UNIVERSITY OF BIRMINGHAM

Fluorescent Anthracene Tags for Detection of Base Pair Mismatches in DNA Duplexes

By

Sarah Didem Turker

Supervisor: Dr J. H. R. Tucker

School of Chemistry
University of Birmingham
October 2009

Abstract.

Non-nucleosidic anthracene-functionalised phosphoramidites **5a** and **5b** have been synthesised. **5b** has been incorporated into oligonucleotide sequences (**PrA**, **PrB**, **PrC**, **PrD**, **PrE**, **PrF** and **PrG**) as a DNA base analogue using automated synthesis.

Melting temperature (T_m) studies of double-stranded 15-mer / 14-mer deletion sequences compared to 15-mer / 15-mer analogues demonstrated that the incorporation of anthracene serves to stabilise the duplexes by improved intercalation in the former case.

PrA, was used as a probe for detection of base pair mismatches in target DNA sequences using fluorescence spectroscopy. Titration of **PrA** versus a range of *matched* target oligonucleotides (**TarAx** where **x** = 1 - 5) demonstrated an enhancement in anthracene fluorescence when adenine was positioned adjacent to the anthracene tag on the opposite strand for 15-mer / 15-mer duplexes . Titration of **PrA** versus a range of *mismatched* target oligonucleotide sequences (**TarBx**, **TarCx**, **TarDx**, **TarEx**, **TarFx** and **TarGx** where **x** = 1 - 2) demonstrated enhancements in emission for C-A, C-T and C-C mismatches, which were dependent on mismatch position (5' or 3') in the target strand. The latter effect is possibly due to the chiral nature of **5b**. Probes such as **PrA** show potential for use as sensors for single nucleotide polymorphisms (SNPs) in DNA sequences.

Acknowledgements

I would like to firstly thank my supervisor, Dr Jim Tucker for giving me the opportunity to work within his group and for the help and support I have had. I would also like to thank Jean-Louis Duprey for all his help throughout the year and Dr John Zhao for his help with DNA synthesis, also thanks to Joe Vyle for useful discussion.

I would like to thank Dr Neil Spencer and Peter Ashton, Graham Burns, Nick May and Lianne Hill for the analytical services.

I would like to thank the rest of the Tucker group especially Huy for the laughs and chats, Luciana for the great cakes she makes for us and her help with VT NMR.

On a more serious note, I have absolutely enjoyed my time in the group and will remember the memories and carry them into my new experiences. I would like to say a big thank you for making my experiences so enjoyable.

I would like to thank my family who are the sole reason for me being where I am, without their support I could not have achieved this.

Abbreviations

DNA	Deoxyribonucleic acid
RNA	Ribobucleic acid
AIDS	Acquired immune deficiency syndrome
HIV	Human immunodeficiency virus
BACE	β-site APP cleaving enzyme
AD	Alzheimer's disease
SNP	Single nucleotide polymorphism
A	Adenine
G	Guanine
C	Cytosine
T	Thymine
U	Uracil
^{m}RNA	Messenger RNA
^{t}RNA	Transfer RNA
BC	Beaucage-Caruthers
DMTr	Dimethoxytrityl
BDF	Base-discriminating fluorescent-
PyBOP	Benzotriazol-1yl-oxytripyrrolidinophosphonium hexafluorophosphate
HOBt	1-hydroxybenzotriazole
HBTU	*O*-benzotriazole-*N,N,N',N'*-tetramethyuronium hexafluorophosphate
NMR	Nuclear magnetic resonance
δ	Chemical shift (NMR)
m/z	Mass-to-charge ratio
DMAP	Dimethylaminopyridine
HSQC	High-spin quantum correlation
HPLC	High-performance liquid chromatography
UV-Vis	Ultraviolet-visible
VT-NMR	Variable temperature NMR
ppm	Parts per million
MS	Mass spectrometry
HRMS	High-resolution mass spectrometry
ES	Electrospray ionisation
T_m	Melting temperature (nucleic acid duplex)
IR	Infra-red
R_f	Retention factor
DCM	Dichloromethane
MeOH	Methanol
DMF	*N,N'*-dimethyl formamide

NMR multiplicities

s	Singlet
d	Doublet
t	Triplet
q	Quartet
dd	Doublet of doublets
m	Multiplet

Contents.

1.0. Introduction

1.1. Background.

Nucleic acids are arguably the most important class of biomolecules that exist within the cells of living organisms. Governing the expression of proteins within a cell, it is nucleic acids that provide the blueprint for an individual organism's biochemistry. Only a single copy of the blueprint exists per cell, and it is this genetic *code* that is replicated and passed on by parent to daughter cells during cell division. When organisms engage in sexual reproduction to produce offspring, half the genetic material from one parent and half from the other parent is passed to the progeny. In this way descendants inherit the genetic code with equal contribution from their parent ancestors.

Nucleic acids are also crucial in the development of a species as a whole. Sections of nucleic acids called *genes* encode for larger regions that determine the outward physical characteristics of an organism. Hence such qualities as skin and eye colour and athletic / intelligence potential are partially controlled by the sequence of nucleic acids contained within a cell – some characteristics require exposure to the correct environmental input, e.g. physical training or learning, to fulfil the genetic potential. Mutations within the nucleic acid palette can therefore produce new traits in an organism. If these new qualities are desirable, they can be passed down from generation to generation until only the strongest organisms remain to flourish. Nucleic acids are therefore the underlying basis of the idea of Darwinian '*survival of the*

fittest'. It is the *survive-inherit* process that is directly responsible for evolution of species, of which there are many examples e.g. the white fur of polar bears, immunity to pesticide in parasitic aphids, resistance of bacteria to an antibiotic, the active sites of proteins (which catalyse perfect chemical reactions) or the ultimate example of the human brain.

The examples above are of nucleic acids in working order and functioning normally, and as described above, mutation can be a benefit to an organism in certain cases. However, the mutation of nucleic acids[1] can also lead to disease. There are various types of lesions that can occur in nucleic acids; most are caused by non-innate external stimuli, such as mutagenic chemicals (e.g. the intercalating hydrocarbons found in cigarette smoke), ionising radiation (e.g. UV light, X-rays or nuclear radiation), radical oxygen species or through pathogenic retroviruses such as the human immuno-deficiency virus (HIV). This category of mutation is known as *somatic mutation*,[2] and is passed on from parent to daughter cells during cell division (mitosis), *via* DNA replication. Exposure of an individual's genome to the first three examples primarily results in diseases such as cancer, a state in which the normal pathways for controlling cell senescence and proliferation have broken down; cells divide rapidly and uncontrollably to produce tumours which eventually overtake and kill the organism if left untreated. Examples are the common skin cancer *malignant melanoma* which is primarily caused by exposure to the UV rays in sunlight and lung cancer in cigarette smokers caused by the polycyclic aromatic hydrocarbons ('tar') contained within the inhaled smoke. It is a miracle of existence that the body is extremely efficient

at repairing somatic mutations on an almost daily basis. The previous two centuries have seen the first real leap in life expectancy due to better sanitation and healthcare, but perversely it has also led to the indomitable rise of mortality from cancer, due to breakdown of the DNA repair pathways in the latter stages of an individual's life due to old age. Similarly, individuals exposed to HIV have their genetic material changed to incorporate the viral genetic material by the virus' nucleic acids and an enzyme known as *reverse transcriptase*. The result, after an incubation period, is a condition known as acquired immune deficiency syndrome (AIDS), where an individual elicits no immune response to pathogenic infection.

In some unfortunate individuals, diseases or susceptibility to diseases are simply inherited from the genetic material from their ancestors.[3] Diseases such as Huntington's disease,[4] Down's syndrome,[5] sickle-cell anaemia,[6], Swedish variant Alzheimer's disease,[7] diabetes,[8] juvenile glaucoma,[9] and colon cancer,[10] are all examples of diseases where the condition (or a susceptibility to the disorder in the latter four examples) is inherited by the offspring. Genetic susceptibility to cancer is determined by regions of nucleic acids sequence popularly described as *oncogenes*. Mutated DNA inherited by the progeny from the parent are called *germ-line mutations*.

It is clear then that a deep understanding of nucleic acid chemistry will be useful for the future of medicine in the identification of new and novel drugs, therapies and treatments for disease. We will therefore discuss the

chemistry of nucleic acids and then consider how the genetic blueprint is used to create proteins, and in doing so we will shed light on why nucleic acids are the ultimate controllers of cellular chemistry.

1.2. Nucleic acid structure

There are two components to nucleic acid structure: the heterocyclic bases and the sugar-phosphate diester 'backbone'. There are five natural bases: the purines guanine (G) and adenine (A), and the pyrimidines cytosine (C), thymine (T), and uracil (U). The first four bases exist in deoxyribonucleic acid (DNA), and in ribonucleic acid (RNA), uracil replaces thymine (uracil differs from thymine by a hydrogen in place of the methyl group on the pyrimidine ring). Certain pairs of these heterocyclic bases are able to make complementary hydrogen bonds to each other (Figure 1.1), G—C and A—T (U in RNA). It is worth noting that the G—C base pair is stronger than the A—T base pair due to the number of hydrogen bonds.

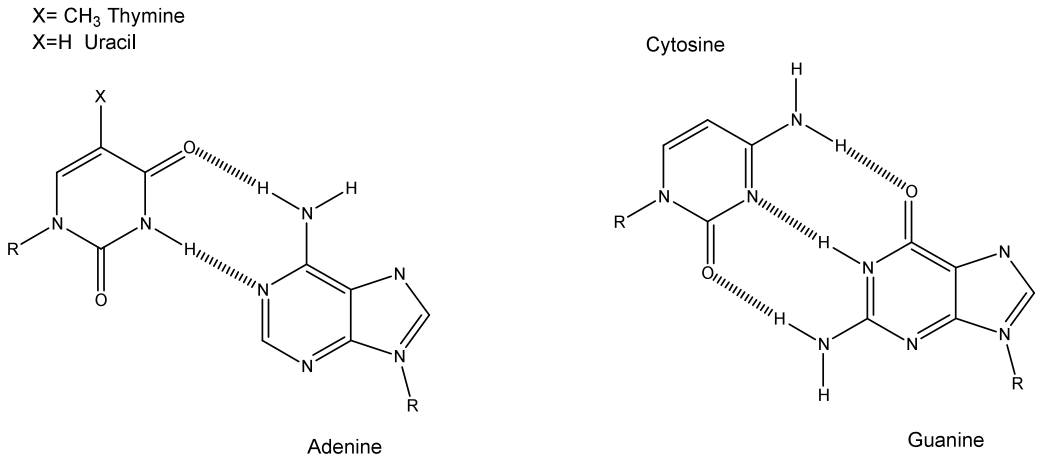


Figure 1.1: Watson- Crick base pairing in nucleic acids.

When the bases are incorporated into a sugar-phosphate backbone made up, not surprisingly of phosphate esters of sugar (the sugar in DNA is deoxyribose, whereas in RNA it is ribose), the effect is to create a 'zipper' of bases that can bind to a complementary strand through hydrogen bonding. The binding effect is cooperative, and leads to a stable double-stranded structure that is right-handed helical under normal base-pairing conditions (sometimes known as Watson and Crick base pairing), the helix is known as the B-form, and contains major and minor grooves. The sense strand is said to run from 3' to 5' based on the numbering of the sugar atoms in the backbone. The complementary antisense strand is then said to run from 5' to 3' if the same direction is considered. It is worth noting that nucleic acids behave as macromolecular polyanions in aqueous solution due to the negatively charged phosphate groups in the backbone of both strands. (Figure 1.2).

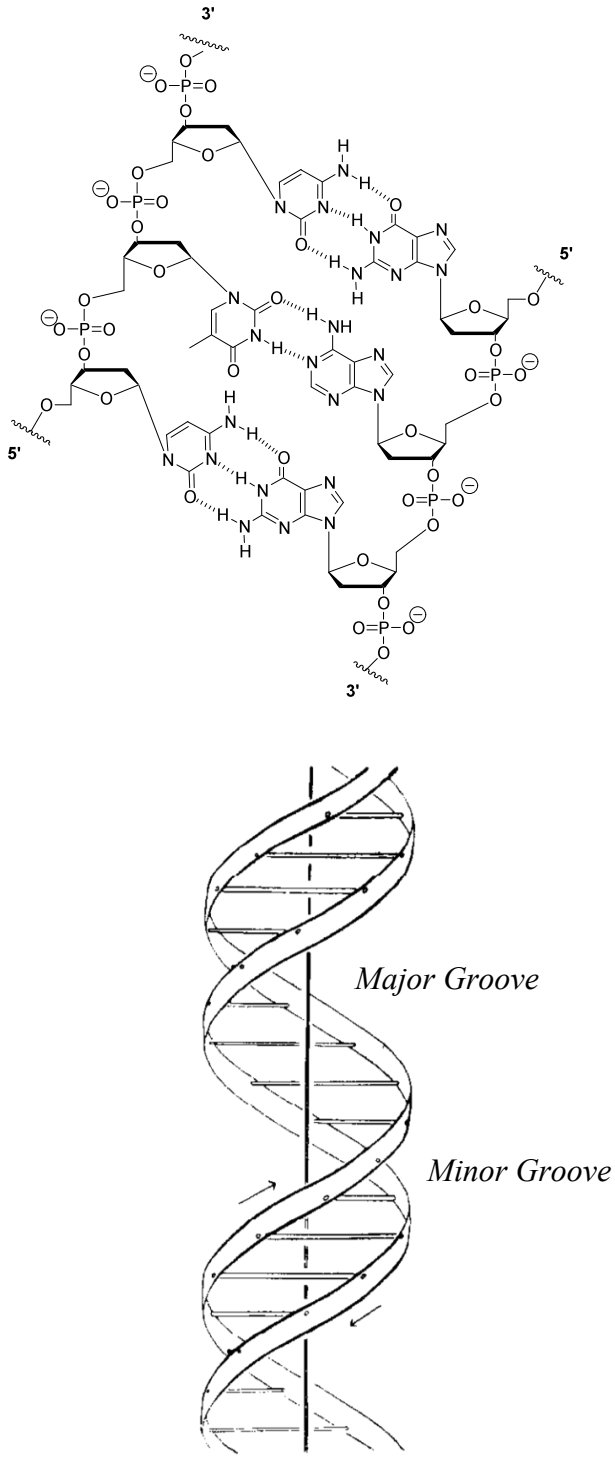


Figure 1.2: Top: Simplified molecular structure of single strand DNA sequence CTC (3'→ 5') paired with a complementary single strand DNA sequence of GAG (5'→ 3') to make a double stranded DNA oligomer. Bottom: Helical B-DNA structure proposed by Watson and Crick.[11]

1.3. The genetic code.

The role of nucleic acids in chemical biology is vital: the DNA contained within the cell nucleus (genotype) is paramount in deciding the characteristics that an individual organism exhibits (phenotype). This is because the DNA is a molecular blueprint for the proteins produced by every cell, and as proteins determine the reactions that a cell can carry out, determines chemistry *in vivo*. The DNA of the cell is therefore said to carry the *genetic code* of an organism. The DNA within the cell is packaged by histone and scaffold proteins; the final macromolecular structures are called *chromosomes*.

The conversion of DNA to protein[12] however is not straightforward, requiring intermediate molecules to relay chemical information in a process known as *transcription*. During this process, the DNA base pairs unzip within the nucleus and the enzyme *RNA polymerase* makes complementary RNA strands to the DNA sense strands using the RNA nucleotides. The single stranded molecule that is formed is known as *messenger RNA* (mRNA). The mRNA then leaves the nucleus and proceeds into the cytoplasm, where the protein synthesis process known as *translation* begins. During translation, the mRNA sequence is converted into protein. This takes place on structures known as *ribosomes*, which are made up of protein and ribosomal RNA. The ribosomes 'read' the mRNA genetic code as a triplet frame of bases – these are known as *codons*. Each triplet codon is then paired up with its complementary triplet *anticodon* contained within a molecule known as *transfer RNA* (tRNA). The pairing up is catalysed by the ribosome, which acts as a matrix for the reaction. The small tRNA molecules each bear a single

amino acid corresponding to a certain codon/anticodon combination; some combinations encode for 'start' or 'stop' functions so that the translation can be controlled (Table 1.1). In this way the 20 amino acids are coded for by the 4 DNA/RNA bases. Catalytic condensation of the amino acids into a polypeptide chain followed by hydrolysis of the growing chain's bonds to the tRNA molecule then follows. This completes the *in vivo* process of protein synthesis from a DNA sequence.

				2nd base							
				U		**C**		**A**		**G**	
1st base	**U**	3rd base	**U**	UUU	*F*	UCU	*S*	UAU	*Y*	UGU	*C*
			C	UUC	*F*	UCC	*S*	UAC	*Y*	UGC	*C*
			A	UUA	*L*	UCA	*S*	UAA	*STOP*	UGA	*STOP*
			G	UUG	*L*	UCG	*S*	UAG	*STOP*	UGG	*W*
	C		**U**	CUU	*L*	CCU	*P*	CAU	*H*	CGU	*R*
			C	CUC	*L*	CCC	*P*	CAC	*H*	CGC	*R*
			A	CUA	*L*	CCA	*P*	CAA	*Q*	CGA	*R*
			G	CUG	*L*	CCG	*P*	CAG	*Q*	CGG	*R*
	A		**U**	AUU	*I*	ACU	*T*	AAU	*N*	AGU	*S*
			C	AUC	*I*	ACC	*T*	AAC	*N*	AGC	*S*
			A	AUA	*I*	ACA	*T*	AAA	*K*	AGA	*R*
			G	AUG	*M (START)*	ACG	*T*	AAG	*K*	AGG	*R*
	G		**U**	GUU	*V*	GCU	*A*	GAU	*D*	GGU	*G*
			C	GUC	*V*	GCC	*A*	GAC	*D*	GGC	*G*
			A	GUA	*V*	GCA	*A*	GAA	*E*	GGA	*G*
			G	GUG	*V*	GCG	*A*	GAG	*E*	GGG	*G*

Table 1.1: mRNA triplet codon sequences and the corresponding amino acid produced post-translation.[1]

1.4. DNA: the blueprint for cell chemistry.

Many of the proteins contained within the genetic code are enzymes which orchestrate cell chemistry. Enzymes are biological catalysts, which in

the traditional definition means an entity that speeds up a chemical reaction without itself being used up. From the selectivity demonstrated by terpene cyclases[13] in plant secondary metabolism on relatively featureless hydrocarbon substrates to the superlatively rapid catalytic activity demonstrated by human carbonic anhydrase (over 10^9 turnovers per second and hence diffusion controlled),[14] enzymes perform perfect chemical reactions within the cell, honed by millions of years of evolution. The way in which proteins are folded, which in itself is a direct result of the amino acid sequence (and that is determined by DNA base sequence) determines which reactions can be catalysed. Sometimes co-factors, such as metals, are involved. The centres of chemical reactivity contained within a protein's three-dimensional structure are known as *active sites*. Within the active site, reactions that usually are recalcitrant become facile, and slow reactions become rapid, usually at around 37 °C for all warm-blooded organisms (this latter fact is usually cited as evidence of a common ancestor between species).

An important topical example of protein pathogenicity through mutation of DNA is that of Alzheimer's disease (AD). β – site APP cleaving enzyme (BACE) catalyses the hydrolysis of amide bonds *in vivo*. In its normal biological role BACE is thought to be responsible for control of myelination of axons. However, BACE is also inadvertently responsible for causing AD in humans.[15] The active site of BACE, when mutated, can also catalyse the cleavage of the membrane-bound amyloid precursor protein (APP) *in vivo*; subsequent catabolism of the C99 peptide from this reaction produces the Aβ42 polypeptide which is primarily responsible for generating the neural

plaques that cause AD. In the hereditary Swedish variant of AD it has been proven that mis-sense mutations (see later) to DNA are to blame for the final pathogenic activity of the protein.[16] There are now potent inhibitors of BACE being developed to prevent onset of the disease,[17] and identification of an individual's susceptibility *via* screening for the mutated DNA sequences will be crucial to the success of any forthcoming therapies.

The latter example quite clearly demonstrates that it is DNA that, in reality, controls the chemistry of the cell, as the protein sequences expressed by a cell are determined *solely* by the DNA contained within every cell nucleus. One thing is also clear, in that DNA provides a much better target for drug molecules etc. as only one copy of the genetic code exists within the cell nucleus of *every* cell – target this effectively and a disease may be successfully cured/treated; this is the basis of *gene therapy*.

1.5. Genetic Mutations.

Structural changes to DNA are known as *genetic mutations*. Naturally, these exist to give variation in the gene pool. As we have seen, some genetic mutations can be advantageous, while others are the basis of disease or lead to evolutionary throwback, with natural selection rapidly eliminating the weak gene.

There are a range of genetic mutations which can occur to disrupt the sequence of cellular DNA. Point mutations, often referred to as single nucleotide polymorphisms (SNPs), are mutations in which a single nucleotide

in a DNA sequence is exchanged for another nucleotide, containing a different base.[18] Purine – purine (A, G) or pyrimidine – pyrimidine (C, T) exchange type mutations are most common simply due to the hydrogen bonding differences of the base pairs in DNA; although purine – pyrimidine exchange is possible, it is much less common as a result. It can be seen from Table 1 that the base exchanged may or may not influence the amino acid produced by an mRNA codon. The third (wobble) position in the codon is far more tolerant to change than for mutation at positions one or two in the triplet sequence. Any mutation that changes an amino acid in a protein sequence is known as *mis-sense mutation*. Mutations that code for the same amino acid are called *silent mutations*, as no effect can be discerned; proteins are unchanged by the mutation. The final type is the *non-sense mutation*, where the mutation codes for the UAA, UAG or UGA mRNA stop codons which then cut the protein sequence short at that point.

Single nucleotide polymorphisms are the basis of many medical conditions. The protein *tumour suppressor P53* is integral to regulating the senescence and apoptosis (cell death) of any potential tumour cells. Breakdown of the P53 repair pathway is associated with uncontrolled cell cycles and hence almost every type of cancer in humans. The TP53 gene which encodes for protein P53 is susceptible to SNPs, with over 200 sites for potential SNPs identified. Two of these SNPs produce the R72P, and P47S mutant P53 proteins which are associated with the onset of tumour growth (tumourigenesis).[19] Similarly, an SNP of the mIR-146 gene was identified recently in a number of ovarian cancer patients.[20]

The second category of DNA mutations arise from *frameshift mutations*. These are caused by the insertion or deletion of a DNA base or large sections of DNA bases within the sequence and hence, because the process of transcription imposes a three-base reading frame on the DNA sequence, the whole gene sequence encodes for completely new amino acids and therefore mutated proteins of unknown or no function. In some cases, this can be detrimental to normal cellular function: returning to the example of the p53 tumour suppressor protein, a 16 base pair insertion into intron 3 (where an intron is a section of DNA that encodes for protein) of the TP53 gene is associated with tumourigenesis.[19]

It is crucial, therefore, that new methods are to be developed for screening of mutated genotype to determine disease phenotype.[21] The fruition of the human genome project[22] has allowed the mapping of the DNA-containing chromosomes enclosed within our cells, and this has subsequently enabled researchers to target the genes responsible for diseases.

1.6. Screening for polymorphisms.

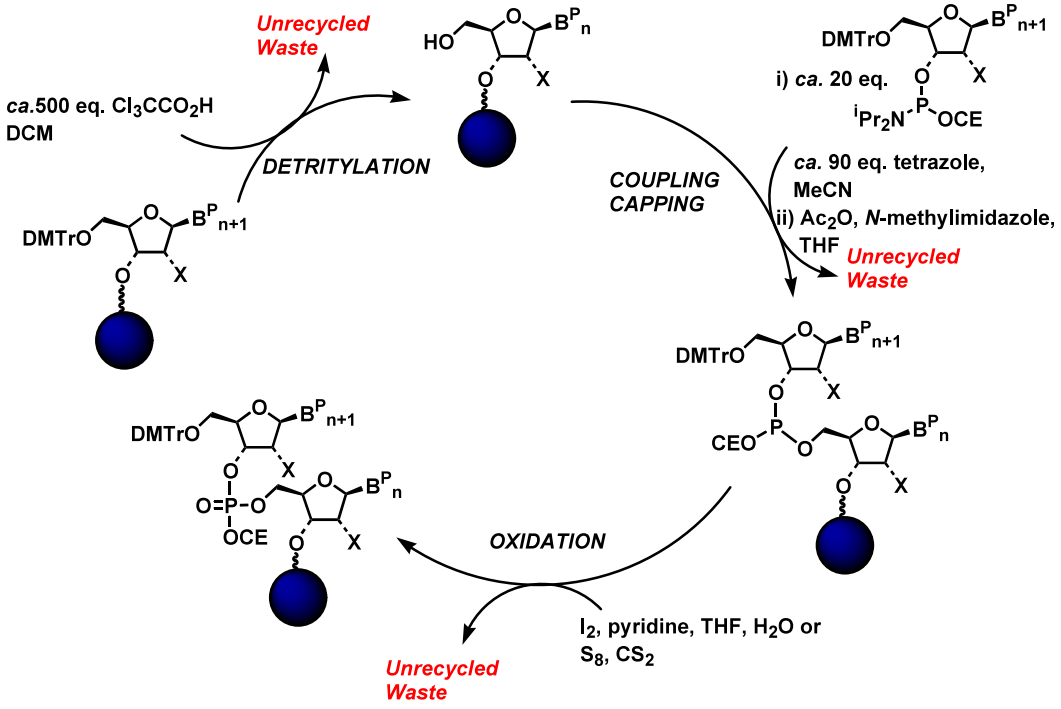
SNPs are a good objective for genotype-phenotype studies due to the relative simplicity of the mutation, as detection of a polymorphism in one base pair is much easier on a molecular scale compared to detection of a full frameshift mutation, which may have large areas of mutated or deleted DNA. As we have also seen, they are responsible for a range of diseases relevant to morbidity in the modern age, so they are attractive targets for research: the

challenge remains to find the best ways to analyse for these genetic mutations. We can, in theory, sequence an entire individual's DNA using traditional gene-sequencing techniques such as Sanger sequencing, which is based on DNA chain termination with fluorescent dideoxynucleotides that reveal the identity of bases in a sequence by four colour outputs corresponding to each base.[23] However, despite the widespread use of this type of sequencing for biochemical research purposes worldwide, this simply would take too much time to perform on an entire population. Any potential analyses must be rapid, selective and sensitive enough to be used routinely, without incurring a large cost.

Designing fluorescent tags for incorporation into biologically relevant molecules has been a popular research topic for many years. Due to the quantitative nature of fluorescence, i.e. measurable energy input and output, it has been used to great effect as a structural probe in biomolecules. Both DNA and proteins/polypeptides, have been labelled with fluorescent labels. Usually in such studies, a photoactive molecule is appended to a biomolecule either by a site-specific chemical reaction that forms a covalent bond between the biomolecule and the label itself (usually referred to as *bioconjugate chemistry*), or the biomolecule itself can be synthesised artificially to contain the label / sensor using a protection-deprotection type approach, either in solution or on solid-phase supports.

The Beaucage-Caruthers (BC) approach to the synthesis of oligonucleotides[24] is one of these stepwise approaches which has been used

to synthesise artificial oligonucleotides *in vitro*. The approach uses nucleotide diols in the synthetic cycle, though in theory any diol can be incorporated into the growing DNA chain. The nucleotides can be reacted with certain directionality as one of the alcohol oxygens in the diol group is usually mono *O*-protected with a dimethoxytrityl (DMTr) protecting group, while the other hydroxyl is used in the form of an activated phosphoramidite, which can undergo nucleophilic substitution at phosphorus. The synthetic cycle is performed on solid supports and is summarised in Scheme 1.1.



Scheme 1.1: A typical cycle in a Beaucage-Caruthers (BC) type *in vitro* synthesis of oligonucleotides (X= H for DNA, X = OH for RNA).[25]

Once incorporated, the label, which must retain its fluorescent character, is then irradiated at the specific wavelength so that the chromophore absorbs a large cross-section of photons. Once the excited

state is populated, a number of processes can happen to release the energy absorbed. One such process, fluorescence, leads to the emission of a photon at a higher wavelength (Stokes shift). Strictly speaking, the latter process must arise from the transition from an excited singlet state to the ground singlet state.[26] The molecular fluorescence described is usually highly dependent on factors such as solvent polarity and molecular orientation; as such, detection of minute changes in the local environment around a fluorescent tag is possible when fluorescent groups are incorporated into a biomolecule. In some cases, computational methods can be used either to predict or interpret results. Modern fluorimeters allow the selection of wavelengths of light appropriate for the excitation of labels by the use of highly accurate monochromators. Benchtop machines are now available due to the miniaturisation of computer parts. The increase in the use of affordable laser instrumentation allows use of intense pulsed or continuous-wave monochromatic light in everyday experiments. Finally, as fluorescence is essentially a dark-background technique, it offers superior sensitivity to many other spectroscopic techniques, with detection of signal possible well into the picomolar concentration range.

1.7. Fluorescent sensors for single nucleotide polymorphisms.

There have been a myriad of reports of sensors which use fluorescence to detect single nucleotide polymorphisms in DNA sequences. A review of the scientific literature is appropriate to look at the evolution of such probes in terms of their selectivity and sensitivity, and to highlight the various approaches which can be used to modify DNA to incorporate such useful reporter groups. Although a number of approaches[18] have been developed toward this end, including fluorescent sequence-specific DNA primers,[27] post-synthetic site-specific labelling of DNA,[28] molecular beacons,[29] dual probe assays[30] and hybridisation probes.[27] This review is limited to consideration of hybridisation probes where modification or substitution of nucleotides with fluorescent groups on single probe sequences is used to detect single base pair mismatches with double stranded DNA due to the relevance to the work performed in our laboratories described later.

Oligonucleotides are short lengths of single-stranded DNA, incorporating usually less than 30 bases. One of the earliest examples of the incorporation of a fluorescent hydrocarbon into an oligonucleotide sequence was reported by Yamana and co-workers,[31] who, by using the BC phosphoramidite method were able to label short oligonucleotides with methylpyrene groups at the 3' and 5' termini which remained fluorescent after modification. Upon binding of the labelled oligonucleotides with complementary sequences, it was demonstrated that the pyrene fluorescence was perturbed.

These early studies demonstrated the usefulness of the aromatic hydrocarbon groups as fluorescent reporters due to their sensitivity to local environment; pyrene emission is known to be enhanced in polar environments but quenched in non-polar environments.[32] It also has the ability to form an *excimer* (an excited state dimer) which drastically change the pyrene emission spectrum, giving output of light at much higher wavelengths (lower energies). A study was reported by Balakin *et al.* on the incorporation of pyrene groups into oligonucleotides[33] Dual pyrene chromophores (Figure 1.3) were incorporated in the centre of an oligonucleotide sequence, 5'-CCCCTAG(Per)$_2$CCTGTA-3'. A second non-labelled complementary sequence was also synthesised, 5'-TACAGGTCCTAGGGG-3'. Drastic changes to the pyrene fluorescence were observed upon hybridisation of the complementary sequences, which were ascribed to excimer formation between the neighbouring fluorescent groups. The structural framework imposed by hybridisation of the strands caused the formation of the excimer and the output of light as the signal. Although such hybridisation probes had not yet addressed base-pair mismatches as such, it was clear at this point that slight modifications to the systems could lead to very useful applications.

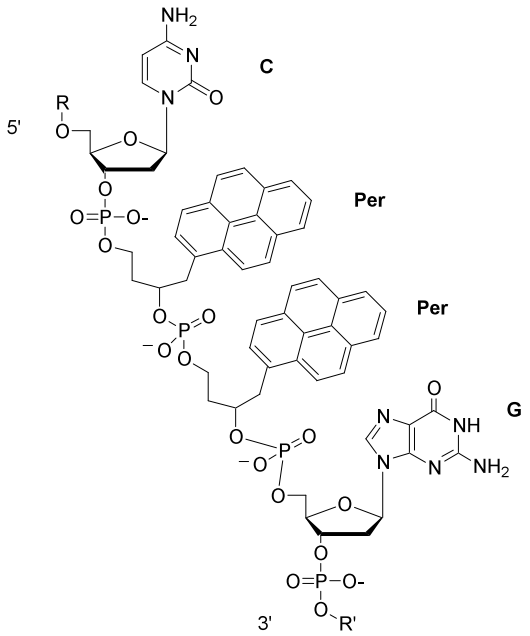


Figure 1.3: Pyrene groups incorporated into an oligonucleotide strand.[33]

Modification of bases with pyrene chromophores has since been a popular synthetic approach to base-discriminating fluorescent (BDF) nucleotides. Such artificial nucleotides display changes in fluorescence output when paired with non-matching bases upon hybridisation to duplex. Saito and co-workers performed a chemical modification of 7-de-aza-adenosine under Sonagashira conditions to incorporate a pyrene moiety directly attached to the adenosine nucleobase analogue (Figure 1.4).[34] Hydrogenation of the alkynyl bond, then protection at the 5'-OH position on the deoxyribose ring with dimethoxytrityl, followed by the formation of the activated phosphoramidite at the 3'-OH allowed the fluorescent base analogue ^{Py}**A** to be used in the DNA synthesis cycle, incorporated into the oligonucleotide sequence 5'-CGCAAT^{Py}**A**TAACGC-3'. Four complementary sequences were also synthesised during the study with the general sequence 5'-GCGTTAXATTGCG-3', where X = T, C, A or G.

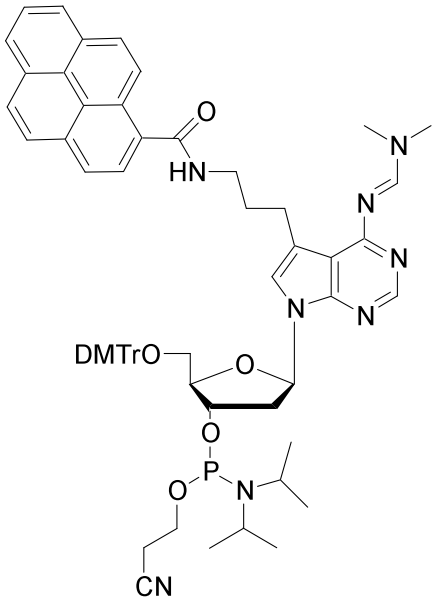


Figure 1.4: Synthetic nucleotide ^{Py}**A** for distinction of thymine in protected form prior to DNA synthesis.[34]

It was found that, upon hybridisation of the ^{Py}**A** strand to the latter four sequences, striking changes occurred in the fluorescence of the pyrene. When X = A, C or G it was found that the pyrene emission at 390 nm was not perturbed by any reasonable amount compared to the 5'-CGCAAT^{Py}**A**TAACGC-3' strand. However, upon hybridisation with the strand with correct Watson-Crick base pair (T) opposite to the ^{Py}**A** residue i.e. a fully complementary sequence, that the emission was effectively switched off, being 100-fold less efficient than the fluorescence of the single stranded 5'-CGCAAT^{Py}**A**TAACGC-3' sequence alone. The mechanism of the quenching was believed to arise from intercalation (the incorporation of planar, usually aromatic molecules, into DNA by their slotting between the hydrogen bonded bases perpendicular to the 5' to 3' sugar phosphate *z*-axis) of the pyrene unit into the double stranded DNA. Under mismatch conditions i.e. when X = A, C

and G the pyrene did not intercalate, probably due to steric blocking of the mismatched base pair. As the signal is switched on under the condition of base mismatch the potential of such BDF nucleotides as sensors for SNPs is clear; less clear from first glance were the driving forces behind the remarkable behaviour. As the studies were carried out in aqueous systems, the pyrene, if not intercalated is exposed to the polar bulk solvent, and remains fluorescent. However, with a matched base pair, order is restored to the helix and a site for intercalation is produced. When slotted between the bases, the pyrene is in effectively a hydrophobic and non-polar environment, and as such the fluorescence was effectively turned off. Computational modelling also supported the intercalation argument, with the intercalated pyrene in theory conferring additional stability to the duplex. This was backed up by melting temperature experiments on the duplexes and by the UV absorption spectrum. The intercalated pyrene unit showed a red-shift in its absorbance spectra due to π stacking interactions with the bases above and below. The melting temperature of a DNA duplex (T_m), i.e. the temperature at which the DNA duplex is half-dissociated, is indicative of the stability of the double stranded oligonucleotide. In this case, the 5'-CGCAAT^{Py}**A**TAACGC-3' / 5'-GCGTTATATTGCG-3' duplex was more stable than the wild type 5'-CGCAATATAACGC-3' / 5'-GCGTTATATTGCG-3' duplex, supporting the intercalation argument. Further work by Saito and co-workers used similar strategies to that described presently. A whole range of base-discriminating fluorescent nucleosides were synthesised[35] to complement the original ^{Py}**A** nucleoside, again using pyrene as the organic reporter group in various guises (Figure 1.5). By changing the structure of the fluorescent nucleoside,

varying base-discriminating properties were conferred. ^{Py}**U**, based on a pyrene attached to a uracil nucleobase was found to give a sizeable increase in fluorescence quantum yield for an adenine (A) at the adjacent position in the antisense strand of an oligonucleotide duplex.[32c] Likewise, the cytosine analogue ^{Py}**C** gave a substantial and highly selective increase in fluorescence when the adjacent base was guanine (G). The fluorescence behaviour of these sensors was opposite to the behaviour of ^{Py}**A**. ^{Py}**U** and ^{Py}**C** acted as 'turn-on' sensors for a base-pair *match*, whereas ^{Py}**A** acted as a 'turn-on' sensor for a base-pair *mismatch*. The difference in behaviour arose from the rigid propargyl linker in ^{Py}**U** and ^{Py}**C**. The molecular conformations meant that it was impossible for the pyrene group to intercalate under normal base-pairing conditions, being effectively extruded from the interior of the DNA duplex by the rigid propargyl group, and so fluorescence was turned on in the polar environment of the bulk solvent; the authors deliberately chose the position of the labelling for this reason alone. The base-discriminating nucleosides ^{8Py}A and ^{MePyd}A displayed the same behaviour with thymine (T) and cytosine (C), thus giving the authors the ability to detect SNPs for any mismatch type. [35]

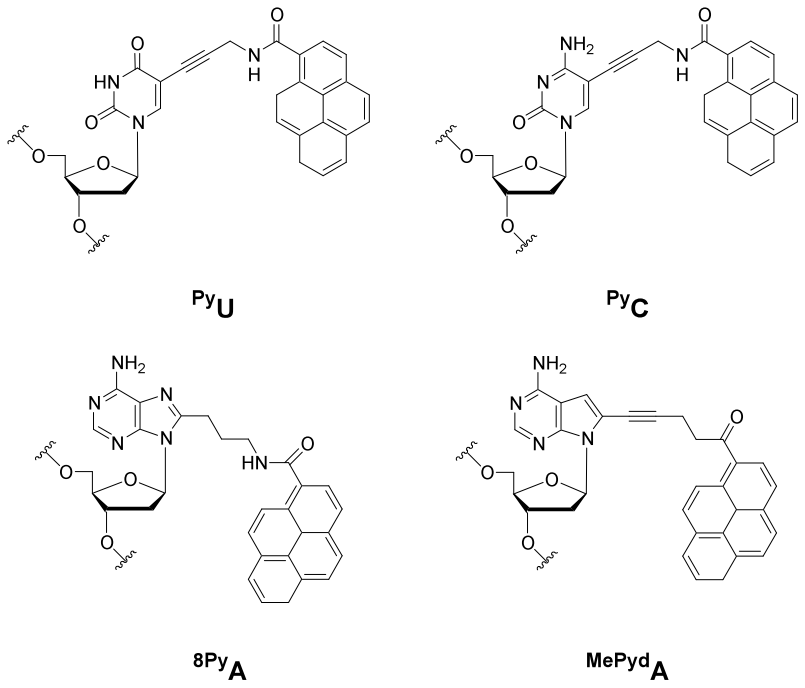
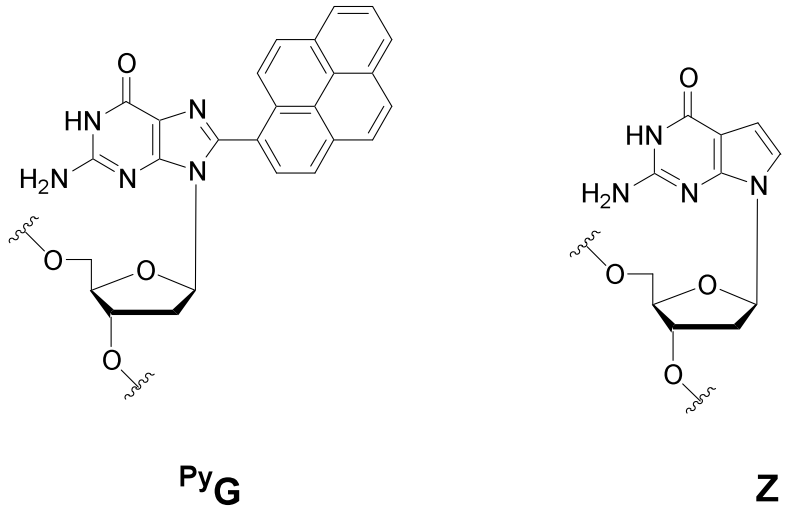


Figure 1.5: Structures of base-discriminating fluorescent nucleosides.[35]

The BDF nucleoside probes **^{Py}C** and **^{Py}U** were later described by Okamoto *et al*. for use in a simple fluoroassay for detection of SNPs.[36] The human gene sequence BRCA1, an oncogene associated with breast cancer, is known to contain a G to A point mutation SNP. The BRCA1 gene was firstly amplified by polymerase chain reaction (PCR – a method to amplify pieces of DNA see Saiki *et al.*[37] for further details) then the amplicon mixed with oligonucleotides complementary to the oncogene SNP sequence 5'-GAGBCGATTTCAT-3' where B = **^{Py}C** or **^{Py}U**. The results using the probe showed the same fidelity as traditional sequencing techniques, effectively identifying the point mutation in much less time and effort, and presumably at a much reduced cost and with better sensitivity.

Wagenknecht and co-workers have also used pyrene as a fluorescent sensor for SNP detection.[38] The pyrene was attached with no spacer linking it to the guanine base, giving fluorobase ^{Py}**G**; also synthesised was a 7-deazaguanine base, **Z (**Figure 1.6). The fluorescence of the ^{Py}**G** base was quenched in the presence of a flanking SNP compared to the matched sequence. It was found that the quenching could be made even more dramatic if **Z** was incorporated in an [5' (*n*+3) 3'] position in the oligonucleotide.



^{Py}**G** **Z**

Figure 1.6: Synthetic nucleosides employed by Wagenknecht and co-workers.[38]

In a similar fashion, Saito and co-workers have also used perylene (^{Per}**U**) and anthracene (^{2Ant}**U** , ^{9Ant}**U**) based tags as SNP sensors (Figure 1.7).[39] In both cases, as before, the fluorescent group was directly attached to the uracil base under Sonagashira conditions. The perylene version of the tag was found to be useful for detection of mismatches in short oligonucleotides, displaying an increase in fluorescence at 496 nm for a ^{Per}**U**-C mismatch compared to the matched ^{Per}**U**-A base pair. It was proposed that

behaviour was due to the expulsion of the intercalating perylene into the bulk polar solvent when a base pair match occurs. When mismatch occurs as the perylene gains limited flexibility to intercalate into the non-polar double-helix and fluorescence is switched off.

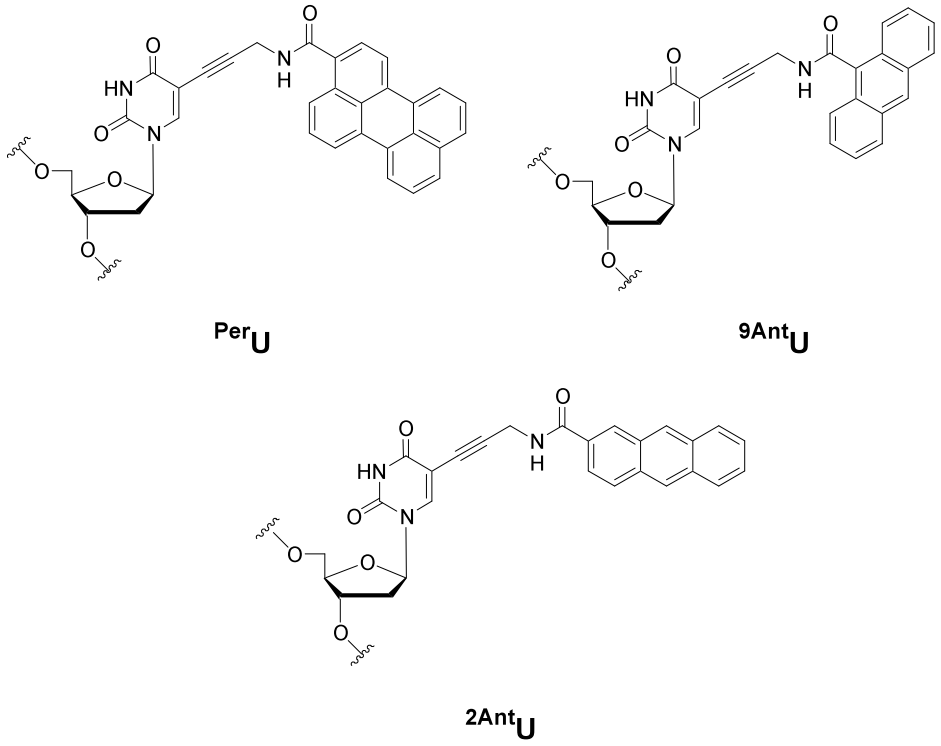


Figure 1.7: Perylene and anthracene based SNP sensors.[39]

Likewise, the anthracene tags were found to be useful as SNP sensors. Sharing the same ethylene linker as the perylene and pyrene versions of the tag, the anthracene fluorescence was observed to increase under mismatch conditions. ^{2Ant}**U** was able to effectively discriminate between A and C in the anti-sense strand, giving a large increase in fluorescence for A, and quenching for C. Little difference in fluorescence intensities were found for oligonucleotide duplexes containing G and T compared to the ^{2Ant}**U** tagged single strand DNA. ^{2Ant}**U** was also found to be

much more suited than ^{9Ant}**U** to the role due to its greater relative fluorescence increase at longer wavelengths, even more so than the pyrene derivatives discussed previously.

Brown and co-workers reported the use of the thymine analogues ^{Ne}**T**, ^{Ae}**T**, and ^{Aee}**T** which utilised anthracene and naphthalene fluorescent groups as SNP sensors (Figure 1.8).[40] Each thymine analogue was found to behave the same, undergoing fluorescence enhancement for a matched target sequences and quenching for T-G mismatch. It was also found that if thymine tagged with fluorescein was also incorporated into the same oligonucleotide with the hydrocarbon probes then the fluorescein fluorescence was also enhanced upon duplex formation in all cases. This led the authors to speculate on the design of probes with dual function.

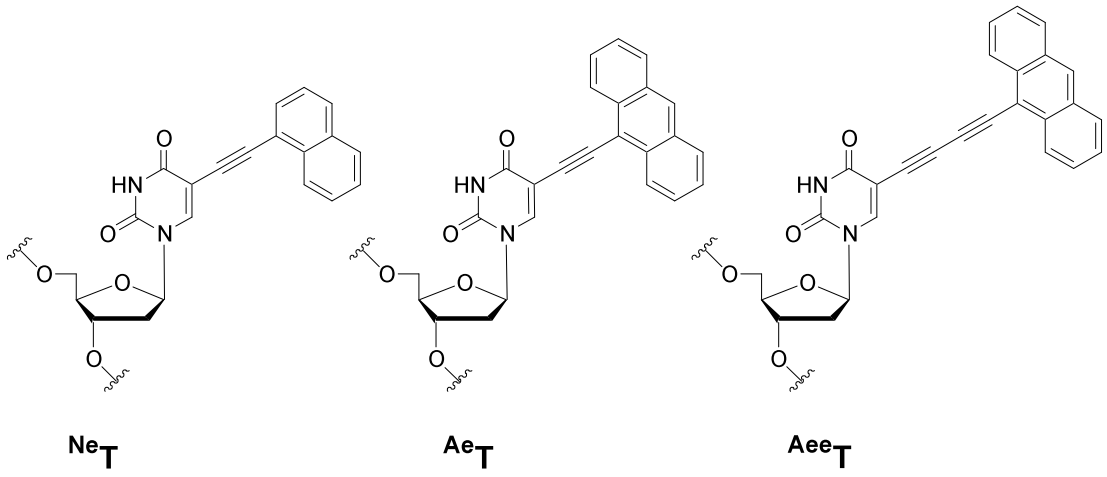


Figure 1.8: Naphthlanene and anthracene thymine analogues for SNP detection.[40]

Tucker and co-workers have also found success in the development of novel SNP sensors based on fluorescent anthracene dyes, though the approach varies to that of Saito in that the tag is not actually appended to a

base. Rather, a non-nucleosidic diol-containing anthracene is first synthesised and then modified to be incorporated into the DNA *via* the BC phosphoramidite method (Figure 1.9).[41]

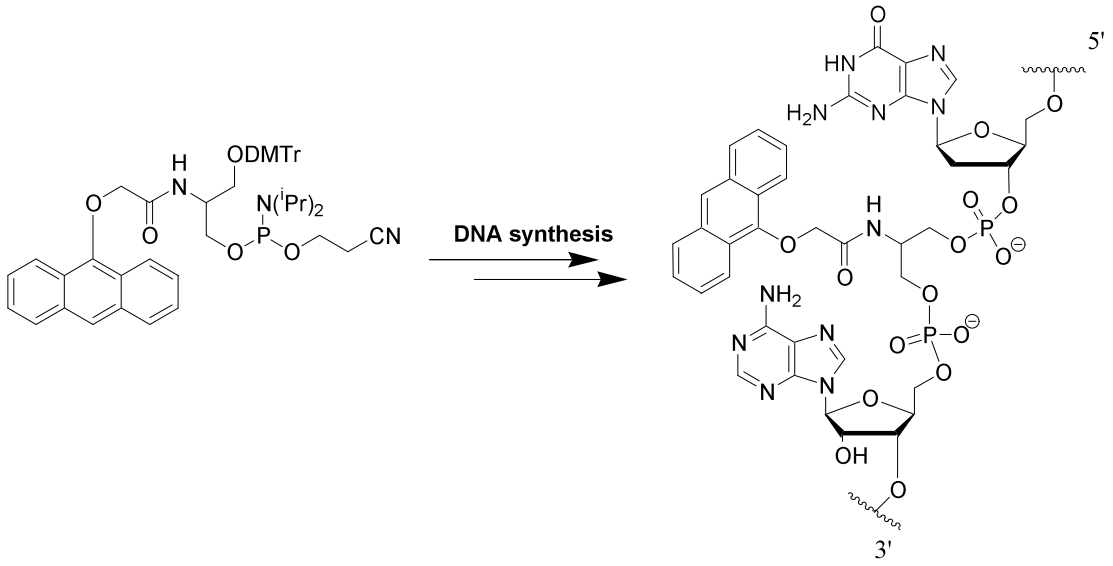
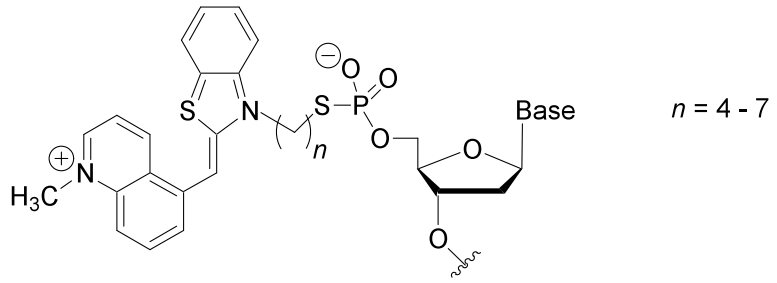


Figure 1.9: Anthracene tag used by Moran *et al.* for SNP detection.[41]

The fluorescence intensity was quenched nearly 100% upon duplex formation with a fully complementary strand (target C-G base pair flanking the anthracene position) of DNA compared to the single stranded tagged oligonucleotide. Upon target mismatch (target C-A base pair flanking the anthracene position), the fluorescence was increased by almost 200%, allowing stark definition between matched and mismatched flanking base pairs. The fluorescence quenching observed for the matched target sequences was attributed to intercalation of the anthracene into the non-polar interior of the duplex. This was backed up by the higher melting temperatures for the anthracene tagged matched duplex compared to the analogous non-tagged matched duplex, the intercalated anthracene serving to stabilise the

helix in the former case. The increase in fluorescence upon flanking C-A mismatch could not easily be explained, though it was postulated that expulsion of the anthracene from the interior of the duplex could be a very likely possibility, due to the relative instability of the mismatch sequences by melting temperature. The authors state that a range of linker and spacer lengths will be used to investigate the remarkable fluorescence behaviour of the probe.

Asseline *et al*. recently unveiled a sensor for the detection of SNPs at the 5'-termini in oligonucleotides, ***n*-TO** (where *n* = 4 – 7), incorporating the organic dye thiazole orange attached to the sugar-phosphate backbone by a thiophosphate ester (Figure 1.10). [42] It was found that the sensor was not only able to sense SNPs by an increase in fluorescence under mismatch conditions, but it was also able to discriminate between G-A, A-G, T-T, C-C, C-T and T-C mismatches; the authors state that these SNP mutations are reported to be among the hardest for cellular repair mechanisms to distinguish from matched base pairs, citing the work by Kramer *et al.*[43] The mechanism of the sensing was again believed to be due to differences in local environment polarity in and around the oligonucleotides, with conformational changes induced by mismatch leading to perturbed fluorescence.



***n*-TO**

Figure 1.10: Terminal SNP sensor incorporating thiazole orange.[42]

Tor and co-workers synthesised an SNP sensor based on thieno[3,4-*d*]pyrimidine, (Figure 1.11), which was incorporated into RNA oligonucleotides by enzymatic catalysis using T7 RNA polymerase.[44] The incorporation of labels into oligonucleotides by enzymatic methods offered several advantages, for instance not having to engage in protecting group chemistry, extremely high efficiency, production of large amounts of oligonucleotides and being able to change the label easily. Once incorporated into the RNA sequences, the fluorescent properties were tested with duplex formation to place the four conventional DNA bases opposite to the fluorescent synthetic base. The emission where the thieno[3,4-*d*]pyrimidine probe was opposite to C gave the largest emission increase, over 7-fold compared to the to the perfect duplex where the thieno[3,4-*d*]pyrimidine is paired with A. Mismatches with T and G opposite to the thieno[3,4-*d*]pyrimidine base, although enhanced, were not as strong as the enhancement observed for the thieno[3,4-*d*]pyrimidine – A mismatch. It was concluded that the sensor was specific for the latter SNP.

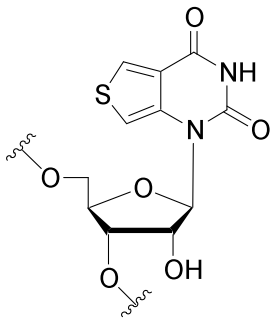
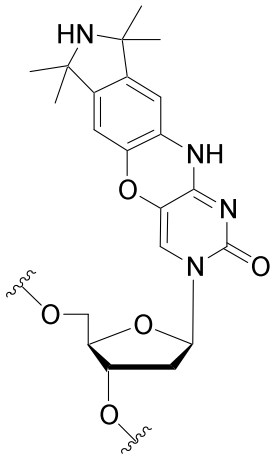


Figure 1.11: Thiopyrimidine base-discriminating fluorescent ribonucleoside.

A slightly different approach has been used by Sigurdsson and co-workers. A cytosine analogue was synthesised by reduction of a spin label to give the highly fluorescent nucleoside **Ç^{f}**, with quantum efficiency in excess of 30% (Figure 1.12). It was found upon incorporation into oligonucleotide sequences that the modified base could successfully identify each of the four bases in the adjacent position on the opposite strand as the fluorescence readout was so different for each base pair upon duplex formation. The **Ç^{f}**-A mismatch gave the brightest fluorescence, with the fully matched **Ç^{f}**-T sequence the least fluorescent. Of the other two remaining combinations, **Ç^{f}**-G and **Ç^{f}**-C could be distinguished due to the striking green-blue colour of the **Ç^{f}**-G fluoresence. It was also demonstrated that the fluorescence behaviour was independent of the flanking base pair in the sequence, which is important for developing robust SNP sensors for universal use.



Ç^{f}

Figure 1.12: Highly fluorescent SNP sensor developed by Cekan *et al.*[45]

1.8. Conclusion.

It is clear from this brief introduction and review of some highlights from the literature that this is an exciting area of research with the aim of medical diagnosis as the major driving force in the area. The discovery of new SNP sensors for rapid and cheap analysis of point mutations in genes is extremely important for the future of this field of medicine. In the next chapter, the synthesis of a new nucleotide analogue **5a** for future incorporation into DNA oligonucleotides will be described, as well as the incorporation of another synthetic nucleoside analogue **5b** into DNA oligonucleotides and the subsequent studies into detection of SNPs in a range of target oligonucleotide sequences.

2.0. Results and Discussion.

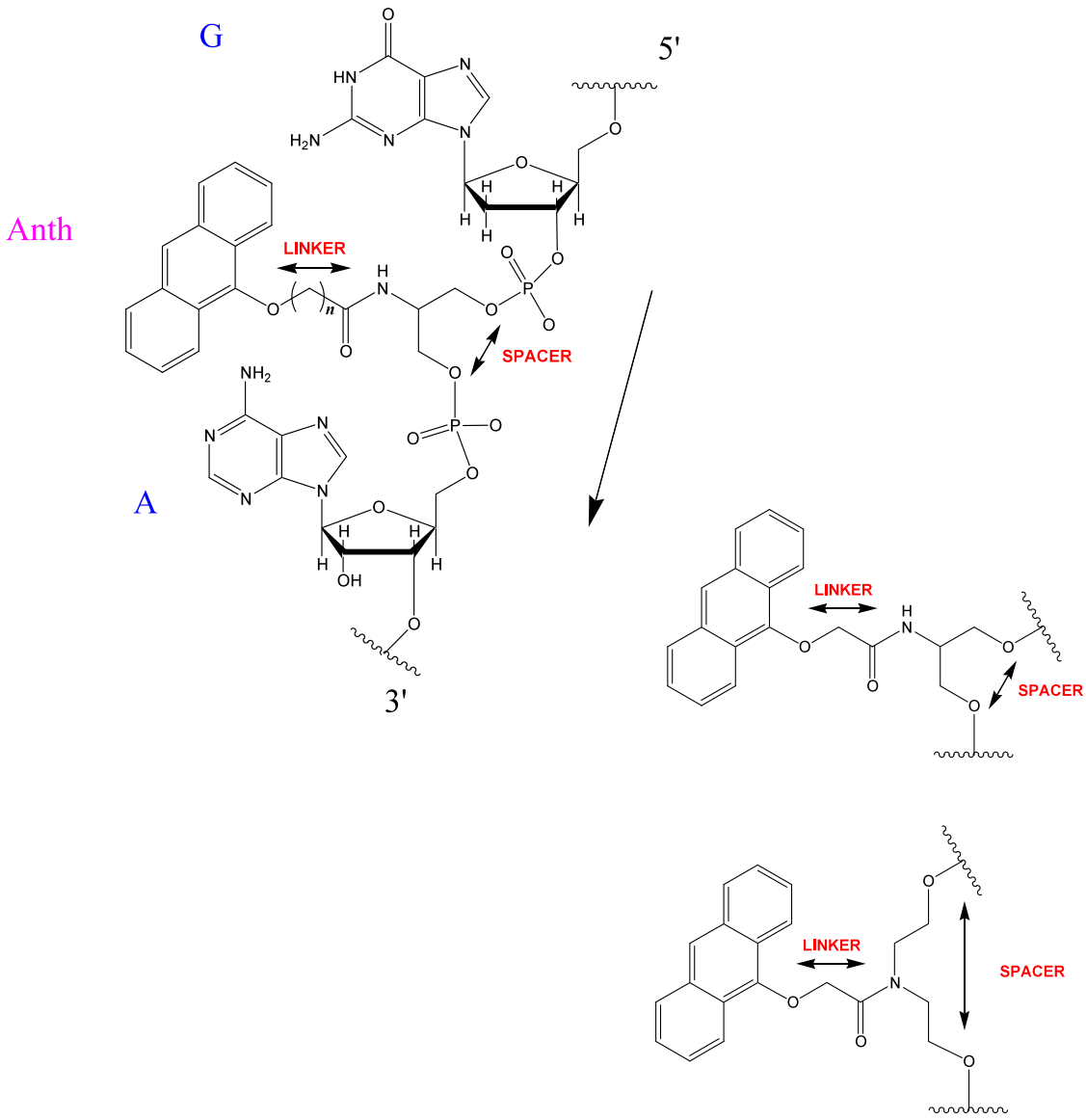


Figure 2.1: Definitions of 'spacer' and 'linker' in anthracene SNP probes.

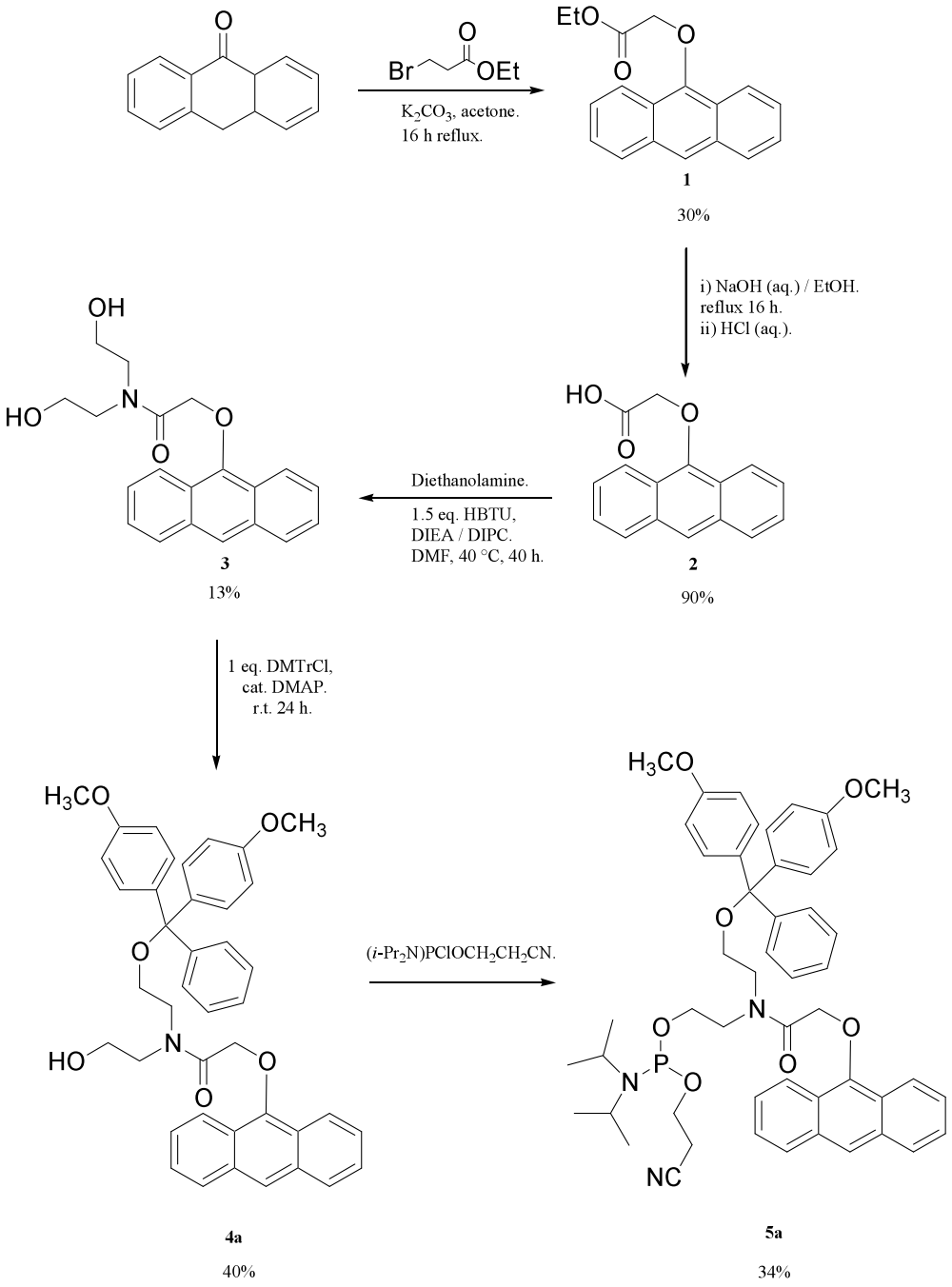
It was decided that a new anthracene-based SNP sensor for use in small DNA oligonucleotides should be synthesised. In a previous communication on sensing SNPs in short oligonucleotide sequences,[41] the anthracene group was appended directly to the sugar phosphate backbone using the Beaucage-Caruthers synthetic cycle. Such sensors have two

variables which can be modified. These are, firstly, the length of the 'linker', which corresponds to the number of carbon centres (*n*) between the anthracene and the amide bond, and, secondly, the nature of the spacer that is attached *via* a phosphodiester linkage upon incorporation into a DNA oligonucleotide (Figure 2.1). Clearly as discussed in the introduction section, the change in the fluorescence in probes of this type is highly dependent on the geometry that the probe finds itself in under match and mismatch conditions. Changing the length of the linker and the nature of the spacer in a systematic manner would be key to understanding the remarkable behaviour that was exhibited by the originally communicated anthracene probe.

2.1. Synthesis of anthracene tag 5a for mismatch detection in DNA duplexes.

The synthetic scheme which was used to produce the *n* = 1 phosphoramidite **5a** is shown in Scheme 2.1. Briefly, there are five synthetic steps in the scheme; the first stage uses an ether synthesis to install the ester group, the second stage is a simple saponification to the corresponding carboxylic acid, the third stage employs traditional solution-phase peptide chemistry with coupling reagents to produce the diol *via* amide linkage, the fourth stage is the mono *O*-protection of the diol with the dimethoxytrityl (DMTr) group. and the fifth and final stage in the chemical synthesis transforms the remaining free hydroxyl group into an activated phosphoramidite. The final molecule can then be entered into a round of

automated DNA synthesis to install it as a synthetic base in an oligonucleotide sequence of choice.



Scheme 2.1: Synthesis of SNP probe precursor **5a**.

The reaction between anthrone and ethyl bromoacetate in the presence of potassium carbonate proceeded as expected, giving the crude product which was purified using flash column chromatography to give the product ester **1** in respectable yield. ^{1}H NMR spectroscopy (Figure 2.2) was used to confirm the successful synthesis of the purified product with the appearance of the ethyl ester quartet and triplet peaks at 4.29 and 1.29 ppm, and the signal for the methylene protons appearing as a 2H singlet at 4.72 ppm. These peaks in the aliphatic region integrated correctly against the aromatic region which confirmed that the appendage was in the correct ratio. Electrospray mass spectrometry gave the expected RMM for **1** as the monosodium adduct; high resolution mass spectrometry was also used to determine the exact mass of the compound which, when compared to a theoretical calculation of exact mass for **1** gave an almost identical mass, within 2 ppm.

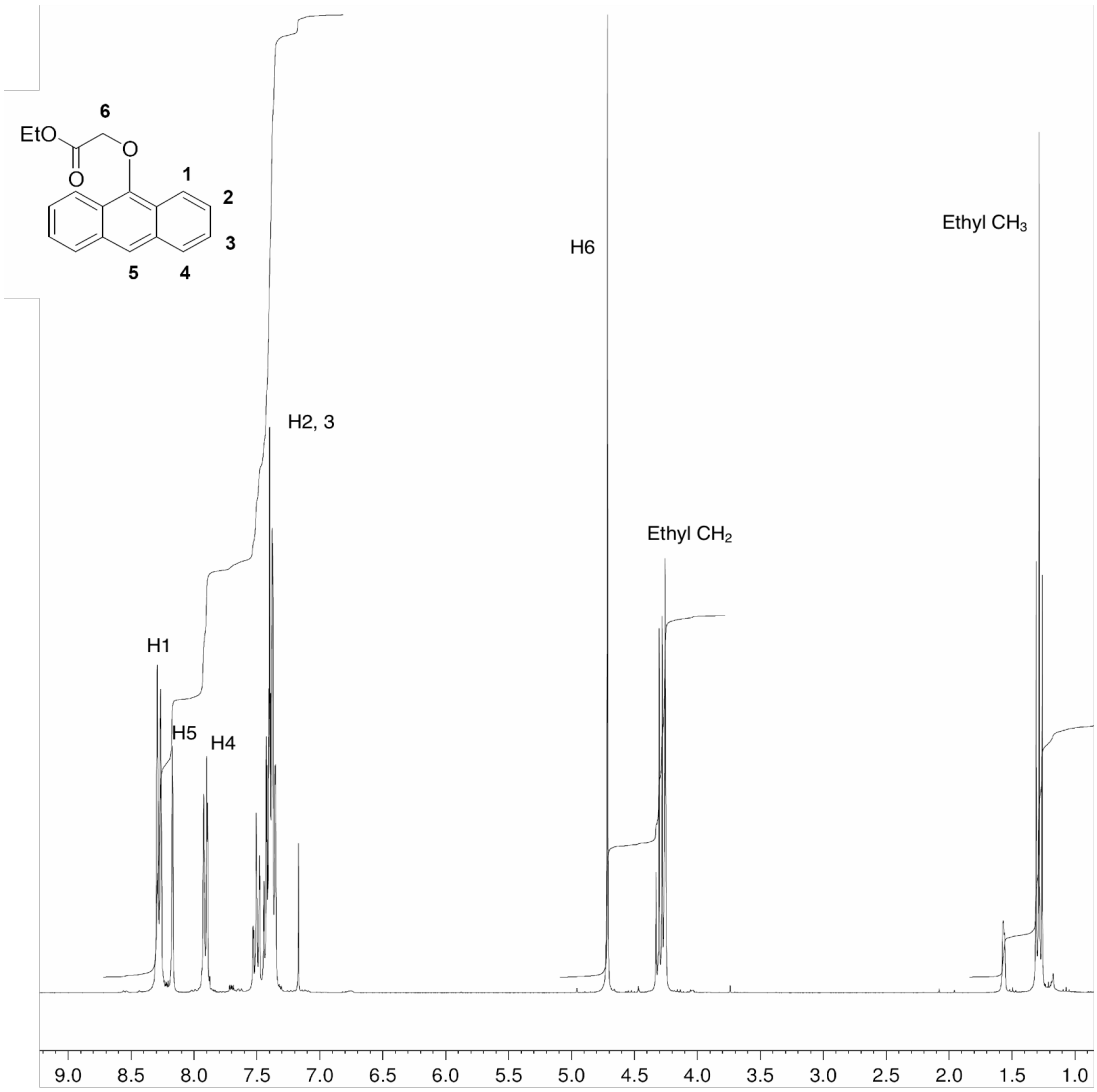


Figure 2.2: ^{1}H NMR spectrum of **1** (300 MHz, $CDCl_3$).

Ester **1** was then converted to the carboxylic acid **2** by saponification. A mixture of ethanol-water was used to aid the solubility of the starting material, but it was still a slurry to begin with. The reaction was monitored by TLC using the solvent system employed for the synthesis of the ester, with the disappearance of the ester peak used to gauge the end of reaction. The reaction mixture became a complete solution by the end of the reaction due to the formation of the carboxylate sodium salt of **2**, the latter being soluble in polar solvents. The crude product was isolated by precipitation after

acidification to pH 1 with concentrated hydrochloric acid to ensure complete protonation of the carboxylate (pK_a ~ 4.5). ^{1}H NMR spectroscopy was useful in the identification of the product due to the simple nature of the transformation; the only changes in the spectrum of **2** compared to **1** were the disappearance of the quartet and triplet ester CH_2 and CH_3 peaks, previously seen at 4.29 ppm and 1.29 ppm. The signal for the methylene protons remained intact and in the correct ratio to the aromatic peaks by integration. Mass spectrometry verified the correct mass of the molecule. Due to the unduly clean ^{1}H NMR trace it was decided that further purification was not required. Consequently, the reaction proceeded in excellent overall yield (90%).

The chemistry to produce diol **3**, which involved the condensation of diethanolamine with **2** proved more problematic. The success of the reaction seemed to hinge on finding a suitable peptide coupling agent for the reaction; a number of failed reactions in the laboratory occurred where only starting material was recovered from the reaction using the coupling agents benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and hydroxybenzatriazole (HOBt). Use of varying equivalents of these coupling reagents also did not seem to help matters. Eventually our attention was turned to using the coupling reagent *O*-Benzotriazole-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) at a molar ratio of 1.5:1 coupling agent to **2**. We were able to isolate a pure product in poor yield (13%) after column, though with acceptable purity. Attempts were made to improve the yield of the reaction, using longer reaction times but this did not seem to have any effect. Mass spectrometry was used to confirm that the

condensation occurred successfully, with a peak for the monosodium adduct at *m/z* 362. The high resolution mass of this peak was also an excellent match to the calculated exact mass, within 1.5 ppm. We also used elemental analysis to confirm the ratio of C, H and N in the product; the results showed that we had isolated the product as a dihydrate, which made sense in terms of the hydrogen bonding capabilities of the diol pendant arms, one water bound to each. The ^{1}H NMR spectrum (Fig 2.3) of compound **3** was the most interesting result of all. Evidence of product formation came from the appearance of the four triplet peaks in the range 3.84 – 3.29 ppm. These peaks correspond to the four CH_2 centres on the diethanolamide arms. Each triplet integrated to 2H, in the correct ratio to both the aromatic peaks originating from anthracene group and the aliphatic singlet corresponding to the resonance of the methylene linker CH_2. Free rotation around the amide bond therefore has to be slower than the NMR timescale at 298 K, which explains why there are four triplets as opposed to two; the protons on each of the diethanolamide arms can be discerned from each other spectroscopically. This effect is particularly common for amides, as rotation around the N-C bond is restricted due to partial π character. ^{13}C NMR gave the expected number of carbon environments required at the expected chemical shifts.

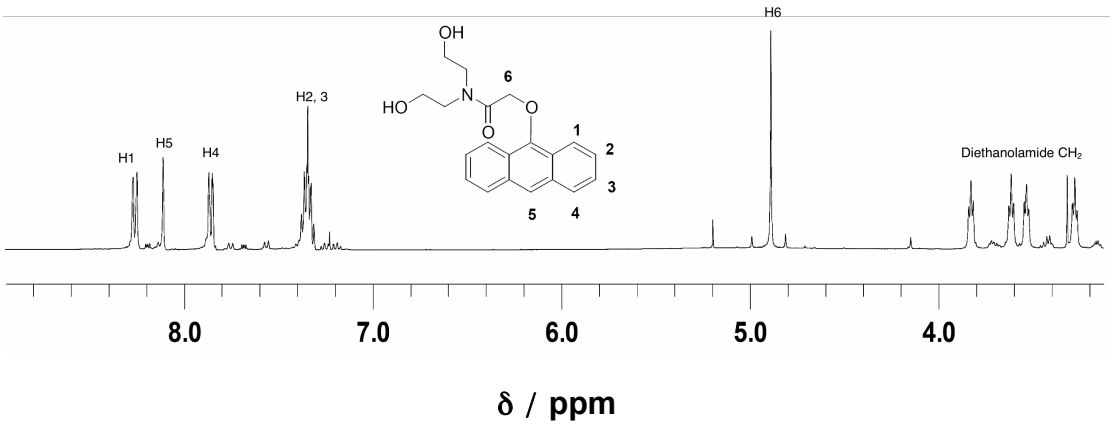


Figure 2.3: ^{1}H NMR spectrum of **3** (300 MHz, 1:1 CD_3CN / $CDCl_3$).

Protection of the first hydroxyl group was achieved using dimethoxytrityl chloride (DMTrCl) in the presence of a catalytic amount of dimethyl aminopyridine (DMAP). In order to give a predominantly mono *O*-protected product, only one molar equivalent of the DMTrCl was used. Even so, the product required purification by chromatography to remove what was presumably a polyfunctionalised by-product, though the yield (40%) was considered to be excellent for a mono-protection in the presence of a competing group. Formation of the product was confirmed by electrospray mass spectrometry, giving a single peak with *m/z* 663, the monosodium adduct of the molecule. High resolution mass spectrometry also confirmed the exact mass experimentally matched the theoretical mass. ^{1}H NMR of **4a** again showed evidence of restricted rotation around the amide bond observed in **3**, in this case due to the signal for the methylene linker protons (H6) being now split into two separate signals, each integrating to effectively half of the original H6 signal. This was ascribed to the mono *O*-functionalisation of **4a** with DMTr, and it was investigated by 2D ^{1}H-^{13}C HSQC NMR spectroscopy.

Cross peaks are observed in this spectrum for protons that are linked to carbon centres by 1 bond i.e. C-H bonds. The two split signals for H6 were observed to show a cross peak to only one carbon centre (Figure 2.4).

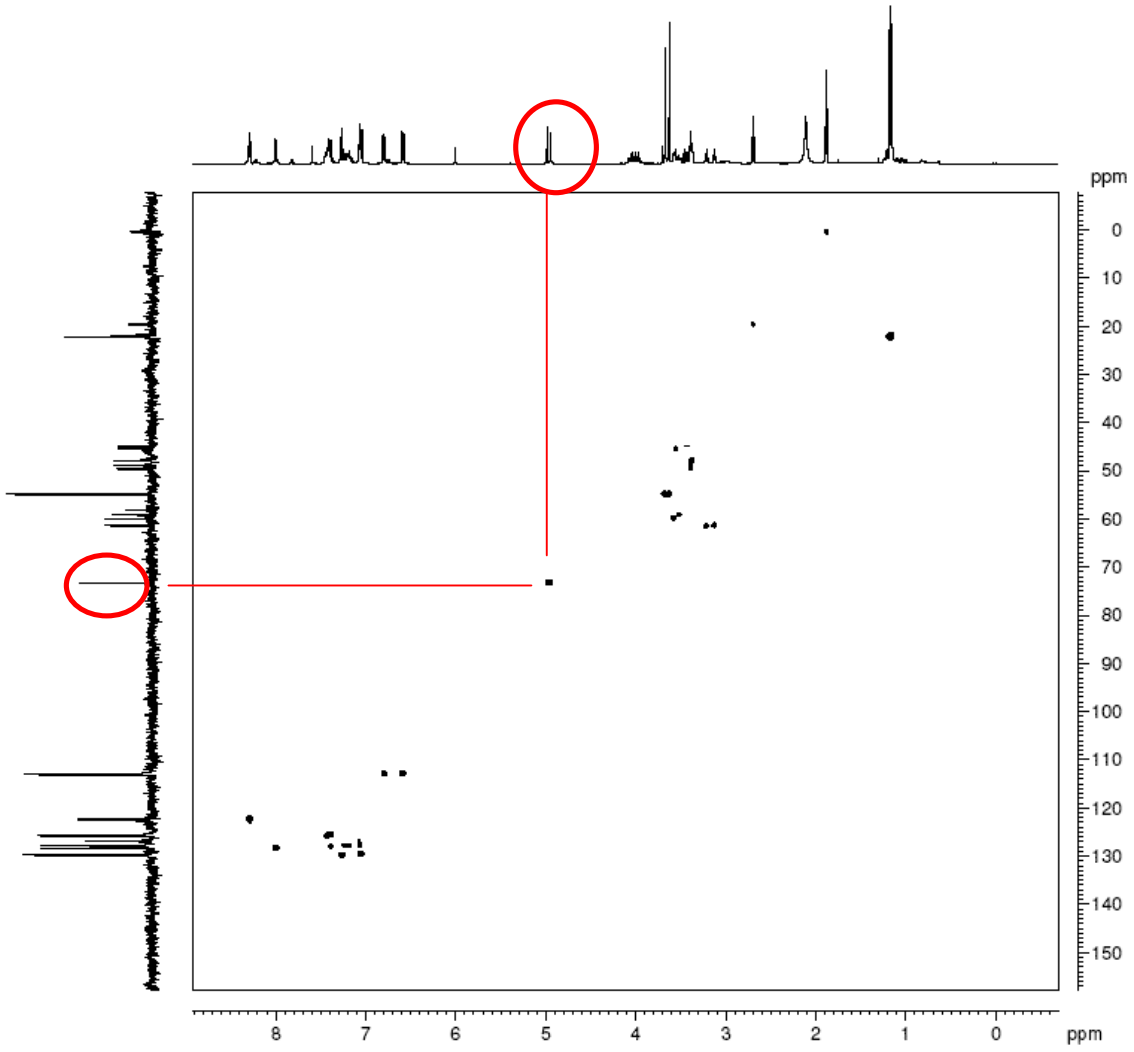


Figure 2.4: HSQC spectrum of **4a** showing the C6-H6 cross-peak (1:1 $CD_3CN/CDCl_3$).

To confirm that the restricted rotation around the amide bond was the reason for the splitting seen in the ^1H NMR spectrum, variable temperature

NMR (VT-NMR) was used to probe the phenomena, in conjunction with Dr L. Giordano. Firstly, a ^{1}H NMR spectrum of a sample containing **4a** was taken at 300 K. This was followed by heating to a range of temperatures; successive ^{1}H NMR spectra were taken at 318 K and 333 K. The results confirmed that the splitting of the peaks was due to restricted rotation around the amide bond, as the peaks merged into one peak as the temperature was successively increased (Figure 2.5).

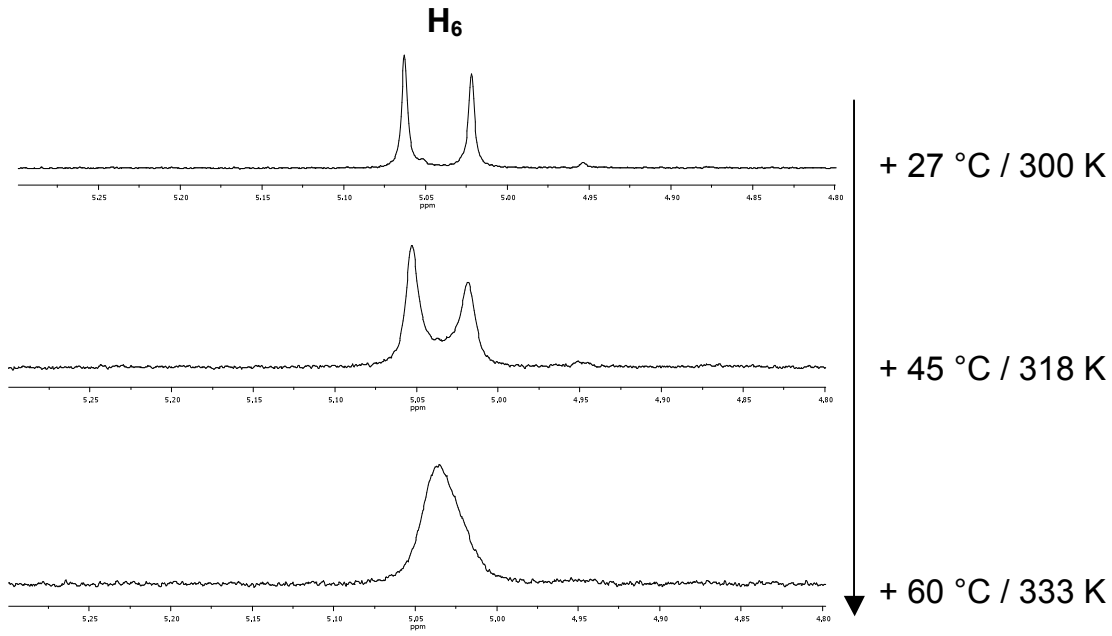


Figure 2.5: VT-NMR experiment monitoring H6 in product **4a** (300 MHz, CD_3CN).

The magnitude of the energy barrier with which had to be overcome to ensure rotation of the bond was then calculated as follows. First of all, the exchange rate between the two conformations at the coalescence temperature (T_c, estimated to be 333 K) was calculated using the following equation:[46]

$$k_{ex} = (\pi \cdot \Delta\nu) \ / \ (\sqrt{2}) \qquad \text{(eq. 1)}$$

Where $\Delta\nu$ is the difference in frequency for the separate sharp line signals at +27 °C. For our signals, $\Delta\nu$ = 2015.1 Hz – 1998.6 Hz = 16.4 Hz.

Hence,

$k_{ex} = 35.5\ s^{-1}$ at 333 K.

The free energy of activation, $\Delta G_{ex}^{\ddagger}$, was also calculated by using the following equation:

$$\Delta G_{ex}^{\ddagger} = R \times T_c(K) [\ 23 + \ln (T_c(K) / \Delta\nu)\ ] \qquad \text{(eq. 2)}$$

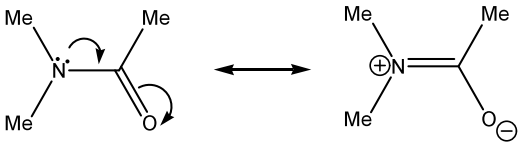
R is the gas constant (8.31×10^{-3} kJ K^{-1} mol^{-1}).

Therefore,

$\Delta G_{ex}^{\ddagger} = 72\ kJ\ mol^{-1}$ at 333 K.

In comparison, T_c = 331 K for the rotation of the amide bond in *N,N*-dimethylacetamide, due to delocalisation of the nitrogen lone pair.[47] *N,N*,-dimethyl formamide also has a similar coalesence temperature of 337 K.[46]

These latter two standard are similar to the calculated value here. It can therefore be concluded that the results for **4a** are fully consistent with restricted rotation around the amide bond due to partial π character (Figure 2.6). It is worth noting that as stated earlier, the coalescence temperature is an estimate, in this case, only two NMR spectra were taken at elevated temperatures, and the T_c temperature of 333K is likely to be an upper limit.



$\Delta G_{ex}^{\ddagger}$ ~ 90 kJ mol^{-1}

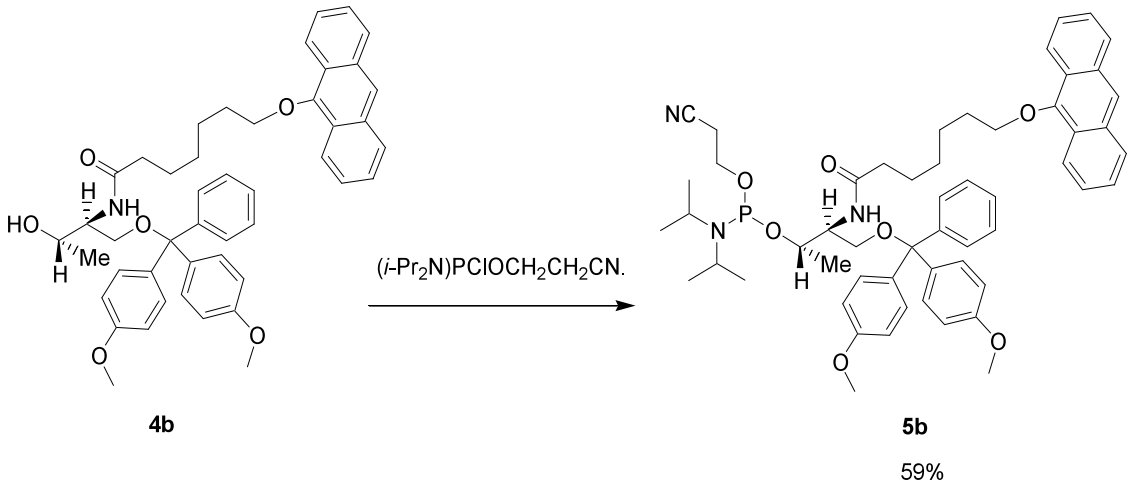
Figure 2.6: Partial π character of the C-N bond in *N, N*-dimethylformamide.

The purified product **4a** was then reacted further to make the activated phosphoramidite precursor to be incorporated into DNA oligonucleotides. Briefly, **4a** was reacted with 2-cyanoethyl-diisopropylchlorophosphoramidite, in the presence of di-isopropylamine. The reaction was very fast, taking only 1 hour to go to completion as evidenced by thin layer chromatography. The product, **5a**, was produced in acceptable yield (34%). ^{31}P {^{1}H, ^{13}C} NMR was extremely useful in demonstrating the incorporation of phosphorus in the molecule, with two peaks appearing at 149 ppm corresponding to the incorporation of the phosphoramidite; although a singlet would be expected, it seems that the effect of the restricted rotation around the amide bond also splits the phosphoramidite peak into two separate signals even in this product.[48] Mass spectrometry of the reaction product gave the correct mass for **5a**, at *m/z* 864, with high resolution mass giving close to the calculated

value. Due to time constraints, no further work was performed to incorporate the label into DNA oligonucleotides, though it is expected that in the near future this task will be completed, and the fluorescence properties of the probe sequences examined when paired with matching and mismatching sequences, the latter containing SNPs. The results will be very interesting to compare against the original anthracene probe discussed previously: although the linker is of the same length in **5a** as the original probe, the spacer distance is slightly larger in **5a**, hence the effect of this change should give significant insight into the importance of this variable.

2.2. Anthracene tag 5b for single base-pair mismatch detection in DNA duplexes: synthesis, incorporation into DNA oligonucleotides and photochemistry with target sequences.

It was also decided to investigate the behaviour of a tag containing a chiral spacer group due to the chiral nature of DNA itself, B-DNA existing as a right handed helix. It was anticipated that probes derived from **5b** (Scheme 2.2), containing a D-threoninal linker with the two chiral centres in (2*S*, 3*S*) configuration, would interact with single-stranded and duplex DNA in a extremely different manner than to the (2*S*, 3*R*), (2*R*, 3*S*) or (2*R*, 3*R*) variants, resulting in different fluorescence profiles both before and after hybridisation. In addition, the linker length was relatively large and would allow comparison with probes containing a shorter linker length.



4b **5b**

59%

Scheme 2.2: Synthesis of SNP probe precursor **5b**, containing a D-threoninal linker.

The mono *O*-protected product **4b** was supplied by Mr J. -L. Duprey from this research group, and was fully characterised at the point of use.[25] **4b**

was reacted with 2-cyanoethyl- *N, N* -diisopropylchlorophosphoramidite (in an identical manner to that discussed previously for the synthesis of **5a**) to give the activated product **5b**. The reaction product was characterised by ^{31}P {^{1}H, ^{13}C} NMR which demonstrated the phosphorus incorporation. Electrospray mass spectrometry gave the correct mass for the sodium adduct of **5b** at *m/z* 935. Again, reaction time was very short as evidenced by thin layer chromatography, ~ 1 hour, and occurred in fair yield (59%).

2.2.1. Oligonucleotide synthesis.

5b was then incorporated into oligonucleotide sequences using an automated DNA synthesiser (see experimental for further details). A range of oligonucleotides containing the fluorescent probe were synthesised. The tagged sequences were fourteen natural bases in length plus the single synthetic anthracene **5b**. Only the base pairs flanking the anthracene tag were changed between different probes. The probe sequences were purified by HPLC, followed by desalting using size-exclusion chromatography (Nap-10 column, GE Healthcare). The incorporation of the probe was verified using negative ion mode electrospray mass spectrometry, which gave excellent matches for every probe sequence compared to the calculated value. A summary of probe sequences and calculated *vs.* experimental mass can be found in Table 2.1 (see appendix for mass spectra). The concentration of each synthesised probe (not shown) was determined by UV-Vis absorption spectroscopy (Beer-Lambert law) using the value of the absorption value at 260 nm and molar absorptivities. The latter was calculated by adding together

the molar absorptivities of the individual bases in the sequence and that of anthracene.

Table 2.1: Summary of probe sequences and their experimentally determined masses.

Designation	Sequence	Calc. RMM	*m/z* (obs.)
PrA	5'-TGGACT***C*****5b*****C***TCAATG-3'	4710.3	4711.0
PrB	5'-TGGACT***C*****5b*****G***TCAATG-3'	4750.3	4751.0
PrC	5'-TGGACT***C*****5b*****A***TCAATG-3'	4734.3	4735.0
PrD	5'-TGGACT***C*****5b*****T***TCAATG-3'	4725.3	4726.0
PrE	5'-TGGACT***G*****5b*****C***TCAATG-3'	4750.3	4751.0
PrF	5'-TGGACT***A*****5b*****C***TCAATG-3'	4734.3	4735.0
PrG	5'-TGGACT***T*****5b*****C***TCAATG-3'	4725.3	4727.0

In the same way, a range of target DNA sequences were synthesised using automated solid phase synthesis followed by purification by HPLC and desalting using size exclusion chromatography. These sequences were comprised of both 15-mer and 14-mer strand lengths. The 14-mer strand length targets were included to test the effect of base deletion on the target strand adjacent to the anthracene unit, allowing the latter group to intercalate more effectively between bases in the DNA duplex. The latter effect, which stabilises the DNA duplex, has been elegantly illustrated in the article by Fukui *et al.* on intercalation of acridine into oligonucleotide sequences (Figure 2.7). [49]

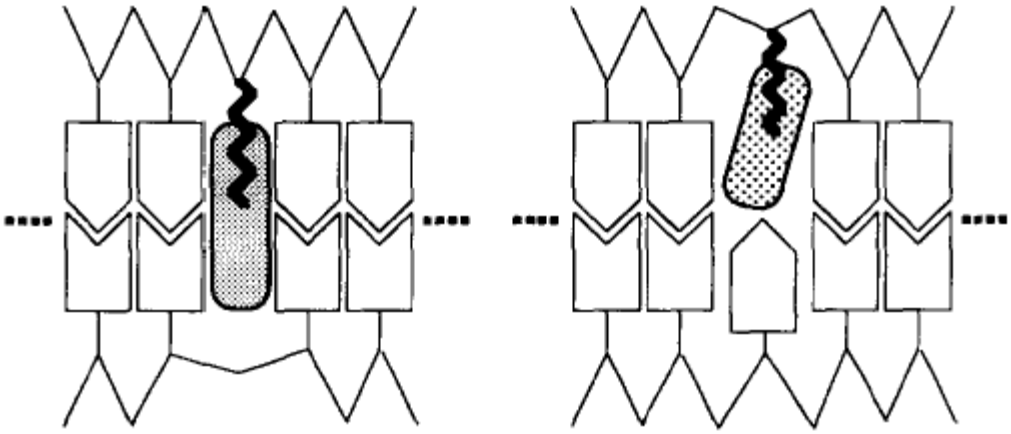


Figure 2.7: Intercalative geometries with deletion sequences (left) and sequences with a base in the adjacent strand (right).[49]

A summary of the target sequences is shown in Table 2.2. The concentrations of the synthesised target sequences (not shown) were determined by UV-Vis absorption spectroscopy using the absorption value at 260 nm, as for the probe sequences.

Table 2.2: Summary of target sequences. "_" refers to a base deletion.

Target	Sequence	Length	'Matched' Probe	Opp. Base‡
TarA1	3'-ACCTGA***GAG***AGTTAC-5'	15-mer	**PrA**	A
TarA2	3'-ACCTGA***G_G***AGTTAC-5'	14-mer	**PrA**	del
TarA3	3'-ACCTGA***GTG***AGTTAC-5'	15-mer	**PrA**	T
TarA4	3'-ACCTGA***GCG***AGTTAC-5'	15-mer	**PrA**	C
TarA5	3'-ACCTGA***GGG***AGTTAC-5'	15-mer	**PrA**	G
TarB1	3'-ACCTGA***GAC***AGTTAC-5'	15-mer	**PrB**	A
TarB2	3'-ACCTGA***G_C***AGTTAC-5'	14-mer	**PrB**	del
TarC1	3'-ACCTGA***GAT***AGTTAC-5'	15-mer	**PrC**	A
TarC2	3'-ACCTGA***G_T***AGTTAC-5'	14-mer	**PrC**	del
TarD1	3'-ACCTGA***GAA***AGTTAC-5'	15-mer	**PrD**	A
TarD2	3'-ACCTGA***G_A***AGTTAC-5'	14-mer	**PrD**	del
TarE1	3'-ACCTGA***CAG***AGTTAC-5'	15-mer	**PrE**	A
TarE2	3'-ACCTGA***C_G***AGTTAC-5'	14-mer	**PrE**	del
TarF1	3'-ACCTGA***TAG***AGTTAC-5'	15-mer	**PrF**	A
TarF2	3'-ACCTGA***T_G***AGTTAC-5'	14-mer	**PrF**	del
TarG1	3'-ACCTGA***AAG***AGTTAC-5'	15-mer	**PrG**	A
TarG2	3'-ACCTGA***A_G***AGTTAC-5'	14-mer	**PrG**	del

‡ *'Opp. base' refers to the base opposite anthracene unit in the probe-target duplex.*

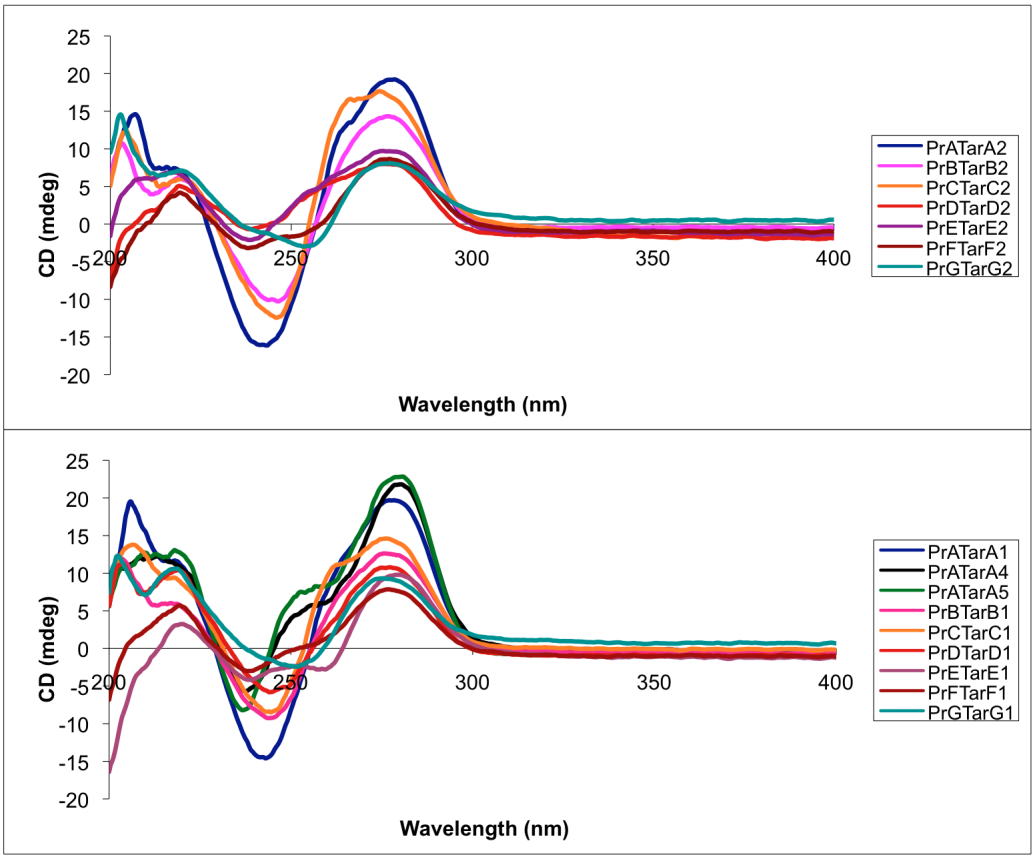


Figure 2.8: CD spectra of duplexes with matched sequences. Top: 14-mer / 15-mer target-probe duplexes. Bottom: 15-mer / 15-mer target-probe duplexes.

2.2.2. Circular dichroism studies.

Circular dichroism (CD) spectra of the matched and mismatched duplexes were recorded to give insight into the solution structure of the sequences (Figures 2.8 and 2.9, see pull-out sheet for sequence details). CD signals arise from the difference in absorbance between incident left- and right-handed circularly polarised radiation by a chiral molecule. Due to its right-handed helicity, B-DNA is inherently chiral, and so produces a CD signal at 250 nm, which mirrors its UV-Vis absorbance spectrum. Changes to the helicity of the DNA, i.e. transformation to other polymorphs such as the Z-form in solution, result in alterations to the profile of the CD spectrum. All of the matched and mismatched 14-mer / 15-mer and 15-mer / 15-mer target-probe

duplexes show the typical circular dichroism expected from DNA in solution in the B-form, with a negative peak at 250 nm corresponding to the CD signal of the aromatic base pairs i.e. the duplexes are still helical in structure. The cause of the differences in peak amplitudes is possibly due to minor differences in the helicity of the probe-target duplexes. The incorporation of the anthracene into the DNA duplexes cannot be discerned at these concentrations. At higher concentrations, one would expect CD signals at 300 - 400 nm corresponding to the well known finger-like absorbances of the vibronic fine structure in anthracene.

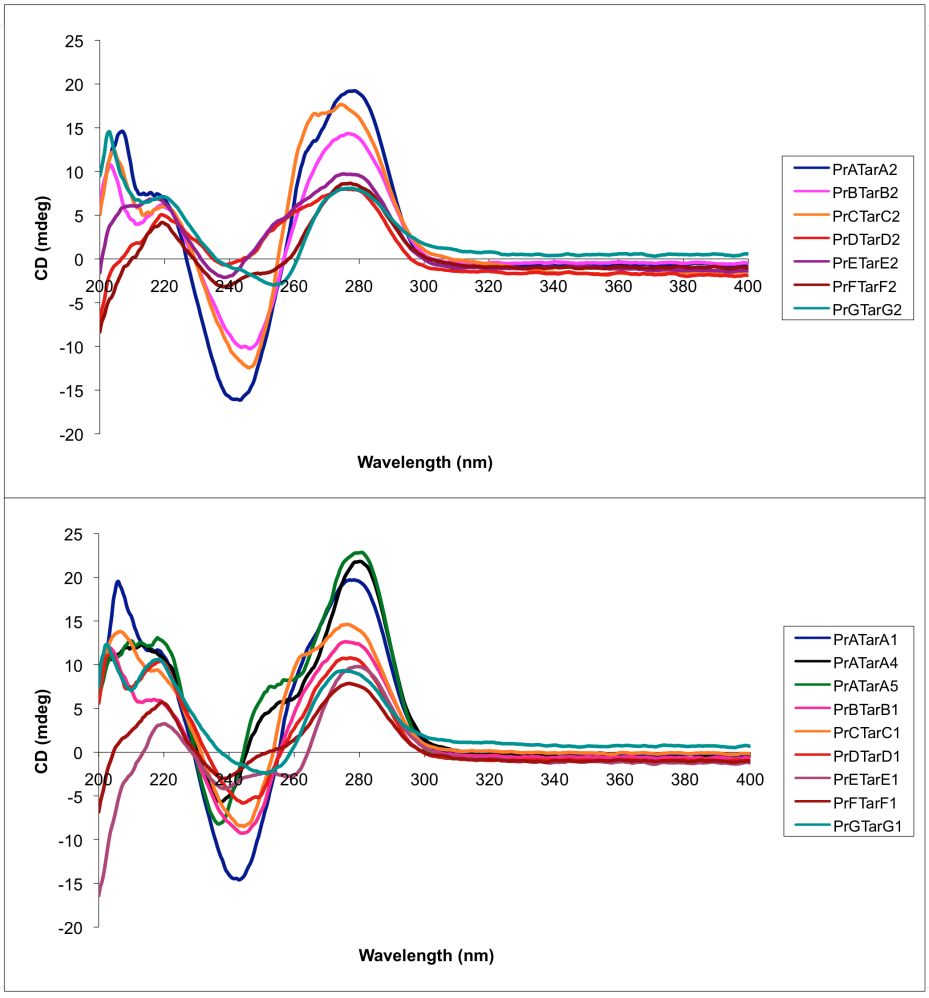


Figure 2.9: CD spectra of mismatch duplexes with probe A. Top: 15-mer / 15-mer probe-target duplexes. Bottom: 14-mer / 15-mer target-probe duplexes.

2.2.3. DNA melting experiments.

The melting temperatures (T_m) of matched probe-target duplexes were investigated by UV-Vis absorption spectroscopy. *Melting* refers to the temperature at which the DNA duplex dissociates into two single strands, which is accompanied by a change in the UV-Visible spectrum of the sample under scrutiny. The same measurements were also performed on a range of mismatched sequences using probe A (**PrA**) alone on all the target sequences, B*x* – G*x* (where *x* = 1 – 2). The UV-Vis instrument (Varian Cary 5000) used in this study allowed the automated recording of melting temperatures (using a ramp from 15 – 85 °C; T_m values reported are an average of two runs on the same sample, the second run preceded by an annealing cool-down step to re-hybridise the double stranded DNA duplex), which are reported in Table 2.3. Note the 3' / 5' designation in the polymorphism type column refers to the location of the mismatch on the flanking bases with respect to the probe strand, rather than the target strand. Taking the example of **PrA** hybridised with **TarB1**:

5'-TGGACTC5b**C**TCAATG-3' **PrA**
3'-ACCTGAGA**C**AGTTAC-5' **TarB1**

would lead to a shorthand description of the mismatch as 3' C-C. Conversely, the example of **PrA** hybridised with **Tar E1:**

5'-TGGACT**C**5bCTCAATG-3' **PrA**
3'-ACCTGA**C**AGAGTTAC-5' **TarE1**

would lead to the mismatch descriptor 5' C-C.

Table 2.3: DNA melting temperatures for matched and mismatched probe-target duplexes.

'Matched' sequences				*'Mismatched' sequences*			
Probe	**Target**	**T_m (°C)**		**Probe**	**Target**	**T_m (°C)**	**Mismatch type**
PrA	TarA1	52.5		PrA	TarB1	34.0	3' C-C
PrA	TarA2	58.5		PrA	TarB2	40.5	3' C-C
PrA	TarA3	55.0		PrA	TarC1	40.5	3' C-T
PrA	TarA4	54.5		PrA	TarC2	43.0	3' C-T
PrA	TarA5	55.0		PrA	TarD1	41.0	3' C-A
PrB	TarB1	50.0		PrA	TarD2	43.0	3' C-A
PrB	TarB2	57.0		PrA	TarE1	44.0	5' C-C
PrC	TarC1	49.0		PrA	TarE2	44.5	5' C-C
PrC	TarC2	56.0		PrA	TarF1	40.5	5' C-T
PrD	TarD1	46.0		PrA	TarF2	43.5	5' C-T
PrD	TarD2	50.0		PrA	TarG1	41.0	5' C-A
PrE	TarE1	51.0		PrA	TarG2	43.5	5' C-A
PrE	TarE2	56.5					
PrF	TarF1	47.0					
PrF	TarF2	53.0					
PrG	TarG1	47.0					
PrG	TarG2	51.5					

The melting temperature of a DNA double helix is indicative of its inherent stability. It is well known that intercalating molecules such as anthracene increase the stability of the DNA double helix, and as such, the melting temperature of a duplex containing an intercalator is usually higher compared to the same duplex without the intercalator present.

The effect of anthracene intercalation in matched sequences is demonstrated nicely by considering the **PrATarA*x*** (where x = 1 – 5) duplex melting temperatures (Figure 2.10, see pull-out sheet for sequence details). It can be seen from the graph that PrATarA2 (14-mer deletion target) gives the largest melting temperature. This is due to the extra space afforded by the deletion in the target strand in the position adjacent to the anthracene on the probe strand, as anthracene can intercalate more effectively in this case. The same can also be said for all of the deletion duplexes (Pr: 15-mer / Tar: 14-mer, x =2) versus the 15-mer duplexes (Pr: 15-mer / Tar: 15-mer, x = 1) e.g. **PrBTarB2** *vs.* **PrBTarB1** (ΔT_m = +7 °C) ; again, the increase in the melting temperature observed are due to improved intercalation from the anthracene tag caused by base deletion (refer to Figure 2.7).

The effect of changing the base in the target strand opposite to the anthracene in the probe strand is not as pronounced an effect as base deletion. This can be seen by comparing the T_m of the **PrATarA1** against the **PrATarA*x*** (where x = 3, 4, 5) sequences. The melting temperature is slightly increased for the sequences where T, C and G are opposite the anthracene though the effect is small and for these 15-mer duplexes it may be

concluded that the degree of anthracene intercalation is probably very similar; hence the effect of the base opposite to anthracene does not influence the intercalative ability of the anthracene tag in any significant way.

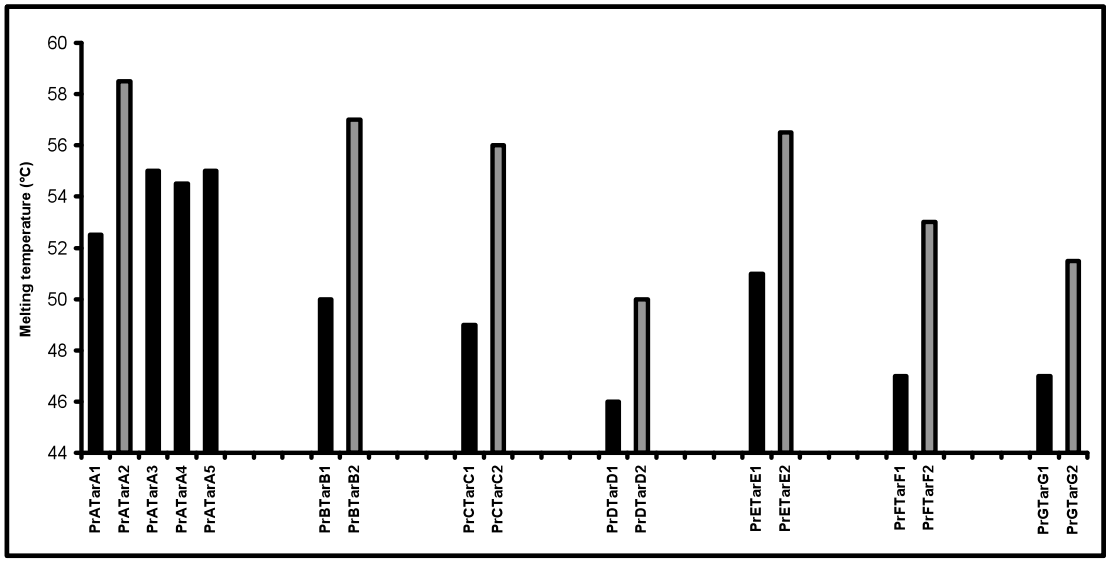


Figure 2.10: Melting temperatures of matching probe-target duplexes. Data for 15-mer / 14-mer deletion duplexes are coloured grey.

The melting temperatures of the mismatched sequences compared to the matched sequences are dramatically depressed, due to the loss of a matched base pair. In addition to this, the effect of anthracene intercalation in the sequences containing **PrA** hybridised with various mismatching target sequences is similar to that observed for the matched sequences, though is less marked (Figure 2.11, see pull-out sheet for sequence details); in general, the duplexes containing deletion target strands (Pr: 15-mer / Tar: 14-mer) are of slightly higher melting point than their 15-mer duplex counterparts. This again is due to the improved ability of the anthracene to intercalate through base deletion on the target strand. However, as these sequences contain

mismatches (see Table 2) both in upstream and downstream flanking positions, the hydrogen bonding in the cavity is imperfect and the intercalation of the anthracene is obstructed by the base pair mismatch, hence the melting temperature increase is not as prominent compared to the fully matched sequences. The trend in melting temperature for 15-mer / 14-mer sequences is also general for 3' and 5' flanking SNPs, which demonstrates the effect is purely due to the increased anthracene intercalation found in these duplexes.

.

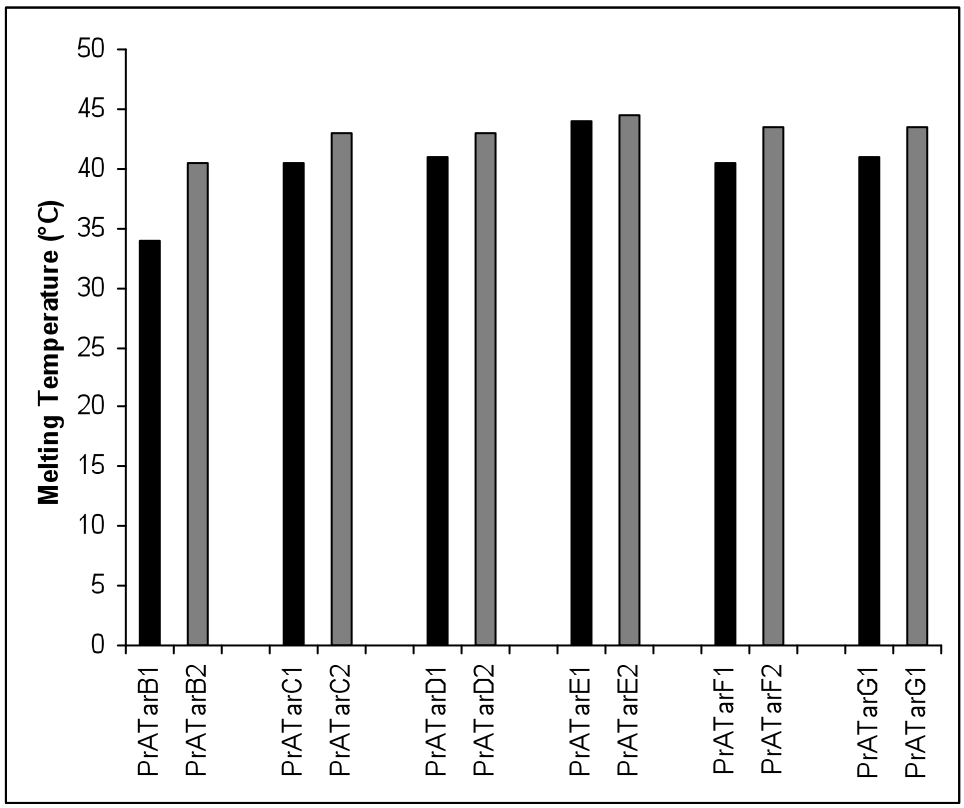
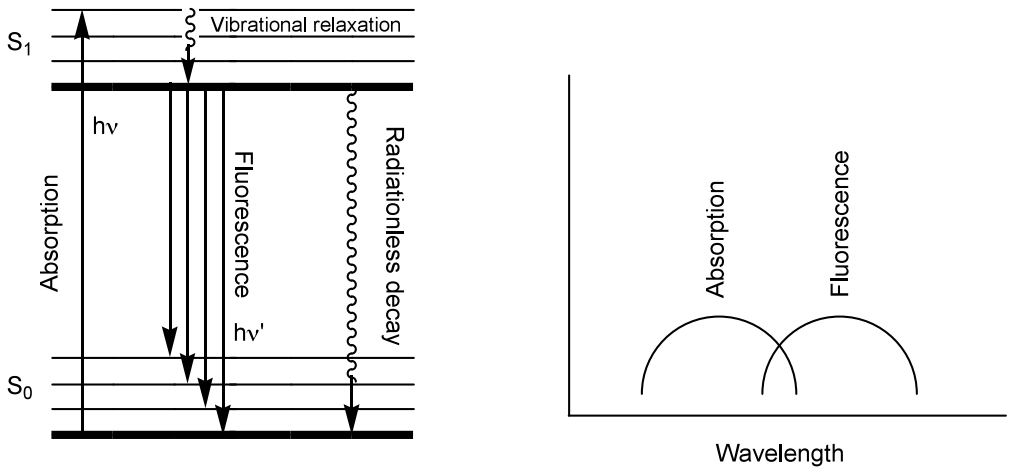


Figure 2.11: Melting temperatures of mismatch duplexes incorporating probe A. Data for 15-mer / 14-mer deletion duplexes are coloured grey.

2.2.4. Fluorescence studies.

Fluorescence involves the relaxation from the lowest level of an excited singlet state to the singlet ground state, ejecting a photon of energy. This is usually preceded by excitation into the excited state by absorbance of a photon of higher energy than that emitted. The difference between the absorbance maximum and the emission maximum, called the Stokes shift, arises from vibrational relaxation in the upper state. The process can be summarised[26] in simplified schemes known as Jablonski diagrams (Scheme 2.3).



Scheme 2.3: Jablonski diagram showing the typical process involved in anthracene fluorescence (left) and a cartoon of an absorption-fluorescence profile (right).

The excited state may also be deactivated (quenched) by radiationless transitions, such as collisions with other molecules. The extent of the quenching is often dependent on the environment in which the excited fluorescent molecule finds itself in. It is well known that polar environments

tend to enhance the molecular fluorescence in polyaromatic hydrocarbons (e.g. anthracene), whereas non-polar environments (such as anthracene intercalated between the aromatic bases of DNA) quench the fluorescence as the extent of radiationless decay is greater. [34]

The quenching or enhancement of anthracene fluorescence when appended to DNA oligonucleotides is the basis of the SNP sensors designed in this laboratory. We decided to investigate the changes in anthracene fluorescence by titration of probes sequences containing appended anthracene against mismatched target sequences at 0, 0.5, 1.0, 1.5 and 2.0 equivalents. A selection of probe-target sequences were selected from the set of duplex sequences which were used in the melting temperature experiments for comparative purposes.

Table 2.4: Summary of results for matching duplexes with probe A.

Probe	Target	T_m (°C)	Opposite base (target strand)	Final Relative Fluorescence Change
PrA	TarA1	52.5	A	+171%
PrA	TarA2	58.5	Delete	-18%
PrA	TarA3	55.0	T	-41%
PrA	TarA4	54.5	C	+31%
PrA	TarA5	55.0	G	+110%

The nature of the matched probe-target sequences for probe A (**PrATarA*x***, where x = 1 – 5) were investigated by fluorescence titration (Figure 2.12, see pull-out sheet for sequence details). Converse to the melting temperature results, it was found that varying, or indeed deletion of, the base in the opposite strand in the position adjacent to the anthracene

probe had a direct influence over the relative fluorescence of the anthracene in the duplex compared to the single strand probe A. If a larger purine base (G, A) is used opposite the anthracene in the duplex, as in **PrATarA1** and **PrATarA5**, the fluorescence of the anthracene increases dramatically upon duplex formation, over 110% in both cases. The smaller pyrimidine bases (T, C) tend to give opposite effects to each other, with a C base opposite giving a 31% increase (**PrATarA4**), and a T base giving a 41% decrease (**PrATarA3**). The duplex sequence where a deletion was made in this position (**PrATarA3**) gave a smaller 18% reduction in relative fluorescence compared to the tagged single strand oligonucleotide, **PrA** (Table 2.4). We concluded from the melting temperature experiments that the anthracene unit is intercalated, and in the duplexes with a base opposite the anthracene the degree of intercalation is similar (apart from **PrATarA1**, which displays a depressed T_m) , so these results imply that the fluorescence behaviour of **5b** within the probe A strand is dependent on the base opposite in the target strand if the flanking base pairs remain the same (in this case the flanking base pairs are GC in both upstream and downstream directions).

By consideration of the fluorescence process, it is apparent that the non-radiative decay of the excited state of the anthracene is either suppressed (in the case of the larger purine bases A and G and also to a lesser extent the smaller pyrimidine, C) or enhanced (in the case of T and deletion). It may be postulated that purine bases, due to their larger size may prevent full anthracene intercalation, and that the anthracene fluorescence is enhanced by partial expulsion into the polar environment of the bulk solvent.

We may say that this is the case for **PrATarA1**, which shows a depressed T_m compared to the other matched sequences with **PrA**. However, this would certainly be an incorrect conclusion to make; the result from **PrATarA4** disagrees with the result from **PrATarA1** , as this also has a purine base opposite anthracene, but has a similar melting temperature to matched duplexes with a pyrimidine base adjacent to anthracene, hence intercalation should not be a factor. The behaviour of the anthracene unit in **PrA** upon duplex formation may therefore simply rely on how efficient the base opposite either quenches or enhances the fluorescence from the excited state. The latter effect has been communicated by Netzel and co-workers[50] who reported the magnitude of fluorescence quenching by the four DNA bases to be as follows: C > T > G > A. This is mirrored in our results, where the C and T bases in the opposite strand lead to relatively reduced fluorescence compared to their A and G counterparts. It is reasonable therefore to suggest that a second factor is in play: the expulsion of the anthracene unit into a binding pocket. This effect has been observed in the binding of *aptamers* (which are short nucleic acid sequences designed to sequester target sequences or fluorescent dyes), to fluorescent organic dye molecules such as malachite green leading to either dramatic quenching or enhancement of the fluorescence signal. [51]

In the case of the deletion (14-mer/15-mer) duplex **PrATarA2**, the anthracene unit in **PrA** is intercalated between two GC base pairs, which are known to quench anthracene fluorescence, [41] hence the decrease observed.

The discovery of this behaviour may be important of the fine tuning of SNP sensors based on anthracene in future applications, in particular tuning the selectivity of the probe.

Table 2.5: Summary of results for mismatch duplexes with probe A.

Probe	Target	T_m (°C)	Duplex type	Mismatch type	Final Relative Fluorescence Change
PrA	TarB1	34.0	15 – 15	3' C-C	+5%
PrA	TarB2	40.5	15 - 14 (del)	3' C-C	-18%
PrA	TarC1	40.5	15 - 15	3' C-T	+19%
PrA	TarC2	43.0	15 - 14 (del)	3' C-T	-31%
PrA	TarD1	41.0	15 - 15	3' C-A	+44%
PrA	TarD2	43.0	15 - 14 (del)	3' C-A	-25%
PrA	TarE1	44.0	15 - 15	5' C-C	+38%
PrA	TarE2	44.5	15 - 14 (del)	5' C-C	-1%
PrA	TarF1	40.5	15 - 15	5' C-T	+2%
PrA	TarF2	43.5	15 - 14 (del)	5' C-T	-32%
PrA	TarG1	41.0	15 - 15	5' C-A	+43%
PrA	TarG2	43.5	15 - 14 (del)	5' C-A	-2%

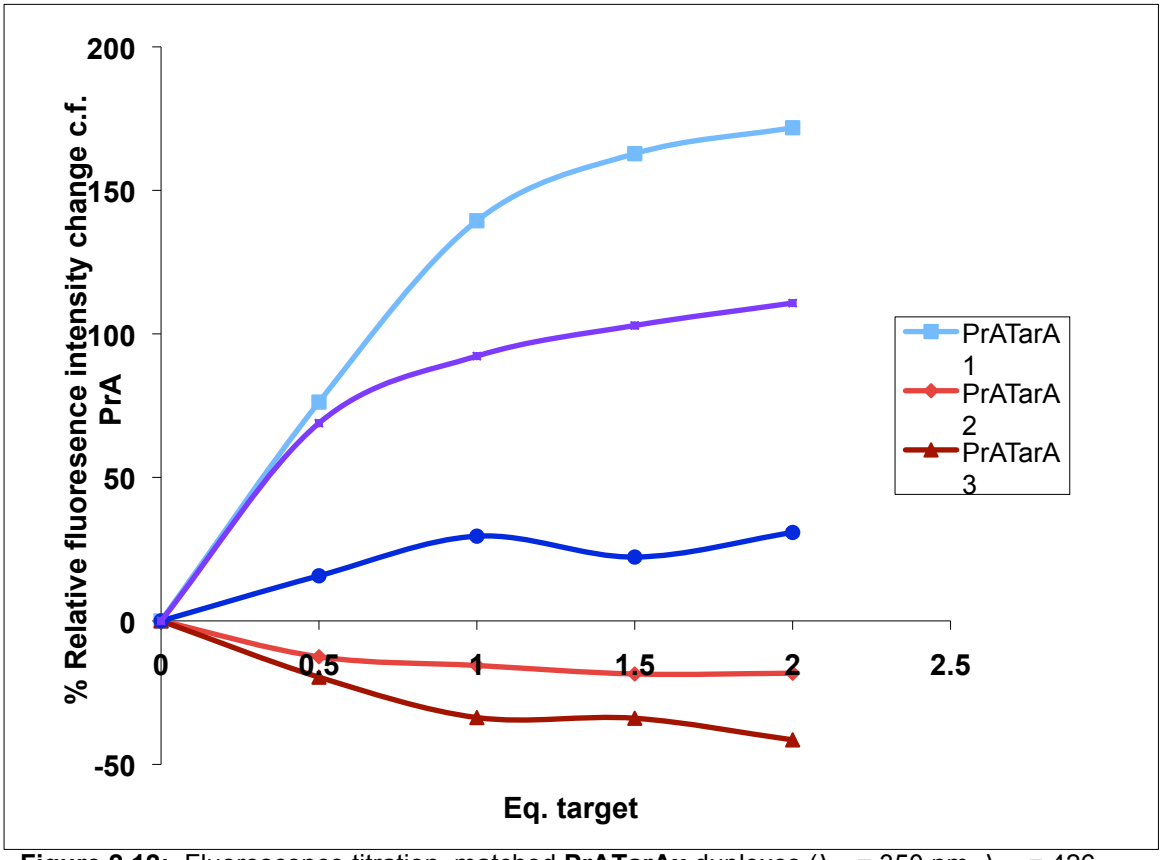


Figure 2.12: Fluorescence titration, matched **PrATarAx** duplexes (λ_{ex} = 350 nm, λ_{em} = 426 nm).

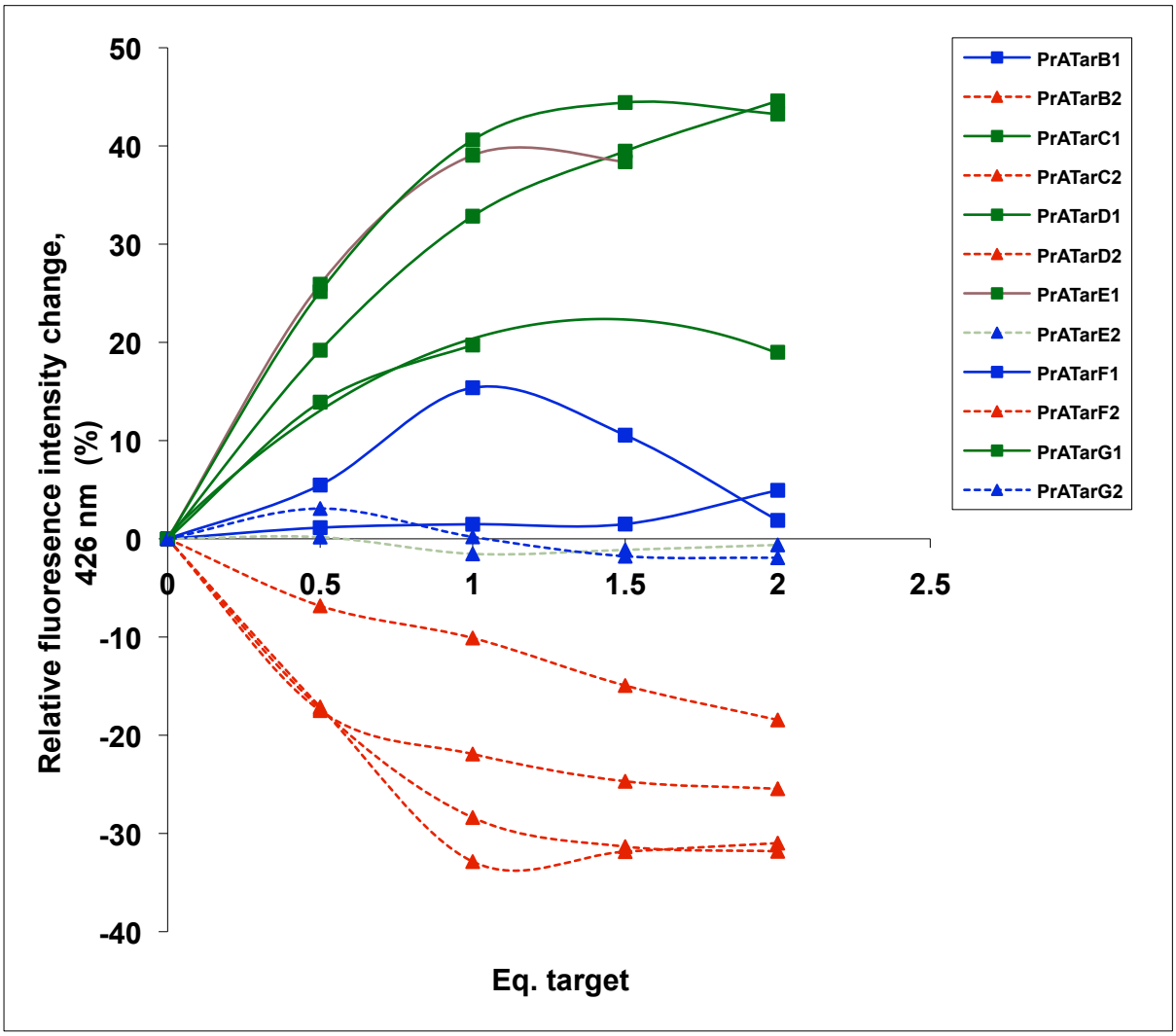


Figure 2.13: Fluorescence titration, mismatched **PrATar** duplexes (λ_{ex} = 350 nm, λ_{em} = 426 nm). Hash lines / triangles: 14-mer / 15-mer deletion duplexes. Solid lines / squares: 15-mer / 15-mer duplexes.

To determine whether **PrA** had potential to be an SNP sensor, the mismatched **PrA**-target duplexes were investigated by fluorescence titration (Figure 2.13, also see pull-out sheet for sequence details). **PrA** was titrated against a number of target sequences **TarB*x*** - **TarG*x***, each with a mismatch (*x* = 1 targets) or a mismatch plus base deletion in the target strand (14-mer / 15 mer duplex, target *x* = 2). In the *x* = 1 sequences, the base

opposite to the anthracene in the target strand was adenine (A). The results fell into three categories. Target sequences were either found to i) enhance, ii) quench or iii) have little effect on the anthracene unit's fluorescence. The results are summarised in Table 2.5.

It was found that for mismatches in the 14-mer / 15-mer sequences (*x* = 2) on the whole led to quenching of fluorescence compared to **PrA** alone, or simply no effect from hybridisation. From the melting temperature experiments in the mismatch sequences, it was clear that in general the 14-mer / 15-mer duplexes gave better anthracene intercalation from the higher melting temperatures compared to the 15-mer duplexes. Hence in these sequences the anthracene residue is intercalated in the non-polar environment of the DNA helix. As a result, the anthracene emission is quenched or simply not enhanced at all. The results from **PrATarC2** and **PrATarF2**, where quenching is of a similar magnitude, demonstrates that the position of C-T mismatches, either upstream or downstream compared to anthracene, does not influence the probe response (3' C-T plus deletion, -31% *vs.* 5' C-T plus deletion, -32%). The result for **PrATarF1** is probably erroneous due to the shape of the curve, fluorescence decreasing after 1 eq. of target.

Conversely, for the sequences with the adjacent base to anthracene included (**PrA** *vs.* **TarB*x*** – **G*x*** where *x* = 1) the anthracene fluorescence was either enhanced or unaffected. The signal of the anthracene in **PrA** was enhanced greatly with the flanking mismatches created upon duplex formation

with in **TarC1** (3' C-T , +19%), **TarD1** (3' C-A, +44%), **TarE1** (5' C-C, +38%) and **TarG1** (5' C-A, +43%). The probe response for C-A mismatches in **PrATarD1** and **PrATarG1** duplexes is particularly striking, giving a sizeable enhancement in fluorescence intensity that is independent of upstream or downstream position of the mismatch relative to the anthracene unit. The flanking position seems to play a much more pivotal role in the case of C-C mismatches, where mutation in the 5' direction (on the probe strand) leads to a distinct fluorescence enhancement, compared to the same mutation 3' direction (**PrATarB1** *vs.* **PrATarE1**). The reverse effect is true for C-T SNPs, where fluorescence is enhanced for the SNP in the 3' direction, and unaffected for the same SNP in the 5' direction (**PrATar C1** *vs.* **PrATarF1**). The reasons for the differences in the latter two examples i.e. the dependence on the flanking bases, could be due to the chiral nature of the anthracene probe and its relative ability (or inability) to interact with the DNA due to the steric demands imposed by certain base pairs.

The fluorescence enhancements again are linked to the melting temperatures found for the 15-mer (*x = 1*) duplexes, which in general show decreased stability compared to their 15-mer / 14-mer duplex counterparts (*x* = 2). It is expected that there is less prominent anthracene intercalation in the duplexes, and as such the anthracene is more likely to be expelled into the polar bulk solvent outside the double helix, or placed into an aptamer-like binding pocket;[51] the latter have been known to enhance fluorescence of organic molecules upon binding. Computer modelling of the latter interactions should be considered to shed more light on the interesting effects

discovered, and should specifically explore the spatial location of the anthracene upon duplex formation to shed light on the mechanism of fluorescence enhancement.

2.3. Conclusions

In conclusion, firstly, a new fluorescent anthracene unit, **5a**, has been synthesised for incorporation into DNA oligonucleotides and use in SNP sensing.

Secondly, we synthesised and incorporated a chiral anthracene unit, **5b**, into DNA oligonucleotides. We found that there were differences in the melting temperature (T_m) of 14-mer/15-mer and 15-mer/15-mer duplexes, almost certainly due to improved anthracene intercalation into the double helix in the former case. Fluorescence titrations of **PrA** versus a range of matched targets demonstrated the effect of base deletion in the adjacent position to the anthracene on the target strand, and the dependence of anthracene fluorescence on the base opposite in the target strand; adenine leads to the strongest enhancement of the anthracene fluorescence signal. Fluorescence titrations of **PrA** versus a range of mismatched targets demonstrate a turn-on effect, especially for C-A mismatches. This latter effect is independent of upstream/downstream (3' / 5') direction with respect to the anthracene unit. The same turn-on effect was seen for C-T and C-C mismatches, though the

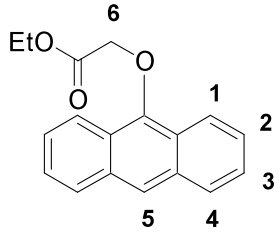
turn-on for both was highly dependent on the SNP position (5' / 3') relative to the probe, probably due to the chiral nature of the probe.

3.0. Experimental

General

Reagents were purchased from Aldrich or Acros. Solvents were purchased from Fisher or Aldrich. NMR solvents were purchased from Aldrich. UV-Vis spectra were recorded on a Varian Cary 5000 spectrometer with a peltier temperature controller for DNA melting experiments. Fluorescence experiments were carried out using a Shimadzu fluorimeter. NMR spectra were recorded on Bruker AVIII300 and Bruker AVIII400 spectrometers. Solid state IR spectra were recorded using a crystal stage. Electrospray mass spectra were recorded on a LC-TOF Micromass instrument. Elemental analysis was performed using a Carlo Erba EA1110. All synthetic procedures were carried out under dry nitrogen.

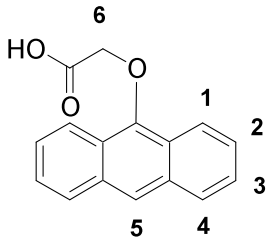
Anthracen-9-yloxy acetic acid ethyl ester (1).



Anthrone (5.83g, 0.03 mol) and potassium carbonate (4.15g, 0.03 mol) were dissolved in acetone (200 ml) and stirred under nitrogen at room temperature in the absence of light for 15 minutes. Ethyl bromoacetate (3.3 ml, 0.03 mol) was added and reaction mixture heated under reflux for 16 hours. The solution was filtered and dichloromethane (100 ml) and water (50 ml) added. After vigorous shaking, the aqueous layer was run off and the

organic layer dried over magnesium sulfate. The dichloromethane was removed under vacuum to leave an orange oil, which solidified on standing. The ester (R_f = 0.58) was purified on silica eluting hexane with 10% ethyl acetate to give a fluffy pale yellow solid (2.50 g, 30%). Analytical data agreed with that previously published by Moran *et al.*[52] TLC (Chloroform on silica): R_f = 0.58 . ^{1}H NMR (300 MHz, $CDCl_3$), δ ppm: 8.29 (2H, d, *J* = 1.0, H1), 8.18 (1H, s, H5), 7.91 (2H, d, *J*= 7.5, H4), 7.41 (4H, m, H2 / H3), 4.72 (2H, s, H6), 4.29 (2H, q, *J* = 7.2, $C\underline{H_2}CH_3$), 1.29, (2H, t, *J* = 7.2, $CH_2C\underline{H_3}$). MS (ES) *m/z*: 303 (M+Na). HRMS (ES) calcd for $C_{18}H_{16}O_3Na$: 303.0997, found: 303.0991. IR (ν_{max}, solid state) cm^{-1}: 1764 (s, C=O, ester), 1204 (s, C-O, ester).

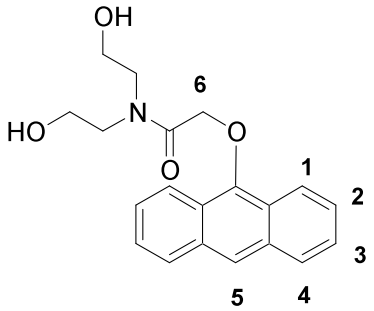
Anthracen-9-yloxy acetic acid (2).



Anthracen-9-yloxy acetic acid ethyl ester (**1**, 4.50 g, 0.01 mol) was dissolved in a 1:1 solution of 10% aqueous sodium hydroxide and ethanol (200 ml) and heated at 90 °C overnight in the absence of light. The solution was concentrated under reduced pressure and the residue dissolved in water (500 ml). 36% hydrochloric acid was added drop-wise to the solution to precipitate a solid which filtered. The solid was washed with water (20 ml) and dried to

give the desired product as a pearlescent, cream solid (90%. 3.65 g) No further purification was required. Analytical data agreed with that previously published by Moran *et. al.*[52]). Elemental analysis calcd for $C_{16}H_{12}O_3$: C 76.18%, H 4.79%, found: C 76.18%, H 4.44%. ^{1}H NMR (300 MHz, CD_3CN / $CDCl_3$), δ ppm: 8.28 (1H, s, H5), 8.26 (2H, d, *J* = 6.7, H1), 8.00 (2H, d, *J* = 6.7, H4), 7.48 (4H, m, H2 / H3), 4.77 (2H, s, H6). MS (ES) *m/z*: 251 (M-H). HRMS (ES) calcd for $C_{16}H_{11}O_3$: 251.0708 , found: 251.0710. IR (ν_{max}, solid state) cm^{-1}: 3100 (w, CO<u>OH</u>), (s, w, <u>CO</u>OH).

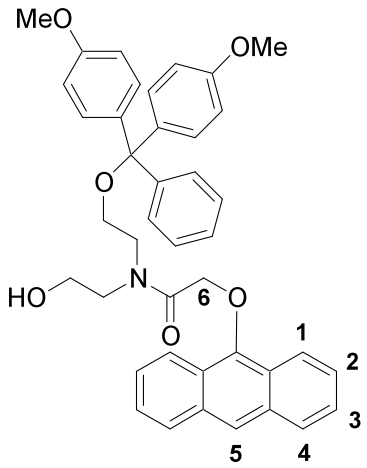
2-(Anthracen-9-yloxy)-*N*-(1-hydroxymethyl-2-hydroxy-ethyl)-acetamide (3).



Anthracen-9-yloxy acetic acid (**2,** 2.00 g, 7.9 mmol, 1.0 eq.) was dissolved in anhydrous *N, N*- dimethylformamide (55 ml). *O*-benzotriazole-*N,N,N′N′*-tetramethyl uronium hexafluorophosphate ("HBTU", 4.51 g, 11.9 mmol, 1.5 eq.) was added to the solution followed by di-isopropylethylamine (1.26 ml, 0.93 g, 7.2 mmol, 0.9 eq.) and the solution stirred under N_2 at room temperature in the absence of light for 15 minutes. Diethanolamine (0.76 g, 7.2 mmol, 0.9 eq.) was added and the reaction was heated to 40 °C and stirred for 40 hours, after which time, the solution was diluted into a mixture of 1:2 methanol/dichloromethane (100 ml) which was washed with water (3 x

50ml) and dried over magnesium sulfate. The solvent was removed under reduced pressure to give the crude product as yellow oil. Purification by silica column chromatography (Gradient from DCM with 1% MeOH via DCM with 3% MeOH to finally DCM with 5% MeOH) gave the desired product as a pale yellow solid (0.31 g, 13%). TLC (5% MeOH in DCM on silica): R_f = 0.15. ^{1}H NMR (300 MHz, $CD_3CN/CDCl_3$), δ ppm: 8.26 (2H, d, *J* = 6.0, H1), 8.11 (1H, s, H5), 7.86 (2H, d, *J* = 6.0, H4), 7.35 (4H, m, H2 / H3), 4.89 (2H, s, H6), 3.84 (2H, t, *J* = 3.6, H8), 3.62 (2H, t, *J*= 3.0 H7), 3.54 (2H, t, *J* = 3.6, H8'), 3.28 (2H, t, *J*= 3.0, H7'). ^{13}C PENDANT NMR (100 MHz, $CD_3CN/CDCl_3$) δ ppm: 169.5 (C=O, amide), 152.0 (C-O of $C_{14}H_9$), 131.5 (C of $C_{14}H_9$), 127.5 (CH of $C_{14}H_9$), 124.9 (CH of $C_{14}H_9$), 124.8 (CH of $C_{14}H_9$),123.6 (C of $C_{14}H_9$), 121.8 (CH of $C_{14}H_9$), 121.5 (C of $C_{14}H_9$), 72.3 (CH_2O), 59.6 (CH_2 ethanolamide), 58.6 (CH_2 ethanolamide), 49.6 (CH_2 ethanolamide), 48.7 (CH_2 ethanolamide). MS (ES) *m/z*: 362 $(M+Na)^+$. HRMS (ES) calcd for $C_{20}H_{21}NO_4Na$: 362.1368, found: 362.1372. IR (ν_{max}, solid state) cm^{-1}: 3354 (br, OH), 1624 (s, C=O amide).

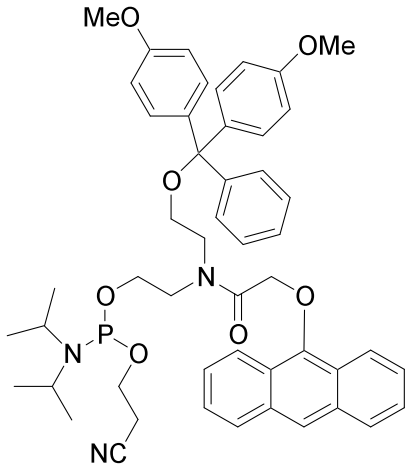
2-(Anthracen-9-yloxy)-*N*-(1-dimethoxytrityl-2-hydroxy-ethyl)-acetamide (4a).



2-(Anthracen-9-yloxy)-*N*-(1-hydroxymethyl-2-hydroxy-ethyl)-acetamide (**3**, 0.15 g, 0.44 mmol, 1.0 eq.) was dissolved in anhydrous pyridine (3 ml) and to it added dimethoxytrityl chloride (0.15 g, 0.44 mmol, 1.0 eq.) and 4-dimethylamino pyridine (5.3 mg, 43 µmol, 10 mol%). The mixture was stirred for 24 hours at room temperature, then poured into water (25 ml). The aqueous phase was washed with dichloromethane (2 × 25 ml) and the combined organic extracts dried over magnesium sulfate, filtered then the solvent removed under reduced pressure to give the orange-yellow crude product, which was purified by chromatography on a silica column eluting 99% hexane : 1% triethylamine, eluant increasing in polarity after sample loading was complete to 50% hexane : 49% ethylacetate, 1% triethylamine, to give the **4a** as a pale yellow solid (0.11 g, 40%) . TLC (5% MeOH in DCM, silica) : R_f = 0.35. ^{1}H NMR (300 MHz, CD_3CN), δ ppm: 8.29 (3H, m, H5 /H1), 7.98 (2H, d, H4), 7.40 (6H, m, H2/3, trityl), 7.25 (2H, m, trityl), 7.18 (1H, m, trityl), 7.05 (4H, m, trityl), 6.79 (2H, m trityl), 6.57 (2H, m, trityl), 4.97 (1H, s,

H6), 4.93 (1H, s, H6'), 3.67 (3H, s, trityl), 3.63 (3H, s, trityl), 3.53 (2H, m), 3.36 (2H, t, *J* = 3.0, CH_2 ethanolamide,), 3.19 (1H, t, *J* = 3.6, CH_2 of ethanolamide), 3.11 (1H, t, *J* = 3.6, CH_2 ethanolamide), 3.00 (1H, t, J = 3.0, CH_2 of ethanolamide), 2.93 (1H, t, J = 3.0, CH_2 ethanolamide). ^{13}C PENDANT NMR (100 MHz, CD_3CN) δ ppm: 170.0 (C=O, amide), 159.0 (C-O of $C_{14}H_9$), 146.4 (C, trityl), 146.2 (C, trityl), 138.4 (C, trityl) 137.8 (C, trityl), 135.0 (CH, trityl) 132.1 (C of $C_{14}H_9$), 129.6 (CH, trityl), 129.4 (CH, trityl), 128.0 (CH, trityl), 127.5 (CH of $C_{14}H_9$) 127.4 (CH, trityl), 126.5 (CH, trityl), 125.4 (CH of $C_{14}H_9$), 125.3 (CH of $C_{14}H_9$), 122.3 (C of $C_{14}H_9$), 122.2 (CH of $C_{14}H_9$), 112.7 (CH, trityl) 112.6 (CH, trityl), 73.0 (CH_2O), 61.1 (CH_2 ethanolamide), 61.1 (CH_2 ethanolamide), 61.0 (CH_2 ethanolamide), 59.6 (CH_2 ethanolamide), 59.6 (CH_2 ethanolamide), 58.8 (CH_2 ethanolamide) 54.5 (OMe, trityl), 49.3 (CH_2 ethanolamide), 48.4 (CH_2 ethanolamide), 47.5 (CH_2 ethanolamide), 45.2, (CH_2 ethanolamide). MS (ES) *m/z*: 664 (M+Na). HRMS (ES) calcd for $C_{41}H_{39}NO_6Na$: 664.2675, found: 664.2705. IR (ν_{max}, solid state) cm^{-1}: 3280 (br, OH), (s, C=O amide).

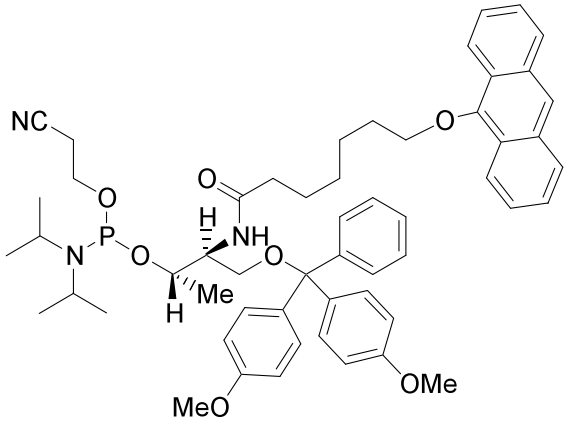
2-(Anthracen-9-yloxy)-*N*-(1-dimethoxytrityl-2-(2-cyanoethyl diisopropylamino phosphoramidite))-acetamide (5a).



2-(Anthracen-9-yloxy)-*N*-(1-dimethoxytrityl-2-hydroxy-ethyl)-acetamide (**4a**, 0.11 g, 0.17 mmol, 1.0 eq.) was dissolved in anhydrous dichloromethane (5 ml) and diisopropylamine (0.09 ml, 0.07 g, 0.51 mmol, 3.0 eq.) added ,followed by stirring for 15 minutes. 2-cyanoethyl-*N,N*-diisopropyl chlorophosphoramidite (0.04 ml, 0.04 g, 0.18 mmol, 1.1 eq.) was added dropwise to the solution followed by stirring at room temperature for 1 hour in the absence of light. Polymer bound benzyl alcohol (0.1 g) was then added and stirring was continued for a further 30 minutes. The solids were filtered off then the solvent removed under reduced pressure. To the residue was added ethyl acetate (5 ml) and 2 M aqueous sodium bicarbonate (5 ml) with agitation, and the aqueous layer run off. The organic layer was washed with brine (5 ml) and dried using magnesium sulfate. The ethyl acetate was removed under reduced pressure to leave a crude residue which was purified on activated basic alumina with methanol as the eluant to give the final

product , **5a** (0.05 g, 34%) after removal of solvent under reduced pressure. ^{31}P NMR (121 MHz, CD_3CN), δ ppm: 148.0 (s , phosphoramidite), 147.6 (s , phosphoramidite). *m/z* (ES): 864 $(M+Na)^+$. HRMS (ES) *m/z* calcd for $C_{50}H_{56}N_3O_7PNa$: 864.3754, found 864.3745.

7-(Anthracen-9-yloxy)-N-((2*S*,3*S*)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl)heptanamide (5b).



2-(Anthracen-9-yloxy)-*N*-(1-dimethoxytrityl-2-hydroxy-ethyl)-acetamide (**4b**, 0.30 g, 0.42 mmol, 1.0 eq.) was dissolved in anhydrous dichloromethane (15 ml) and diisopropylamine (0.21 ml, 0.17 g, 1.3 mmol, 3.0 eq.) added, followed by stirring for 15 minutes. 2-cyanoethyl-*N,N*-diisopropyl chlorophosphoramidite (0.10 ml, 0.10 g, 0.45 mmol, 1.1 eq.) was added dropwise to the solution followed by stirring at room temperature for 1 hour in the absence of light. Polymer bound benzyl alcohol (0.25 g) was then added and stirring was continued for a further 30 minutes. The solids were filtered off then the solvent removed under reduced pressure. To the residue was added ethyl acetate (10 ml) and 2 M aqueous sodium bicarbonate (10 ml) with

agitation, the aqueous layer was then run off. The organic layer was washed with brine (10 ml) and dried using magnesium sulfate. The ethyl acetate was removed under reduced pressure to leave a crude residue which was purified on activated basic alumina with methanol as the eluant to give the final product , **5b** (0.23 g, 59%) after removal of solvent under reduced pressure. ^{31}P NMR (121 MHz, CD_3CN), δ ppm: 147.6 (s , phosphoramidite), 146.9 (s , phosphoramidite) MS (ES) *m/z*: 935 (M+Na). HRMS (ES) calcd for $C_{55}H_{66}N_3O_7PNa$: 934.4536 , found: 934.4573.

General procedure for oligonucleotide synthesis.

Synthesis of probe and target sequences were performed using an automated ABI 394 DNA/RNA synthesiser running Oligonet 1.0 software. Final base deprotection was achieved by heating the oligonucleotides (released from the machine into solution containing ammonia), at 55 °C in a sealed vial for 6 hours.

The solvents were removed under reduced pressure then the residue re-dissolved in water (1 ml). The absorbance at 260 nm was measured by taking 10 µl of the oligonucleotide and mixing with 990 µl water, and then used to calculate oligonucleotide concentrations and hence yields. Preparative HPLC purification of the oligonucleotides was performed on a reversed phase column eluting a 0.1 M triethylammmonium acetate / acetonitrile gradient. Oligonucleotides were desalted on Nap-10 size

exclusion columns (GE Healthcare) to give the final products. Desalted probe sequences were characterised by electrospray mass spectrometry.

Probe sequences synthesised.

Designation	Sequence	Calc. RMM	*m/z* (obs.)
PrA	5'-TGGACT***C*5b*C***TCAATG-3'	4710.3	4711.0
PrB	5'-TGGACT***C*5b*G***TCAATG-3'	4750.3	4751.0
PrC	5'-TGGACT***C*5b*A***TCAATG-3'	4734.3	4735.0
PrD	5'-TGGACT***C*5b*T***TCAATG-3'	4725.3	4726.0
PrE	5'-TGGACT***G*5b*C***TCAATG-3'	4750.3	4751.0
PrF	5'-TGGACT***A*5b*C***TCAATG-3'	4734.3	4735.0
PrG	5'-TGGACT***T*5b*C***TCAATG-3'	4725.3	4727.0

Target sequences synthesised.

Target	Sequence
TarA1	3'-ACCTTGA*GAG*GAGTTAC-5'
TarA2	3'-ACCTTGA*G_G*GAGTTAC-5'
TarA3	3'-ACCTTGA*GTG*GAGTTAC-5'
TarA4	3'-ACCTTGA*GCG*GAGTTAC-5'
TarA5	3'-ACCTTGA*GGG*GAGTTAC-5'
TarB1	3'-ACCTTGA*GAC*GAGTTAC-5'
TarB2	3'-ACCTTGA*G_C*GAGTTAC-5'
TarC1	3'-ACCTTGA*GAT*GAGTTAC-5'
TarC2	3'-ACCTTGA*G_T*GAGTTAC-5'
TarD1	3'-ACCTTGA*GAA*GAGTTAC-5'
TarD2	3'-ACCTTGA*G_A*GAGTTAC-5'
TarE1	3'-ACCTTGA*CAG*GAGTTAC-5'
TarE2	3'-ACCTTGA*C_G*GAGTTAC-5'
TarF1	3'-ACCTTGA*TAG*GAGTTAC-5'
TarF2	3'-ACCTTGA*T_G*GAGTTAC-5'
TarG1	3'-ACCTTGA*AAG*GAGTTAC-5'
TarG2	3'-ACCTTGA*A_G*GAGTTAC-5'

References

(1) Isaacs, A.; Dainteth, J.; Martin, E. A. *Oxford Concise Science Dictionary*; 3rd ed.; Oxford University Press: Oxford, 1996.

(2) Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G. L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Edkins, S.; O'Meara, S.; Vastrik, I.; Schmidt, E. E.; Avis, T.; Barthorpe, S.; Bhamra, G.; Buck, G.; Choudhury, B.; Clements, J.; Cole, J.; Dicks, E.; Forbes, S.; Gray, K.; Halliday, K.; Harrison, R.; Hills, K.; Hinton, J.; Jenkinson, A.; Jones, D.; Menzies, A.; Mironenko, T.; Perry, J.; Raine, K.; Richardson, D.; Shepherd, R.; Small, A.; Tofts, C.; Varian, J.; Webb, T.; West, S.; Widaa, S.; Yates, A.; Cahill, D. P.; Louis, D. N.; Goldstraw, P.; Nicholson, A. G.; Brasseur, F.; Looijenga, L.; Weber, B. L.; Chiew, Y. E.; Defazio, A.; Greaves, M. F.; Green, A. R.; Campbell, P.; Birney, E.; Easton, D. F.; Chenevix-Trench, G.; Tan, M. H.; Khoo, S. K.; Teh, B. T.; Yuen, S. T.; Leung, S. Y.; Wooster, R.; Futreal, P. A.; Stratton, M. R. *et al. Nature* **2006**, *446*, 153.

(3) Chakravarti, A.; Little, P. *Nature* **2003**, *421*, 412.

(4) Walker, F. O. *Lancet* **2007**, *369*, 218.

(5) Nagalla, S. R.; Canick, J. A.; Jacob, T.; Schneider, K. A.; Reddy, A. P.; Thomas, A.; Dasari, S.; Lu, X.; Lapidus, J. A.; Lambert-Messerlian, G. M.; Gravett, M. G.; Roberts, C. T.; Luthy, D.; Malone, F. D.; D'Alton, M. E. *J. Proteome Res.* **2007**, *6*, 1245.

(6) Guo, W.; Yuan, J.; Dong, Q.; Wang, E. *J. Am. Chem. Soc.* **2009**, *132*, 932.

(7) Wolfe, M. S. *J. Med. Chem.* **2001**, *44*, 2039.

(8) Hochstrasser, D. *J. Proteome Res.* **2008**, *7*, 5071.

(9) Sugrue, M. F. *Journal of Medicinal Chemistry* **1997**, *40*, 2793.

(10) Samowitz, W. S. *Exp. Mol. Pathol.* **2008**, *85*, 64.

(11) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(12) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*; 7th ed.; W. H. Freeman: New York, 2007.

(13) Cane, D. E. *Chem. Rev.* **1990**, *90*, 1089.

(14) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. *Chem. Rev.* **2008**, *108*, 946.

(15) Pastorino, L.; Lu, K. P. *Eur. J. Pharm.* **2006**, *545*, 29.

(16) Haas, C.; Lemerre, C. A.; Capell, A.; Citron, M.; Seubert, P.; Schenk, D.; Lannfelt, L.; Selkoe, D. J. *Nature Medicine* **1995**, *1*, 1291.

(17) John, V.; Beck, J. P.; Bienkowski, M. J.; Sinha, S.; Heinrikson, R. L. *J. Med. Chem.* **2003**, *46*, 4625.

(18) Nakatani, K. *ChemBioChem* **2004**, *5*, 1623.

(19) Whibley, C.; Pharoah, P. D. P.; Hollstein, M. *Nat. Rev. Cancer* **2009**, *9*, 95.

(20) Fasching, P. A.; Gayther, S.; Pearce, L.; Schildkraut, J. M.; Goode, E.; Thiel, F.; Chenevix-Trench, G.; Chang-Claude, J.; Wang-Gohrke, S.; Ramus, S.; Pharoah, P.; Berchuck, A. *Mol. Oncol.* **2009**, *3*, 171.

(21) Suh, Y.; Vijg, J. *Mutat. Res.* **2005**, *573*, 41.

(22) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.;

LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M. *Nature* **2001**, *409*, 860.

(23) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463.

(24) Beaucage, S. L.; Caruthers, M. H. *Tett. Lett.* **1981**, *22*, 1859.

(25) Duprey, J.-L.; Tucker, J. H. R. *Unpublished Data* **2009**.

(26) Atkins, P. W. *Physical Chemistry*; 6th ed.; Oxford University Press: Oxford, 1998.

(27) Ranasinghe, R. T.; Brown, T. *Chem. Commun.* **2005**, 5487.

(28) Weisbrod, S. H.; Marx, A. *Chem. Commun.* **2008**, 5675.

(29) Venkatesan, N.; Seo, Y. J.; Kim, B. H. *Chem. Soc. Rev.* **2008**, *37*, 648.

(30) Umemoto, T.; Hrdlicka, P. J.; Babu, B. R.; Wengel, J. *ChemBioChem* **2007**, *8*, 2240.

(31) Yamana, K.; Nunota, K.; Nakano, H.; Sangen, O. *Tett. Lett.* **1994**, *35*, 2555.

(32) (a) de Silva, A. P.; Gunaratne, H. Q. N.; Gunnlaugsson, T.; Huxley, A. J. M.; McCoy, C. P.; Rademacher, J. T.; Rice, T. E. *Chem. Rev.* **1997**, *97*, 1515(b) Kalyanasundaram, K.; Thomas, J. K. *J. Phys. Chem.* **1977**, *81*, 2176(c) Okamoto, A.; Kanatani, K.; Saito, I. *J. Am. Chem. Soc.* **2004**, *126*, 4820.

(33) Balakin, K. V.; Korshun, V. A.; Mikhalev, II; Maleev, G. V.; Malakhov, A. D.; Prokhorenko, I. A.; Berlin, Y. A. *Biosens. Bioelectron.* **1999**, *14*, 597.

(34) Saito, Y.; Miyauchi, Y.; Okamoto, A.; Saito, I. *Chem. Commun.* **2004**, 1704.

(35) Saito, Y.; Hanawa, K.; Hayashi, K.; Motegi, K.; Okaoto, A.; Saito, I. *Nucleic Acids Symp Ser (Oxf)* **2005**, 153.

(36) Okamoto, A.; Tainaka, K.; Ochi, Y.; Kanatani, K.; Saito, I. *Mol. Biosyst.* **2006**, *2*, 122.

(37) Saiki, R. K.; Gelfland, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. *Science* **1988**, *239*, 487.

(38) Wanninger-Weiss, C.; Valis, L.; Wagenknecht, H. A. *Bioorg. Med. Chem.* **2008**, *16*, 100.

(39) (a) Sekhar Bag, S.; Saito, Y.; Hanawa, K.; Kodate, S.; Suzuka, I.; Saito, I. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 6338(b) Saito, Y.; Motegi, K.; Sekhar Bag, S.; Saito, I. *Biorg. Med. Chem.* **2008**, *16*, 107.

(40) Xiao, Q.; Ranasinghe, R. T.; Tang, A. M. P.; Brown, T. *Tetrahedron* **2007**, *63*, 3483.

(41) Moran, N.; Bassani, D. M.; Desvergne, J. P.; Keiper, S.; Lowden, P. A. S.; Vyle, J. S.; Tucker, J. H. R. *Chem. Commun.* **2006**, 5003.

(42) Asseline, U.; Chassignol, M.; Aubert, Y.; Roig, V. *Org. Biomol. Chem.* **2006**, 1949.

(43) Kramer, B.; Kramer, W.; Fritz , H. J. *Cell* **1984**, *38*, 879.

(44) Srivatsan, S. G.; Weizman, H.; Tor, Y. *Org. Biomol. Chem.* **2008**, *6*, 1334.

(45) Cekan, P.; Sigurdsson, S. T. *Chem. Commun.* **2008**, 3393.

(46) Williams, D. H.; Fleming, I. *Spectroscopic Methods in Organic Chemistry, 5th Ed.*; McGraw-Hill: New York, 1995.

(47) Gasparro, F. P.; Kolodny, N. H. *J. Chem. Ed.* **1977**, *54*, 258.

(48) Vu, H.; Majlessi, M.; Adelpour, D.; Russel, J. *Tett. Lett.* **2008**, *50*, 737.

(49) Fukui, K.; Iwane, K.; Shimidzu, T.; Tanaka, K. **1996**, *37*, 4983.

(50) Manoharan, M.; Tivel, K. L.; Zhao, M.; Nafisi, K.; Netzel, T. L. *J. Phys. Chem.* **1995**, *99*, 17461.

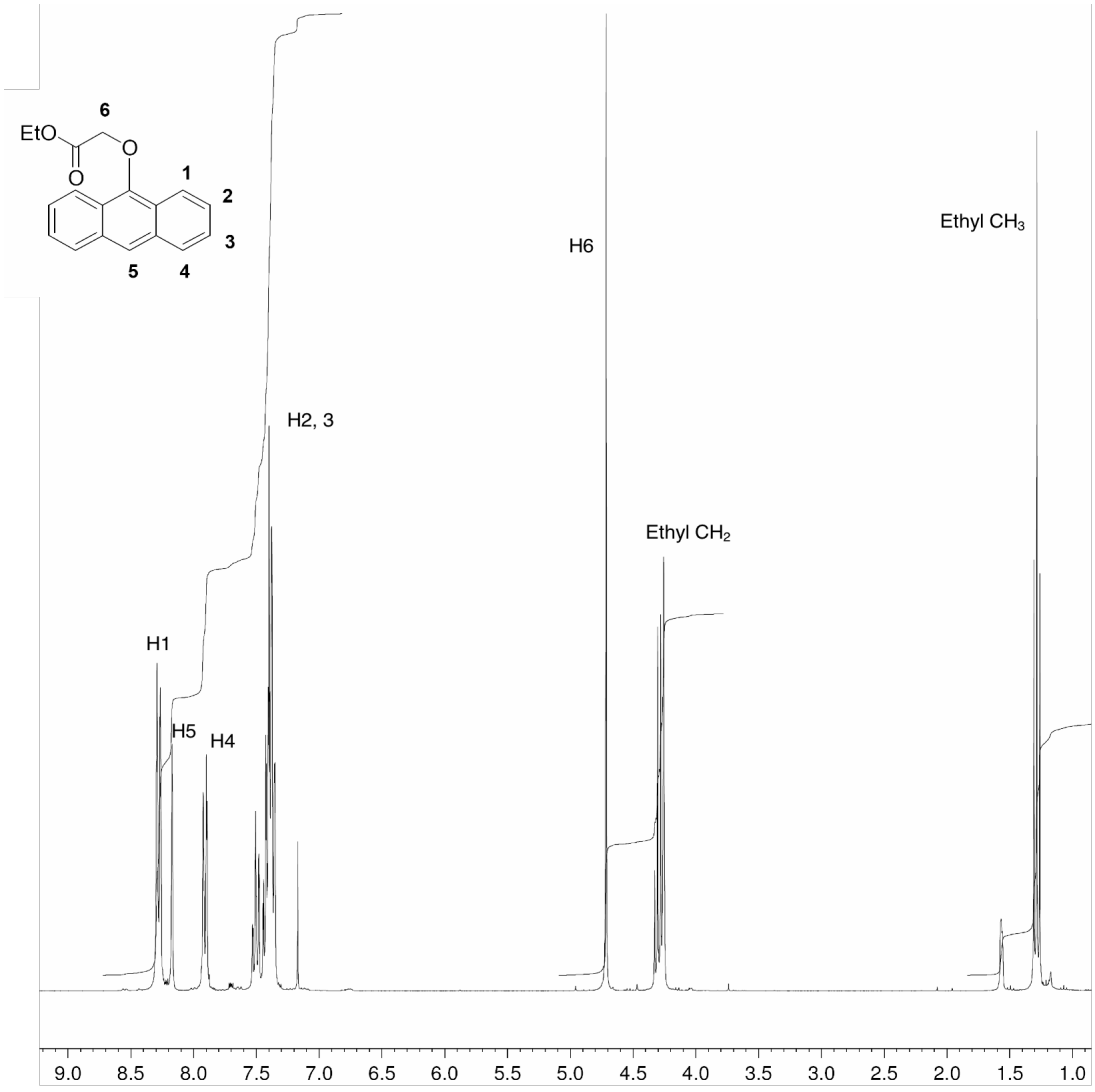
(51) Babendure, J. R.; Adams, S. R.; Tsien, R. Y. *J. Am. Chem. Soc.* **2003**, *125*, 14716.

(52) Moran, N. Ph.D. Thesis, Exeter University, 2006.

Appendix

Product: 1.

^{1}H NMR (300 MHz, $CDCl_3$).

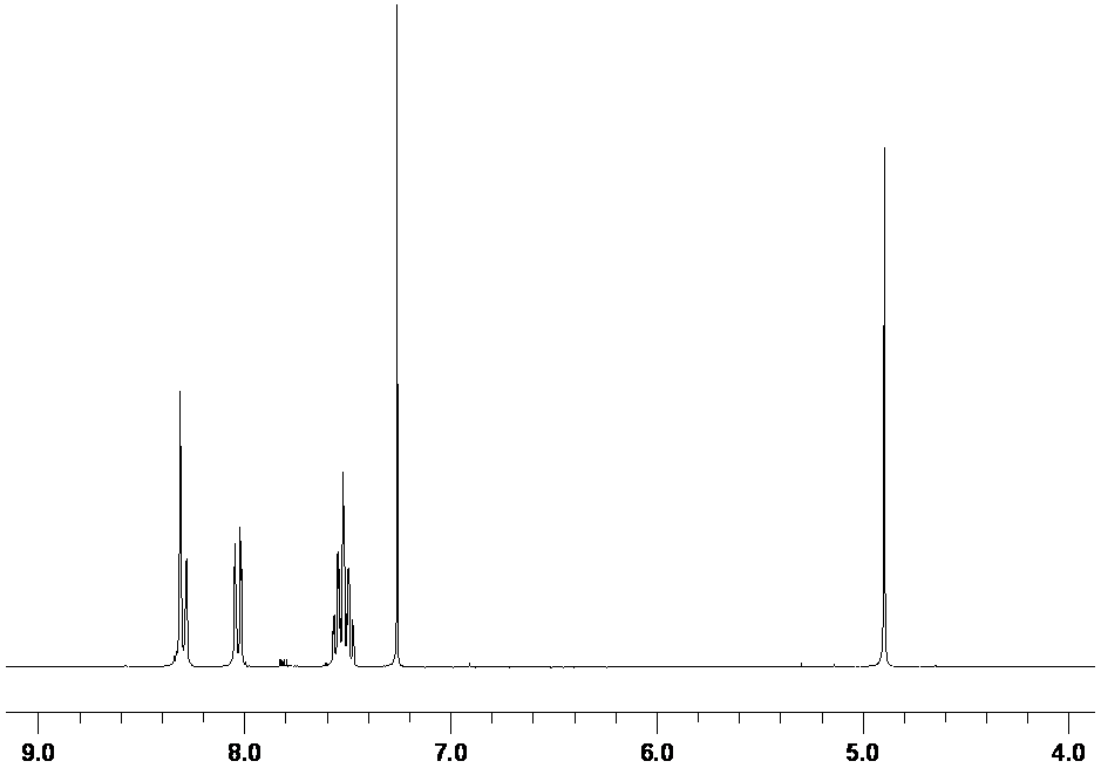


δ / ppm

Product: 2.

^{1}H NMR (300 MHz, CD_3CN / $CDCl_3$).

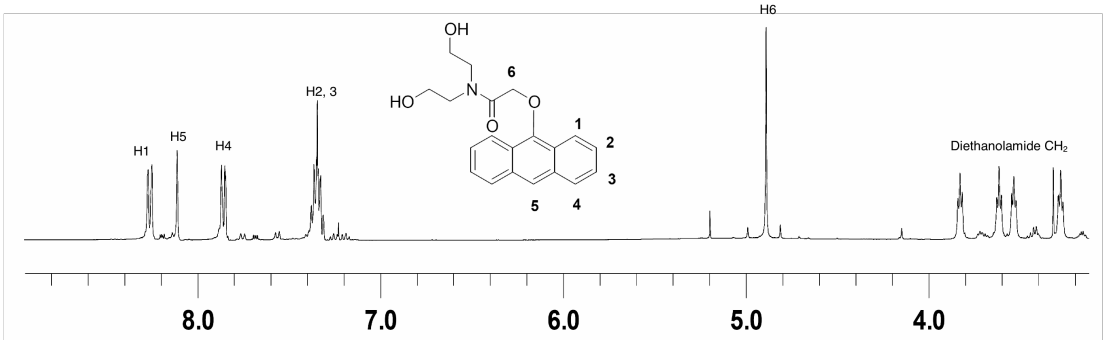
Aromatic assignments identical to product **1**.



δ / ppm

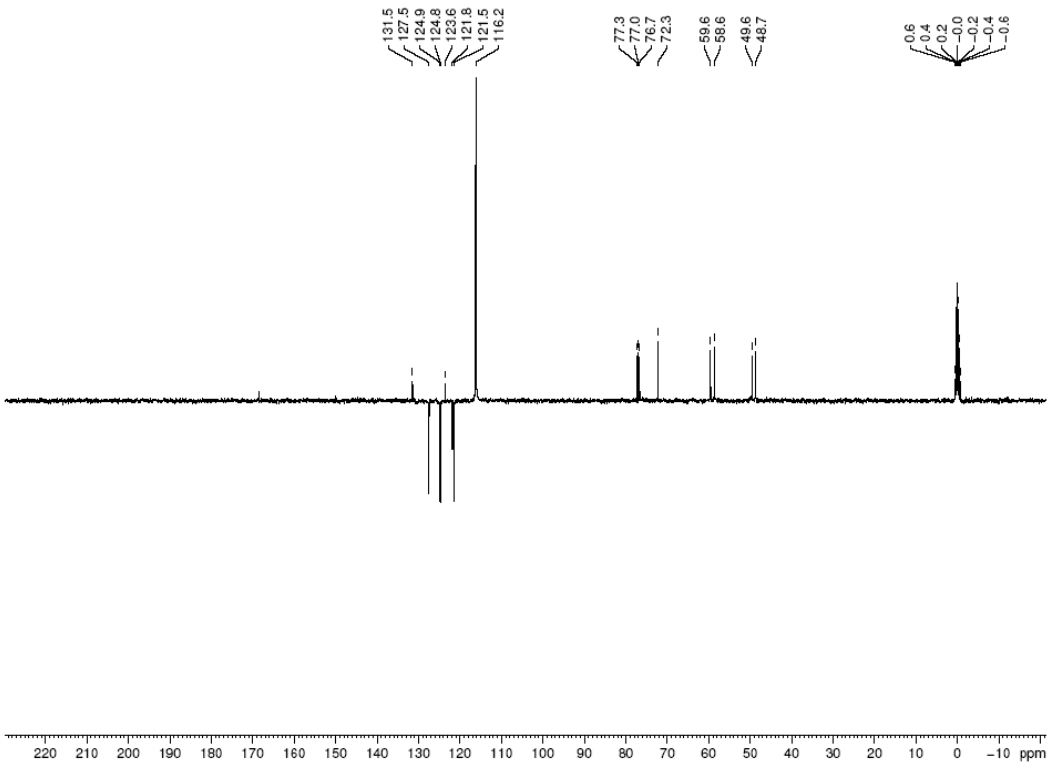
Product: 3.

^{1}H NMR (300 MHz, CD_3CN / $CDCl_3$).



δ / ppm

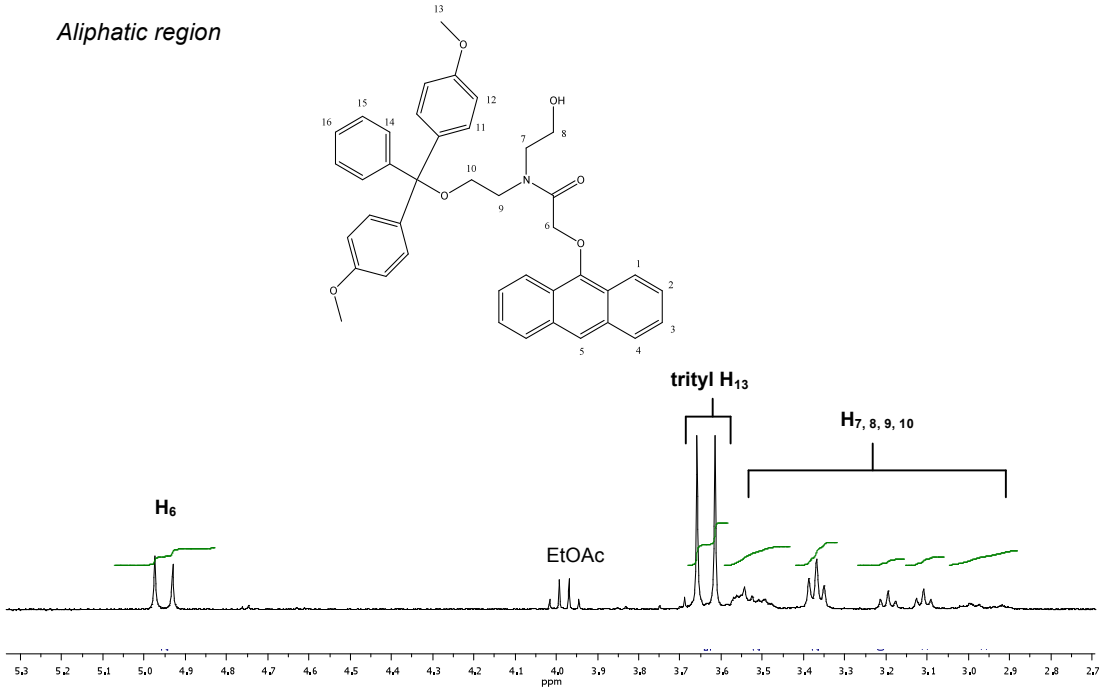
^{13}C PENDANT NMR (100 MHz, $CD_3CN/CDCl_3$).



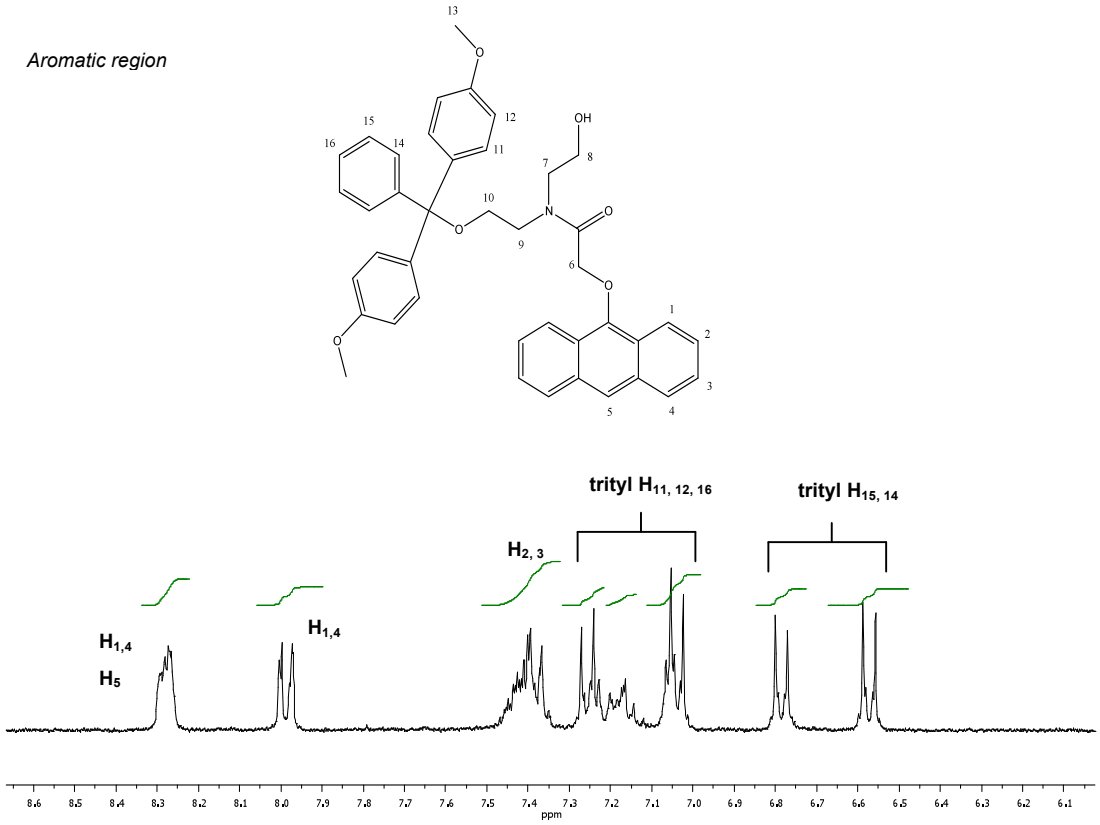
Product: 4a.

^{1}H NMR (300 MHz, CD_3CN).

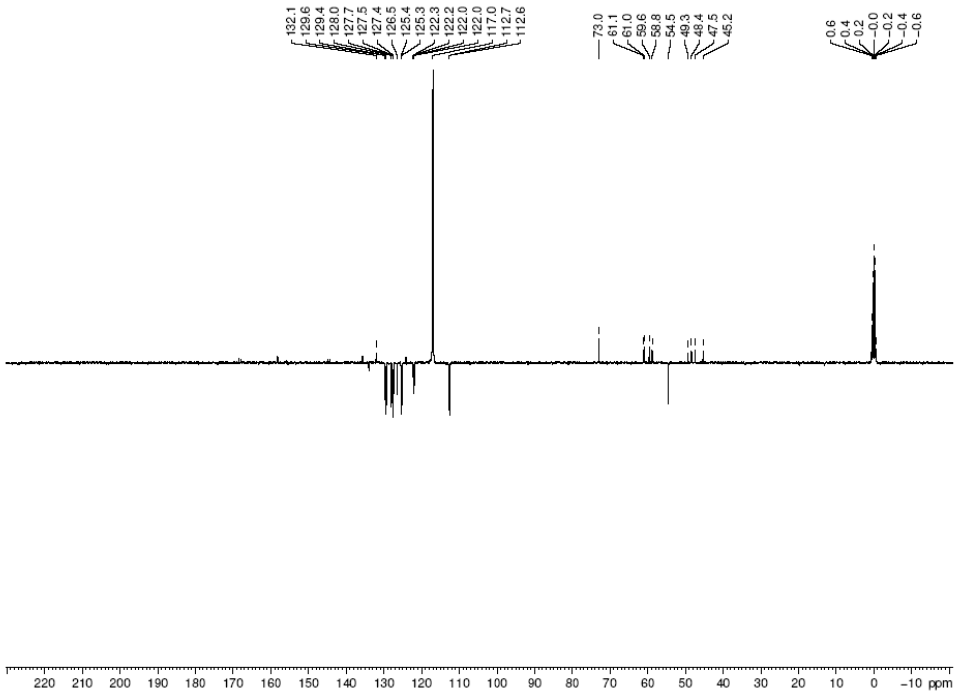
Aliphatic region



Aromatic region

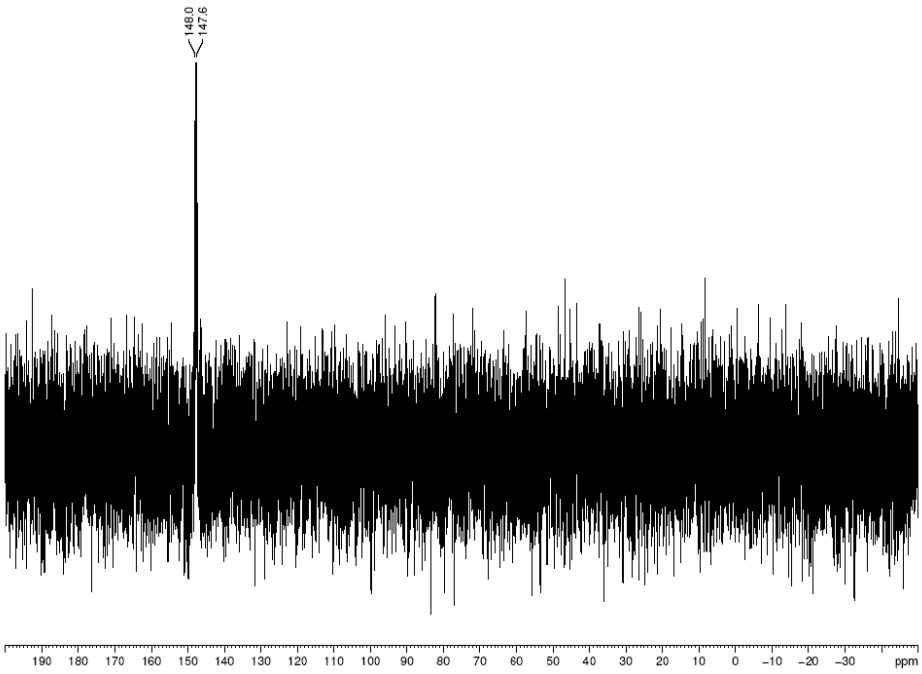


^{13}C PENDANT NMR (100 MHz, CD_3CN).



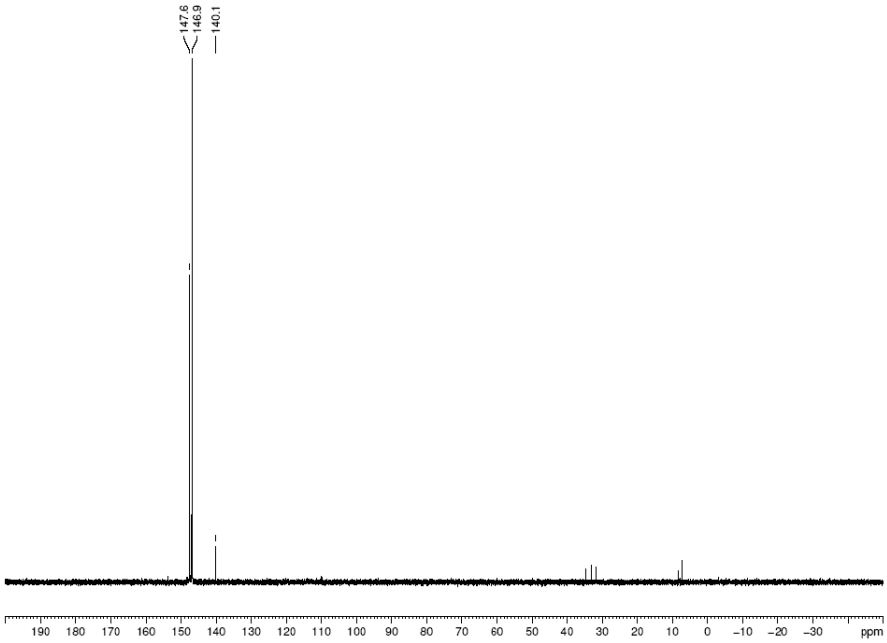
Product: 5a.

^{31} P NMR (121 MHz, CD_3CN).



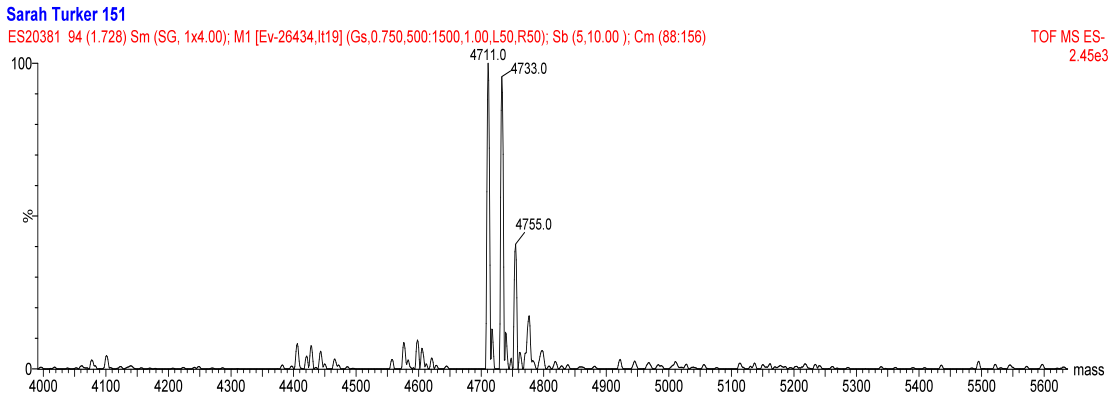
Product: 5b.

^{31} P NMR (121 MHz, CD_3CN).



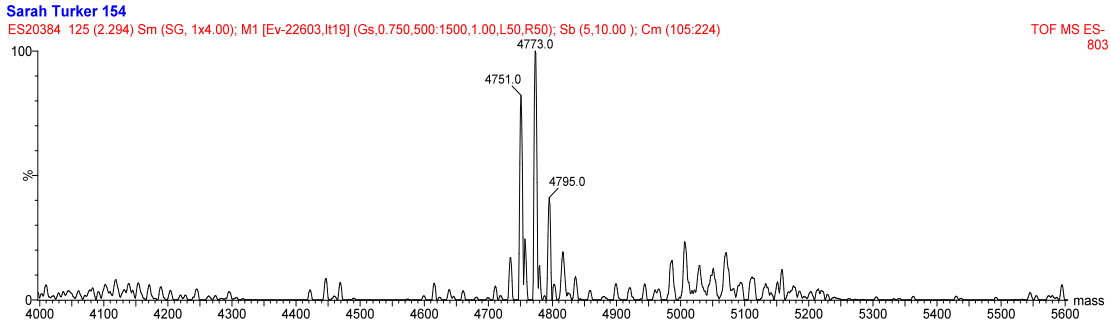
Probe A

ES (-)



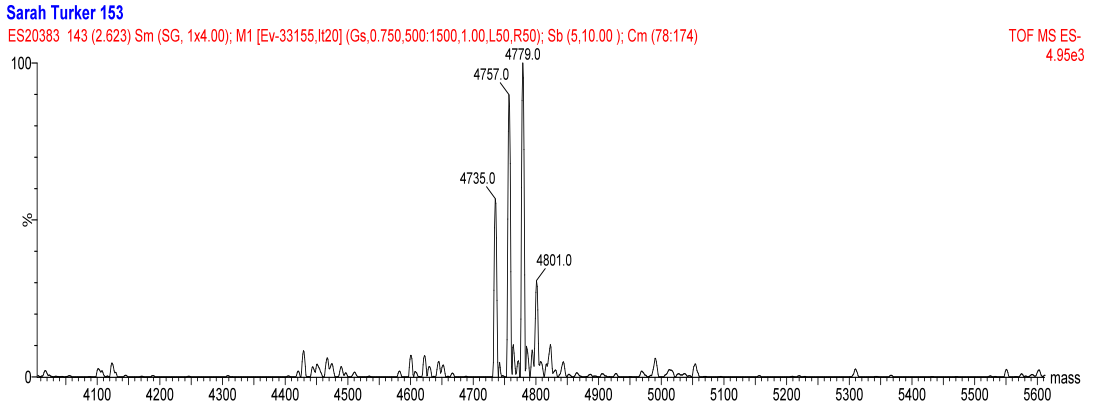
Probe B

ES (-)



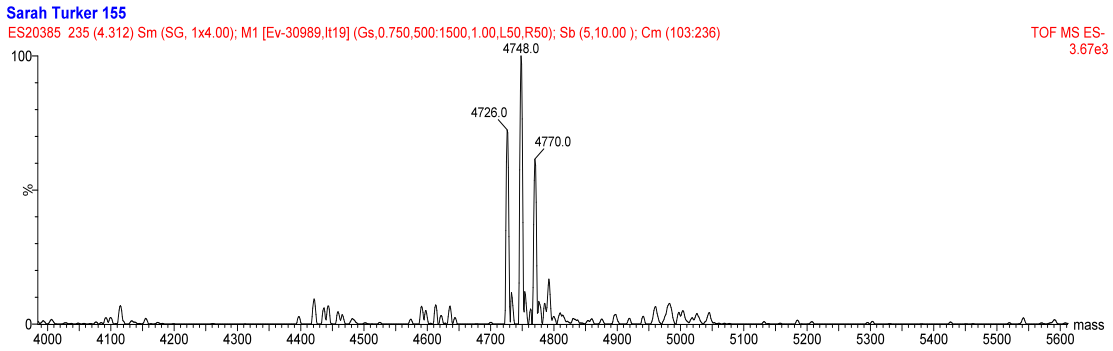
Probe C

ES (-)



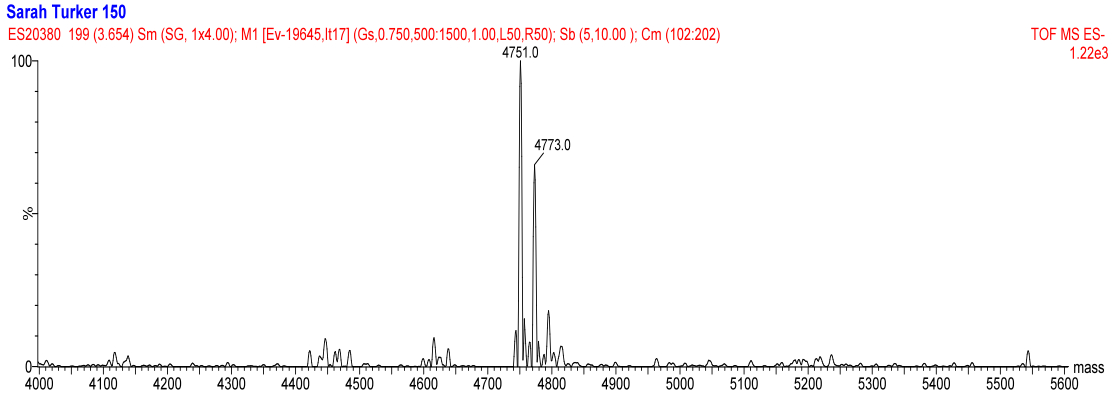
Probe D

ES (-)



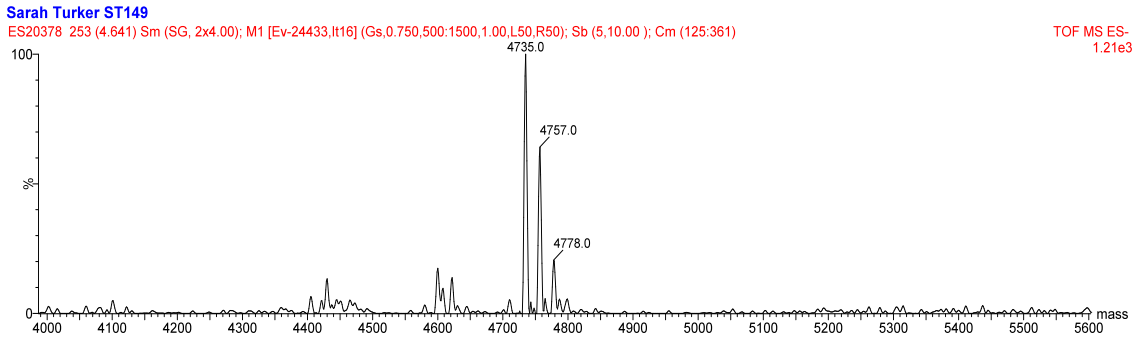
Probe E

ES (-)



Probe F

ES (-)



Probe G

ES (-)