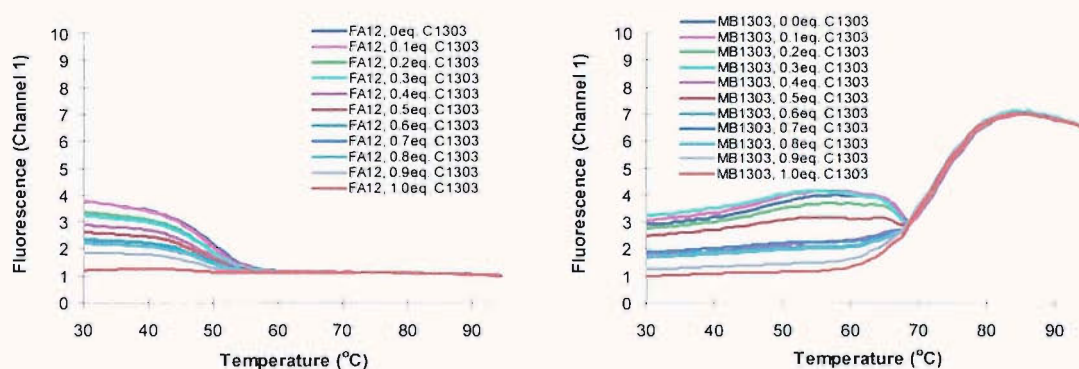
Figure 4.2 Fluorescence melting curves showing displacement of fluorophore-intercalator probe (left) and Molecular Beacon (right) by the competing strand of amplicon. One equivalent of target strand T1303 was included in each mixture.

4.2 Solutions to probe displacement by competing strand

The strand displacement problem demonstrated in Section 4.1 is likely to be a general limitation in the detection of PCR products by short oligonucleotide probes. If solutions to it could be found, the signal obtained from many genetic analysis methodologies could be enhanced, leading to either more sensitive or more cost-effective assays, requiring lower concentrations of fluorescently-labelled probes.

One solution may lie in using probes that have a higher affinity for their target sequence, shifting the equilibrium in Figure 4.1 to the left. The use of longer oligonucleotides as probes presents problems. Most importantly, the sensitivity of probes to mismatches reduces with probe length, making genotyping more difficult with longer probes. Also, the potential for undesirable probe secondary structure is increased, as is the risk of partial complementary with either the primer sequences or other genomic sequences. For these reasons, the use of longer probes is not a viable solution.

However, if the T_m of the probe-target duplex can be increased without increasing its length, then the problem could be overcome. Chemical modification of the probe is the most obvious way to achieve this. The use of modified nucleotides in the probe, such as substitutions of diaminopurine[211] and/or propyne dU[212] for dA and T residues respectively, or by the use of peptide nucleic acid (PNA)[213] or locked nucleic acid (LNA)[214] backbone modifications should enhance probe affinities without the need for longer probes.

Chemical modifications of oligonucleotide probes were not explored here, but several modifications of the real time PCR designed to overcome the probe displacement problem were investigated.

4.2.1 Primer-probe approach

Primer-probes are commonly used in genetic analysis (Section 1.3.2.2). This approach obviates the need for a separate hybridisation probe, since the primer is used to generate a fluorescent signal. A primer-probe version of the fluorophore-intercalator probe (Figure 4.3) described in Chapter 2 was designed and used in a real time PCR assay. The ARMS primer was designed to facilitate genotyping of the N1303K locus (Table 4.2) of the ABCC7 gene.[180]
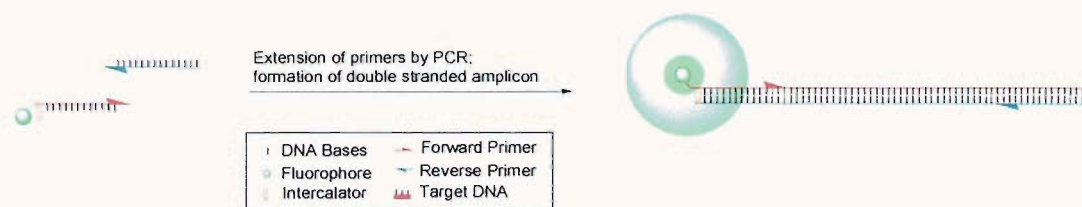
Figure 4.3 Expected mode of action of dual-labelled primer-probes

Oligonucleotide	Description	Sequence (5'-3')
PRP1303	Primer-probe	**FAM-ACD**-ACTGTTCATAGGGATCCAAG
FP1303	Forward primer	ATTTCTTGATCACTCCACTGTT

Table 4.2 Oligonucleotides used in primer-probe amplification of the N1303K locus. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine.
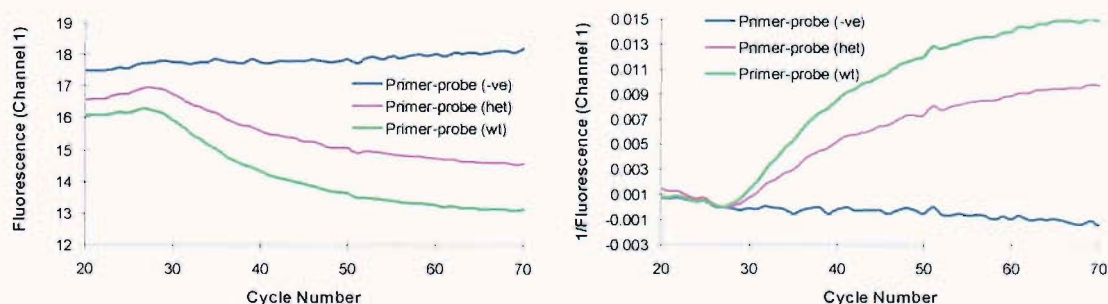
Figure 4.4 Uncorrected (left), and normalised and inverted (right) real time PCR traces obtained with allele-specific primer-probe PRP1303. wt = wild type, het. = heterozygous, -ve = negative control.

Although the primer-probe was expected to increase its fluorescence upon extension and hybridisation to the complementary strand of the amplicon (as the hybridisation probes had done, Table 2.2), real time PCR with the primer-probes led to a decrease in fluorescence with accumulation of PCR product (Figure 4.4). Nevertheless, the real time data could be normalised and inverted to produce a trace that displayed allelic discrimination between wild type and heterozygote genomic DNA samples.

Closer inspection of the duplexes formed by the primer-probe and the hybridisation probe allows the difference in properties to be rationalised. The hybridisation probe binds to the central region of the amplicon, leading to an overhang on either side of the probe, allowing 'three-quarter intercalation' of the acridine moiety, which in turn leads to increased fluorescence from fluorescein. The extended primer-probe hybridises to the unlabelled strand of the amplicon, forming a blunt-ended duplex. The acridine moiety can only stack on the end of the terminal base pair, which appears to lead to greater quenching of the fluorophore than in the single stranded form. It should be noted that this mode of action has not been tested in other stacking environments and therefore could be strongly sequence dependent.

4.2.2 Scorpion approach

As outlined in Chapter 1, adding a PCR stopper and primer sequence to a probe leads to intramolecular probing of the amplicon, a strategy used in the Scorpion,[111] Duplex Scorpion,[115] Angler[®130] and intramolecular TaqMan[®95] probe formats. Intramolecular probing should lead to enhanced probe hybridisation and therefore performance, so this strategy was applied to the fluorophore-intercalator probe format (Figure 4.5, Table 4.3).

	DNA Bases	~ PCR Stopper
	Fluorophore	▲ PCR Primer
	Intercalator	■ Probe Sequence
	Target DNA	~ Opposite PCR Primer
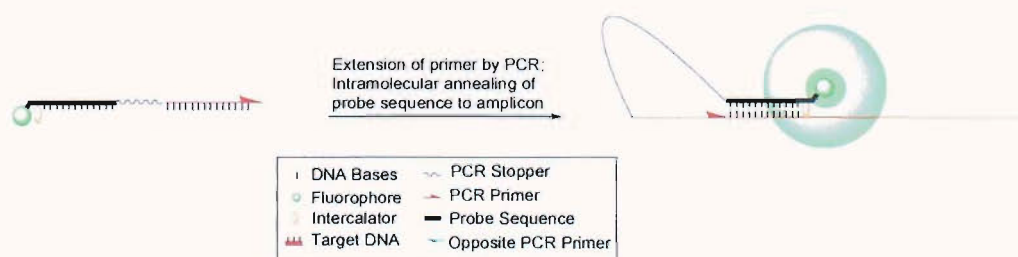
Figure 4.5 Fluorophore-intercalator Scorpion.

Oligonucleotide	Description	Sequence (5'-3')
FAS1303	Fluorescein-acridine scorpion	**FAM-ACD**-GGAACATTTAGAA**AAAACTTGG**AT CCC-**HEG**-TT**TCTTGATCA**CTCCACTGTTC
FP1303	Forward primer	ATTTCTTGATCACTCCACTGTT

Table 4.3 Oligonucleotides used in primer-probe amplification of the N1303K locus. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine, **HEG** = hexaethylene glycol (PCR stopper).

Although the PCR product could be successfully detected, the amount of signal generated was only similar to that obtained from the hybridisation probe, and discrimination of heterozygote from wild type genomic DNA was not possible (Figure 4.6). The reason for the low signal obtained may be that interaction of the acridine with the nucleobases at the 5'-end of the probe presents a kinetic impediment to hybridisation, which even intramolecular probing is unable to overcome.
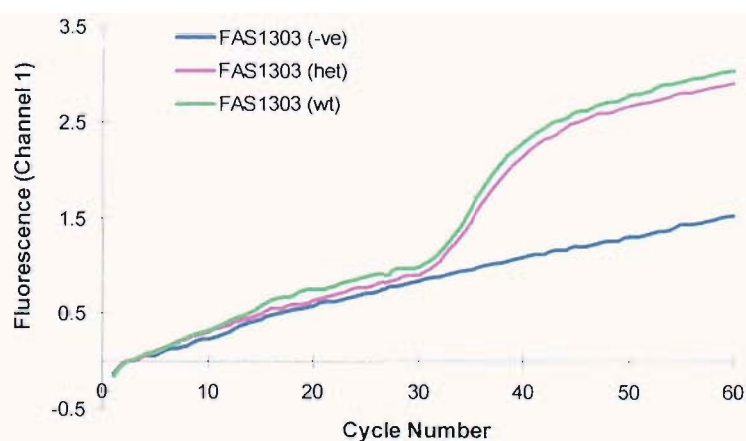
Figure 4.6 Real time PCR trace obtained using the fluorescein-acridine scorpion as the probe. wt = wild type, het. = heterozygous, -ve = negative control.

4.2.3 Asymmetric PCR

The competing strand is necessarily generated as a result of PCR – for a chain reaction to occur, two primers must be used, and so two amplicon strands are produced. Asymmetric PCR uses one primer in excess of the other, so that the initial chain reaction is followed by a linear phase that leads to production of an excess of

one of the amplicon strands.[215] The use of this method to produce an excess of target strand has been reported to result in an increase in sensitivity for Molecular Beacons in real time PCR, with a 50:1 ratio of target strand primer:competing strand primer giving the best results.[216] However, this approach was not successful with the dual-labelled probe FA12 and the standard primers FP1303 and RP1303 (Table 4.4) - the greatest signal was obtained from a 1:1 ratio of the two primers, i.e. under standard conditions (Figure 4.7). These conditions would have yielded the largest quantity (albeit double stranded) of amplicon.

Oligonucleotide	Description	Sequence (5'-3')
FP1303	Forward primer	ATTTCTTGATCACTCCACTGTT
RP1303	Reverse primer	CATACTTTCTTCTTCTTTTCTTT

Table 4.4 Sequences primers used for asymmetric PCR amplification of N1303K locus.
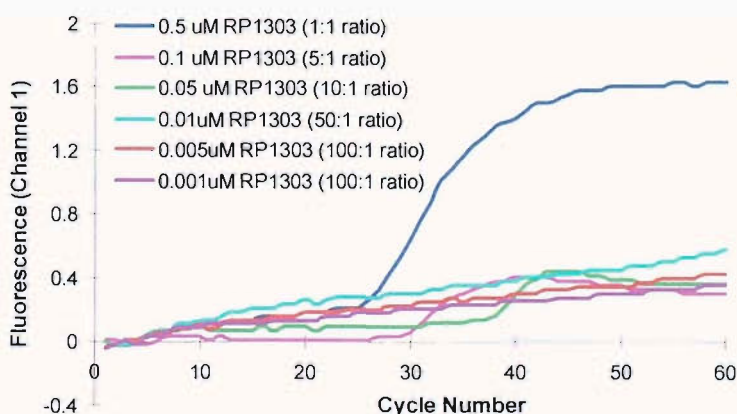
Figure 4.7 Real time asymmetric PCR traces detection of N1303K locus amplicons with 0.5 μM dual-labelled probe FA12, 5 μM of target strand primer FP1303 and varying amounts of competing strand primer RP1303. Negative controls omitted for clarity.

4.2.4 Incorporation of a destabilising nucleotide in PCR

Another possible solution to the hybridisation problem could be to reduce the affinity of the two strands of the amplicon for one another. This would require use of modified nucleotide triphosphates in the PCR, to selectively reduce the affinity of the competing strand for the target strand.

Some work has been done in this area by Thuong and Southern,[217, 218] who have used *N*4-ethyl-2'-deoxycytidine (d^{4Et}C) to reduce the sequence-dependence of duplex stability, the result of G:C base pairs forming three H-bonds and A:T base pairs forming only two. However, they have had difficulties in incorporating the *N*-4-ethyl-2'-cytidine triphosphate (d^{4Et}CTP) with thermostable *Taq* polymerase, a requirement for the preparation of PCR products containing the modification.

Another possibility is to use deoxyinosine triphosphate (dITP) in the PCR in place of dGTP. Inosine only forms two hydrogen bonds with cytosine, destabilising the resulting duplex (Figure 4.8).
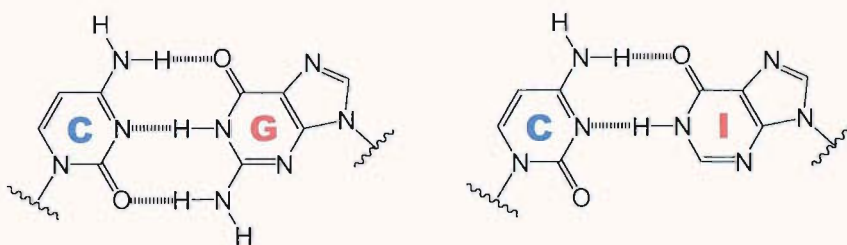
Figure 4.8 Base pairs formed by guanine (left) and hypoxanthine, the nucleobase component of inosine (right) with cytosine.

Since the probe sequence will be a synthetic oligonucleotide, it will contain the natural base (G), and should form more stable duplexes than the competing strand of the amplicon. Incorporation of dITP by DNA polymerases has long been established,[6] and partial replacement of dGTP by dITP has been used extensively to produce PCR products with reduced secondary structure for DNA sequencing.[219-221]

It is also possible that primers could continue to bind to the template strands when their concentration is further depleted, allowing more PCR product to be synthesised before the reaction plateaus.

4.2.4.1 Investigation of the effect of complete inosine substitution on fluorescence melting of a single stranded probe with synthetic amplicons

The potential benefits of inosine incorporation in real time PCR were demonstrated by fluorescence melting (Figure 4.9). The sequence chosen for study was the human MTHFR (Methylene Tetrahydrofolate Reductase) locus. Sequences of functional

Molecular Beacons,[102] as well as TaqMan® probes and Scorpions,[111] for this locus have been reported. This locus was chosen because the probe sequences contain a high proportion of G residues, and thus would potentially hybridise preferentially against an inosine-containing competing strand. The TaqMan® probe, TQM (Table 4.5), which shows a modest fluorescence enhancement upon duplex formation, was used to monitor probe hybridisation.

Oligonucleotide	Description	Sequence (5'-3')
TQM	TaqMan® probe	**FAM**-TGCGGGAGCCGATTT-**TAMRA**
CMG	Competing strand of amplicon	CTGACCTGAAGCACTTGAAGGAGAAGGTGTCTGC GGGAGCCGATTTCATCATCACGCAGCTTTTCTTTG AGGCTGACACATTCTTCCGCTTTGTGAAGGCATG CACCGACAT
TMG	Target strand of amplicon	ATGTCGGTGCATGCCTTCACAAAGCGGAAGAATG TGTCAGCCTCAAAGAAAAGCTGCGTGATGATGAA ATCGGCTCCCGCAGACACCTTCTCCTTCAAGTGCT TCAGGTCAG
CMI	Competing strand of amplicon (I- substituted)	CTIACCTIAAICACTTIAAIIAIAAIITITCTICIIIAICCIA TTCATCATCACICAICTTTTCTTTIAIICTIACACATTC TTCCICTTTITIAAIICATICACCIACAT
TMI	Target strand of amplicon (I- substituted)	ATITCIITICATICCTTCACAAAICIIAAIAATITITCAIC CTCAAAIAAAAICTICITIATIATIAAATCIICTCCCICA IACACCTTCTCCTTCAAITICTTCAIITCAI

Table 4.5 Sequences of MTHFR oligonucleotides used for fluorescence melting. **FAM** = fluorescein, **TAMRA** = tetramethylrhodamine, I = 2'-deoxyinosine.

It is immediately apparent that hybridisation to the single stranded I-containing target strand (TMI) results in an approximately 67 % greater fluorescence increase than hybridisation to the G-containing target strand (TMG). This is a result of fluorescence quenching by guanine, a well-known phenomenon.[122]

As with similar experiments performed with fluorophore-intercalator probes (Figure 4.2), increasing the concentration of the competing strand reduces the initial fluorescence, by causing less probe to be hybridised. Assuming 100% hybridisation

with the target strand alone, the amount of hybridised probe in the presence of one equivalent of each amplicon strand (the result of a symmetric PCR) at 30 °C can be estimated from:

$$[P]_b = (F_A - F_P)/(F_T - F_P)$$

Where $[P]_b$ = proportion of bound probe at 30 °C; F_A = Fluorescence observed at 30 °C in the presence of a 1:1 mixture of amplicon strands; F_P = Fluorescence of probe alone at 30 °C and F_T = Fluorescence of probe in the presence of 1 equivalent of target strand at 30 °C.

This is 32% with the G-containing amplicons, and 43% with the I-containing amplicons. The combination of the increased proportion of hybridised probe, and the reduction of quenching by the PCR product means that the fluorescence increase $(F_A - F_P)$ is approximately doubled when hybridising the probe to a 1:1 mixture of I-substituted target and competing strands. It was therefore concluded that this approach warranted further study.
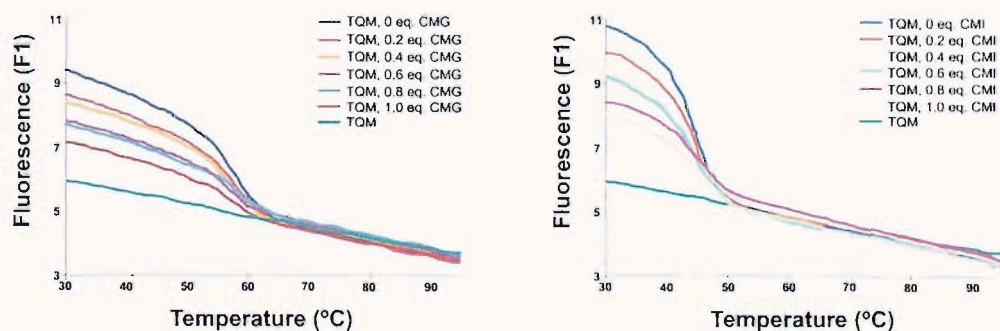
Figure 4.9 Fluorescence melting curves obtained for TaqMan® probe (TQM) in the presence of G- (left graph) or I-containing (right graph) synthetic amplicon strands (TMG/CMG or TMI/CMI). One equivalent of the appropriate target strand is included in all cases except for the melting of the TaqMan® probe alone.

4.2.4.2 Optimisation of PCR conditions for incorporation of dITP

If a destabilising nucleoside is to be incorporated in real time PCR assays, it is important to completely replace the natural dNTP in the reaction mixture. Not only will this produce the largest destabilising effect, but it will also ensure homogeneity of the PCR product, which is essential if genotyping relies on thermodynamic differentiation of mutant and wild type sequences (e.g. by selective probe hybridisation or melting curve analysis). Although dITP incorporation is used in DNA sequencing, partial replacement of dGTP is used. This is acceptable because the reaction relies on enzymatic selection to elucidate the sequence.

The use of dITP in PCR has been the subject of some interest in the literature. It has been reported that the introduction of dITP in PCR leads to decreased PCR efficiency and increased misincorporation, leading to frequent chain terminations.[222] Deoxyinosine triphosphate has also been used in PCR to induce random mutagenesis in the presence of a limiting concentration of one of the natural dNTPs,[223] but partial replacement of dGTP by dITP has been used to produce mutation-free PCR products with reduced secondary structure for sequencing.[223, 224] The only description of a non-mutagenic PCR with dITP totally replacing dGTP reported a 15% yield compared to when dGTP was used.[225] It was therefore important to establish conditions under which dITP could selectively replace dGTP in the PCR mixture and lead to a high yield of PCR product.

Initial attempts to amplify a 112bp region of the MTHFR gene with the reported primer pair (Table 4.6)[111] under standard conditions with dITP failed to give any product detectable on agarose gel (data not shown). Reasoning that inosine-containing PCR product would form a hybrid with the primers with lower affinity, and therefore lower T_m than the guanosine-containing product, the annealing temperature was reduced by 10 °C (from 55 °C to 45 °C). This gave weak PCR products when dITP replaced dGTP in the reaction mixture, as visualised on agarose gel (data not shown). The electrophoretic mobility of these products was unchanged relative to the products obtained with dGTP, implying that only full-length amplicons had been produced.

94

Oligonucleotide	Description	Sequence (5'-3')
FPM	Forward primer	CTGACCTGAAGCACTTGAAGG
RPM	Reverse primer	ATGTCGGTGCATGCCTTCAC

Table 4.6 Sequences of primers used for PCR amplification of MTHFR locus.
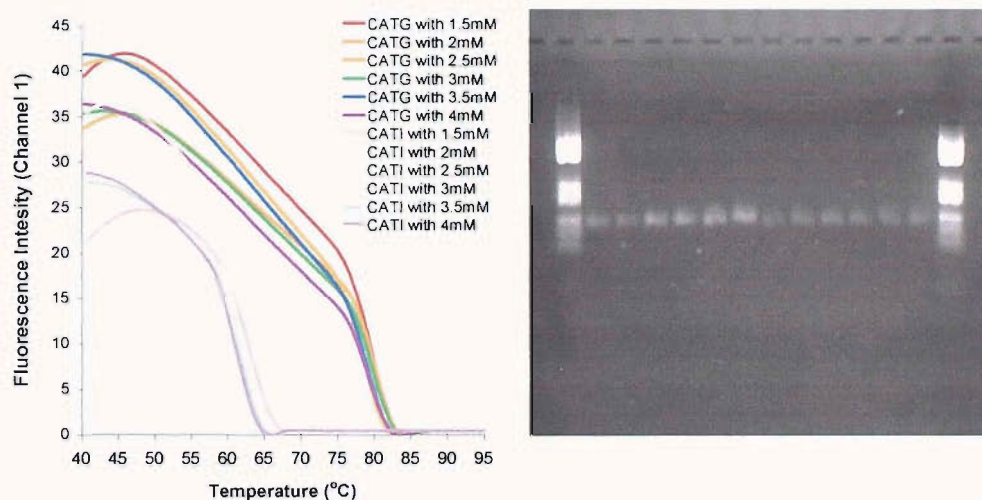
Figure 4.10 SYBR Green fluorescence melting curves (left) and agarose gel electrophoresis (right) of PCR products obtained with 2.5 U *Taq* polymerase at varying MgCl$_2$ concentrations. Lane 1. Molecular weight markers; 2. dGTP, 1.5 mM MgCl$_2$; 3. dGTP, 2.0 mM MgCl$_2$; 4. dGTP, 2.5 mM MgCl$_2$; 5. dGTP, 3.0 mM MgCl$_2$; 6. dGTP, 3.5 mM MgCl$_2$; 7. dGTP, 4.0 mM MgCl$_2$; 8. dITP, 1.5 mM MgCl$_2$; 9. dITP, 2.0 mM MgCl$_2$; 10. dITP, 2.5 mM MgCl$_2$; 11. dITP, 3.0 mM MgCl$_2$; 12. dITP, 3.5 mM MgCl$_2$; 13. dITP, 4.0 mM MgCl$_2$; 14. Molecular weight markers.

Further encouraging the PCR by addition of another equivalent of HotStarTaq polymerase gave much stronger bands for I-containing amplicons on the agarose gel, although still weaker than G-containing products (Figure 4.11). Again the electrophoretic mobility of the products obtained with dITP were identical to those obtained with dGTP, suggesting full-length products had been obtained in both cases. Post-PCR melting curves obtained by addition of SYBR Green to an aliquot of the reaction mixtures showed one distinct melting transition for each product, with T_m ~60 °C for I-containing PCR amplicons and ~80 °C for G-containing products (Figure 4.11). This, in conjunction with the electrophoresis results shows that dITP has been selectively incorporated in place of dGTP, yielding a single

95

product with greatly reduced melting temperature ($\Delta T_m \sim 20$ °C). A high rate of misincorporation of dITP, and opposite to I-residues in the template would be expected to produce a diverse mixture of PCR products, with different T_ms. This would lead to a broad melting transition corresponding to melting of many different species. These were not seen, only sharp transitions, presumably corresponding to melting of single products. Shortened products, corresponding to terminations as reported by Innis et al.[222] were not seen on the agarose gel.

Further optimisation of dITP incorporation was attempted by varying the magnesium chloride concentration from 3.0-7.5 mM. Divalent cations such as Mg^{2+} are known to significantly stabilise duplexes by relieving electrostatic repulsion of the polyanionic backbone of DNA, as well as being essential for polymerase activity. It might be expected that increasing $MgCl_2$ concentration could increase the T_m of primer-template hybrids and allow increased accumulation of PCR product, just as reducing the annealing temperature had. However, visual comparison of the intensity of bands on the agarose gel (Figure 4.11) and of the SYBR Green fluorescence intensity at 40 °C (Figure 4.11) suggest that the amount of PCR product obtained is more or less independent of $MgCl_2$ concentration, except that at high concentrations (>7.0 mM) the efficiency is significantly reduced.
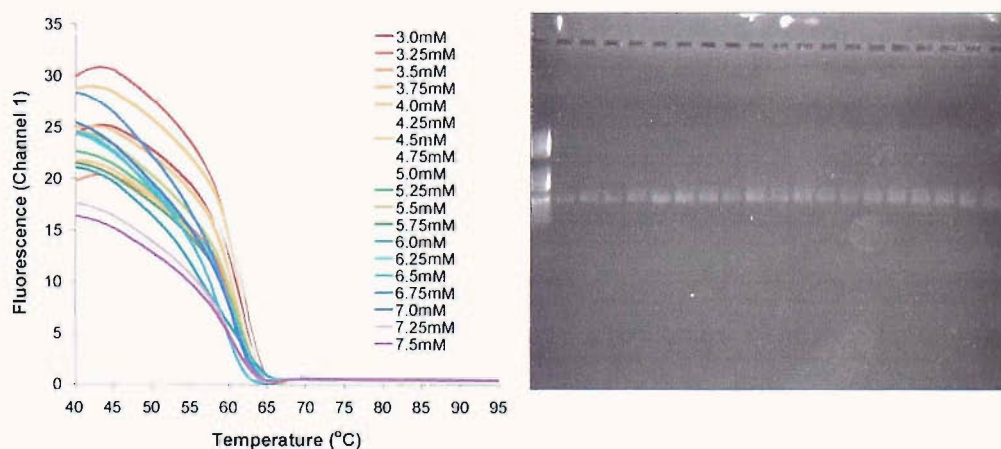
Figure 4.11 SYBR Green fluorescence melting curves (left) and agarose gel electrophoresis (right) of PCR products obtained with dITP at varying $MgCl_2$ concentrations. Lane 1. Molecular weight markers; 2. 3.0 mM $MgCl_2$; 3. 3.25 mM; 4. 3.5 mM; 5. 3.75 mM; 6. 4.0 mM; 7. 4.25 mM; 8. 4.5 mM; 9. 4.75 mM; 10. 5.0 mM; 11. 5.25 mM; 12. 5.5 mM; 13. 5.75 mM; 14. 6.0 mM; 15. 6.25 mM; 16. 6.5 mM; 17. 6.75 mM; 18. 7.0 mM; 19. 7.25 mM; 20. 7.5 mM.

However, whilst incorporation at low annealing temperature was working relatively efficiently, reducing the annealing temperature increases the probability of mispriming and primer-dimer formation in a real genetic analysis assay, both of which give rise to spurious amplicons which cannot be probed with a specific oligonucleotide probe, reducing signal. As a result, it is desirable to use the normal annealing temperature in such assays. This was achieved by tailing G-rich 'clamps' to the 5'-terminus of each primer (Figure 4.12, Table 4.7).
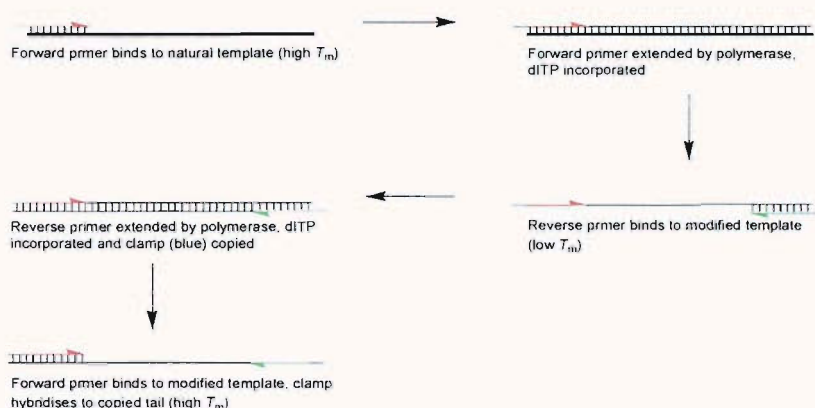
Figure 4.12 Mode of action of 'clamped' PCR primers.

Using the clamped PCR primers, it was possible to obtain high levels of PCR product at the normal annealing temperature, 55 °C. In fact, the intensity of the bands on agarose gel (Figure 4.13), and the fluorescence of SYBR Green at 40 °C (Figure 4.13) were roughly equivalent to that with dGTP, indicating the PCR had proceeded with the similar efficiency. The relative fluorescence of SYBR Green and ethidium bromide interacting with G- or I-containing duplexes are not known, so no absolute quantification can be made on the basis of either the gel or the melting curves. Importantly, the T_m of the whole amplicon was not significantly increased by the presence of the tail.

Oligonucleotide	Description	Sequence (5'-3')
CFPM	Forward primer	GGAGCTGACCTGAAGCACTTGAAGG
CRPM	Reverse primer	GAGGAGATGTCGGTGCATGCCTTCAC

Table 4.7 Sequences of clamped primers used for PCR amplification of MTHFR locus. The tailed G-rich 'clamp sequences' are underlined.
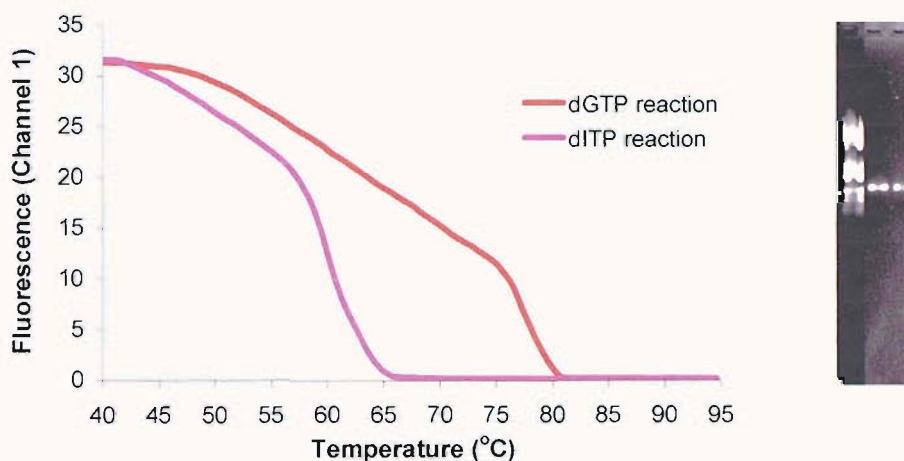
Figure 4.13 SYBR Green I fluorescence melting curves (left) and agarose gel electrophoresis (right) of PCR products obtained with clamped primers. Lane 1. Molecular weight markers; 2. dGTP reaction; 3. dITP reaction.

The T_ms of the 123 bp PCR products were also of interest. In addition to generating low T_m amplicons in real time PCR, it would also simplify the design of highly multiplexed methods (such as microarrays) if their stability were dependent on length alone. The T_m of the I-containing amplicon is almost identical to that estimated by substituting A/T for G/C, i.e. the I:C base pair is very close in stability to the A:T base pair. As the T_m of I-containing amplicon is close to that of the amplicon composed entirely of A/T nucleotides, its T_m is dependent only on length.

dNTP used	Estimated T_m (°C)	Observed T_m (°C)
dGTP	77	77
dITP	62	60

Table 4.7 Estimated and observed T_ms of PCR products obtained with clamped primers. Estimated T_ms are the result of 'nearest neighbour' calculations made using the parameters of Breslauer *et al.*[226] with the modifications described by Sugimoto *et al.*[227] Calculations were made by the program available online at http://www.basic.nwu.edu/biotools/oligocalc.htm using estimated concentrations of 0.4 µM for each amplicon strand (this assumes 80 % primer conversion) and 20 mM for Na⁺ ions. T_ms of I-containing amplicons were calculated by substituting A for G and T for C in the primer extension sequences.

4.2.4.3 dITP incorporation in real time PCR

Having developed strategies for efficient incorporation of dITP in place of dGTP in PCR to give low T_m amplicons, the potential for signal enhancement in real time PCR assays was investigated. The first assay investigated was the TaqMan® assay. The TaqMan® probe for the MTHFR locus, TQM (Table 4.5) was used with the PCR primers FPM and RPM (Table 4.6). Initial real time SYBR Green traces (Figure 4.14) indicate that the yield of PCR product was reduced by approximately 50 % when dITP was used in place of dGTP under these conditions. SYBR Green melting curves (data not shown) again indicated that the resultant amplicon had a reduced T_m compared to the G-containing product (58 °C as opposed to 80 °C).
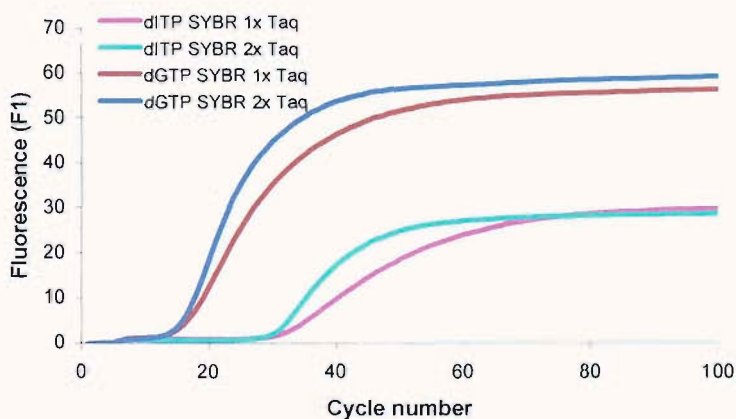
Figure 4.14 SYBR Green traces obtained for the TaqMan PCR with 0.5 U (1x) and 1 U (2x) *Taq* polymerase.

Despite this, almost no signal was obtained in the corresponding TaqMan® assay (Figure 4.15). This is likely due to the TaqMan® probe/dITP-containing amplicon being a less efficient substrate for the 5'-3' exonucleolysis catalysed by *Taq* polymerase. This is not unusual; modified oligonucleotide/RNA duplexes are frequently not substrates for RNaseH. When the concentration of *Taq* polymerase was doubled, although no increase in quantity of PCR product obtained was seen (Figure 4.14), TaqMan® signal generation occurred, giving 63% of the fluorescence obtained when dGTP was used (Figure 4.15). In these assays, doubling the amount of *Taq* polymerase had not increased the amount of PCR product obtained with dITP significantly. Previously, in simple PCRs where SYBR Green and oligonucleotide

probes had been omitted, doubling the concentration of *Taq* polymerase had given higher amounts of PCR product than seen here. This may be due to some inhibition of the dITP PCR by the presence of TaqMan® probes and/or SYBR Green.
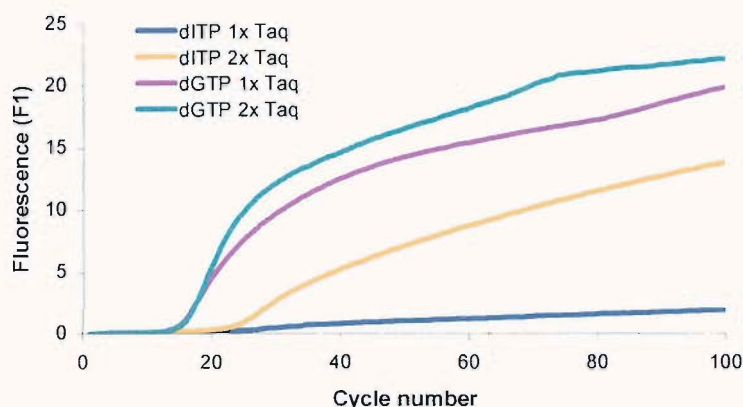
Figure 4.15. Real time PCR traces for TaqMan® assays in the presence of 0.5 U (1x) and 1 U (2x) *Taq* polymerase.

In order to further encourage the PCR and the TaqMan® cleavage, 5U of *Taq* polymerase, ten times the standard amount, were used. Although the amount of PCR product obtained was unchanged (SYBR Green traces not shown), the signal obtained from the TaqMan® assay was increased again, to similar levels obtained in the dGTP reactions with 0.5 U or 1 U of *Taq* polymerase and to 84% of that obtained in the corresponding dGTP reaction, despite producing ~40% (estimated from SYBR Green fluorescence) of the amount of PCR product (Figure 4.16). This indicates that the probe is indeed better able to hybridise to the I-containing amplicon.
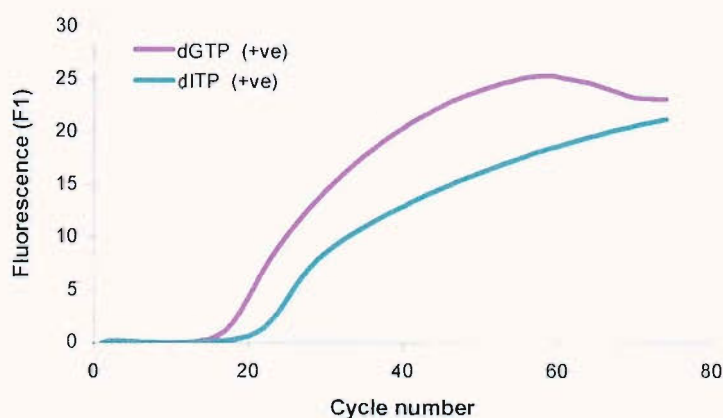
Figure 4.16 Real time TaqMan PCR traces obtained with 5 U *Taq* polymerase.

The use of dITP in real time PCR was also investigated in the context of a Molecular Beacon assay, which had previously been described for the MTHFR locus.[228]

Oligonucleotide	Description	Sequence (5'-3')
MBM	Molecular Beacon	FAM-<u>GCGAG</u>TGCGGGAGCCGATTT<u>CTCGC</u>-MR

Table 4.7 Sequence of MTHFR Molecular Beacon. **FAM** = fluorescein, **MR** = Methyl Red dR. Molecular Beacon stem sequence is underlined.

The SYBR Green traces (data not shown) showed a similar amount of PCR product was obtained for dGTP and dITP reactions as in the TaqMan® assay, with the expected decreased T_m for I-containing amplicons compared to G-containing ones. However, the Molecular Beacon assay with dITP failed to give any signal (Figure 4.17), although the assay with dGTP was successful. This is presumably due to the stem-loop (closed) structure being thermodynamically favoured over the open conformation. This is likely because, although the probe sequence is G-rich, it must form three C:I base pairs with the I-containing target strand, which may lead to a less favourable ΔH for the opening process.
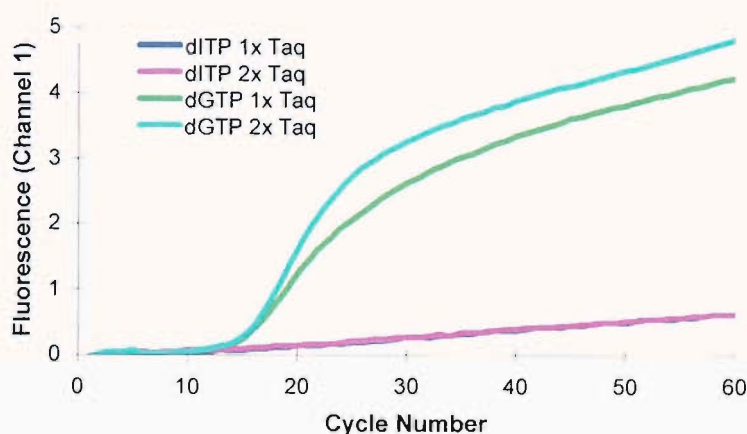
Figure 4.17 Real time Molecular Beacon PCRs with 0.5 U (1x) and 1 U (2x) *Taq* polymerase.

4.3 Conclusions

A general limitation to genetic analysis has been demonstrated and several possible solutions to it investigated. The tendency of Molecular Beacons, fluorescein-acridine probes and TaqMan® probes to be out-competed for their target strand of the amplicon by the other strand produced by PCR was demonstrated by fluorescence melting (Figure 4.2).

Three previously described methods were used in an attempt to increase the signal obtained from fluorescein-acridine probes. The first, using a fluorophore-intercalator labelled primer-probe did allow successful amplicon detection in real time PCR, but the subtle difference in the fluorophore/intercalator environment led to a fluorescence decrease on hybridisation (Figure 4.4). In any case, the primer-probe approach is not really a solution to the hybridisation problem, but a method that avoids hybridisation. The disadvantages associated with primer-probes outlined in Section 1.3.2.2, i.e. that they are liable to score false positives if mispriming or primer-dimer formation occurs are inherent to the method.

Attaching one of the primer sequences to the fluorophore-acridine probe *via* a PCR stopper, thus causing unimolecular probing to occur, was unsuccessful. The amount of signal obtained and the discrimination properties of the 'fluorescein-acridine Scorpion' were poorer than those observed for the bimolecular probe (Figure 4.6). The reasons for this are unclear. Greater binding of the probe to a mismatched target might be expected since the probing is intra- rather than intermolecular. However, this has not been observed in previous studies of Scorpions, which generally display excellent discrimination of mismatches.[111, 115] The use of an ARMS primer conjugated to a fluorophore-intercalator probe could facilitate allelic discrimination. The lack of signal increase suggests that there is a kinetic impediment to hybridisation resulting from the intercalator interacting with nucleobases at the 5'-end of the probe, in addition to the short nature of the probe meaning that hybridisation is accompanied by a small gain in enthalpy relative to formation of the amplicon duplex.

Asymmetric PCR was used in an attempt to produce a surplus of the target strand. This gave a smaller signal increase than the symmetric PCR, which is most likely due to asymmetric PCR yielding a lower concentration of target strand (Figure 4.7). The use of the limiting (competing strand) primer at very low concentration will also

reduce its T_m, and could lead to an inefficient exponential phase. Very recently, a modification to asymmetric PCR, Linear-After-The-Exponential PCR (LATE-PCR) has been described, which applies design criteria to the limiting primer sequence that ensure a high yield of amplicon from the exponential phase.[229] The length and sequence of the limiting primer is adjusted so its T_m is greater than or equal to the T_m of the excess (target strand) primer once the relative concentrations have been taken into account. The method has been validated in Molecular Beacon assays (leading to an increase in signal change of 80-250 %) and is likely to be applicable to the covalent and non-covalent fluorophore-intercalator probes described here.

A novel approach to facilitate probe binding, based on the incorporation of a destabilising base into PCR amplicons was studied. Proof of principle was obtained from fluorescent melting, which showed that a ~100 % increase in signal strength would result from replacement of all guanosine residues in the PCR product by inosine (Figure 4.9). The increase in fluorescence is due to a combination of reduced quenching by the nucleobases in the amplicon, and increased levels of probe hybridisation. Incorporation of dITP in PCR was found to be efficient under conditions of reduced annealing temperature and higher concentration of *Taq* polymerase (Figure 4.10). Alternatively, 'clamped' primers with G-rich tails could be used at the normal annealing temperature (Figure 4.13). In general, though, less PCR product was obtained with dITP than with dGTP. Using other polymerases that accept modified dNTPs more readily could perhaps increase the yield of I-substituted amplicons.

Application of this methodology to TaqMan[®] or Molecular Beacon assays did not improve their performance, for specific reasons related to their mode of action. TaqMan[®] relies on exonucleolytic cleavage of the probe, which is less efficient if the probe-target duplex contains modified nucleotides (Figures 4.15 and 4.16). Molecular Beacons have a hairpin loop structure that must be carefully designed to allow efficient hybridisation and discrimination. The stem sequence must not be too stable or hybridisation to the target sequence will be impaired, and must not be too unstable, or background fluorescence will be high and mismatch discrimination poor. Molecular Beacons designed to hybridise to normal, G-containing amplicons probably have to be redesigned with lower T_m stems to probe I-containing PCR products. This could be investigated further.

Perhaps what the results in this Chapter reinforce most of all though, is the need for single stranded probes. The fluorescein-acridine probes described here (although functional) display surprisingly poor hybridisation properties, given that they contain an intercalator that should stabilise probe/target duplexes, and were not studied in I-substituted real time PCR because it seemed unlikely that they could be a viable alternative to other available technologies. Other linear probes that fluoresce upon hybridisation such as the non-covalent fluorophore-intercalator probes (Chapter 3) or others described in the literature (Section 1.3.2.5) could be tested in conjunction with I-substituted PCR. Truly non-competitive hybridisation probes, which do not require disruption of the amplicon duplex for binding, such as those based on hairpin polyamides[139] or triplex-forming oligonucleotides could be the best way of obtaining the greatest possible signal.

5. Synthesis of long oligonucleotides with hydrophobic tags for use in genetic analysis

5.1 The synthesis of long oligonucleotides

5.1.1 Oligonucleotide probes used in MLPA

The ligation-based MLPA assay (multiplex ligation-dependent probe amplification, Section 1.3.3.2) is a very useful tool in genetic analysis.[153] In the two years since its first appearance in the literature, 22 further papers have been published describing its use in determining gene dosage,[155] SNP genotyping,[153] large genomic deletions,[157] genomic rearrangements[158, 230] and gene regulation analysis.[231]

The technique relies on the use of two probe oligonucleotides (Figure 5.1), which are ligated together, then amplified by PCR (Figure 1.32). The PCR products are separated by gel or capillary electrophoresis. The varying length of the 'stuffer sequence' allows identification of the original probe from the electrophoretic mobility of the PCR product. In this way, multiplexing of up to 79 probe sets has been achieved.[232] Modern electrophoresis instruments allow detection of different dyes in separate channels, and two-colour MLPA has already been described.[232] Two (or more) colour detection allows for greater multiplexing without increasing the range of probe lengths.

Short MLPA probe:
Synthetic oligonucleotide
40-49nt

Long MLPA probe: 5'-phosphorylated
oligonucleotide derived from plasmid
80-420nt

5'_____3'

Universal primer sequence (19nt) | Hybridising sequence (21-30nt)

5' PO_4^{2-} _____ 3'

Hybridising sequence (25-43nt) | "stuffer" sequence (19-370nt) | Universal primer sequence (36nt)
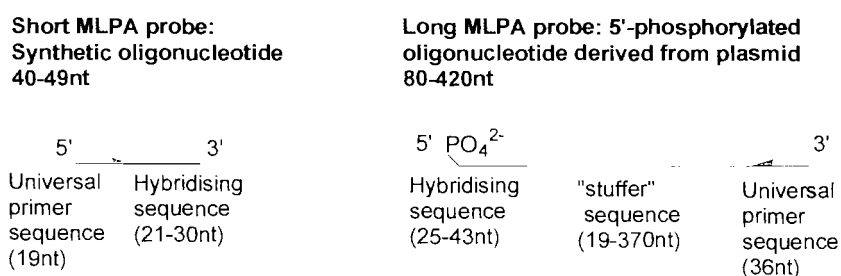
Figure 5.1 Representation of the two probe oligonucleotides used in MLPA.

The long phosphorylated probe is produced by laborious cloning methods. In the original paper by Schouten *et al.*,[153] it was stated that "chemically synthesised oligonucleotides of this length (80-440 nt) are not commercially available in the

105

quality needed for MLPA." Subsequently, smaller probe sets have been composed of synthetic oligonucleotides,[232] but highly multiplexed assays use plasmid-derived probes.

Oligonucleotides of up to 150 nt can be produced by automated solid-phase oligonucleotide synthesis. Assuming the universal primer sequence and the hybridising sequence comprise 50 nt of each ligated probe, this leaves 100 nt for the longest stuffer sequence. If a stuffer sequence is included in both ligation probes, the total stuffer sequence length can therefore be varied between 0 and 200 nt. With an electrophoretic resolution of 2 nt available, a 100 member probe set is therefore available, per colour.

5.1.2 Available methods for synthesis of long oligonucleotides

Although automated DNA synthesis can be used to produce oligomers of >100 nt, separation of the full-length product from truncated failure sequences is problematic. Since oligonucleotides are usually purified by reversed-phase HPLC (RP HPLC),[233] introduction of a hydrophobic group in the last step of DNA synthesis ensures full-length sequences are more effectively retained on the HPLC column than failure sequences, allowing them to be obtained in high purity. Typically, the 4,4'-dimethoxytrityl (DMT) protecting group of the last nucleotide is left on after oligonucleotide synthesis, providing a hydrophobic handle for purification that is later removed by acid treatment, yielding oligonucleotides unmodified at the 5'-terminus (the 'trityl-on' method, Scheme 5.1).
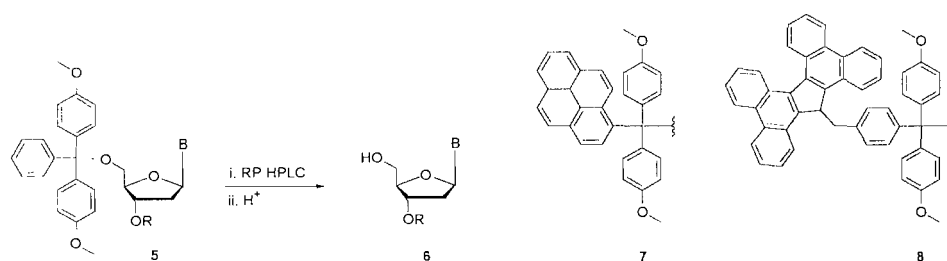
Scheme 5.1 The 'trityl-on'strategy for oligonucleotide synthesis.

Unfortunately, the DMT group traditionally used in DNA synthesis[234] is insufficiently hydrophobic to allow effective separation of products >50 nt in length.

106

As a result, modified protecting groups that are more lipohilic have been developed. Initially, Goertz and Seliger introduced longer alkyloxy- groups at the 4 and 4' positions of the trityl group,[235] but these required deprotection with strong acid. Later, extended aromatic carbocycles were introduced in place of the phenyl ring of DMT. Fourrey et al.,[236] and Ramage and Wahl,[237] subtituted pyrenyl (7) and 17-tetrabenzo[a,c,g,i]fluorenylmethyl (8) groups for the phenyl ring, facilitating greater separation by RP HPLC.

5.1.4 Available methods for the synthesis of 5'-phosphorylated oligonucleotides

Several monomers for incorporating a 5'-phosphate into synthetic oligonucleotides have been presented (Figure 5.2). As with unmodified oligonucleotides, it is difficult to separate failure sequences from the full-length, phosphorylated product for sequences longer than 50 nt. It is therefore desirable to incorporate a transient hydrophobic tag to facilitate purification by RP HPLC, which is later removed and a phosphate generated.

Of the previously described phosphate synthons, sulfonylethyl derivative 9,[238] and the bis-cyanoethoxy- functionalised monomer, 10,[239] generate phosphate in the initial deprotection, and the (p-nitrophenyl)ethyl phosphoramidite 11[240] is only moderately hydrophobic. The remaining monomers 12,[241] 13[242] and 14[243] are of similar lipophilicity. Of these, the acid labile DMTO- and TrtS- groups in 13 and 14 offer the mildest deprotection. Compounds 9 and 13 are used commercially for phosphorylated oligonucleotide synthesis in the 'trityl-off' and 'trityl-on' modes respectively.
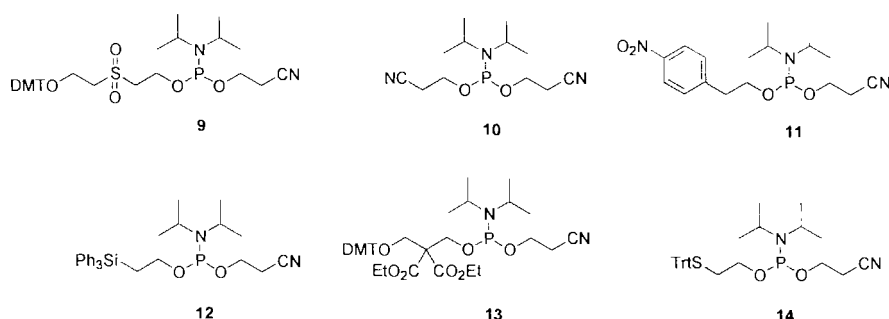
Figure 5.2. Phosphoramidite monomers used for incorporation of a 5'-phosphate in oligonucleotide synthesis.

107

For synthesis of MLPA probes, >100mer oligonucleotides incorporating a 5'-phosphate would have to be synthesised. To obtain pure products, the transient protecting group must be very hydrophobic, deprotected under very mild conditions and generate phosphate efficiently. Acidic deprotections are to be avoided if possible, since there is a risk of depurination and strand scission.

5.2 Strategy for sythesis of long MLPA probes

The proposed method for oligonucleotide synthesis utilises a separate hydrophobic tagging monomer for each MLPA probe. For the probe whose 3'-hydroxyl group is to be ligated, a permanent lipophilic tail, based on hexadecanol, is incorporated at the 5'-end. The hexadecyl group should provide a handle for RP HPLC, and ensure the 5'-terminus is chemically and biologically inert, which could be important in highly multiplexed assays. For 5'-phosphorylation, a modified version of Connolly's 2-(triylthio)ethyl phosphoramidite monomer **14** is to be used. The more labile 4,4'-dimethoxytrityl (DMT) group is used for thiol protection, in order to allow milder deprotection conditions to be used. Instead of a 2-thioethanol backbone, 2,3-dimercaptopropanol is used, providing a 'dummy' thiol to which a second hydrophobic DMT group can be attached. The synthetic targets were therefore phosphoramidite monomers **15** and **16**.
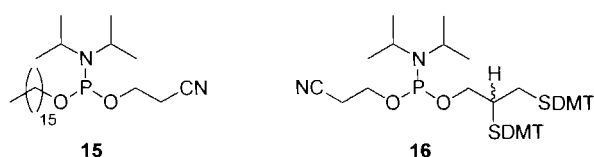
Figure 5.3. Hydrophobic tagging monomers for MLPA probe synthesis.

5.2.1 Synthesis of hydrophobic tagging monomers

2-Cyanoethyl-*N,N*-diisopropyl-(hexadecyl) phosphoramidite, **15**, was obtained directly from phosphitylation of hexadecanol, **17**. Synthesis of the *bis*-DMT protected **19** required some optimisation. Attempted protection of the thiol groups of 2,3-dimercapto-1-propanol, **18**, by treatment with two equivalents of 4,4'-

dimethoxytrityl chloride in pyridine yielded a mixture of *O-* and *S*-protected products, presumably to the reduced reactivity of the secondary thiol. Use of the stronger base *N,N*-diisopropylethylamine to effect thiol deprotonation allowed selective *S*-tritylation, affording **19** in 72 % yield. The structure of alcohol **19** could be assigned unambiguously by examination of its ^{13}C NMR spectrum. The chemical shifts of the tertiary sp^3 carbons of the DMT groups were 65 and 66 ppm, demonstrating that the adjacent heteroatom was sulfur, and not oxygen. The chemical shift of this carbon when adjacent to oxygen is ~85 ppm. Phosphitylation gave the phosphoramidite monomer **16** for use in automated DNA synthesis.
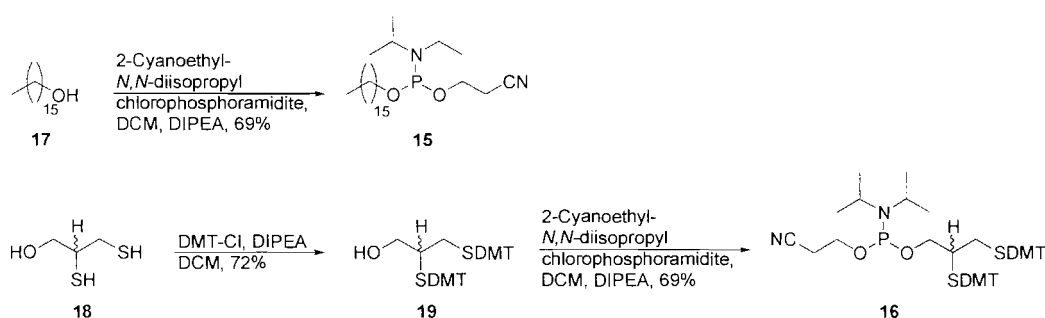
Scheme 5.2 Synthesis of hydrophobic tagging monomers **15** and **16**.

5.2.2 Initial use of the *bis*-DMT monomer in oligonucleotide synthesis

With the *bis*-DMT protected monomer **16** in hand, an oligothymidine nucleotide was synthesised with modification at the 5'-end, omitting a final detritylation step to allow 'trityl-on' HPLC purification. However, retention time on RP HPLC was unexpectedly short, less than that for an authentic sample of unmodified T_{12} (Figure 5.4). This suggested *S*-detritylation had occurred during oligonucleotide synthesis, leading to generation of phosphate in oligonucleotide deprotection. *S*-detritylation during the iodine oxidation, followed by deprotonation in concentrated aqueous ammonia would give the thiolate **21**. This could then γ-eliminate phosphorylated oligonucleotide **22** as a leaving group to give the episulfide **23** (Scheme 5.3), analogous to the mechanism established by Connolly.[243] Electrospray mass spectrometry confirmed the identity of the oligonucleotide obtained (Table 5.1).
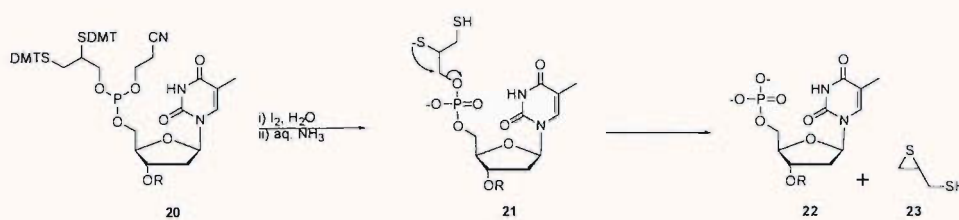
Scheme 5.3 Generation of 5'-phosphate by *bis*-DMT dimercaptopropanol monomer after detritylation in oligonucleotide synthesis.
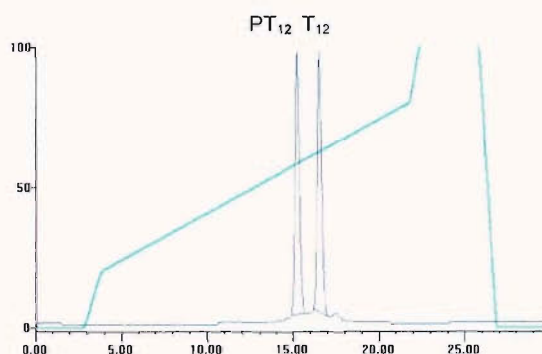
Figure 5.4 RP HPLC chromatogram of T_{12} and phosphorylated T_{12} (co-injection). X-axis: Time (minutes); Y-axis: Relative OD.

Oligonucleotide	Sequence (5'-3')	ESMS	
		Observed	Expected
PT12	p-TTTTTTTTTTTT	3668.0	3688.4

Table 5.1 Sequence and mass spectrometry data for 5'-phosphorylated oligothymidine; **p** – phosphate.

This demonstrates that the monomer can be used to generate 5'-phosphorylated oligonucleotides *via* 'trityl-off'synthesis, the preferred method for short oligomers, and that the DMT group is cleaved under milder conditions than the Trt group used in the 2-(triylthio)ethyl phosphoramidite **14**, which withstands standard oxidation conditions. This introduces an extra level of versatility over Connolly's phosphorylating reagent, which cannot be removed in automated DNA synthesis, (even by the detritylation step), so cannot be used for 'trityl-off' synthesis.

110

5.2.3 Short oligonucleotide synthesis with hydrophobic tagging monomers

The two hydrophobic tagging monomers **15** and **16**, and the commercially used chemical phosphorylation reagent II, **13**, were evaluated in the synthesis of short oligomers (Table 5.2). Deprotection of S-DMT groups during oxidation was avoided by using the alternative oxidising reagent $(1S)$-$(+)$-$(10$-camphorsulfonyloxaziridine)[244] for the final oxidation step.

This allowed assessment of the relative lipophilicities of the three reagents. The retention times decrease in the order **16** (bis-DMT dimercaptopropanol)>**15** (hexadecanol)>**13** (chemical phosphorylation reagent) (Figure 5.5 A, C and F, Table 5.2). As a result, the two new hydrophobic tagging monomers displayed the maximal separation between the full-length, tagged oligonucleotide and failure sequences.

Also of interest are the relative amounts of tagged and untagged sequences present in the chromatograms. Unsurprisingly, since the hexadecyl- modification is chemically inert, it is retained to the highest degree (Figure 5.5A). The bis-tritylated BDT12 (Figure 5.5C) compares favourably to the $mono$-tritylated CP12 (Figure 5.5F) in this respect too. In general, S-DMT groups are more stable to acidic conditions than their O- counterparts, so differing amounts of detritylation in the work-up is the probable cause of this disparity.

Oligonucleotide	Sequence (5'-3')	HPLC Retention time (minutes)		
		'Trityl-on'	Detritylated	Phosphorylated
HT12	**H**-TTTTTTTTTTTT	24.8	-	-
BDT12	**BD**-TTTTTTTTTTTT	25.5	13.1	12.5
CP12	**CP**-TTTTTTTTTTTT	22.5	n.d.	12.5

Table 5.2 Sequences and HPLC retention times of T_{12} oligonucleotides functionalised with hydrophobic residues. **H** = Hexadecanol, **BD** = Bis-DMT dimercaptopropanol, **CP** = Chemical Phosphorylation reagent II.

Deprotection of the bis-DMT groups was accomplished under exceptionally mild deprotection conditions. Heating in 2 M TEAA buffer (pH 6.5) at 55 °C overnight was sufficient for complete S-detritylation (Figure 5.5D). This is in contrast to its stability in work-up of the oligonucleotide. However, it is also true that though deprotection conditions for chemical phosphorylation reagent II are more acidic, the

low temperature (25 °C) and short time required mean that depurination and other side reactions are fairly unlikely. The conditions for elimination in both cases are mild and rapid.

Mass spectrometry confirmed the mechanisms of phosphate generation (Scheme 5.4). ESMS indicates that the result of the TEAA deprotection is the disulfide **25** (Table 5.3, Scheme 5.4), so the oxidative nature of the deprotection may explain the weakly acidic conditions required. It should be noted that several disulfides with one or two S-S bonds are possible, and these could not be distinguished by mass spectrometry. These results confirm that both hydrophobic tags developed are excellent handles for RP HPLC and produce oligonucleotides of the required structure for use as MLPA probes.

Oligonucleotide	MS Trityl-on		MS detritylated		MS Phosphorylated	
	Observed	Expected	Observed	Expected	Observed	Expected
HT12	3894.0*	3893.8*	-	-	-	-
BDT12	4380.0	4379.3	7543.0	7545.1	3668.0	3688.4
CP12	4171.0	4173.0	3870.0	3870.6	3668.4*	3689.4*

Table 5.3 Mass spectrometry data obtained for oligonucleotides in 'trityl-on' study. Data were obtained by either ESMS or MALDI-TOF MS. *Indicates MALDI-TOF data, where [M+H]' ions are obtained. Expected masses are corrected to reflect this.
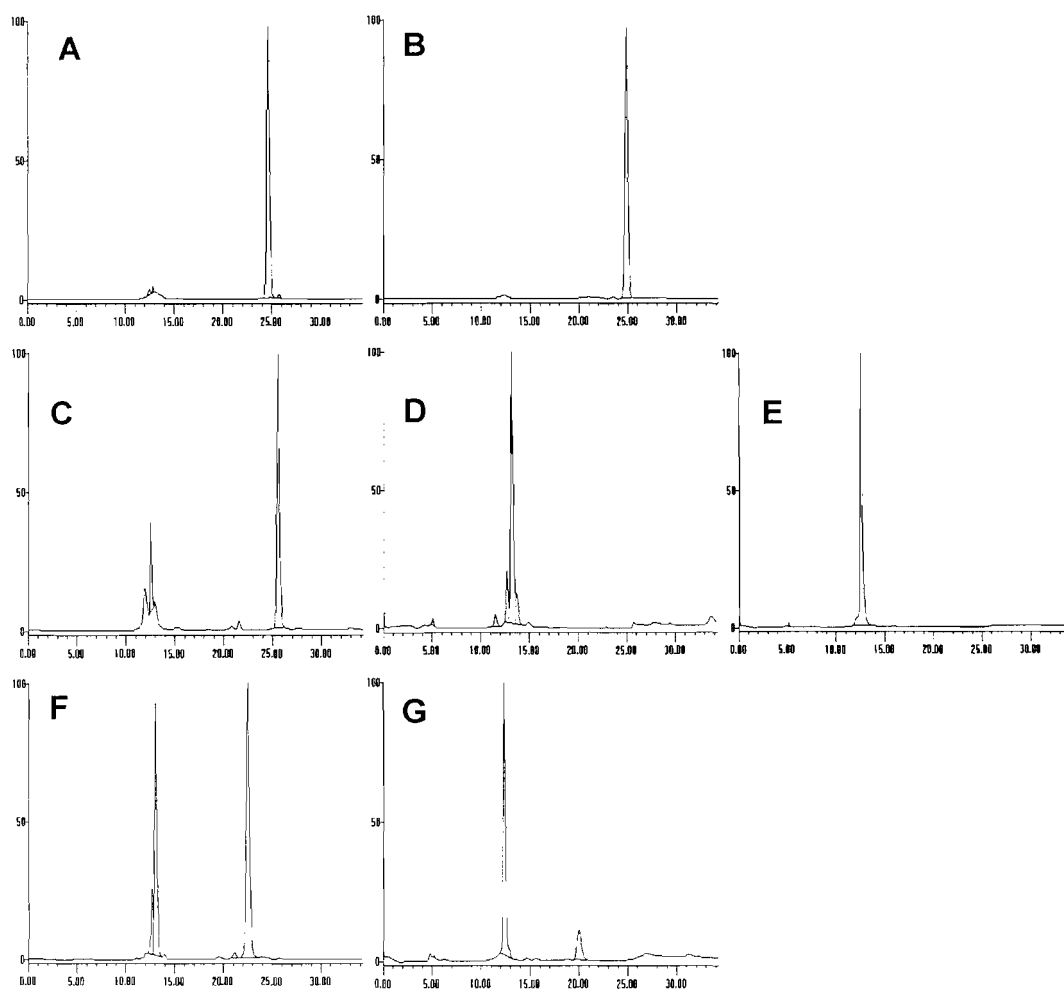
Figure 5.5 RP HPLC traces of short oligonucleotides in 'trityl-on' study. X-axis: Time (minutes); Y-axis: Relative OD. A. HT12 (crude); B. HT12 (after purification); C. BDT12 (crude); D. BD12 (after purification and deprotection in 2 M TEAA, pH 6.5, 55 °C, 12 h.); E. BD12 (after purification, deprotection and elimination in 15 mg.mL^{-1} DTT in conc. NH$_3$ (aq.) for 1.5 h.; F. CPT12 (crude); G. CPT12 (after detritylation in 4:1 AcOH:H$_2$O, RT, 1 h., then elimination in 2:1 conc. NH$_3$ (aq.):H$_2$O, 15 minutes).
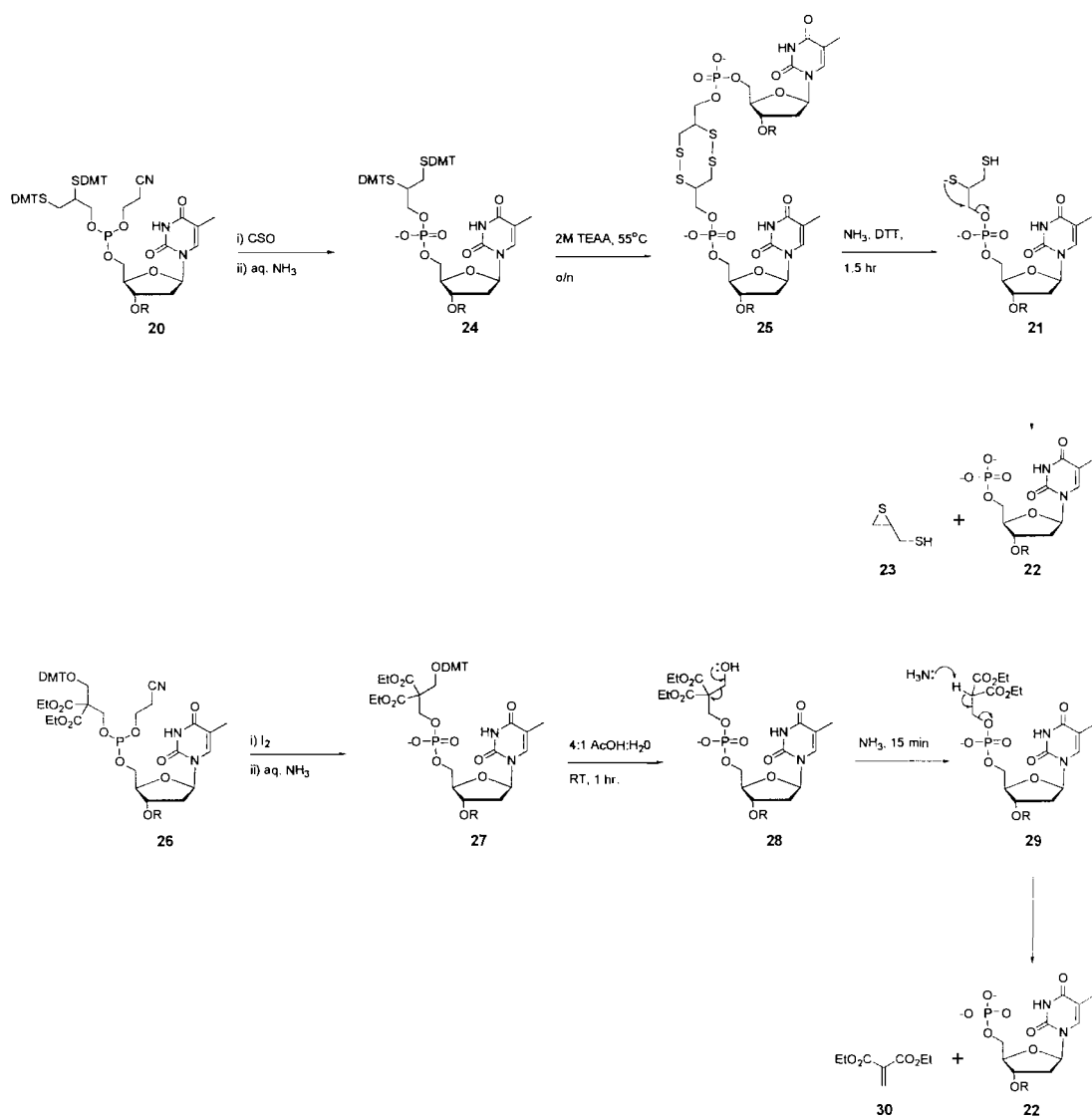
Scheme 5.4 Mechanisms of phosphate generation by *bis*-DMT dimercaptopropanol monomer and chemical phosphorylation reagent II suggested by mass spectrometry.

114

5.2.4 Long oligonucleotide synthesis with the hydrophobic tagging monomers

Given the encouraging results obtained from the synthesis of short oligomers, the hydrophobic tagging monomers were evaluated in the synthesis of two long oligonucleotides, HL1 (100 nt) and BDL2 (101 nt) (Table 5.4).

Oligonucleotide	Sequence (5'-3')	HPLC Retention time (minutes)		
		'Trityl-on'	Detritylated	Phosphorylated
HL1	H-GAGTTGTTCCTACCCAT GTGTGTTAGACTATCTCTC GCGTCGTCGGTCAGTCTAG TCGAGTTGTTCATACCCAT TTGCAAGACTTGTGTTTAG ACGTCCA	22.3	-	-
L2	GATCTGTTGACAATTATCA TCGGCTCGTATAATGTGTG GAATTGGTCGACGATCTGT TGACAATTAATCATCGGCT CGTATAATGTATGGAATTG GTCGAC	-	12.1	-
BDL2	BD-GATCTGTTGACAATTA TCATCGGCTCGTATAATGT GTGGAATTGGTCGACGATC TGTTGACAATTAATCATCG GCTCGTATAATGTATGGAA TTGGTCGAC	22.5	11.9	11.9

Table 5.4 Sequences and HPLC retention times of long oligonucleotides functionalised with hydrophobic residues. **H** = Hexadecanol, **BD** = *bis*-DMT dimercaptopropanol.

As expected, the retention times of tagged sequences HL1 and BDL2 were reduced relative to HLT12 and BDT12, reflecting the weaker influence of the lipophilic moieties on the overall polarities of long oligonucleotides. However, both full-length tagged sequences were well separated from truncated failure sequences, with the *bis*-DMT protected **16** again slightly more effective than the hexadecyl- tailed **15** (Table 5.4, Figure 5.6A and C). The relative ratio of tagged:untagged species in the HPLC

chromatograms show that significant *S*-detritylation had occurred. Despite this, the purified fraction could be subjected to detritylation and elimination to produce the 5'-phosphorylated sequence.

Since the hexadecyl- chain is not removed, RP HPLC could be used to evaluate the purity of HL1 (Figure 5.6B). While the purity of HL1 was high, some untagged sequences could be seen in the chromatogram. This is unlikely to be due to post-purification loss of the tag, but is probably due to some co-elution with the tagged sequence, due either to some partial sequence-complementarity or interaction between failure sequences and the hexadecyl- chain of the tag (similar to the interaction between oligonucleotides and the octyl- chain of the reverse phase column).
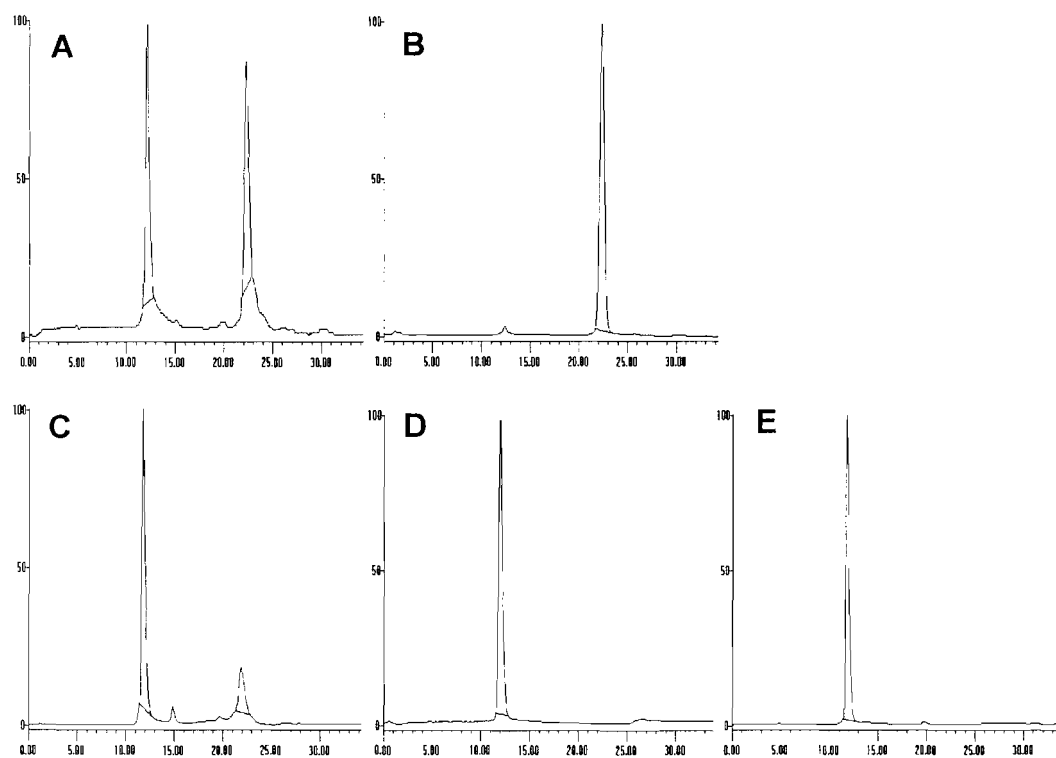
Figure 5.6 RP HPLC traces of long oligonucleotides in 'trityl-on' study. X-axis: Time (minutes); Y-axis: Relative OD. A. HL1 (crude); B. HL1 (after purification); C. BDL2 (crude); D. BDL2 (after purification and deprotection in 2 M TEAA, pH 6.5, 55 °C, 12 h.); E. BDL2 (after purification, deprotection and elimination in 15 mg.mL^{-1} DTT in conc. NH$_3$ (aq.) for 1.5 h).

116

As the phosphorylated BDL2 cannot be distinguished from the failure sequences by HPLC, agarose gel electrophoresis was used to assay its purity. Although a small amount of failure sequences can be seen (appearing as a slight streak above the product band), it is much cleaner than L2, which has no hydrophobic handle for purification (Figure 5.7). The presence of truncated sequences in the purified BDL2 may be due to a combination of some co-elution in 'trityl-on' HPLC and some incomplete capping, which would lead to truncated sequences incorporating the tag.
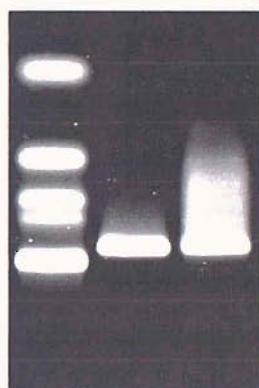
Figure 5.7 Agarose gel electrophoresis of synthetic 101mers. Lane 1. Molecular weight markers; 2. BDL2; 3. L2.

5.3 Conclusions

Two hydrophobic tagging monomers have been developed for synthesis of MLPA probes. The first, **15**, can be synthesised by simple phosphitylation of commercially available hexadecanol, and is used to permanently tag the 5'-terminus of the probe whose 3'-OH is to be ligated. This method is less versatile than the traditional 'trityl-on' method, which produces unmodified oligonucleotides, but is specifically tailored to the synthesis of MLPA probes. The blocking of the 5'-end ensures it cannot participate in any unwanted chemical or biological processes. The hydrophobicity of the hexadecyl- chain was sufficient to allow separation of a full-length tagged 100mer from truncated failure sequences (Figure 5.6A).

5'-Phosphorylation of the other ligation probe was achieved by the use of the *bis*-DMT protected monomer **16**. This monomer is more versatile than the previously described 2-(tritythio)ethyl phosphoramidite **14**, as the protecting groups can be

removed either on the DNA synthesiser or after HPLC purification. The presence of two hydrophobic trityl groups enhances its hydrophobicity relative to the commercially available monomer **13** used for 'trityl-on' synthesis of 5'-phosphorylated oligonucleotides (Figure 5.5A and F). The detritylation step, although time consuming, is exceptionally mild and minimises side reactions. Again, the applicability of the monomer to long oligonucleotide synthesis was demonstrated by the synthesis of a 101mer. The purity of the oligonucleotide was high (Figure 5.7), but HPLC suggested significant *S*-detritylation had occurred, presumably in the work-up steps between deprotection and purification (Figure 5.6A).

A modified procedure using the hexadecanol monomer **15**, and commercially available reverse 5'-3' amidites (bearing 4,4'-dimethoxytritylated 3'-hydroxyl groups, and phosphitylated 5'-hydroxyl groups)[245] and the phosphate column **31** removes the need for protecting groups, and might be the best way to produce pure 5'-phosphorylated MLPA probes (Scheme 5.5). This method would have the additional benefit that the 3'-terminus is blocked, so cannot participate in any unwanted ligation reactions.
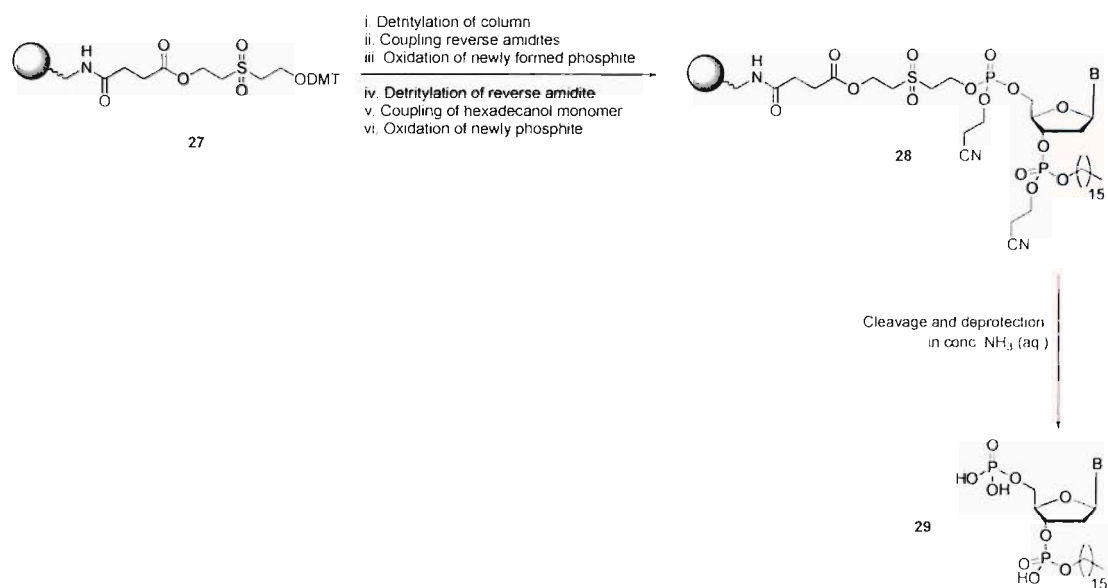
Scheme 5.5 Proposed synthesis of 5'-phosphorylated oligonucleotides using reverse 5'-3' amidites and the hexadecanol monomer.

6. Experimental

6.1 Preparation of Compounds

6.1.1 General methods

All reactions requiring anhydrous conditions were performed in dry glassware under an atmosphere of argon. Pyridine, N,N-diisopropylethylamine and dichloromethane were distilled from calcium hydride. Methanol was distilled from magnesium and iodine. Anhydrous N,N-dimethylformamide was purchased from Aldrich. Other reagents were purchased from Aldrich, Lancaster, Cruachem, Avocado, Link Technologies or Fluka. Column chromatography was carried out under pressure on Fisher Scientific DAVISIL 60A (35-70 micron) silica. TLC analysis was performed using Merck Kieselgel 60 F_{24} (0.22 mm thickness, aluminium backed). Compounds were visualised by irradiation at 254 nm, staining with iodine:silica (1:1) or phosphomolybdic acid: ethanol (1:10). Proton, carbon and phosphorus NMR spectra were recorded at 300 or 400 MHz, 75.5 or 100 MHz and 121 MHz respectively. Spectra were recorded on a Bruker DPX 400 or AC 300 spectrometer in deuterated chloroform or dimethylsulfoxide and calibrated to the residual solvent peak. Low-resolution mass spectra were recorded using electrospray ionisation on a Waters ZMD quadrupole mass spectrometer or a Fisons VG platform instrument in methanol or acetonitrile. High-resolution mass spectra were recorded on a Bruker APEX III FT-ICR mass spectrometer in methanol or acetonitrile using an Apollo electrospray ionisation source. Infrared spectra were recorded on a BIORAD FT-IR instrument using a Golden Gate adaptor and BIORAD WIN-IR software. Absorptions are described as strong (s), medium (m), broad (br), or weak (w), and are measured in units of cm^{-1}. Ultraviolet spectra were measured on a Perkin Elmer UV/Vis Lambda 2 spectrometer. Melting points were recorded on a Gallenkamp electrothermal melting point apparatus and are uncorrected.

(2*R*,3*R*)-2-[(6-chloro-2-methoxy-9-acridinyl)amino]butane-1,3-diol (2)
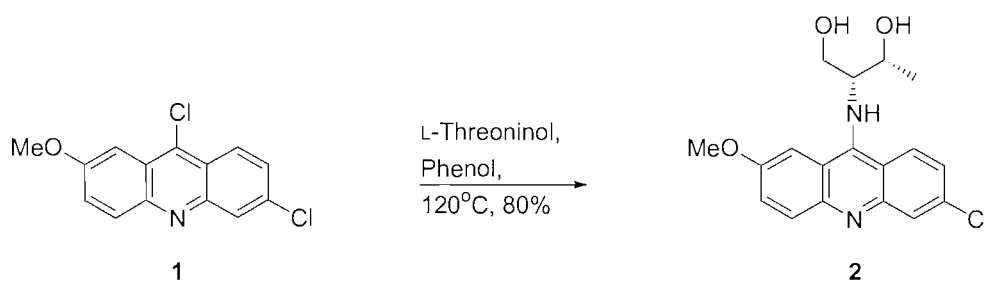
To compound **1** (3 g, 10.8 mmol) was added phenol (25 g) and L-threoninol (1.11 g, 10.8 mmol), and the mixture heated to 120 °C for 12 h. At this time, MeOH (20 mL) and toluene (100 mL) were added and the mixture filtered while hot. This afforded a yellow solid, which was recrystallised from MeOH/toluene to yield the title compound as yellow crystals (2.99 g, 8.64 mmol, 80 %).

R_f (9:1 DCM:MeOH): 0.16

δ_H (300 MHz, DMSO): 8.84 (1H, d, J = 9.5 Hz, Ar$\underline{H}$), 8.26 (1H, d, J = 2.0 Hz, Ar$\underline{H}$), 8.01 (1H, d, J = 2.0 Hz, Ar$\underline{H}$), 7.94 (1H, d, J = 9.5 Hz, Ar$\underline{H}$), 7.64 (1H, dd, J = 9.5, 2.0 Hz, Ar$\underline{H}$), 7.49 (1H, dd, J = 9.5, 2.0 Hz, Ar$\underline{H}$), 5.32 (1H br s, N$\underline{H}$), 4.43 (1H, m, HOCH$_2$C$\underline{H}$(NHR)CH(CH$_3$)OH), 4.07 (1H, apparent quintet, J = 6.5 Hz, C$\underline{H}$(CH$_3$)OH), 3.89-3.96 (4H, m, OC$\underline{H}_3$, 1 x C$\underline{H}_2$OH), 3.80 (1H, dd, J = 11.5, 4.0 Hz, 1 x C$\underline{H}_2$OH), 1.09 (3H, d, J = 6.5 Hz, CH(C$\underline{H}_3$)OH)

δ_C (75.5 MHz, DMSO): 158.45, 155.41, 140.04, 138.50 (5 x Ar$\underline{C}$), 128.43, 126.97, 123.18, 120.78 (5 x Ar$\underline{C}$H), 117.60 (2 x Ar$\underline{C}$), 105.61 (1 x Ar$\underline{C}$H), 68.01 (HOCH$_2$$\underline{C}$H(NHR)CH(CH$_3$)OH), 66.21 ($\underline{C}$H(CH$_3$)OH), 61.12 ($\underline{C}H_2$OH), 56.18 (O$\underline{C}H_3$), 20.79 (CH($\underline{C}H_3$)OH)

λ_{max} (MeOH): 427 nm (7,320 mol^{-1}.dm^3.cm^{-1})

ν_{max} (neat): 3380 (br), 3264 (br) 3104 (m), 3075 (m), 2943 (w), 1629 (s), 1581 (s), 1377 (m), 1272 (m), 1246 (s)

LRMS (ES$^+$): *m/z* 347.2, 349.2 (3:1) [M+H]$^+$

M.Pt.: 199-203 °C

HRMS (ES$^+$) *m/z*: C$_{18}$H$_{20}$ClN$_2$O$_3$ requires 347.1162, found 347.1147

120

(2R,3R)-3-[(6-chloro-2-methoxy-9-acridinyl)amino]-4-[(4,4'-dimethoxytrityl) oxy] butan-2-ol (3)
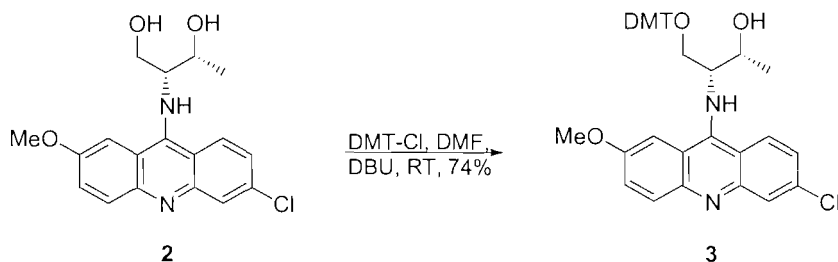
To diol **2** (1.09 g, 3.15 mmol), dissolved in dry DMF (20 mL) was added DBU (0.66 mL, 4.41 mmol), and the mixture stirred for 15 minutes at room temperature under an inert atmosphere. 4,4'-Dimethoxytrityl chloride (1.50 g, 4.43 mmol) was added, and the mixture stirred for a further 3 h., after which TLC (9:1 DCM:MeOH) showed consumption of starting material. The volatiles were removed *in vacuo*, the residue redissolved in EtOAc (75 mL) and washed with water (2 x 100 mL) and sat. KCl (3 x 100 mL). The organic extract was dried over anhydrous MgSO$_4$, dried *in vacuo*, and the residual yellow foam purified by flash column chromatography on silica gel, eluting with DCM:MeOH (19:1) to yield the title compound as a yellow foam (1.52 g, 2.34 mmol, 74 %).

R$_f$ (9:1 DCM:MeOH): 0.48

δ_H (*300 MHz, DMSO*): 8.48 (1H, d, *J* = 9.0 Hz, ArH), 7.96 (1H, d, *J* = 2.0 Hz, ArH), 7.90 (1H, d, *J* = 9.0 Hz, ArH) 7.62 (1H, d, *J* = 2.5 Hz, ArH), 7.43 (1H, dd, *J* = 9.0, 2.5 Hz, ArH), 7.34 (1H, dd, *J* = 9.0, 2.0 Hz, ArH), 7.12-7.14 (3H, m, ArH), 6.99-7.03 (2H, m, ArH), 6.89-6.93 (4H, m, ArH), 6.70-6.73 (4H, m, ArH), 5.16 (1H, d, *J* = 4.0 Hz, NH), 3.94-4.05 (2H, m, DMTOCH$_2$CH(NHR)CH(CH$_3$)OH, CH(CH$_3$)OH), 3.82 (3H, s, OCH$_3$), 3.70 (6H, s, 2 x OCH$_3$), 3.30-3.33 (2H, m, CH$_2$ODMT), 1.00 (3H, d, *J* = 6.0 Hz, CH(CH$_3$)OH)

δ_C (*75.5 MHz, DMSO*): 157.76, 155.06, 151.05, 147.88, 146.35, 144.60, 135.32, 135.20, 133.18 (10 x ArC), 130.83, 129.38, 127.40, 127.28, 126.30, 126.06, 124.22, 123.04 (14 x ArCH), 118.13, 116.02 (2 x ArC), 112.75, 100.56 (5 x ArCH), 85.49 (OC(Ar)$_3$), 66.72 (CH(CH$_3$)OH), 65.12 (CH$_2$ODMT), 64.95 (DMTOCH$_2$CH(NHR)CH(CH$_3$)OH), 55.33 (OCH$_3$), 54.81 (2 x OCH$_3$), 20.35 (CH(CH$_3$)OH)

v_{max} (neat): 3381 (br), 2958 (w), 2934 (w), 2835 (w), 1631 (m), 1607 (m), 1508 (s), 1463 (s) 1365 (m), 1302 (m), 1233 (s)

λ_{max} (MeOH): 423 nm (7,430 $mol^{-1}.dm^3.cm^{-1}$)

LRMS (ES$^+$): m/z 649.5, 651.5 (3:1) [M+H]$^+$

HRMS (ES$^+$) m/z: C$_{39}$H$_{38}$ClN$_2$O$_5$ requires 649.2464, found 649.2457

2-Cyanoethyl-*N*,*N*-diisopropyl-[(2*R*,3*R*)-3-[(6-chloro-2-methoxy-9-acridinyl) amino]-4-[(4,4'-dimethoxytrityl)oxy]but-2-yl] phosphoramidite (4)

2-Cyanoethyl-
N,*N*-diisopropyl
chlorophosphoramidite,
DCM, DIPEA, RT,
74%

3   4
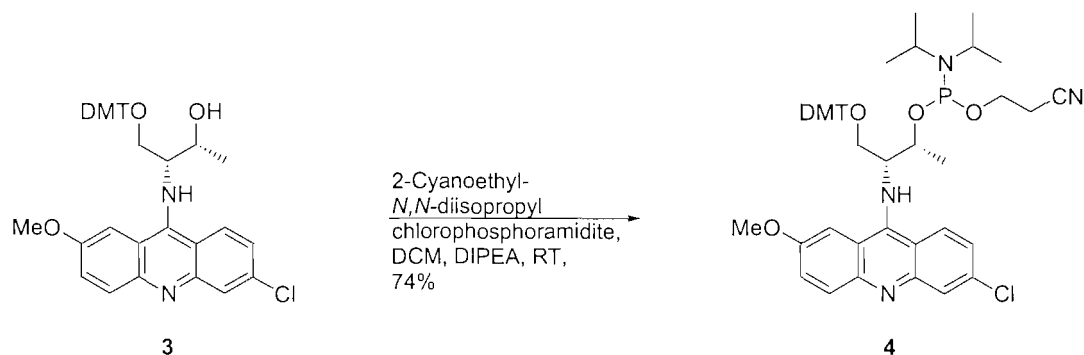
To alcohol **3** (1.09 g, 1.68 mmol), dissolved in dry, degassed DCM (20 mL) was added 2-cyanoethyl-*N*,*N*-diisopropyl chlorophosphoramidite (0.44 mL, 1.84 mmol) and dry DIPEA (0.71 mL, 4.2 mmol). The reaction mixture was stirred for 2 h., at which time TLC (9:1 DCM:MeOH) showed no further reaction. The reaction was reduced to dryness *in vacuo*, and the residue partitioned between degassed EtOAc (35 mL) and degassed sat. KCl (35 mL). The organic layer was recovered, dried over anhydrous MgSO$_4$ and the solvent removed under reduced pressure. The resultant yellow foam was purified by flash silica column chromatography, eluting with NEt$_3$:DCM:MeOH (1:94:5), filtered and evaporated to dryness to give a mixture of diastereoisomers of the title compound as a yellow foam (1.46 g, 1.24 mmol, 74 %).

δ_H (300 MHz, DMSO): 8.46 (1H, m, Ar$\underline{H}$), 7.99 (1H, d, *J* = 2.0 Hz, Ar$\underline{H}$), 7.92 (1H, d, *J* = 9.0 Hz, Ar$\underline{H}$), 7.57 (1H, m, Ar$\underline{H}$), 7.45 (1H, dd, *J* = 9.5, 2.5 Hz, Ar$\underline{H}$), 7.34 (1H, m, Ar$\underline{H}$), 7.07-7.18 (3H, m, Ar$\underline{H}$), 6.85-7.07 (6H, m, Ar$\underline{H}$), 6.67-6.78 (4H, m, Ar$\underline{H}$), 4.07-4.30 (2H, m, DMTOCH$_2$C$\underline{H}$(NHR)CH(CH$_3$)OH, C$\underline{H}$(CH$_3$)OP), 3.86 (3H, s, OC$\underline{H}_3$), 3.65-3.75 (8H, m, 2 x OC$\underline{H}_3$, NCCH$_2$C$\underline{H}_2$OP), 3.44-3.60 (4H, m, C$\underline{H}_2$ODMT, NC$\underline{H}$(CH$_3$)$_2$), 2.72 (2H, t, *J* = 6.0 Hz, NCC$\underline{H}_2$CH$_2$OP), 1.20 (3H, d, *J* = 6.5 Hz, NCHC$\underline{H}_3$), 1.09 (3H, d, *J* = 6.5 Hz, NCHC$\underline{H}_3$), 1.07 (3H, d, *J* = 6.5 Hz, NCHC$\underline{H}_3$), 1.00 (3H, d, *J* = 7.0 Hz, NCHC$\underline{H}_3$), 0.85 (3H, d, *J* = 6.5 Hz, CH(C$\underline{H}_3$)OP)

δ_P (121 MHz, CDCl$_3$): 148.50, 148.27

LRMS (ES$^+$): *m/z* 849.6, 851.6 (3:1) [M+H]$^+$

2-Cyanoethyl-*N*,*N*-diisopropyl-(hexadecyl) phosphoramidite (15)
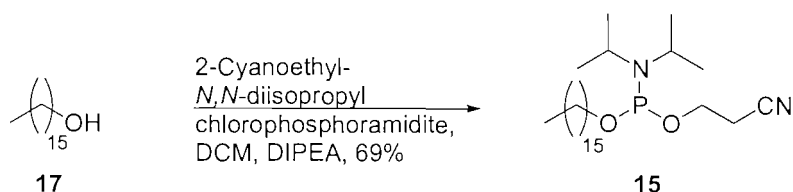
To hexadecanol, **17**, (1.0 g, 4.12 mmol), dissolved in dry, degassed DCM (20 mL) was added 2-cyanoethyl *N*,*N*-diisopropyl chlorophosphoramidite (1.04 mL, 4.33 mmol) and dry DIPEA (1.79 mL, 10.3 mmol). The reaction mixture was stirred for 4 h., at which time TLC (2:1 hexane:EtOAc) showed no further reaction. The reaction was reduced to dryness *in vacuo*, and the residue partitioned between degassed EtOAc (20 mL), and degassed sat. KCl (20 mL) The organic layer was recovered, dried over anhydrous $MgSO_4$ and the solvent removed under reduced pressure. The was purified by flash silica column chromatography, eluting with NEt_3:EtOAc:hexane (1:49:50), filtered and evaporated to dryness to give the title compound as a clear oil (1.26 g, 2.84 mmol, 69 %).

δ_H (400 MHz, $CDCl_3$): 3.78-3.90 (2H, m, $NCCH_2C\underline{H}_2OP$), 3.53-3.68 (4H, m, 2 x $NC\underline{H}(CH_3)_2$, $POC\underline{H}_2(CH_2)_{14}CH_3$), 2.64 (2H, t, J = 6.5 Hz, $NCC\underline{H}_2CH_2OP$), 1.60 (2H, apparent quintet, J = 7.0 Hz, $POCH_2C\underline{H}_2(CH_2)_{13}CH_3$), 1.26-1.36 (26H, m, $POCH_2CH_2(C\underline{H}_2)_{13}CH_3$), 1.19 (6H, d, J = 7.0 Hz, $NCH(C\underline{H}_3)_2$), 1.18 (6H, d, J = 7.0 Hz, $NCH(C\underline{H}_3)_2$), 0.88 (3H, t, J = 7.0 Hz, $POCH_2(CH_2)_{14}C\underline{H}_3$)

δ_P (121 MHz, $CDCl_3$): 147.77

LRMS (ES$^+$): *m/z* 443.6 [M+H]$^+$

The phosphoramidite monomer was used to synthesise a test oligonucleotide of sequence (**15**)TTTTTTTTTT; MALDI-TOF MS (ES$^+$) *m/z*: Requires 3893.8, found 3894.0.

2,3-*Bis*-(4,4'-dimethoxytrityl)sulfanyl-1-propanol (19)
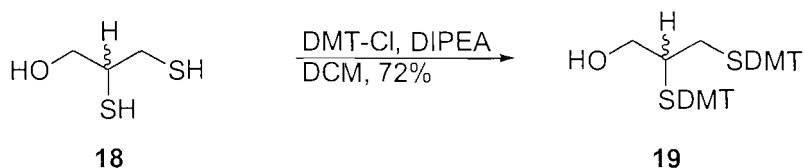
2,3-Dimercapto-1-propanol (1.0 g, 8.0 mmol), DIPEA (3.5 mL, 20 mmol) and 4,4'-dimethoxytrityl chloride (5.18 g, 15.3 mmol) were dissolved in dry DCM (50 mL). The mixture was stirred overnight (16 h.) at room temperature, after which time TLC (2:1 hexane:EtOAc) showed consumption of starting material. The volatiles were removed *in vacuo*, the residue redissolved in EtOAc (200 mL) and washed with water (2 x 200 mL) and sat. KCl (3 x 200 mL). The organic extract was dried over anhydrous $MgSO_4$, dried *in vacuo*, and the residual pale yellow foam purified by flash column chromatography on silica gel, eluting with hexane:EtOAc (3:1) to yield the title compound as a white foam (4.20 g, 5.76 mmol, 72 %).

R_f (Hexane:EtOAc, 2:1): 0.27

δ_H (300 MHz, DMSO): 7.10-7.30 (18H, m, Ar$\underline{H}$), 6.59-6.81 (8H, m, Ar$\underline{H}$), 4.69 (1H, br, exchanges with D_2O, O$\underline{H}$), 3.71 (12H, s, 4 x OC$\underline{H}_3$) 3.01-3.10 (2H, m, simplifies with D_2O, C$\underline{H}_2$OH), 2.05-2.32 (3H, m, OCH_2(C$\underline{H}$S)C$\underline{H}_2$S)

δ_C (75.5 MHz, DMSO): 157.65, 145.47, 145.29, 136.87, 136.71 (Ar$\underline{C}$) 130.37, 128.98, 128.90, 127.93, 126.54, 113.17 (Ar$\underline{C}$H), 66.10, 65.00 (S$\underline{C}$(Ar)$_3$), 62.21 ($\underline{C}$H_2O), 55.09 (O$\underline{C}$H_3), 46.02 (HOCH_2($\underline{C}$HS)CH_2S), 33.87 ($\underline{C}$H_2S)

ν_{max} (neat): 3512 (br), 3004 (w), 2932 (w), 2835 (w), 1734 (w), 1606 (m), 1504 (s), 1466 (m), 1443 (m), 1296 (m), 1246 (s)

LRMS (ES$^+$): *m/z* 303.1 [DMT]$^+$, 751.5 [M+Na]$^+$

125

## 2-Cyanoethyl *N,N*-diisopropyl [2,3-*bis*-(4,4'-dimethoxytrityl)sulfanylpropyl] phosphoramidite (16)
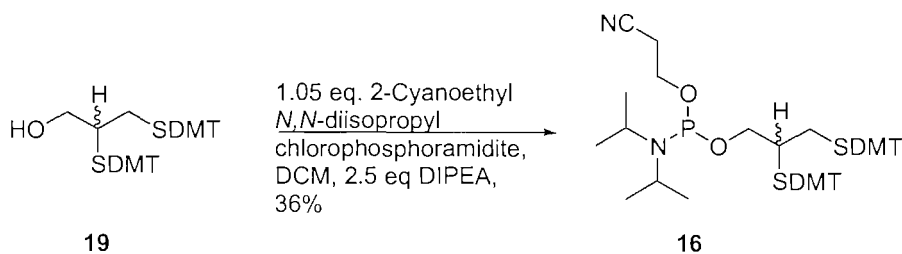
To alcohol **19** (1.0 g, 1.37 mmol) dissolved in dry, degassed DCM (15 mL) was added 2-cyanoethyl-*N,N*-diisopropyl chlorophosphoramidite (0.34 mL, 1.44 mmol) and dry DIPEA (0.59 mL, 3.43 mmol). The reaction mixture was stirred for 3 h., at which time TLC (2:1 hexane:EtOAc) showed no further reaction. The reaction was reduced to dryness *in vacuo*, and the residue partitioned between degassed EtOAc (15 mL), and degassed sat. KCl (15 mL) The organic layer was recovered, dried over anhydrous $MgSO_4$ and the solvent removed under reduced pressure. The resultant off-white foam was dissolved in dry, degassed DCM (5 mL) and precipitated by dropwise addition into degassed hexane (500 mL) at −78 °C. The supernatant was discarded, and the residue redissolved in dry, degassed DCM (5 mL), filtered and evaporated to dryness to give a mixture of diastereoisomers of the title compound as a white foam (0.45 g, 0.49mmol, 36 %)

δ_H (300 MHz, $CDCl_3$): 7.10-7.30 (18H, m, Ar$\underline{H}$), 6.59-6.81 (8H, m, Ar$\underline{H}$), 3.71 (12H, s, 4 x OC$\underline{H}_3$) 3.42-3.75 (4H, m, POC$\underline{H}_2$(CHS)CH$_2$S, NCCH$_2$C$\underline{H}_2$OP), 3.19-3.26 (2H, m, NC$\underline{H}$(CH$_3$)$_2$), 2.33-2.55 (5H, m, NCC$\underline{H}_2$CH$_2$O, OCH$_2$(C$\underline{H}$S)C$\underline{H}_2$S), 1.16 (6H, d, *J* = 7.0 Hz, NCH(C$\underline{H}_3$)$_2$), 1.09 (3H, d, *J* = 7.0 Hz, NCH(C$\underline{H}_3$)CH$_3$), 1.05 (3H, d, *J* = 7.0 Hz, NCH(CH$_3$)C$\underline{H}_3$)

^{31}P NMR ($CDCl_3$) δ 148.51, 148.18

LRMS (ES$^+$) *m/z* 951.6 [M+Na]$^+$, 967.5 [M+K]$^+$

The phosphoramidite monomer was used to synthesise a test oligonucleotide of sequence (**16**)TTTTTTTTTTT; ESMS (ES$^-$) *m/z*: Requires 4379.3, found 4380.0.

6.2 Oligonucleotide Synthesis and purification

6.2.1 Oligonucleotide synthesis and deprotection

All oligonucleotides were synthesised on either an Applied Biosystems ABI 394 or Expedite automated DNA/RNA synthesiser on the 0.2 μM scale using the standard cycles of acid-catalysed detritylation, coupling, capping and iodine oxidation procedures, using phosphoramidite monomers and other reagents purchased from Applied Biosystems, Cruachem or Link Technologies. (1S)-(+)-(10-camphorsulfonyloxaziridine) was purchased from Aldrich, and was used in place of iodine oxidation for the final oxidation step in the synthesis of oligonucleotides BDT12 and BDL2.[244] All β-cyanoethyl phosphoramidite monomers were dissolved in anhydrous acetonitrile to a concentration of 0.1 M immediately prior to use. Stepwise coupling efficiencies (where applicable) were determined by the trityl cation conductivity monitoring facility, and were >95 %.

Oligonucleotides were deprotected in concentrated aqueous ammonia (55 °C, 5.5 h.), apart from oligonucleotides labelled with the acridine-labelling monomer, 1-(4,4'-dimethoxytrityloxy)-2-(N-acridinyl-4-aminobutyl)-propyl-3-O-(2-cyanoethyl)-(N,N-diisopropyl) phosphoramidite, which were deprotected in 0.4 M methanolic sodium hydroxide at room temperature for 17 h., neutralised with 2 M TEAA, (pH 6.0) and desalinated using a Sephadex® containing NAP-10 column (Pharmacia Biotech) before purification. This protocol requires the use of 5'-(4,4'-dimethoxytrityl)-N4-acetyl-2'-deoxycytidine-3'-(2-cyanoethyl-N,N-diisopropyl) phosphoramidite (Ac-dC CE phosphoramidite).[246] Unlabelled PCR primers were used without further purification.

Oligonucleotides containing chemical phosphorylation reagent II, [3-(4,4'-dimethoxytrityloxy)-2,2-dicarboxyethyl] propyl-(2-cyanoethyl)-(N,N-diisopropyl) phosphoramidite (Glen Research), were detritylated after HPLC purification in AcOH:H$_2$O (4:1, 25 °C, 1 h.), reduced to dryness, and redissolved in conc. NH$_3$ (aq.):H$_2$O (2:1, 25 °C, 15 minutes) to generate 5'-phosphate.[246]

Oligonucleotides containing the *bis*-DMT phosphorylation reagent, 2-Cyanoethyl N,N-diisopropyl [2,3-*bis*-(4,4'-dimethoxytrityl)sulfanylpropyl] phosphoramidite, were detritylated after HPLC purification in 2M TEAA (pH 6.5, 55 °C, 12 h.),

desalinated using a Sephadex® containing NAP-10 column, reduced to dryness and redissolved in a solution of DTT (15 mg.mL^{-1}) in conc. NH$_3$ (aq.), (25 °C, 1 h.) to generate 5'-phosphate.

6.2.2 Purification of synthetic oligonucleotides

Deprotected oligonucleotides were purified by RP HPLC on a Gilson system at the Oswel Laboratories or at the Department of Chemistry, using an ABI Aquapore column (C8), 8 mm x 250 mm, pore size 300 Å. The system was controlled by Gilson 7.12 software. The following protocol was used except for 'trityl-on' oligonucleotides: Run time 30 minutes, flow rate 4 mL.minute^{-1}, binary system, gradient: Time in minutes (% buffer B); 0 (0); 3 (0); 5 (20); 21 (100); 25 (100); 27 (0); 30 (0). Elution buffer A: 0.1 M NH$_4$OAc, pH 7.0, buffer B: 0.1 M NH$_4$OAc (35 % MeCN) pH 7.0.

For 'trityl-on' oligonucleotides a different protocol was used: Run time 35 minutes, flow rate 3 mL.minute^{-1}, binary system, gradient: Time in minutes (% buffer B); 0 (0); 2.5 (0); 20 (80); 21 (100); 30 (100); 32.5 (0); 35 (0). Elution buffer A: 0.1 M NH$_4$OAc, pH 7.0, buffer B: 0.1 M NH$_4$OAc (50 % MeCN) pH 7.0.

Elution of oligonucleotides was monitored by UV absorption at 295 nm. Purified oligonucleotides were desalted using Sephadex® containing NAP-10 columns (Pharmacia Biotech), and their optical densities recorded on a Lambda 15 UV/Vis Spectrophotomer (Perkin Elmer). The optical densities recorded were used to calculate their concentrations. Electrophoresis of oligonucleotides BDL2 and L2 was performed on a 4% agarose gel (running buffer 0.5x TBE), which was stained with ethidium bromide, and photographed under a UV-transilluminator.

6.2.3 Mass spectrometry of oligonucleotides

The molecular masses of oligonucleotides were determined by either ESMS or MALDI-TOF MS on a Fisons VG platform instrument or a ThermoBioAnalysis Dynamo MALDI-TOF mass spectrometer.

For ESMS, oligonucleotides were prepared in H$_2$O:MeCN (1:1) containing 0.01 % tripropylamine. Data were collected in negative ion mode and deconvoluted using the MaxEnt electrospray function.

MALDI-TOF data were obtained in positive ion mode using delayed extraction and an initial accelerating voltage of 20 kV. Spectra were recorded from a matrix containing 4:1 HPA:PA in 1:1 acetonitrile:water in the presence of Dowex 50WX8-200 ion exchange beads according to the method of Langley *et al.* [247]

6.2.4 Sequences of synthetic oligonucleotides

Code	Sequence (5' to 3')
FA12	**FAM-ACD**-AAACTTGGATCC-**OCT**
FA14	**FAM-ACD**-AAAAACTTGGATCC-**OCT**
FA16	**FAM-ACD**-GAAAAAACTTGGATCC-**OCT**
FA18	**FAM-ACD**-TAGAAAAAACTTGGATCC-**OCT**
ST1303	GGGATCCAAGTTTTTTCTAAATGTTCC
P0	**FAM-ACD**-ATTTAGAAAAAACTTGGATCCC
P1	**FAM-A-ACD**-TTTAGAAAAAACTTGGATCCC
P2	**FAM-AT-ACD**-TTAGAAAAAACTTGGATCCC
P4	**FAM-ATTT-ACD**-AGAAAAAACTTGGATCCC
P6	**FAM-ATTTAG-ACD**-AAAAAACTTGGATCCC
P9	**FAM-ATTTAGAAA-ACD**-AAACTTGGATCCC
P11	**FAM-ATTTAGAAAAA-ACD**-ACTTGGATCCC
A1	ATT-**ACD**-TAGAAAAAAC
F1	**FAM**-ATTTAGAAAAAACTTGGATCCC
FA14	**FAM-ACD**-AAAAACTTGGATCC-**OCT**
FA16	**FAM-ACD**-GAAAAAACTTGGATCC-**OCT**
UP12	AAACTTGGATCC
UP14	AAAAACTTGGATCC
UP16	GAAAAAACTTGGATCC
UP18	TAGAAAAAACTTGGATCC
MM1	GGGATCCAACTTTTTTCTAAATGTTCC
MM2	GGGATCCAATTTTTTTCTAAATGTTCC
MM3	GGGATCCAAATTTTTTCTAAATGTTCC
MB1303	**FAM**-<u>CCCGCGCGG</u>-AACATTTAGAAAAAACTTGGA-<u>TCCCGCGCGGG</u>-**MR**
FP1303	TTTCTTGATCACTCCACTGTTC
RP1303	CATACTTTCTTCTTCTTTTCTTT
RFP1303	ATTTCTTGATCACTCCACTGTT
RRP1303	CTTTTTTGCTATAGAAAGTATTTA
163R	**FAM**-ACGCAATTGGT
158R	ACCAATTGCGT

164R	**FAM**-AGCGAATTGGT
159R	ACCAATTCGCT
165R	**FAM**-AATTGCGTCAG
160R	CTGACGCAATT
172R	**FAM**-CGCATCGCAATT
173R	AATTGCGATGCG
FP1282	GATGGTGTGTCTTGGGATTCA
RP1282	TGGCTAAGTCCTTTTGCTCAC
PRP1282	**FAM**-CTTTCCTCCACTGTTGC
HP1282	**FAM**-CTTTCCTCCACTGTTGC-**p**
T1303	TTTCTTGATCACTCCACTGTTCATAGGGATCCAAGTTTTTTTCTAAATGTTCCA GAAAAAATAAATACTTTCTATAGCAAAAAAGAAAAGAAGAAGAAAGTATG
C1303	CATACTTTCTTCTTCTTTTCTTTTTTGCTATAGAAAGTATTTATTTTTTCTGGA ACATTTAGAAAAAACTTGGATCCCTATGAACAGTGGAGTGATCAAG AAA
PRP1303	**FAM-ACD**-ACTGTTCATAGGGATCCAAG
FAS1303	**FAM-ACD**-GGAACATTTAGAAAAAACTTGGATCCC-**HEG**- TTTCTTGATCACTCCACTGTTC
TQM	**FAM**-TGCGGGAGCCGATTT-**TAMRA**
CMG	CTGACCTGAAGCACTTGAAGGAGAAGGTGTCTGCGGGAGCCGATTTCATCAT CACGCAGCTTTTCTTTGAGGCTGACACATTCTTCCGCTTTGTGAAGGCATGCA CCGACAT
TMG	ATGTCGGTGCATGCCTTCACAAAGCGGAAGAATGTGTCAGCCTCAAAGAAA AGCTGCGTGATGATGAAATCGGCTCCCGCAGACACCTTCTCCTTCAAGTGCT TCAGGTCAG
CMI	CTIACCTIAAICACTTIAAIIAIAAIITITCTICIIIAICCIATTCATCATCACICAICTTT TCTTTIAIICTIACACATTCTTCCICTTTITIAAIICATICACCIACAT
TMI	ATITCIITICATICCTTCACAAAICIIAAIAATITITCAICCTCAAAIAAAAICTICITIA TIATIAAATCIICTCCCICAIACACCTTCTCCTTCAAITICTTCAIITCAI
FPM	CTGACCTGAAGCACTTGAAGG
RPM	ATGTCGGTGCATGCCTTCAC
CFPM	GGAGCTGACCTGAAGCACTTGAAGG
CRPM	GAGGAGATGTCGGTGCATGCCTTCAC
MBM	**FAM**-<u>GCGAG</u>TGCGGGAGCCGATTT<u>CTCGC</u>-**MR**
PT12	**p**-TTTTTTTTTTTT
HT12	**H**-TTTTTTTTTTTT
BDT12	**BD**-TTTTTTTTTTTT
CP12	**CP**-TTTTTTTTTTTT
HL1	**H**-GAGTTGTTCCTACCCATGTGTGTTAGACTATCTCTCGCGTCGTCGGTCAGT CTAGTCGAGTTGTTCATACCCATTTGCAAGACTTGTGTTTAGACGTCCA
L2	GATCTGTTGACAATTATCATCGGCTCGTATAATGTGTGGAATTGGTCGACGA

	TCTGTTGACAATTAATCATCGGCTCGTATAATGTATGGAATTGGTCGAC
BDL2	**BD**-GATCTGTTGACAATTATCATCGGCTCGTATAATGTGTGGAATTGGTCGA
	CGATCTGTTGACAATTAATCATCGGCTCGTATAATGTATGGAATTGGTCGAC

Table 6.1 Sequences of synthetic oligonucleotides. **FAM** = fluorescein, **ACD** = 9-amino-6-chloro-2-methoxyacridine, **MR** = Methyl Red dR, **OCT** = octanediol, **TAMRA** = tetramethylrhodamine, **HEG** = hexaethylene glycol (PCR stopper), I = 2'-deoxyinosine, **p** = phosphate **H** = Hexadecanol, **BD** = *Bis*-DMT dimercaptopropanol, **CP** = Chemical Phosphorylation reagent II. FA12-18 and UP12-18 are composed of 2'-*O*-methyl ribonucleotides. Molecular Beacon stem sequences are underlined. Chemical structures of non-standard monomers are given in the appendix.

6.3 Fluorimetry

6.3.1 Effect of hybridisation on fluorescence of fluorophore-intercalator probes

The fluorescence spectra were recorded using an LS 50B Luminescence Spectrometer (Perkin Elmer). To 2.95 mL of filtered buffer (100 mM Sodium Phosphate, 1 mM EDTA, 0.1 M NaCl, pH 7.0) was added the dual-labelled probe (FA12-18, to a final concentration of 0.21 µM), and the emission spectra recorded, from fixed excitations at 384 nm and 495 nm. To the solution was then added the synthetic target (ST1303, 50 eq., final concentration 10.5 µM), and the emission spectra from fixed excitation at 384 nm and 495 nm re-recorded.

6.3.2 Fluorescence spectra of acridine orange derivatives

All compounds were provided by Dr. L.J. Brown. The fluorescence spectra were recorded using an LS 50B Luminescence Spectrometer (Perkin Elmer). To 2.95 mL of filtered buffer (100 mM Sodium Phosphate, 1 mM EDTA, 0.1 M NaCl, pH 7.0) was added the dye (to a final concentration of 100 nM), and the excitation and emission spectra recorded, across the 400-800 nm range.

6.3.3 Fluorescence quenching titration of Hoechst 33258 with fluorecently-labelled duplex

The fluorescence spectra were recorded using an LS 50B Luminescence Spectrometer (Perkin Elmer). To 2.95 mL of filtered buffer (10 mM tris-HCl, 50 mM NaCl, 1.5 mM MgCl$_2$, 250 ng.μL^{-1} BSA, pH 8.0 at 25 °C) was added the labelled duplex (00165R/00160R, 0.1 μM), and the emission at 520 nm recorded, from fixed excitation at 495 nm. To the solution was then added Hoechst 33258 (0.01 μM aliquots to 0.2 μM, then 0.1 μM aliquots to 1 μM), and the emission spectra from fixed excitation at 495 nm re-recorded after each addition. Non-linear curve fitting was carried out using the SigmaPlot 8.0 program (SPSS Inc.).

6.4 Fluorescence melting protocols

6.4.1 Mismatch discrimination by fluorophore-intercalator probes

The melting curves obtained on the LightCycler (Roche) were performed at the laboratories of Oswel Research Products Ltd., Boldrewood, University of Southampton. The probe (FA12-18, 0.5 μM) was added to a LightCycler capillary containing a buffer solution (20 mM (NH$_4$)$_2$SO$_4$, 75 mM tris-HCl, 0.1 % tween, 4 mM MgCl$_2$, 250 ng.μL^{-1} BSA, pH 8.0), and the complement (0.5 μM ST1303, MM1-3 or water), final volume 10 μL. Each sample was subjected to denaturation (95 °C, 30 s.), cooled (to 25 °C, 0 s.), and then heated to 95 °C (heating rate 10 °C.minute^{-1}), whilst monitoring fluorescence in Channel 1 of the LightCycler ($\lambda_{obs.}$ ~ 520 nm). Each solution was made up and run in triplicate, and the average reported.

6.4.2 Post-PCR melting curves

Melting curves recorded after each PCR were collected by heating to effect denaturing (95 °C, 0-180 s.) cooled (to 30 °C, 0-180 s.), and then heated to 95 °C (heating rate 10 °C.minute^{-1}), whilst monitoring fluorescence in Channel 1 of the LightCycler ($\lambda_{obs.}$ ~ 520 nm).

For experiments where synthetic template (ST1303) or a further aliquot of probe (FA14) were added to the amplified mixture, the appropriate oligonucleotide was introduced after PCR, to a concentration of 0.5 μM. Each sample was then subjected to denaturation (95 °C, 30 s.), cooled (to 25 °C, 0 s.), and heated to 95 °C (heating rate 10 °C.minute^{-1}), whilst monitoring fluorescence in Channel 1 of the LightCycler ($\lambda_{obs.}$ ~ 520 nm).

6.4.3 LightCycler study of DNA binding agents' interactions with fluorescently labelled oligonucleotides

Mitoxantrone, ethidium bromide, coralyne chloride, 7-aminoactinomycin D, acridine orange and nogalamycin were purchased from Sigma/Aldrich and were of molecular biology grade. Hoescht 33258, daunomycin and actinomycin D were kind gifts of Prof. K. Fox (University of Southampton). 9-(methylsulfanyl)-acridine orange (AO-SMe) was obtained from Dr. L.J. Brown (Oswel Research Products). Disperse Blue 1 was provided in pure form by Dr. J. P. May (University of Southampton). NU:UB 31 was a gift of Dr. David Mincher (Napier University). Stock solutions of DNA-binding agents were prepared to a minimum concentration of 1 mM in water, DMSO or MeOH, such that final assay conditions included no more than 1 % organic solvent.

The fluorescently-labelled oligonucleotides (00163R, 00164R, 00165R or 00172R, 0.5 μM) were added to a LightCycler capillary containing a buffer solution (10 mM tris-HCl, 50 mM NaCl, 1.5 mM MgCl$_2$, 250 ng.μL^{-1} BSA, pH 8.0 at 25 °C), the DNA-binding agent (0, 0.25, 0.5, 1, 2.5, 5 or 10, 25, 50 or 100 μM) and the complement (0.5μM 00158R, 00159R, 00160R, 00173R or water), final volume 10 μL. Each tube was subjected to denaturation (95 °C, 30 s.), cooled (to 25 °C, 0 s.), and then heated to 95 °C (heating rate 10 °C.minute^{-1}), whilst monitoring fluorescence.

6.4.4 Investigation of the hybridisation of probes to synthetic amplicons

The results were obtained using the LightCycler. Each LightCycler capillary contained the probe (FA12, MB1303 or TQM, 0.5 μM), the target strand of the amplicon (T1303, TMG or TMI, 0.5 μM), the competing strand (C1303, CMG or CMI, 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45 or 5 μM), and a buffer solution (20 mM $(NH_4)_2SO_4$, 75 mM tris-HCl, 0.1% tween, 4 mM $MgCl_2$, 250 ng.$μL^{-1}$ BSA, pH 8.0). Each sample was subjected to denaturation (95 °C, 30 s.), cooled (to 25 °C, 0 s.), and then heated to 95 °C (heating rate 10 °C.$minute^{-1}$), whilst monitoring fluorescence in Channel 1 of the LightCycler ($\lambda_{obs.}$ ~ 520 nm).

6.4 UV-Spectroscopy

6.4.1 Mode of action studies of fluorophore-intercalator labelled probes

UV/visible spectra were measured on a Perkin Elmer UV/Vis Lambda 2 spectrometer. To 1 mL of filtered buffer (100 mM Sodium Phosphate, 1 mM EDTA, 0.1 M NaCl, pH 7.0) was added the labelled oligonucleotide (A1, F1, FA14 or FA16, to a final concentration of 7.7 μM), and the absorption spectra recorded, from 230-595 nm. To the solution was then added the synthetic target (ST1303, 2 eq., final concentration 15.4 μM), and the absorption spectra from 230-595 nm re-recorded.

6.4.2 UV Melting

Ultra-violet absorbance was measured on a Lambda 15 UV/VIS Spectrophotometer (Perkin Elmer) at a wavelength of 260 nm (A_{260}). The melting curves were established by heating from 15.0 °C to 80.0 °C, with a 1 °C.$minute^{-1}$ temperature gradient. Melting temperature (T_m) was then determined from the position of the peak in the first derivative plot of the melting curve, generated using the PECSS 2 software (Perkin Elmer).

To 1 mL of filtered buffer (100 mM Sodium Phosphate, 1 mM EDTA, 0.1 M NaCl, pH 7.0) was added the probe (FA12-18, 1.0 μM) and the synthetic target (ST1303,

1.3 eq., 1.3 μM), and the sample heated, (75 °C, 1 minute) cooled (15 °C, 1 minute), then the melting curve collected. The samples were run in triplicate, and the average reported.

6.5 PCR reactions

Human genomic DNA samples were obtained from Roche or Coriell and were stored in 1 % BSA at a concentration of 5 ng.μL^{-1}. Sequence data for all loci were obtained from GenBank (Table 6.1). Primer sequences were designed close to the mutation site using Oligonucleotide 4.0 software (National Biosciences Inc.). Primers and probes were checked for secondary structure using the DNA mfold program (http://www.ibc.wustl.edu.~zuker), using the thermodynamic parameters of Santalucia.[248] BioTaq Polymerase, exo- *Taq* Polymerase and HotStarTaq polymerases were purchased from BioLine, DNAmp and Qiagen Ltd. respectively, and were used with the manufacturers' buffers (1x concentration). All dNTPs were obtained from Roche. SYBR Green was obtained from Eurogentec. Unless otherwise stated, fluorescence intensities of real time PCR reactions were corrected using the arithmetic normalisation function of the LightCycler software (Roche). The average fluorescence during the first five cycles is subtracted from the intensity observed in each cycle.

Locus	GenBank Accession number
N1303K	M55128
W1282X	M55127
MTHFR	AY338232

Table 6.2 GenBank accession numbers for loci amplified by PCR.

6.5.1 N1303K locus

6.5.1.1 Real time symmetric PCR with bimolecular fluorophore-intercalator labelled hybridisation probes

Reactions monitored with bimolecular fluorophore-intercalator hybridisation probes were performed in a LightCycler capillary which contained the probe (FA12 or 14, 0.5 μM), dNTPs (200 μM each of the four nucleotides), each primer (FP1303 and RP1303, 0.5 μM), BioTaq Polymerase (BioLine) or exo- *Taq* Polymerase (DNAmp) (0.5 U) with the manufacturer's buffer (1x) and the template (5 ng heterozygote or homozygote genomic DNA or water), final volume 10 μL. Each tube was subjected to denaturation (95 °C, 5 minutes), cooled to annealing temperature (30 s.), and then subjected to 100 cycles of extension, denaturing and annealing.

Reactions monitored with SYBR Green were performed in the same way, except samples additionally contained SYBR Green (1 μL from a 1/1000 dilution of the stock).

6.5.1.2 Real time PCR with fluorophore-intercalator primer-probe or Scorpion

Reactions with primer-probe PRP1303 and Scorpion FAS1303 were performed in a LightCycler capillary which contained dNTPs (200 μM each of the four nucleotides), each primer (FP1303 and PRP1303 or FAS1303, 0.5 μM) BioTaq Polymerase (BioLine) or exo- *Taq* Polymerase (DNAmp) (0.5 U) with the manufacturer's buffer (1x) and the template (5 ng heterozygote or homozygote genomic DNA or water), final volume 10 μL. Each tube was subjected to denaturation (95 °C, 5 minutes), cooled to annealing temperature (30 s.), and then subjected to 100 cycles of extension, denaturing and annealing.

Reactions monitored with SYBR Green were performed in the same way, except samples additionally contained SYBR Green (1 μL from a 1/1000 dilution of the stock).

6.5.1.1 Real time asymmetric PCR with bimolecular fluorophore-intercalator labelled hybridisation probe

Reactions monitored with bimolecular fluorophore-intercalator hybridisation probes were performed in a LightCycler capillary, which contained the probe (FA12, 0.5 µM), dNTPs (200 µM each of the four nucleotides), each primer (FP1303, 0.5 µM. and RP1303, 0.001, 0.005, 0.01, 0.05, 0.1 or 0.5 µM), BioTaq Polymerase (BioLine) or exo- *Taq* Polymerase (DNAmp) (0.5 U) with the manufacturer's buffer (1x) and the template (5 ng heterozygote or homozygote genomic DNA or water), final volume 10 µL. Each tube was subjected to denaturation (95 °C, 5 minutes), cooled to annealing temperature (30 s.), and then subjected to 100 cycles of extension, denaturing and annealing.

Cycle type	Annealing temperature (°C)	Annealing time (s)	Monitoring temperature (°C)	Monitoring time (s)	Extension temperature (°C)	Extension time (s)	Denaturing temperature (°C)	Denaturing time (s)
Long	44[†]	5	-	-	72	3	95	1
Standard	44[†]	30	-	-	72	30	95	30
Step down	55*[†]	30	-	-	72	20	95	5
Step down (fast)	55* [†]	5	-	-	72	3	95	5
Scorpion	55*	0	62[†]	3	-	-	95	0
Primer-Probe	55* [†]	3	-	-	72	1	95	5

Table 6.3 Cycling conditions used in N1303K PCR reactions. * Indicates that annealing temperatures were stepped down to 44 °C over the first sixteen cycles of PCR. [†] Indicates that fluorescence was monitored during this temperature in the cycle.

6.5.2 W1282X locus

Reactions were performed in a Rotorgene tube (Corbett Research), which contained dNTPs (200 μM each of the four nucleotides), each primer (FP1282 and RP1282 or PRP1282, 0.5 μM), the probe (HP1282, 0 or 0.5 μM) exo- *Taq* Polymerase (DNAmp) (0.5 U) with the manufacturer's buffer (1x), ethidium bromide (10 μM) and the template (5 ng genomic DNA or water), final volume 20 μL. Each tube was subjected to denaturation (95 °C, 5 minutes), cooled to annealing temperature (45 °C, 30 s.), and then subjected to 100 cycles of extension (72 °C, 30 s.), denaturing (95 °C, 30 s.) and annealing (45 °C, 30 s.).

6.5.3 MTHFR locus

6.5.3.1 Optimisation of dITP incorporation

PCR reactions in a final volume of 50 μL contained each primer (FPM or CFPM and BPM or CBPM, 0.25 μM), the template (TMG, 25 ng), dNTPs (dATP, dCTP, dTTP and dGTP or dITP, 200 μM), $MgCl_2$ (1.5 – 7.5 mM) and HotStarTaq Polymerase (Qiagen) (1.25 or 2.5 U) in the manufacturer's buffer (1x). Each tube was subjected to an initial activation (95 °C, 15 minutes), cooled to annealing temperature (45 °C or 55 °C), and then subjected to 30 cycles of annealing (45 or 55 °C, 30 s.) extension (72 °C, 30 s.) and denaturing (95 °C, 30 s.), followed by a final extension phase (72 °C, 10 minutes) using a Progene thermal cycler (Techne). After thermal cycling, 5 μL of each sample was electrophoresed on a 2 % agarose gel (running buffer 0.5x TBE), stained with ethidium bromide, and photographed under a UV-transilluminator.

Post-PCR melting curves were obtained by adding SYBR Green (1 μL from a 1/1000 dilution of the stock) to 9 μL of the crude reaction mixture following thermal cycling in a LightCycler capillary. Each tube was subjected to denaturation (95 °C, 30 s.), cooled (to 40 °C, 0 s.), and then heated to 95 °C (heating rate 10 °C.minute^{-1}, whilst monitoring fluorescence in Channel 1 of the LightCycler ($\lambda_{obs.} \sim 520$ nm).

6.5.3.2 Real time TaqMan® and Molecular Beacon PCRs

Real time PCRs were performed in a LightCycler capillary which contained the probe (TQM or MBM, 0.5 µM), the manufacturer's buffer (1x), $MgCl_2$ (4 mM) each primer (FPM and RPM, 0.5 µM) HotStarTaq polymerase (0.5, 1.0 or 5.0 U), dNTPs (dATP, dCTP, dTTP and dGTP or dITP, 200 µM) and the template (5 ng human genomic DNA), final volume 10 µL. Each tube was subjected to denaturation (95 °C, 15 minutes), cooled to annealing temperature (45 °C, 30 s.), and then subjected to 100 cycles of extension (72 °C, 30 s.), denaturing (95 °C, 30 s.) and annealing (45 °C, 30 s.).

Reactions monitored with SYBR Green were performed in the same way, except samples additionally contained SYBR Green (1 µL from a 1/1000 dilution of the stock).

7. Appendix

FAM (5'-end)

9-amino-6-chloro-2-methoxyacridine (chain form)

Methyl Red dR (3'-end)

2'-deoxyinosine (chain form)

TAMRA (3'-end)

Octanediol (3'-end)

Hexaethylene glycol (chain form)
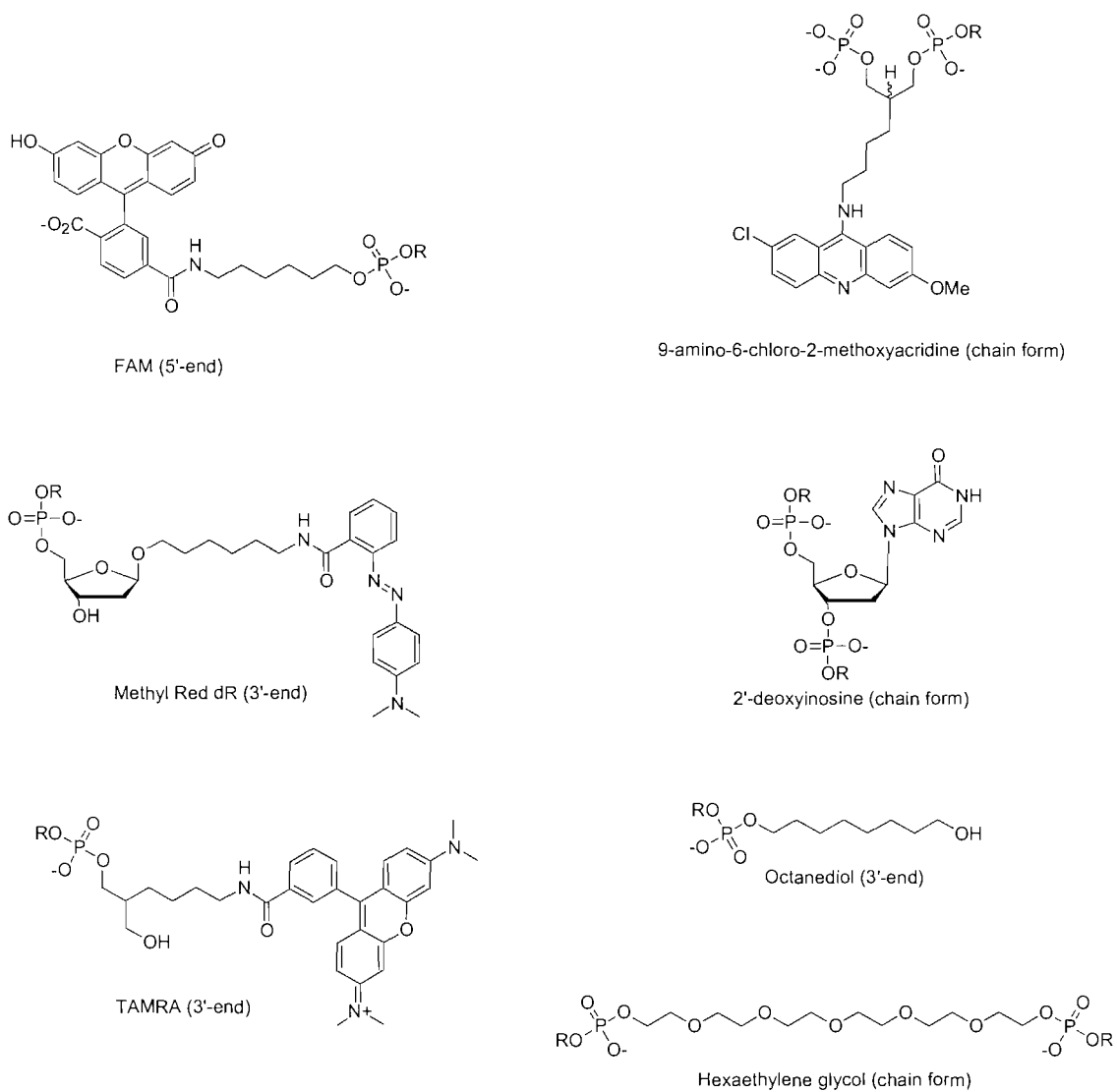
Figure 7.1 Chemical structures of non-standard monomers incorporated into synthetic oligonucleotides.

8. References

1   E. Chargaff, E. Vischer, R. Doniger, C. Green, and F. Misani, *J. Biol. Chem.*, 1951, **177**, 405.

2   J. D. Watson and F. H. C. Crick, *Nature (London)*, 1953, **171**, 737.

3   R. Franklin and R. G. Gosling, *Nature (London)*, 1953, **172**, 156.

4   http://www.urmc.rochester.edu smd grad molecule.htm

5   J. D. Watson and F. H. C. Crick, *Nature (London)*, 1953, **172**, 156.

6   M. J. Bessman, I. R. Lehman, J. Adler, S. B. Zimmerman, E. S. Simms, and A. Kornberg, *Proc. Natl. Acad. Sci. USA*, 1958, **44**, 633.

7   A. Kornberg, *Science*, 1960, **131**, 1503.

8   L. S. Lerman, *J. Mol. Biol.*, 1961, **3**, 18.

9   L. P. G. Wakelin, M. Romanos, T. K. Chen, D. Glaubiger, E. S. Cannelakis, and M. J. Waring, *Biochemistry*, 1978, **17**, 5057.

10   M. J. Waring, *J. Mol. Biol.*, 1970, **54**, 247.

11   H. M. Berman and P. R. Young, *Ann. Rev. Biophys. Bioeng.*, 1981, **10**, 87.

12   D. J. Patel and L. L. Canuel, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 3343.

13   W. Muller and D. M. Crothers, *Eur. J. Biochem.*, 1975, **54**, 267.

14   M. J. Waring, *Annu. Rev. Biochem.*, 1981, **50**, 159.

15   A. H.-J. Wang, G. Ughetto, G. J. Quigley, and A. Rich, *Biochemistry*, 1987, **26**, 1152.

16   Y. C. Liaw, Y. G. Gao, H. Robinson, G. A. Vandermarel, J. H. Vanboom, and A. H. J. Wang, *Biochemistry*, 1989, **28**, 9913.

17   W. Muller and D. M. Crothers, *J. Mol. Biol.*, 1968, **35**, 251.

18   M. Gellert, C. E. Smith, D. Neville, and G. Felsenfeld, *J. Mol. Biol.*, 1965, **11**, 445.

19   H. M. Sobell and S. C. Jain, *J. Mol. Biol.*, 1972, **68**, 21.

20   R. P. Haugland, 'Handbook of Fluorescent Probes and Research Chemicals', Molecular Probes Inc., 1999.

21   U. Asseline, M. Delarue, G. Lancelot, F. Toulme, N. T. Thuong, T. Montenay-Garestier, and C. Hélène, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 3297.

22   M. Grigoriev, D. Praseuth, P. Robin, A. Hemar, T. Saisonbehmoaras, A. Dautryvarsat, N. T. Thuong, C. Hélène, and A. Harelbellan, *J. Biol. Chem.*, 1992, **267**, 3389.

23   J. D. Watson, 'Molecular Biology of the Gene', Addison-Wesley, 1976.

24   P. H. Silverman, *The Scientist*, 2004, **18**, 32.

25   J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, *et al.*, *Science*, 2001, **291**, 1304.

26   R. F. Service, *Science*, 2003, **302**, 1316.

27   R. Kumar, N. Vandegraaff, L. Mundy, C. J. Burrell, and P. Li, *J. Virol. Methods*, 2002, **105**, 233.

28   P. O. Davies and G. L. Ridgway, *Int. J. STD AIDS*, 1997, **8**, 731.

29   T. Ishiguro, J. Saitoh, H. Yawata, H. Yamagishi, S. Iwasaki, and Y. Mitoma, *Anal. Biochem.*, 1995, **229**, 207.

30   M. Varma-Basil, H. El-Hajj, S. A. E. Marras, M. H. Hazbon, J. M. Mann, N. D. Connell, F. R. Kramer, and D. Alland, *Clin. Chem.*, 2004, **50**, 1060.

31   S. A. Bustin, *J. Mol. Endocrinol.*, 2000, **25**, 169.

32   S. A. Bustin, *J. Mol. Endocrinol.*, 2002, **29**, 23.

33   K. Tanimoto, H. Eguchi, T. Yoshida, K. Hajiro-Nakanishi, and S. Hayashi, *Nucleic Acids Res.*, 1999, **27**, 903.

34   F. Borriello, D. S. Weinberg, and G. L. Mutter, *Diagn. Mol. Pathol.*, 1994, **3**, 246.

35   L. M. Kunkel, *Nature*, 1986, **322**, 73.

36   N. Puget, D. Stoppa-Lyonnet, O. M. Sinilnikova, S. Pages, H. T. Lynch, G. M. Lenoir, and S. Mazoyer, *Cancer Res.*, 1999, **59**, 455.

37   J. F. Gusella, N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi, *et al.*, *Nature*, 1983, **306**, 234.

38   G. T. Horn, B. Richards, J. J. Merrill, and K. W. Klinger, *Clin. Chem.*, 1990, **36**, 1614.

39   C. M. Waterfall and B. D. Cobb, *Nucleic Acids Res.*, 2001, **29**, e119.

40   C. H. M. van Moorsel, E. E. van Wijngaarden, I. Fokkema, J. T. den Dunnen, D. Roos, R. van Zwieten, P. C. Giordano, and C. L. Harteveld, *Eur. J. Hum. Genet.*, 2004, **12**, 567.

41   L. Roberts, *Science*, 2000, **287**, 1898.

42   S. M. Weissman, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 8543.

43   Z. Wang and J. Moult, *Hum. Mutat.*, 2001, **17**, 263.

44   K. J. Morrow, *Genet. Eng. News*, 2000, **20**, 18.

45   L. Jin, R. Chakraborty, H. A. Hammond, and C. T. Caskey, *Am. J. Hum. Genet.*, 1991, **49**, 14.

46   E. Trabetti, R. Galavotti, and P. Pignatti, *Mol. Cell. Probes*, 1993, **7**, 81.

47   N. Goedecke, B. McKenna, S. El-Difrawy, L. Carey, P. Matsudaira, and D. Ehrlich, *Electrophoresis*, 2004, **25**, 1678.

48   J. R. Lakowicz, 'Principles of Fluorescence Spectroscopy', Plenum Press, 1983.

49   L. J. Brown, 'Oswel Training Manual', Oswel Research Products, 2003.

50   D. L. Dexter, *J. Chem. Phys.*, 1953, **21**, 836.

51   S. A. E. Marras, F. R. Kramer, and S. Tyagi, *Nucleic Acids Res.*, 2002, **30**, e122.

52   T. Förster, *Annalen der Physik*, 1948, **2**, 55.

53   K. J. Livak, S. A. J. Flood, J. Marmaro, W. Giusti, and K. Deetz, *PCR Methods and Applications*, 1995, 357.

54   M. K. Johansson and R. M. Cook, *Chem.-Eur. J.*, 2003, **9**, 3466.

55   R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich, *Science*, 1988, **239**, 487.

56   K. B. Mullis, *Scientific American*, 1990, 36.

57   C. R. Newton, A. Graham, L. E. Heptinstall, S. J. Powell, C. Summers, N. Kalsheker, J. C. Smith, and A. F. Markham, *Nucleic Acids Res.*, 1989, **25**, 2516.

58   F. Sanger, *Science*, 1981, **214**, 1205.

59   J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jensen, and K. Baumeister, *Science*, 1987, **238**, 336.

60   L. G. Lee, S. L. Spurgeon, C. R. Heiner, S. C. Benson, B. B. Rosenblum, S. M. Menchen, R. J. Graham, A. Constantinescu, K. G. Upadhya, and J. M. Cassel, *Nucleic Acids Res.*, 1997, **25**, 2816.

61   L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, and L. E. Hood, *Nature*, 1986, **321**, 674.

62   C. A. Heid, J. Stevens, K. J. Livak, and P. M. Williams, *Genome Res.*, 1996, **6**, 986.

63   U. E. M. Gibson, C. A. Heid, and P. M. Williams, *Genome Res.*, 1996, **6**, 995.

64   J. Wilhelm and A. Pingoud, *Chembiochem*, 2003, **4**, 1120.

65   D. S. Zarlenga and J. Higgins, *Vet. Parasitol.*, 2001, **101**, 215.

66   C. A. Foy and H. C. Parkes, *Clin. Chem.*, 2001, **47**, 990.

67   A. K. Lockley and R. G. Bardsley, *Trends Food Sci. Technol.*, 2000, **11**, 67.

68   E. Anklam, F. Gadani, P. Heinze, H. Pijnenburg, and G. Van Den Eede, *Eur. Food Res. Technol.*, 2002, **214**, 3.

69   C. T. Wittwer, M. G. Herrmann, A. A. Moss, and R. P. Rasmussen, *Biotechniques*, 1997, **22**, 130.

70   J. Skeidsvoll and P. M. Ueland, *Anal. Biochem.*, 1995, **231**, 359.

71   R. Higuchi, C. Fockler, G. Dollinger, and R. Watson, *Bio-Technology*, 1993, **11**, 1026.

72   S. Y. Tseng, D. Macool, V. Elliott, G. Tice, R. Jackson, M. Barbour, and D. Amorese, *Anal. Biochem.*, 1997, **245**, 207.

73   T. B. Morrison, J. J. Weis, and C. T. Wittwer, *Biotechniques*, 1998, **24**, 954.

74   K. M. Ririe, R. P. Rasmussen, and C. T. Wittwer, *Anal. Biochem.*, 1997, **245**, 154.

75   N. Marziliano, E. Pelo, B. Minuti, I. Passerini, F. Torricelli, and L. Da Prato, *Clin. Chem.*, 2000, **46**, 423.

76   D. Pirulli, M. Boniotto, D. Puzzer, A. Spano, A. Amoroso, and S. Crovella, *Clin. Chem.*, 2000, **46**, 1842.

77   N. von Ahsen, M. Oellerich, and E. Schutz, *Clin. Chem.*, 2001, **47**, 1331.

78   S. Germer and R. Higuchi, *Genome Res.*, 1999, **9**, 72.

79   I. A. Nazarenko, S. K. Bhatnagar, and R. J. Hohman, *Nucleic Acids Res.*, 1997, **25**, 2516.

80   G. J. Nuovo, R. J. Hohman, G. A. Nardone, and I. A. Nazarenko, *J. Histochem. Cytochem.*, 1999, **47**, 273.

81   I. Nazarenko, R. Pires, B. Lowe, M. Obaidy, and A. Rashtchian, *Nucleic Acids Res.*, 2002, **30**, 2089.

82   I. Nazarenko, B. Lowe, M. Darfler, P. Ikonomi, D. Schuster, and A. Rashtchian, *Nucleic Acids Res.*, 2002, **30**, e37.

83   E. R. Kandimalla and S. Agrawal, *Bioorg. Med. Chem.*, 2000, **8**, 1911.

84   Q. Q. Li, G. Y. Luan, Q. P. Guo, and J. X. Liang, *Nucleic Acids Res.*, 2002, **30**, e5.

85   M. J. Fiandaca, J. J. Hyldig-Nielsen, B. D. Gildea, and J. M. Coull, *Genome Res.*, 2001, **11**, 609.

86   O. K. Kaboev, L. A. Luchkina, A. N. Tret'iakov, and A. R. Bahrmand, *Nucleic Acids Res.*, 2000, **28**, e94.

87   D. M. Kong, H. X. Shen, Y. P. Huang, and H. F. Mi, *Biotechnol. Lett.*, 2004, **26**, 277.

88   P. M. Holland, R. D. Abramson, R. Watson, and D. H. Gelfand, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 7276.

89   L. G. Lee, C. R. Connell, and W. Bloch, *Nucleic Acids Res.*, 1993, **21**, 3761.

90   T. Morris, B. Robertson, and M. Gallagher, *J. Clin. Microbiol.*, 1996, **34**, 2933.

91   B. Kimura, S. Kawasaki, T. Fujii, J. Kusunoki, T. Itoh, and S. J. A. Flood, *J. Food Protection*, 1999, **62**, 329.

92   I. Laurendeau, M. Bahuau, N. Vodovar, C. Larramendy, M. Olivi, I. Bieche, M. Vidaud, and D. Vidaud, *Clin. Chem.*, 1999, **45**, 982.

93   J. Nurmi, A. Ylikoski, T. Soukka, M. Karp, and T. Lovgren, *Nucleic Acids Res.*, 2000, **28**, e28.

94   J. Nurmi, T. Wikman, M. Karp, and T. Lovgren, *Anal. Chem.*, 2002, **74**, 3525.

95   A. Solinas, N. Thelwell, and T. Brown, *Chem. Commun.*, 2002, 2272.

96   J. B. de Kok, E. T. G. Wiegerinck, B. A. J. Giesendorf, and D. W. Swinkels, *Hum. Mutat.*, 2002, **19**, 554.

97   http://www.appliedbiosystems.com

98   M. J. Moser, D. J. Marshall, J. K. Grenier, C. D. Kieffer, A. A. Killeen, J. L. Ptacin, C. S. Richmond, E. B. Roesch, C. W. Scherrer, C. B. Sherrill, *et al.*, *Clin. Chem.*, 2003, **49**, 407.

99   J. A. Piccirilli, T. Krauch, S. E. Moroney, and S. A. Benner, *Nature*, 1990, **343**, 33.

100   S. Tyagi and F. R. Kramer, *Nat. Biotechnol.*, 1996, **14**, 303.

101   L. G. Kostrikis, S. Tyagi, M. M. Mhlanga, D. D. Ho, and F. R. Kramer, *Science*, 1998, **279**, 1228.

102 A. S. Piatek, S. Tyagi, A. C. Pol, L. P. Miller, F. R. Kramer, and D. Alland, *Nat. Biotechnol.*, 1998, **16**, 359.

103 M. Taveau, D. Stockholm, M. Spencer, and I. Richard, *Anal. Biochem.*, 2002, **305**, 227.

104 J. Weusten, W. M. Carpay, T. A. M. Oosterlaken, M. C. A. van Zuijlen, and P. A. van de Wiel, *Nucleic Acids Res.*, 2002, **30**, e26.

105 M. M. Klerks, G. O. M. Leone, M. Verbeek, J. van den Heuvel, and C. D. Schoen, *J. Virol. Methods*, 2001, **93**, 115.

106 A. Tsourkas, M. A. Behlke, S. D. Rose, and G. Bao, *Nucleic Acids Res.*, 2003, **31**, 1319.

107 E. Ortiz, G. Estrada, and P. M. Lizardi, *Mol. Cell. Probes*, 1998, **12**, 219.

108 A. Tsourkas, M. A. Behlke, and G. Bao, *Nucleic Acids Res.*, 2002, **30**, 4208.

109 A. Tsourkas, M. A. Behlke, Y. Q. Xu, and G. Bao, *Anal. Chem.*, 2003, **75**, 3697.

110 D. Whitcombe, S. Kelly, J. Mann, J. Theaker, C. Jones, and S. Little, *Am. J. Hum. Genet.*, 1999, **65**, 2333.

111 N. Thelwell, S. Millington, A. Solinas, J. Booth, and T. Brown, *Nucleic Acids Res.*, 2000, **28**, 3752.

112 D. Whitcombe, J. Theaker, S. P. Guy, T. Brown, and S. Little, *Nat. Biotechnol.*, 1999, **17**, 804.

113 B. K. Saha, B. H. Tian, and R. P. Bucy, *J. Virol. Methods*, 2001, **93**, 33.

114 G. H. Cunnick, W. G. Jiang, K. F. Gomez, and R. E. Mansel, *Biochem. Biophys. Res. Commun.*, 2001, **288**, 1043.

115 A. Solinas, L. J. Brown, C. McKeen, J. M. Mellor, J. T. G. Nicol, N. Thelwell, and T. Brown, *Nucleic Acids Res.*, 2001, **29**, e96.

116 D. M. Kong, Y. P. Huang, X. B. Zhang, W. H. Yang, H. X. Shen, and H. F. Mi, *Anal. Chim. Acta*, 2003, **491**, 135.

117 G. Bonnet, S. Tyagi, A. Libchaber, and F. R. Kramer, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6171.

118 N. J. Gibson, H. L. Gillard, D. Whitcombe, R. M. Ferrie, C. R. Newton, and S. Little, *Clin. Chem.*, 1997, **43**, 1336.

119 D. J. French, C. L. Archard, T. Brown, and D. G. McDowell, *Mol. Cell. Probes*, 2001, **15**, 363.

120   N. Dobson, D. G. McDowell, D. J. French, L. J. Brown, J. M. Mellor, and T. Brown, *Chem. Commun.*, 2003, 1234.

121   D. J. French, C. L. Archard, M. T. Andersen, and D. G. McDowell, *Mol. Cell. Probes*, 2002, **16**, 319.

122   C. A. M. Seidel, A. Schulz, and M. H. M. Sauer, *J. Phys. Chem.*, 1996, **100**, 5541.

123   A. O. Crockett and C. T. Wittwer, *Anal. Biochem.*, 2001, **290**, 89.

124   C. P. Vaughn and K. S. J. Elenitoba-Johnson, *Am. J. Pathol.*, 2003, **163**, 29.

125   R. A. Cardullo, S. Agrawal, C. Flores, P. C. Zamecnik, and D. E. Wolf, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 8790.

126   K. S. J. Elenitoba-Johnson, S. D. Bohling, C. T. Wittwer, and T. C. King, *Nature Medicine*, 2001, **7**, 249.

127   S. Sueda, J. L. Yuan, and K. Matsumoto, *Bioconjugate Chem.*, 2000, **11**, 827.

128   G. L. Wang, J. L. Yuan, K. Matsumoto, and Z. Hu, *Anal. Biochem.*, 2001, **299**, 169.

129   S. Sueda, J. L. Yuan, and K. Matsumoto, *Bioconjugate Chem.*, 2002, **13**, 200.

130   M. A. Lee, A. L. Siddle, and R. H. Page, *Anal. Chim. Acta*, 2002, **457**, 61.

131   W. M. Howell, M. Jobs, and A. J. Brookes, *Genome Res.*, 2002, **12**, 1401.

132   T. B. Rasmussen, A. Uttenthal, K. de Stricker, S. Belak, and T. Storgaard, *Arch. Virol.*, 2003, **148**, 2005.

133   N. Svanvik, A. Stahlberg, U. Sehlstedt, R. Sjoback, and M. Kubista, *Anal. Biochem.*, 2000, **287**, 179.

134   K. Yamana, H. Zako, K. Asazuma, R. Iwase, H. Nakano, and A. Murakami, *Angew. Chem.-Int. Edit.*, 2001, **40**, 1104.

135   K. Yamana, T. Iwai, Y. Ohtani, S. Sato, M. Nakamura, and H. Nakano, *Bioconjugate Chem.*, 2002, **13**, 1266.

136   X. F. Wang and U. J. Krull, *Anal. Chim. Acta*, 2002, **470**, 57.

137   A. Okamoto, K. Tanaka, T. Fukuta, and I. Saito, *J. Am. Chem. Soc.*, 2003, **125**, 9296.

138   S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, and P. B. Dervan, *Nature*, 1998, **391**, 468.

139   V. C. Rucker, S. Foister, C. Melander, and P. B. Dervan, *J. Am. Chem. Soc.*, 2003, **125**, 1195.

140   M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, *Science*, 1995, **270**, 467.

141   E. Hitt, *The Scientist*, 2004, **18**, 38.

142   M. Huber, D. Losert, R. Hiller, C. Harwanegg, M. W. Mueller, and W. M. Schmidt, *Anal. Biochem.*, 2001, **299**, 24.

143   F. Erdogan, R. Kirchner, W. Mann, H.-H. Ropers, and U. A. Nuber, *Nucl. Acids. Res.*, 2001, **29**, e36.

144   U. Landegren, R. Kaiser, J. Sanders, and L. Hood, *Science*, 1988, **241**, 1077.

145   F. Barany, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 189.

146   M. Wiedmann, F. Barany, and C. A. Batt, *Appl. Environ. Microbiol.*, 1993, **59**, 2743.

147   D. A. Nickerson, R. Kaiser, S. Lappin, J. Stewart, L. Hood, and U. Landegren, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 8923.

148   M. Samiotaki, M. Kwiatkowski, J. Parik, and U. Landegren, *Genomics*, 1994, **20**, 238.

149   X. N. Chen, K. J. Livak, and P. Y. Kwok, *Genome Res.*, 1998, **8**, 549.

150   Y. Z. Xu, N. B. Karalkar, and E. T. Kool, *Nat. Biotechnol.*, 2001, **19**, 148.

151   Y. Z. Xu and E. T. Kool, *Nucleic Acids Res.*, 1999, **27**, 875.

152   S. Sando and E. T. Kool, *J. Am. Chem. Soc.*, 2002, **124**, 2096.

153   J. P. Schouten, C. J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, and G. Pals, *Nucleic Acids Res.*, 2002, **30**, e57.

154   D. Cockburn, H. Snowden, J. Schouten, and G. Taylor, *J. Med. Genet.*, 2003, **40**, S86.

155   D. Baty, R. Butler, G. Cross, and C. McAnulty, *J. Med. Genet.*, 2003, **40**, S29.

156   E. Sistermans, C. de Kovel, D. Smeets, C. van Ravenswaay, B. de Vries, H. Brunner, and W. Nillesen, *Am. J. Hum. Genet.*, 2003, **73**, 876.

157   A. Erlandson, L. Samuelsson, B. Hagberg, M. Kyllerman, M. Vujic, and J. Wahlstrom, *Genet. Test.*, 2003, **7**, 329.

158   J. Herbergs, C. Van Roozendaal, E. Smeets, D. Tserpelis, C. De Die-Smulders, A. Midro, U. Moog, C. Schrander-Stumpel, and H. Smeets, *Am. J. Hum. Genet.*, 2003, **73**, 2469.

159   A. Errami, G. S. Salomons, S. J. M. van Dooren, J. J. P. Gille, G. Pals, J. P. Schouten, O. Elpeleg, and C. Jakobs, *Am. J. Hum. Genet.*, 2002, **71**, 1486.

160   M. Nilsson, H. Malmgren, M. Samiotaki, M. Kwiatkowski, B. P. Chowdhary, and U. Landegren, *Science*, 1994, **265**, 2085.

161   M. Nilsson, K. Krejci, J. Koch, M. Kwiatkowski, P. Gustavsson, and U. Landegren, *Nature Genet.*, 1997, **16**, 252.

162   V. Lyamichev, A. L. Mast, J. G. Hall, J. R. Prudent, M. W. Kaiser, T. Takova, R. W. Kwiatkowski, T. J. Sander, M. de Arruda, D. A. Arco, *et al.*, *Nat. Biotechnol.*, 1999, **17**, 292.

163   M. Olivier, L. M. Chuang, M. S. Chang, Y. T. Chen, D. Pei, K. Ranade, A. de Witte, J. Allen, N. Tran, D. Curb, *et al.*, *Nucleic Acids Res.*, 2002, **30**, e53.

164   M. Nagano, S. Yamashita, K. Hirano, M. Ito, T. Maruyama, M. Ishihara, Y. Sagehashi, T. Oka, T. Kujiraoka, H. Hattori, *et al.*, *J. Lipid Res.*, 2002, **43**, 1011.

165   Y. Mashima, M. Nagano, T. Funayama, Q. Zhang, T. Egashira, J. Kudho, N. Shimizu, and Y. Oguchi, *Clin. Biochem.*, 2004, **37**, 268.

166   N. Svanvik, G. Westman, D. Y. Wang, and M. Kubista, *Anal. Biochem.*, 2000, **281**, 26.

167   J. Isacsson, H. Cao, L. Ohlsson, S. Nordgren, N. Svanvik, G. Westman, M. Kubista, R. Sjoback, and U. Sehlstedt, *Mol. Cell. Probes*, 2000, **14**, 321.

168   J. Nygren, N. Svanvik, and M. Kubista, *Biopolymers*, 1998, **46**, 39.

169   K. Yamana, S. Kumamoto, and H. Nakano, *Chem. Lett.*, 1997, 1173.

170   A. Mahara, R. Iwase, T. Sakamoto, K. Yamana, T. Yamaoka, and A. Murakami, *Angew. Chem.-Int. Edit.*, 2002, **41**, 3648.

171   K. Yamana, R. Iwase, S. Furutani, H. Tsuchida, H. Zako, T. Yamaoka, and A. Murakami, *Nucleic Acids Res.*, 1999, **27**, 2387.

172   K. Yamana, T. Mitsui, J. Yoshioka, T. Isuno, and H. Nakano, *Bioconjugate Chem.*, 1996, **7**, 715.

173   K. Yamana, R. Aota, and H. Nakano, *Tetrahedron Lett.*, 1995, **36**, 8427.

174   K. Yamana, M. Takei, and H. Nakano, *Tetrahedron Lett.*, 1997, **38**, 6051.

175   K. Shinozuka, Y. Seto, and H. Sawai, *Chem. Commun.*, 1994, 1377.

176   A. Yamane, *Nucleic Acids Res.*, 2002, **30**, e97.

177   D. E. Comings, J. Limon, A. Ledochowski, and K. C. Tsou, *Exp. Cell Res.*, 1978, **117**, 451.

178   L. G. Lee, C. R. Connell, and W. Bloch, *Nucleic Acids Res.*, 1993, **21**, 3761.

179   L. Osborne, G. Santis, M. Schwarz, K. Klinger, T. Dork, I. McIntosh, M. Schwartz, V. Nunez, M.Maczek, J. Reiss, *et al.*, *Hum. Genet.*, 1992, **89**, 653.

180   L. Osborne, G. Santis, M. Schwarz, K. Klinger, T. Dork, I. McIntosh, M. Schwartz, V. Nunes, M. Macek, J. Reiss, *et al.*, *Hum. Genet.*, 1992, **89**, 653.

181   J. Marmur and P. Doty, *Nature*, 1959, **183**, 1427.

182   J. P. May, L. J. Brown, I. Rudloff, and T. Brown, *Chem. Commun.*, 2003, 970.

183   A. M. Boguszewska-Chachulska, M. Krawczyk, A. Stankiewicz, A. Gozdek, A. L. Haenni, and L. Strokovskaya, *FEBS Lett.*, 2004, **567**, 253.

184   H. Asanuma, T. Takarada, T. Yoshida, D. Tamaru, X. G. Liang, and M. Komiyama, *Angew. Chem.-Int. Edit.*, 2001, **40**, 2671.

185   Y. Shi, A. Kuzuya, and M. Komiyama, *Chem. Lett.*, 2003, **32**, 464.

186   K. Fukui, K. Iwane, T. Shimidzu, and K. Tanaka, *Tetrahedron Lett.*, 1996, **37**, 4983.

187   R. T. Ranasinghe, L. J. Brown, and T. Brown, *Chem. Commun.*, 2001, 1480.

188   M. Egli, L. D. Williams, C. A. Frederick, and A. Rich, *Biochemistry*, 1991, **30**, 1364.

189   K. X. Chen, N. Gresh, and B. Pullman, *J. Biomol. Struct. Dyn.*, 1985, **3**, 445.

190   G. Gaudiano, M. Frigerio, P. Bravo, and T. H. Koch, *J. Am. Chem. Soc.*, 1992, **114**, 3107.

191   P. F. Wiley, R. B. Kelly, E. L. Caron, V. H. Wiley, J. H. Johnson, F. A. MacKellar, and S. A. Mizsak, *J. Am. Chem. Soc.*, 1977, **99**, 542.

192   R. E. Jeeninga, H. T. Huthoff, A. P. Gultyaev, and B. Berkhout, *Nucleic Acids Res.*, 1998, **26**, 5472.

193   F. M. Chen, F. Sha, K. H. Chin, and S. H. Chou, *Biophys. J.*, 2003, **84**, 432.

194   F. M. Chen, F. Sha, K. H. Chin, and S. H. Chou, *Nucleic Acids Res.*, 2004, **32**, 271.

195   R. M. Wadkins, E. A. Jares-Erijman, R. Klement, A. Rüdiger, and T. M. Jovin, *J. Mol. Biol.*, 1996, **262**, 53.

196   R. M. Wadkins and T. M. Jovin, *Biochemistry*, 1991, **30**, 9469.

197   H. Yoo and R. L. Rill, *J. Mol. Recognit.*, 2001, **14**, 145.

198   D. L. Boger, B. E. Fink, S. R. Brunette, W. C. Tse, and M. P. Hedrick, *J. Am. Chem. Soc.*, 2001, **123**, 5878.

199   K.-Y. Zee-Cheng, K. D. Paull, and C. C. Cheng, *J. Med. Chem.*, 1974, **17**, 347.

200   D. Makhey, B. Gatto, C. A. Yu, A. Liu, L. F. Liu, and E. J. LaVoie, *Bioorg. Med. Chem.*, 1996, **4**, 781.

201   W. D. Wilson, A. N. Gough, and J. J. Doyle, *J. Med. Chem.*, 1976, **19**, 1261.

202   J. S. Lee, L. J. P. Latimer, and K. J. Hampel, *Biochemistry*, 1993, **32**, 5591.

203   J. Ren and J. B. Chaires, *Biochemistry*, 1999, **38**, 16067.

204   S. Jackson, J. A. Double, P. M. Loadman, D. J. Mincher, A. Turnbull, and M. C. Bibby, *Br. J. Cancer*, 2002, **86**, S112.

205   K. R. Fox, in 'Drug-DNA interaction protocols', ed. J. M. Walker, Humana Press, 1997.

206   F. G. Loontiens, P. Regenfuss, A. Zechel, L. Dumortier, and R. M. Clegg, *Biochemistry*, 1990, **29**, 9029.

207   C. M. McKeen, L. J. Brown, J. T. G. Nicol, J. M. Mellor, and T. Brown, *Org. Biomol. Chem.*, 2003, **1**, 2267.

208   T. C. Jenkins, A. N. Lane, S. Neidle, and D. Brown, *Eur. J. Biochem.*, 1993, **213**, 1175.

209   W. Leupin, W. J. Chazin, S. Hyberts, W. A. Denny, and K. Wütrich, *Biochemistry*, 1986, **25**.

210   R. A. J. Darby, M. Sollogoub, C. McKeen, L. Brown, A. Risitano, N. Brown, C. Barton, T. Brown, and K. R. Fox, *Nucleic Acids Res.*, 2002, **30**, e39.

211   B. L. Gafney, L. A. Marky, and R. A. Jones, *Tetrahedron*, 1984, **40**, 3.

212   B. C. Froehler, S. Wandwani, T. J. Terhorst, and S. R. Gerrard, *Tetrahedron Lett.*, 1992, **33**, 5307.

213   M. Egholm, O. Buchardt, and L. Christensen, *Nature*, 1993, **365**, 566.

214   L. A. Ugozzoli, D. Latorra, R. Pucket, K. Arar, and K. Hamby, *Anal. Biochem.*, 2004, **324**, 143.

215   U. B. Gyllensten and H. A. Erlich, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 7652.

216   S. K. Poddar, *Mol. Cell. Probes*, 2000, **14**, 25.

217 H.-K. Nguyen, E. Bonfiils, P. Auffray, P. Costaglioli, P. Schmitt, U. Asseline, M. Durand, J.-C. Maurizot, D. Dupret, and N. Thuong, *Nucleic Acids Res.*, 1998, **26**, 4249.

218 H.-K. Nguyen and E. M. Southern, *Nucleic Acids Res.*, 2000, **28**, 3904.

219 D. R. Mills and F. R. Kramer, *Proc. Natl. Acad. Sci. USA*, 1979, **76**, 2232.

220 S. Tabor and C. C. Ricardson, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4767.

221 N. Sasaki, M. Izawa, Y. Sugahara, T. Tanaka, M. Watahiki, K. Ozawa, E. Ohara, H. Funaki, Y. Yoneda, S. Matsuura, *et al.*, *Gene*, 1998, **222**, 17.

222 M. A. Innis, K. B. Myambo, D. H. Gelfand, and M. A. D. Brow, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 9436.

223 J. H. Spee, W. M. Devos, and O. P. Kuipers, *Nucleic Acids Res.*, 1993, **21**, 777.

224 S. L. Turner and F. J. Jenkins, *Biotechniques*, 1995, **19**, 48.

225 C. Bailly and M. J. Waring, *Nucleic Acids Res.*, 1995, **23**, 885.

226 K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 3746.

227 N. Sugimoto, S. Nakano, M. Yoneyama, and K. Honda, *Nucleic Acids Res.*, 1996, **24**, 4501.

228 A. S. Piatek, S. Tyagi, A. C. Pol, A. Telenti, L. P. Miller, F. R. Kramer, and D. Alland, *Nat. Biotechnol.*, 1998, **16**, 359.

229 J. A. Sanchez, K. E. Pierce, J. E. Rice, and L. J. Wangh, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1933.

230 P. Hilbert, M. N. Michel, D. Sartenaer, S. Rombout, B. Parmentier, Y. Gillerot, and L. Van Maldergem, *Am. J. Hum. Genet.*, 2003, **73**, 2290.

231 E. Eldering, C. A. Spek, H. L. Aberson, A. Grummels, I. A. Derks, A. F. de Vos, C. J. McElgunn, and J. P. Schouten, *Nucleic Acids Res.*, 2003, **31**, e153.

232 S. J. White, G. R. Vink, M. Kriek, W. Wuyts, J. Schouten, B. Bakker, M. H. Breuning, and J. T. den Dunnen, *Hum. Mutat.*, 2004, **24**, 86.

233 T. Brown and D. J. S. Brown, *Methods Enzymol.*, 1992, **211**, 20.

234 M. Smith, R. H. Rammler, I. H. Goldberg, and H. G. Khorana, *J. Am. Chem. Soc.*, 1963, **84**, 430.

235 H. H. Goertz and H. Seliger, *Angew. Chem.-Int. Edit. Engl.*, 1981, **20**, 681.

236 J. L. Fourrey, J. Varenne, C. Blonski, P. Dousset, and D. Shire, *Tetrahedron Lett.*, 1987, **28**, 5157.

237   R. Ramage and F. O. Wahl, *Tetrahedron Lett.*, 1993, **34**, 7133.

238   T. Horn and M. S. Urdea, *Tetrahedron Lett.*, 1986, **27**, 4705.

239   N. T. Thuong and M. Chassignol, *Tetrahedron Lett.*, 1987, **28**, 4157.

240   E. Uhlmann and J. Engels, *Tetrahedron Lett.*, 1986, **27**, 1023.

241   J. E. Celebuski, C. Chan, and R. A. Jones, *J. Org. Chem.*, 1992, **57**, 5535.

242   A. Guzaev and H. Lonnberg, *Tetrahedron*, 1999, **55**, 9101.

243   B. A. Connolly, *Tetrahedron Lett.*, 1987, **28**, 463.

244   R. I. Hogrefe, 'Glen Report', Glen Research, Virginia, 1996.

245   J. Bogan, L. Ignatovich, and E. Stankevich, *Nucleosides Nucleotides*, 1999, **18**, 1183.

246   'User's Guide to DNA Modification', Glen Research, 2002.

247   G. J. Langley, J. M. Herniman, N. L. Davies, and T. Brown, *Rapid Commun. Mass Spectrom.*, 1999, **13**, 1717.

248   J. Santalucia, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 1460.