Anthracene Tagged Biomolecules for DNA Binding

Gemma Anne Bullen

UNIVERSITY OF BIRMINGHAM

A thesis submitted to the University of Birmingham for the degree of Doctor of Philosophy

School of Chemistry
College of Engineering and Physical Sciences
The University of Birmingham
August 2015

Abstract

Within this thesis, the use of anthracene to perform various applications within biomolecules is assessed. Anthracene displays two interesting photo properties which make it an appealing molecule for incorporation; fluorescence and photodimerisation.

Within the first project presented herein, anthracene is incorporated into an oligonucleotide to create a probe to sense for single nucleotide polymorphisms (SNPs). The intensity of anthracene fluorescence is shown to depend upon the base opposite it on a complementary strand of DNA. This intensity displays a linear relationship with respect to the amount of each base present at the SNP site, allowing for a quantitative assay to be developed. Further to this, the fluorescence output was also shown to allow for the identification of the 8-oxoG base modification.

The unprecedented use of anthracene photodimerisation within a peptide is also explored. It is shown that, within a model system and at low biologically relevant micromolar concentrations, the photodimerisation only occurs when the peptide units are bound to target DNA, which preorganises the units for photodimerisation. Once formed, the photodimer species shows enhanced binding affinity for target DNA.

In addition to this, the photodimerisation properties of anthracene were utilised within oligonucleotides, to control the duplex formation and secondary structure of sequences. It was shown that anthracene could successfully form photodimers when incorporated into oligonucleotides. When this was performed between two strands of a three stranded toe-hold system, it resulted in a shift in the equilibrium and an effective release of an unmodified strand. This process was shown to be both reversible and repeatable, allowing for cycling of the system. When two anthracene moieties were incorporated into a quadruplex forming sequence, anthracene photodimerisation was achieved between the two units within the same strand. This action was shown to alter the folding of the structure, and initial studies towards establishing the effect this has on binding to a target protein, showed promise.

Acknowledgments

There are many people I need to thank for their help, support and guidance within my studies. Firstly, I would like to thank my supervisors, Anna Peacock and Jim Tucker. Jim, your optimism knows no bounds and you have always been able to put a positive spin on results I was disappointed with, re-motivating me to get back in the lab. Your ideas have been invaluable and I appreciate all the time and effort you have put in to help me with my projects. Anna, thank you for all the support over the last five years. Your passion for the area is addictive and it definitely rubbed off on me in my master's year and contributed to me staying in research. Your opinions and criticisms have been priceless and I have learnt a lot from you.

Thanks also to the University of Birmingham and the fantastic staff and students I have had the pleasure of working alongside. A special mention is needed for the analytical staff, past and present, in particular Chi Tsang and Peter Ashton for their continued hard work and dedication. To the support staff, Helen, Stuart, Lynn and Bernard; always friendly, smiley faces to offer an ear to talk to.

The Peacock and Tucker groups have been a pleasure to work in. I am grateful to the past members that taught me lots: John, Jack, Manu, Xun, Andrea, Antoine and Pete; the current members that have been a pleasure to work with: Louise, Sarah, Holly, Haydn, David, Jon and Francia; and those that left, but missed us so much they came back: Jean-Louis and Huy.

I have met many wonderful people along the way that deserve a special note: Rich and Ziggy, thanks for your kindness and valued friendship; Rosie, always there to listen to my moans and offer fantastic advice; Haydn and Maddy, my drinking buddies, always on hand for a Prosecco; Mike and Danielle (and of course Millie), your laid back attitude sometimes scares me but I wish I could be as relaxed as you. And yes, I will look after the cats; James and Jen (and more recently Poppy), you have been great friends since undergraduate and I truly value our friendship. I look forward to being there for Poppy as she grows up and James (as much as I moan) I will miss making you a tea in the morning just so I can have a ten minute catch up with my buddy!

To the Berwick's and extended family, you have welcomed me as one of your own. Your support has been wonderful, always there at the end of the phone and helping in any way you can. I hope to be able to repay you someday.

To my family, Sarah, Chris, Olivia, Scarlett and Mum. I am grateful for your encouragement (despite not understanding what I do!) and I would not be where I am today if it was not for you. I hope I have made you proud.

And finally thank you to Matt. We entered this journey as strangers 8 years ago but we are leaving together. Thanks for putting up with me and always being there. I am eternally grateful for your undying support and words of wisdom! I look forward to planning our future together and cannot wait to be your wife.

Table of Contents

Abbreviations

Abbreviations

2D	2-Dimensional
3D	3-Dimensional
APS	Ammonium Persulphate
ATP	Adenosine Triphosphate
BDF	Base Discriminating Fluorophore
CD	Circular Dichroism
CPG	Controlled Pore Glass
DCM	Dichloromethane
DIPEA	N, N-Diisopropylethylamine
DMT	Dimethoxytrityl
DNA	Deoxyribose Nucleic Acid
dsDNA	Double Stranded DNA
DTNB	5,5'-Dithiobis-(2-Nitrobenzoic Acid)
ELISA	Enzyme-Linked Immunosorbent Assay
EMSA	Electrophoretic Mobility Shift Assay
ESI	Electrospray Ionisation
ETT	5-(Ethylthio)-1H-Tetrazole
FcNA	Ferrocene Nucleic Acid
Fmoc	Fluorenylmethyloxycarbonyl
FRET	Förster Resonance Energy Transfer
GC	Gas Chromatography
H-H	Head-to-Head
H-T	Head-to-Tail
HBTU	N,N,N′,N′-Tetramethyl-O-(1H-Benzotriazol-1-yl)Uronium Hexafluorophosphate
HOMO	Highest Occupied Molecular Orbital
HPLC	High Performance Liquid Chromatography
LC	Liquid Chromatography
LNA	Locked Nucleic Acid
MALDI	Matrix Assisted Laser Desorption Ionisation
MBHA	4-Methylbenzhydrylamine
mRNA	Messenger Ribose Nucleic Acid

MS	Mass Spectrometry
Mtt	Methyltrityl
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PET	Photo-Induced Electron Transfer
PICUP	Photo-Induced Crosslinking of Unmodified Proteins
PNA	Peptide Nucleic Acid
PYP	Photo-Activated Yellow Protein
RNA	Ribose Nucleic Acid
RP-HPLC	Reversed Phase High Performance Liquid Chromatography
Rt	Retention Time
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SMCC	Succinimidyl Trans-4-(Maleimidylmethyl)Cyclohexane-1-Carboxylate
SNP	Single Nucleotide Polymorphism
ssDNA	Single Stranded DNA
T_a	Annealing Temperature
TAMRA	Tetramethylrhodamine
TBA	Thrombin Binding Aptamer
TBE	Tris/Borate/EDTA
TEAA	Triethylammonium Acetate
TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
T_m	Melting Temperature
TOF	Time of Flight
tRNA	Transfer Ribose Nucleic Acid
UAA	Unnatural Amino Acid
UV	Ultraviolet
UV-Vis	Ultraviolet-Visible
VT	Variable Temperature

DNA Bases

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil
8-oxoG	8-oxoguanine
5-MeC	5-methylcytosine

Amino Acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutaimine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
U	Sec	Selenocysteine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Chapter 1

Introduction - The Use of Anthracene within Biomolecules

1.1 Introduction

Within this thesis the use of anthracene within biomolecules is assessed, in particular within DNA and peptides. This chapter details the biological relevance of DNA and proteins, along with modified derivatives which have been utilised for many applications. The uses of anthracene are also explored, focussing on the current applications of anthracene fluorescence and photodimerisation within the literature, particularly within biomolecules.

1.2 DNA

In the last 60 years, since the first proposed correct model of the structure of deoxyribose nucleic acid (DNA)[1] and the determination of its function,[2] knowledge of the area has greatly expanded. DNA is now understood to be the carrier of genetic information for all living organisms and studies on the molecule continue due to its potential as a data carrier and a target for therapeutics.

1.2.1 DNA Structure

DNA is comprised of a sugar-phosphate backbone, from which one of four naturally occurring bases protrude. These bases form hydrogen bonds to bases in a complementary strand, whilst π-stacking with neighbouring bases, all contributing to the stability of the complex, Figure 1.1.

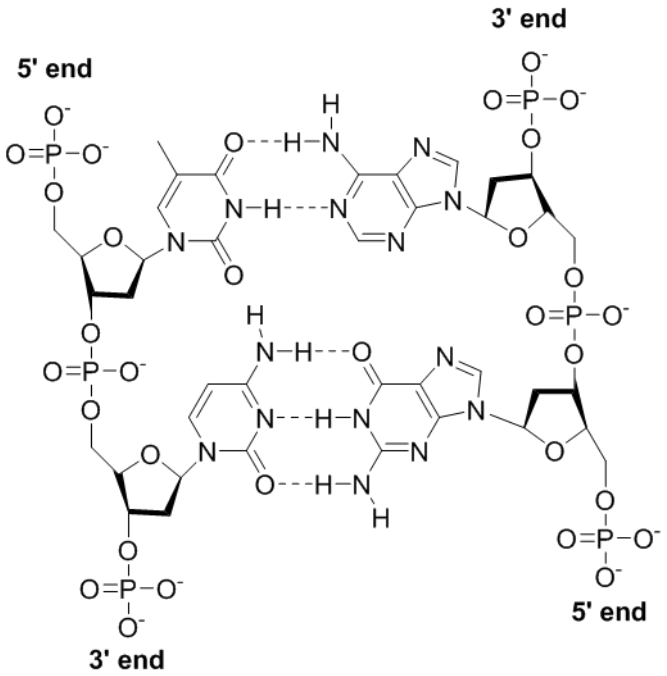


Figure 1.1 DNA structure showing the sugar-phosphate backbone with thymine (top left) hydrogen bonding to adenine (top right) and guanine (bottom right) bonded to cytosine (bottom left).

Each unit within the DNA structure is linked to the next via the 3′ carbon on the deoxyribose sugar, to the 5′ carbon on the next unit *via* a phosphate group. Each base pair consists of a purine (adenine, A and guanine, G) and a pyrimidine (thymine, T and cytosine, C). The A-T pairing is held together by two hydrogen bonds whilst the G-C pairing has three (see Figure 1.1). Due to the specific nature of the base pairings, hybridisation (the self-assembly of two strands of DNA into the duplex structure) usually occurs with a sequence with the correct base pairing, the complementary strand. The double stranded DNA (dsDNA) is formed in an antiparallel nature, and folds into a helix so as to maximise the overlap of the bases and increase π-stacking. This also increases the distance between the negatively charged phosphate groups, which experience electrostatic repulsion. This effect is also minimised by cations which act as counter ions.[3]

Double stranded DNA is known to fold into 3 main forms, A-, B- and Z-DNA, known as the secondary structure, Figure 1.2. Of these the most common is B-DNA, which is made of a right-handed helix with approximately 10 base pairs per turn, a diameter of 24 Å and a pitch of 35.5 Å. Within this structure, each phosphate group is 7 Å apart. In the A-DNA form the phosphate groups are closer together (5.9 Å), with the base pairs being tilted with respect to the backbone. The diameter of the helix has increased to 25.5 Å, and there are approximately 11 base pairs per turn. Z-DNA is considerably different from A- and B-DNA, forming a left-handed helix, and is typically found in regions of alternating purine-pyrimidine bases where the purine residues adopt a *syn*

configuration. This structure has 12 base pairs per turn, a diameter of 18.4 Å and a rise per base of 3.63 Å.[4–6]

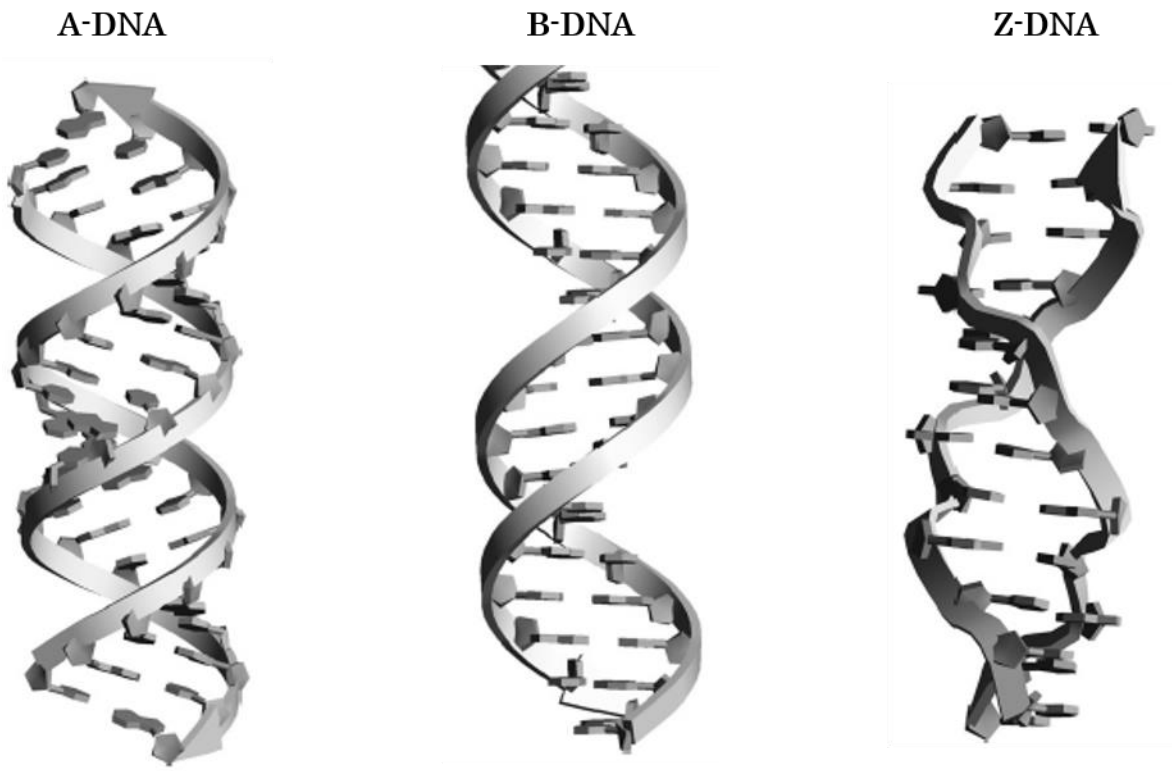


Figure 1.2 Structures of the common forms of DNA. A-DNA (left) is a more compact structure of the most common B-DNA structure (middle). B-DNA is a right-handed helix with base pairs perpendicular to the axis of the helix. Finally Z-DNA (right) is a left-handed helix with the most base pairs per turn. Figure adapted from reference 4.

Single stranded DNA can also fold within itself, to form various conformations such as hairpins. In addition triplex and quadruplex structures of multiple strands are also known. Of particular note are junctions, which form between several strands of DNA; G-Quadruplexes, a four stranded structure which forms in guanine rich sequences and i-motifs, another four stranded structure which forms in cytosine rich sequences.[7]

1.2.2 Transcription and Translation

DNA acts as the genetic blueprint, coding for all proteins and smaller signalling peptides within the body. In order for the proteins to be made from the information contained within DNA, it must first be copied into ribonucleic acid (RNA). RNA is a polymer similar to DNA that consists of ribose sugar units rather than deoxyribose as found in DNA. The bases attached to the sugar units are identical to those found within DNA, except for thymine which is replaced by uridine, which lacks the methyl group on the pyrimidine, Figure 1.3. The two main types of RNA which will be referred to within this work are messenger RNA (mRNA) and transfer RNA (tRNA).

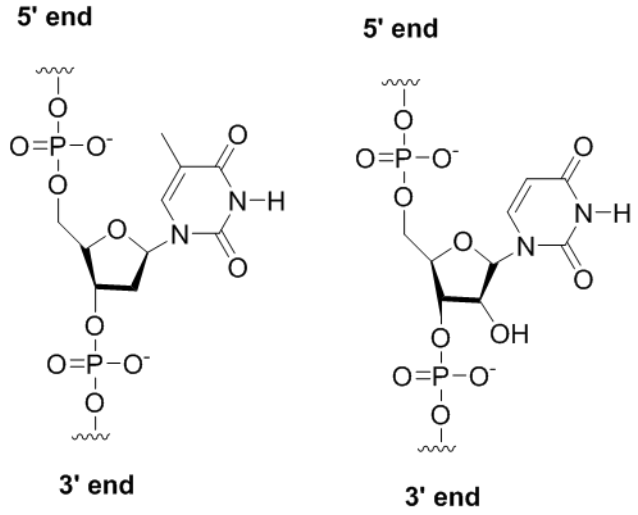


Figure 1.3 Comparison of the structure of DNA and RNA. Within RNA the deoxyribose sugar of DNA is replaced with ribose. All the bases are the same between the two systems apart from thymine (shown in the structure on the left) which is replaced with uracil (right) which lacks the methyl group on the pyrimidine.

The process of DNA being copied into RNA, is known as transcription, the rate of which is controlled by proteins known as transcription factors which can either promote, or inhibit, the transcription process by binding to the DNA strand. When an RNA polymerase binds to the DNA, it initially separates the DNA duplex to create a region of single stranded DNA (ssDNA). Each base within the DNA strand is then paired with the complementary RNA base analogue. As the base interactions are formed, the sugar-phosphate backbone is made by joining the groups with those of the preceding unit. The hydrogen bonds to the complementary DNA are then broken by a helicase enzyme, to produce the single stranded RNA (ssRNA) which can be translated into proteins, Figure 1.4. This RNA is referred to as messenger RNA (mRNA).[8,9]

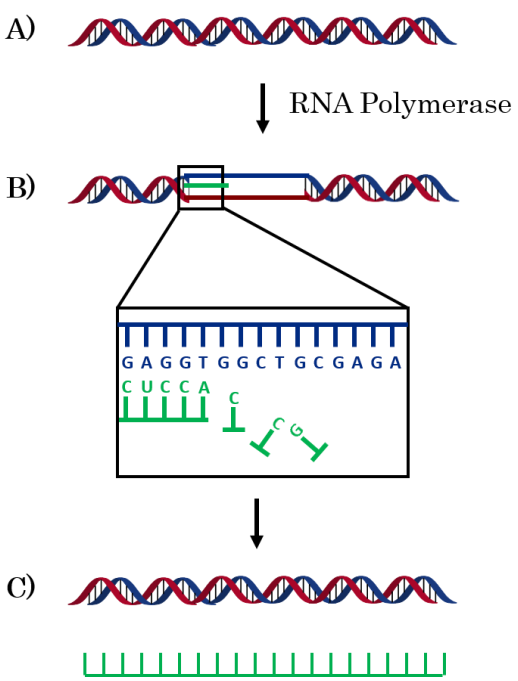


Figure 1.4 Schematic diagram of the transcription process. A) The region of dsDNA to be transcribed is identified and RNA polymerase binds to it. B) The dsDNA is dissociated to expose a region of ssDNA which can be paired with complementary RNA bases. As the base pair is formed between the DNA and the new RNA base, the sugar-phosphate backbone is made. C) Once the process is completed the new mRNA strand is released and the DNA refolds into a double stranded helix.

The RNA formed in the translation process is read by the ribosome, an organelle within the cell which performs protein synthesis. Each set of three bases, known as a codon, details an instruction to either incorporate a specific amino acid or to stop the synthesis. A transfer RNA (tRNA) molecule, consisting of a short oligonucleotide sequence with an anticodon region (complementary to the codon found in mRNA), binds to the codon on the mRNA. The tRNA molecule has the selected amino acid bound to the other end, and hence aligns it within the sequence, so it can be joined with the preceding amino acid, Figure 1.5. The ribosome then moves along the sequence, connecting the next codon in the chain.[4,9]

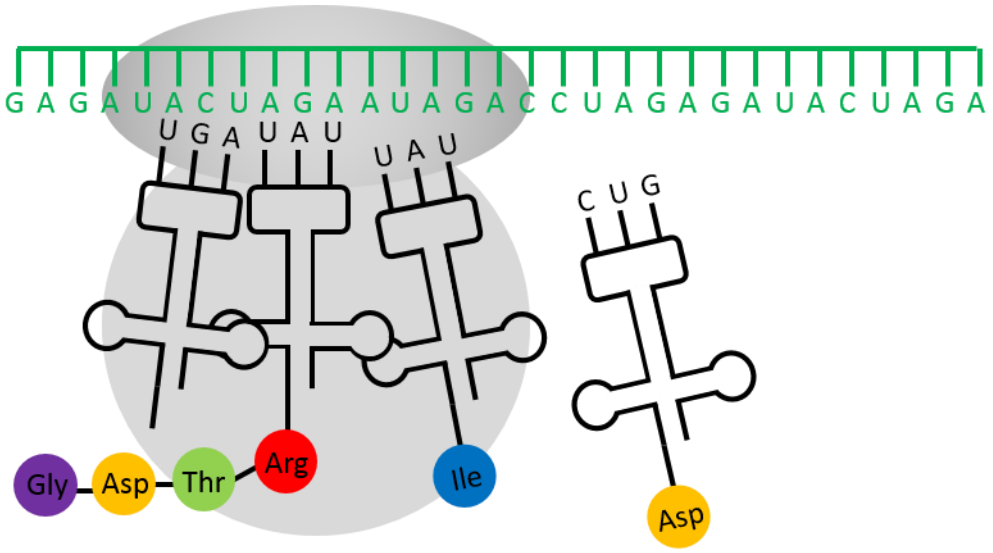


Figure 1.5 A schematic representation of the translation process. The mRNA (green) is read by the ribosome (grey). tRNA molecules are matched with each codon by forming complementary base pairs. The amino acids attached to them are then synthesised into the protein by the ribosome.

1.3 Proteins

Proteins are polymers formed from building blocks called amino acids. There are 21 naturally occurring amino acids, all comprising of a carboxylic acid, an amine and a side chain. All naturally occurring amino acids are referred to as having L stereochemistry (meaning polarised light passed through a solution containing them, is rotated in the same direction as a solution containing L-glyceraldehyde). Using standard chirality labelling, the chiral amino acids are shown to have S chirality, with the exceptions of cysteine and selenocysteine, which are R. Amino acids are linked together through the creation of a peptide bond between the carboxylic acid of one amino acid with the amine of another, Figure 1.6. The side chains of amino acids contain a variety of functional groups shown in Figure 1.7.[10]

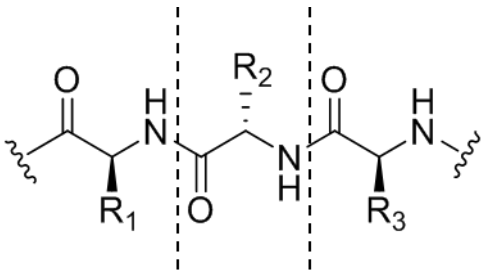


Figure 1.6 The structure of peptide chains. Amino acid units (distinguished by the dashed lines) are linked *via* the amine of one, to the carboxylic acid of the next, to create a peptide bond. Each amino acid has a side chain (R) which varies between the amino acids.

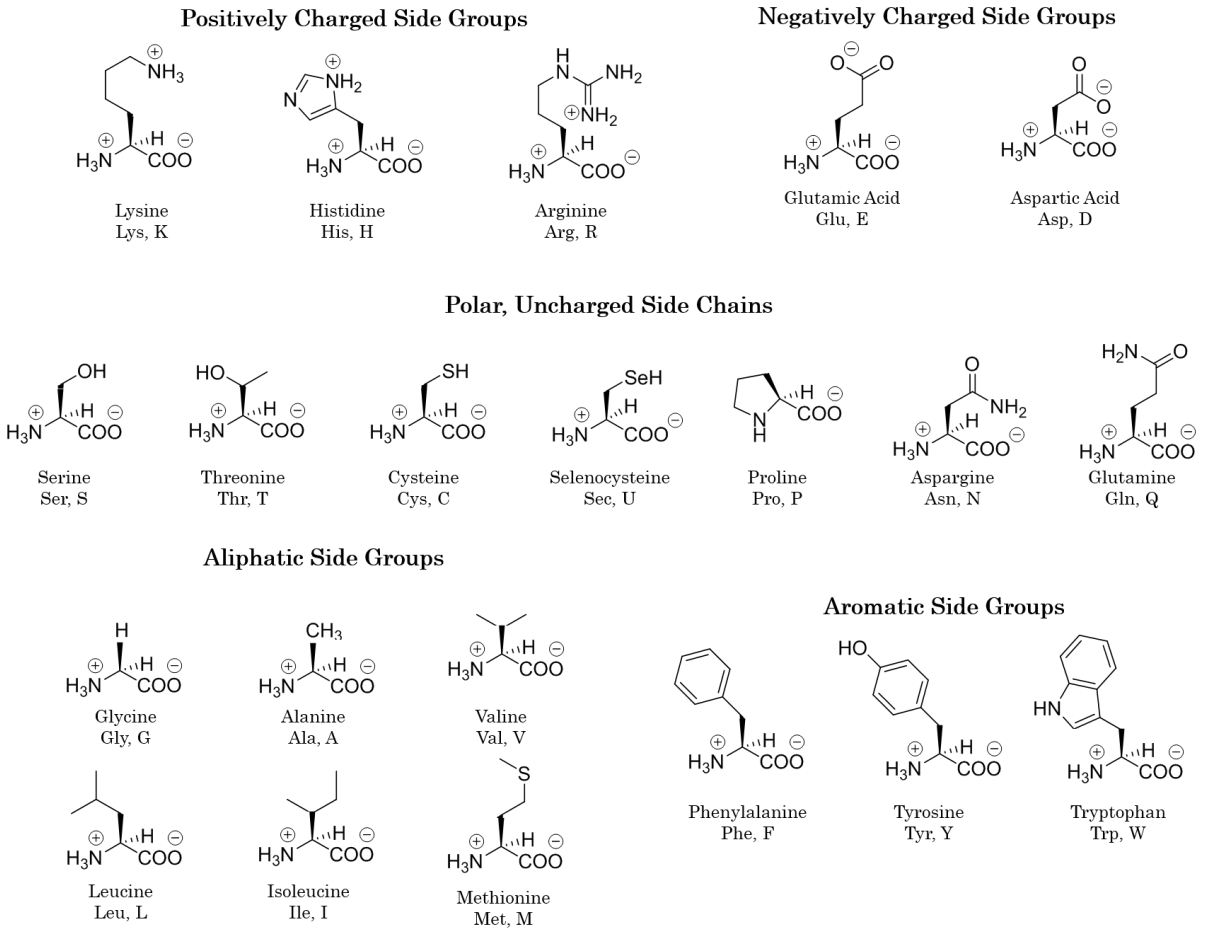


Figure 1.7 Structures of the 21 natural occurring amino acids at physiological pH, grouped by property of the side chain group.

The amino acid sequence, which a protein is made up of, is known as the primary structure. This polymer chain can be folded into a variety of different conformations, such as α-helices or ß-pleated sheets. This is referred to as the secondary structure. A single chain of protein may contain several regions which fold into different secondary structures, and how these associate is known as its tertiary structure. Finally, several separate protein chains can assemble to form large oligomers; this level of assembly is known as the quaternary structure, see Figure 1.8. [10,11]

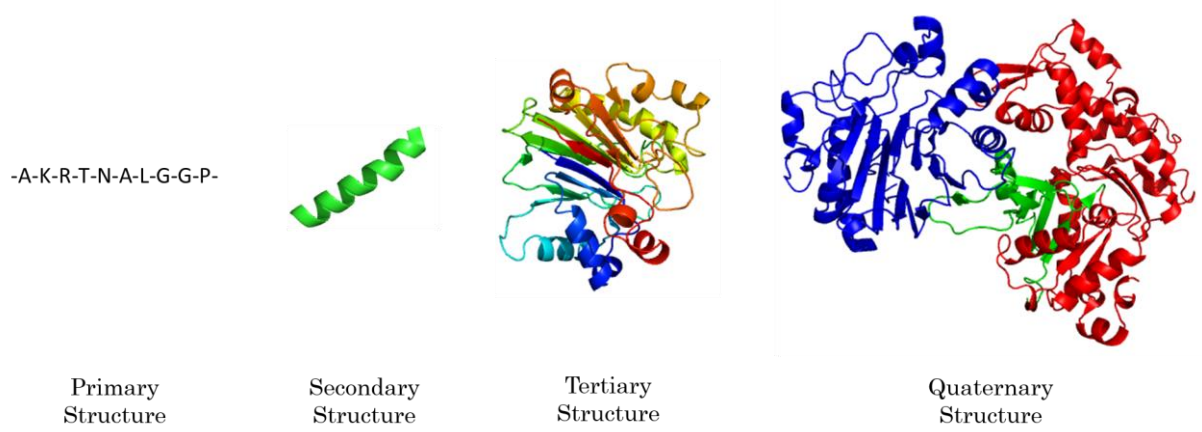


Figure 1.8 The protein structure is separated into four levels. The amino acid sequence is known as the primary structure. This sequence folds into particular conformations and is referred to as secondary structure. The separate regions of folded protein can associate with each other to create the tertiary structure. Finally, several separate protein chains can associated together to form a quaternary structure. All adapted from PDB 1ATN.[11]

Proteins are responsible for many of the processes within the body, and as such, are of great interest for the treatment of diseases and control of processes. As well as biomolecules being extensively explored for targeting by therapeutics, they are also being used for technologies outside of the body.

1.4 Tagging Biomolecules

Nature is limited to the naturally occurring bases and amino acids to produce DNA, RNA and proteins. In order to widen biomolecule applicability and functions, modified derivatives are being created, for example, systems with fluorescent labels or responsive features. Various bioconjugation techniques are now established and widely used within the field and are described in more detailed reviews and summarised in Chapter 5.[12,13] In general, the label is attached *via* a reactive group which targets a specific functional group within the protein or DNA.

1.4.1 Modifying DNA

In addition to the standard techniques used to attach fluorophores and other functionalities to DNA, many research groups have worked to modify the natural units within the DNA structure, e.g. the backbone or bases, as described below.

1.4.1.1 Modifying the Backbone of DNA

Modification of the sugar-phosphate backbone of DNA has the potential to increase the stability of the structure and make it less susceptible to degradation by enzymes, which becomes important when using DNA as a therapeutic or probe. The most notable modification in the field, is that of peptide nucleic acid (PNA). In this derivative, the backbone of DNA is replaced by a *pseudo* peptide unit, Figure 1.9.[14]

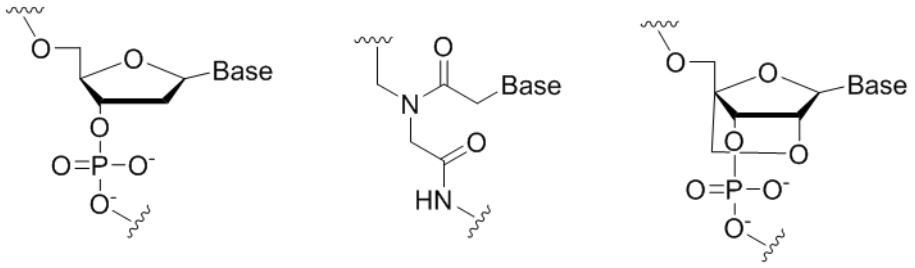


Figure 1.9 Comparison of the structures of DNA (left), PNA (middle) and LNA (right) showing the modified backbone used within PNA, that replaces the sugar-phosphate repeating pattern found in natural DNA and the "locked" sugar of LNA.

PNA has been shown to hybridise with both DNA and RNA, resulting in duplexes with higher T_m values than the corresponding DNA:DNA and RNA:RNA duplexes. It is thought that this is primarily due to the lack of charge on PNA, reducing the same charge repulsion found within the natural oligomers. In

addition to this, PNA is found to not be degraded by the cellular environment. These characteristics make for a very stable "biomolecule", which has been used for many applications.[14] Many other backbone derivatives have also been developed such as locked nucleic acid (LNA),[15] an RNA derivative with a connection between the 2′ oxygen and 4′ carbon of the ribose (Figure 1.9);[15] and ferrocene nucleic acids (FcNA) which incorporates a ferrocene into the backbone, replacing two bases.[15–17]

1.4.1.2 Modifying the Bases of DNA

In addition to modifying the backbone of DNA, the base attached can also be modified, or replaced completely. The majority of these modifications incorporate a fluorescent label,[18] although other derivatives have also been made, such as azobenzene switches[19] and metal binding ligands.[20] More recently artificial base pairs have been made, with the aim of expanding the code that is available to researchers. Romesberg and co-workers developed a hydrophobic nucleotide[21] and later showed that enzymes could successfully replicate DNA containing this modification.[22] In 2015 Benner and co-workers developed the first unnatural, hydrogen bonding bases which, when incorporated into a sequence, increased the stability of DNA.[23,24]

Tucker and co-workers have previously reported on the development of an anthracene nucleobase derivative. Within this derivative the natural sugar-

phosphate backbone is replaced with a theroninol linker; this is attached to anthracene *via* a carbon chain (further detail on the structure can be found in Chapter 3.1.4). This derivative has been successfully incorporated into DNA and has been shown to give a fluorescent output that is dependent on the nature of the surrounding bases.[25–28] The design of the nucleobase is explored further in Chapter 2, and the use of the fluorescent output, in Chapter 3.

1.4.1.3 Applications of Modified and Tagged DNA

1.4.1.3.1 DNA Sensing

An obvious use of DNA, is to use it as a probe to monitor processes based on DNA hybridisation, such as diseases and genetic traits within individuals. Many sensing assays have been developed which can target specific regions of DNA and are detailed in the following chapters. The majority of these use modified DNA with a functional group attached that can generate a measureable output, such as a fluorophore or redox-active moiety, which can subsequently be detected and monitored.

The DNA sensing field was dramatically enhanced by the development of the polymerase chain reaction (PCR), allowing for small amounts of DNA material to be amplified, making them much easier to detect. Dr Kary Mullis first proposed the idea in the early 1980's.[29] He was subsequently awarded a Nobel Prize for the work in 1993. The technique works by taking the sample of DNA, and adding in

two primers (short oligonucleotides that are complementary to the end section of the DNA to be amplified). A primer is used for each of the two DNA strands that form the duplex, as shown in Figure 1.10. Upon heating the dsDNA dissociates, subsequently lowering the temperature allows for the primers to bind to the complementary DNA (known as the template), preventing the reformation of dsDNA, and leaving the remainder of the strand as ssDNA. The enzyme, typically *Taq* polymerase, can then extend the primer in a process known as elongation, to complete the sequence. By thermal cycling the system, the strands dissociate, anneal and elongate several times with the newly made strands becoming templates themselves. This results in exponential growth of the DNA sequences.[30]

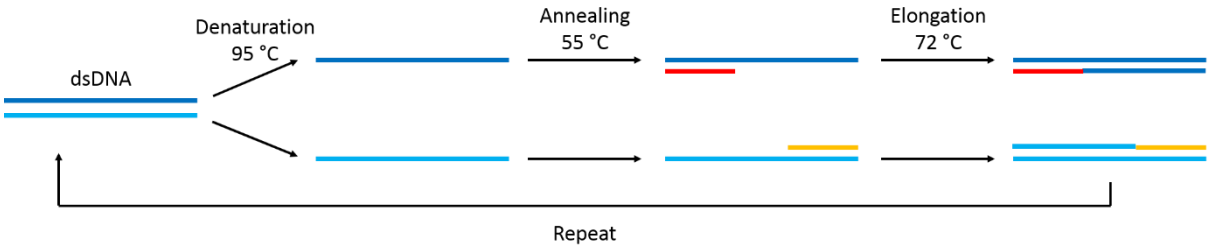


Figure 1.10 Schematic of the PCR techniques showing the stages of denaturation, annealing and elongation. dsDNA (blue) is heated to denature the duplex. Lowering the temperature allows for the primers (red and yellow) to bind to the complementary sequence and *taq* polymerase extends the sequences.

1.4.1.3.2 DNA as a Therapeutic Agent

Many diseases, such as cystic fibrosis, cancer and neurodegenerative diseases, are thought to be partially or wholly caused by a patient's genes. One can imagine that if a gene sequence is changed, the protein the gene codes for will also be altered, potentially resulting in disrupted function and disease. By

replacing the effected region of the patients DNA, or by supplying the affected cells with a "working" DNA sequence, the symptoms can be suppressed and potentially the patient cured.[31] With the development of these therapeutic techniques, there are many ethical and social issues which need to be discussed. Although this technology would allow for many diseases to be cured, it could also allow for the genetic modification of other traits such as sex, personality and physical appearance. Due to this, the currently approved therapeutics do not modify any information that is passed on to future generations.[32] Another potential way to control the production of proteins, is to silence genes. This can be done using small fragments of RNA that bind to mRNA and inhibit the replication, so called anti-sense therapy.[32,33] Due to the RNA binding and creating a double stranded section, the replication process is halted and therefore the target gene is not expressed. The use of modified DNA derivatives which are not as readily broken down in the body as natural DNA or RNA, is now an active area of research.[34]

1.4.1.3.3 DNA Nanostructures and Machines

DNA use is also being explored outside of the obvious medical uses. Due to the highly specific binding of the oligomer, and its ability to hold information, it is being used to create nanomachines and build potentially complex nanostructures.[35]

The building of nanostructures from DNA is referred to as DNA origami, and was pioneered in 1998 by Nadrian Seeman,[36] although it wasn't termed DNA origami until later.[37] Since then more complex designs have been made, with ever increasing functionality incorporated. Work in this area was summarised in a recent review by Kuzuya and Ohya.[38]

Of particular significance to the area, is the work of Turberfield. He has created an array of devices utilising DNA as both a nanostructure, but also as a fuel to run systems.[35,39–42] This technology has been used to make DNA walkers which follow specific set routes on a course.[42] Nir and co-workers have used the same approach to create a bipedal motor which can travel backwards and forwards between two DNA "tiles".[43]

In 2014, Du and co-workers reported on the use of DNA nanostructures as drug delivery vehicles to cancerous tumours.[44] Several DNA origami structures were made and loaded with an intercalating, anti-cancer drug. The authors found that the biodistribution of the nanostructures was dependent upon the shape of the DNA origami. In this case, a triangle was found to display the best localisation properties for the tumour. The release of the anti-cancer drug was achieved by utilising the naturally more acidic environment of the tumour. This increased the release by 15% when compared to physiological conditions.[44]

1.4.2 Modifying Proteins

Proteins are often modified with fluorophores, like DNA, to allow for detection within systems. In addition to this, they are routinely reacted with functional groups that allow for crosslinking and immobilisation.[45,46] These modified proteins can then be used to study the association of proteins, control processes and develop sensing assays. As with DNA, various routes of modification have been developed for proteins. The backbone of the structure, along with the side chains of amino acids can be modified, and these are explored below.

1.4.2.1 Side Chain Modifications

The most widely used route to modify proteins, is to react amine or thiol containing amino acids side chain with reactive molecules, which are attached to the desired functionality.[12] These derivatives comprise of the reactive component, which targets the side chain of interest, a linker and the modification to be incorporated. Typically, maleimides and idoacetamides are used to target thiols and succinimides are used to target amines, Figure 1.11.[47] Thiols are particularly attractive to target due to their low abundance within proteins. This allows for a more controlled incorporation of the derivative of interest.

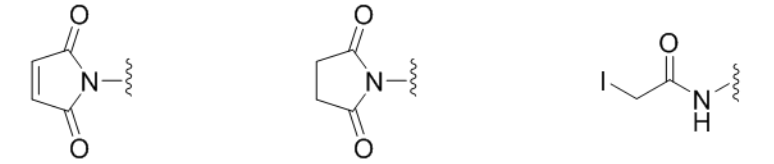


Figure 1.11 Common reactive groups used to derivatise proteins. Maleimides and idoacetamides target thiols and succinimides target amines.

In addition, unnatural amino acids (UAAs) can be incorporated into structures. This has been demonstrated using standard peptide synthesis as well as *in vitro*.[48,49] Hundreds of derivatives have been made with side chains that incorporate fluorescent groups,[50] entities for conjugation[51] and redox-active groups.[52]

Crosslinking of proteins is of particular relevance to the work within this thesis. This is routinely done using the reactive groups shown in Figure 1.11. Two of the reactive groups are joined together and each end is reacted with the protein of interest. A common reagent used for this is succinimidyl trans-4-(maleimidylmethyl)cyclohexane-1-carboxylate (SMCC), Figure 1.12.

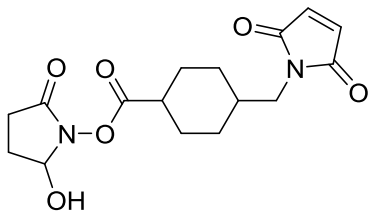


Figure 1.12 Chemdraw structure of SMCC, a crosslinking agent routinely used with proteins. The succinimide part of the molecule (left) reacts with amines and the maleimide part (right) reacts with thiols.

Aryl azides are an alternative approach. These derivatives, when excited with UV light, react with nucleophilic groups in close proximity. In addition, peptide stapling can be use. By using modified amino acids, which can react with an additional modified residue, the peptide can be "stapled" into a particular structure.[53]

1.4.2.2 Backbone Modifications

The peptide backbone has also been modified. A whole array of modifications have been incorporated with various objectives, including allowing for molecular recognition and increasing cell penetration. However, so far, no derivatives have become as widely used as those developed for DNA, with many hindered in their folding ability compared to natural peptide structures.[54–56]

Of particular note is the work performed by Horne *et al.* in 2004. They incorporated a triazole amino acid derivative into a peptide, Figure 1.13. It was shown that a peptide containing this modification at one site, could still fold into the native coiled coil structure. This modification is the same as would be present if two groups were joined using click chemistry, and the group therefore proposed that this modification could be used to link peptide chains.[57]

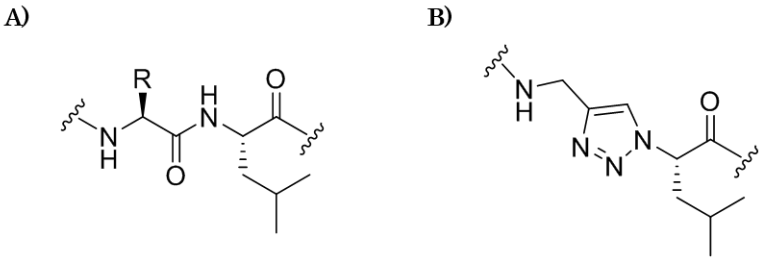


Figure 1.13 The triazole derivative (B) used in the work of Horne *et al.* to modify the peptide backbone shown with the comparative unmodified linkage (A).

1.5 Anthracene

Incorporation of anthracene into biomolecules could have many uses, being a fluorescent unit and able to dimerise upon irradiation with UV light. Other fluorophores are available, many of which emit at more biologically compatible wavelengths and that have higher quantum yields.[58,59] However, anthracene is the only molecule that displays fluorescence within the visible spectrum that can also dimerise upon irradiation, making it a useful tool for incorporation.

Anthracene is an organic molecule comprised of three benzene rings fused in a planar structure, Figure 1.14A. It is a fluorescent molecule that emits in the blue region. Examples of absorption and emission spectra are shown in Figure 1.14B.

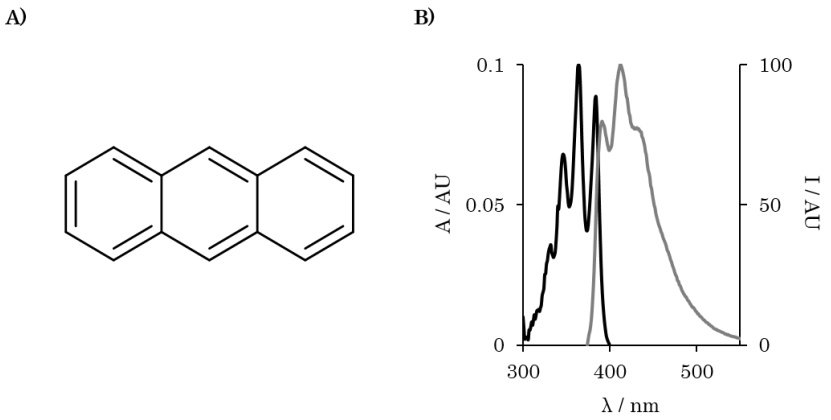


Figure 1.15 A) Chemdraw structure of anthracene and B) its characteristic absorbance (between 300 – 400 nm, black) and fluorescence (between 400 – 500 nm, grey) spectra in 100 mM sodium phosphate buffer at pH 7.

In addition to this property, anthracene is known to form photodimers when irradiated with UV light (around 365 nm), Figure 1.15. This process forms a mixture of head-to-head (H-H) and head-to-tail (H-T) orientations. The photodimer product can be reverted to the monomers, by heating or irradiation with light < 300 nm.[60–62]

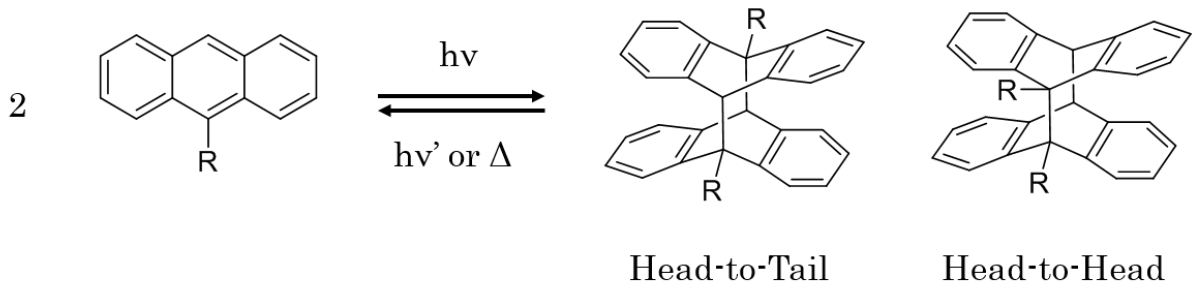


Figure 1.14 Anthracene undergoes a 4π+4π pericyclic addition upon excitation with UV light to form photodimer products with either head-to-head or head-to-tail conformations. The dimer species can be reverted to monomers, by heating or irradiation with a shorter wavelength of light.

Tracking of photodimer formation can be easily achieved by monitoring the absorbance and fluorescence bands shown in Figure 1.14B. These bands arise from the $\pi \rightarrow \pi^*$ transitions of the extended conjugated system. Upon photodimer

formation this conjugation is reduced and therefore the characteristic anthracene bands are no longer observed.

1.5.1 Uses of Anthracene within Supramolecular Chemistry

The photochemical properties of anthracene have been utilised extensively within the supramolecular community. Beyeler *et al.* described the development of a molecular switch using anthracene units as the "switching" entity and a metal binding centre as a fluorescence output. In the monomer forms, the anthracene units quench fluorescence from the bipyridine-metal complex, Figure 1.16. Upon photodimerisation the extended π system is removed and therefore the quenching pathway is no longer available and fluorescence is observed.[63]

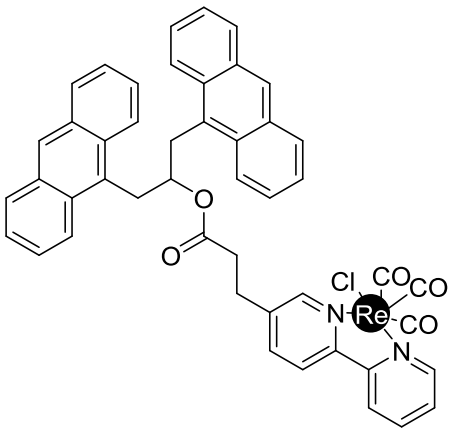


Figure 1.16 Molecular switch developed by Beyeler *et al.* comprised of anthracene units attached to a bipyridine metal binding ligand. Upon irradiation the photodimer product forms, removing the quenching pathway and fluorescence is observed.

Jin demonstrated the use of anthracene to control Na^+ transport across a phospholipid bilayer in 2000.[64] A calix[4]arene ester, a known carrier of Na^+ across membranes, was modified with anthracene units. Prior to exposure to UV light, it was shown that the modified derivative was still able to transport the sodium ion across the membrane, however, upon photodimerisation this was inhibited. McSkimming *et al.* demonstrated a similar principle using an aza-crown ether.[65,66]

The Tucker group have also utilised an anthracene photo-switch within a Hamilton receptor.[67,68] As with the above examples, prior to photodimerisation the guest molecule can bind to the host, however, upon irradiation, the photodimer forms and the guest can no longer bind, Figure 1.17A. This system has also been show to work on a gold surface which will enable a reusable sensor to be developed.[69] Of particular note in these examples, is that the anthracene photodimer formation closes the ring structure. The group has gone on to utilise this property to form a rotaxane by closing the ring using anthracene photodimerisation, Figure 1.17B.[70] The system utilises the Hamilton receptor, with the barbiturate guest acting to preorganise the "thread" within the rotaxane. Bulky groups incorporated onto the thread prevent the rotaxane "slipping off". The group were able to demonstrate cycling of the system with a recovery rate of over 80% per cycle.

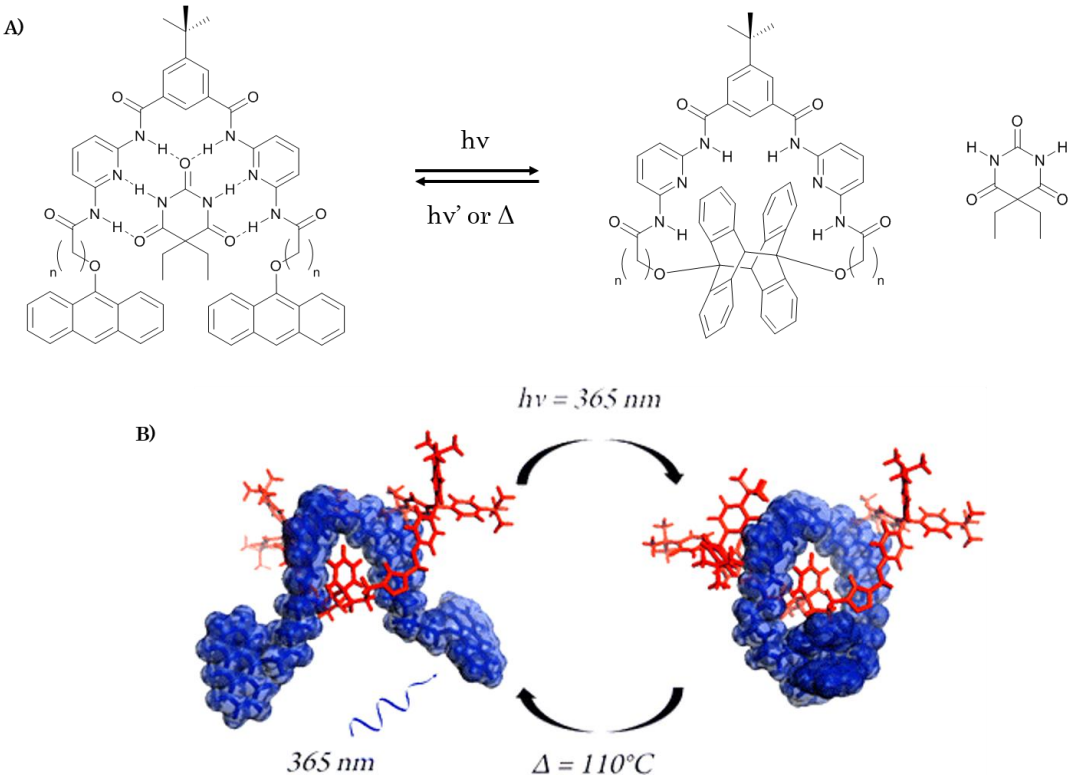


Figure 1.17 Supramolecular systems using anthracene photodimerisation. A) A Hamilton receptor incorporating anthracene moieties. Upon photodimerisation the guest molecule is released. B) Anthracene photodimerisation has been used to form a rotaxane from a Hamilton receptor formed around a thread molecule, which is held in place by a barbiturate guest. Adapted with permission from Tron *et al.*, *J. Org. Chem.*, 2015, **80**, 988-96. Copyright 2015 American Chemical Society.

1.5.2 Uses of Anthracene within Polymers

In addition to small molecule work, anthracene has been utilised within polymer systems. Goldbach *et al.* first demonstrated the potential use of anthracene photodimerisation in polymers in 2002.[71] They incorporated an anthracene derivative on to two polymer building blocks. These units could then be linked together by irradiation with UV light to make a "diblock". Subsequent thermal release by heating to cause the reversion of the anthracene photodimer was achieved, demonstrating the potential polymer breakdown that the system could be utilised for. Unfortunately it was considered that this system could not be

cycled due to the formation of homodimer products upon irradiation of the samples.[71]

A potential use of these systems was demonstrated by Wells *et al.* in 2011.[72] The group utilised a hydrogel, which had anthracene moieties appended to the polymer structure. The gel was soaked in a solution containing a model compound, which was absorbed into the gel structure. The samples were then exposed to UV light to induce the photodimer formation of anthracene, which resulted in the release of the compound from the gel. Although the compound was successfully released, the additional amount released when compared to samples not exposed to light, is modest. However, the group proposed this technology could be used to release drug molecules at specifically irradiated sites.[72]

In recent years, more elegant work has been performed to produce photo-patterns using anthracene modified polymers.[73] By shielding the polymer with a metal template, only certain regions of the sample are exposed to UV irradiation and, therefore, only these regions form the photodimer product. This was then visualised using polarised light, which showed the difference between the amorphous photodimer and the ordered unirradiated phase, Figure 1.18. Subsequent heating to 220 °C reverted the photodimer to monomer species and the sample returned to its crystalline state. This process was found to be repeatable over several cycles.[73]

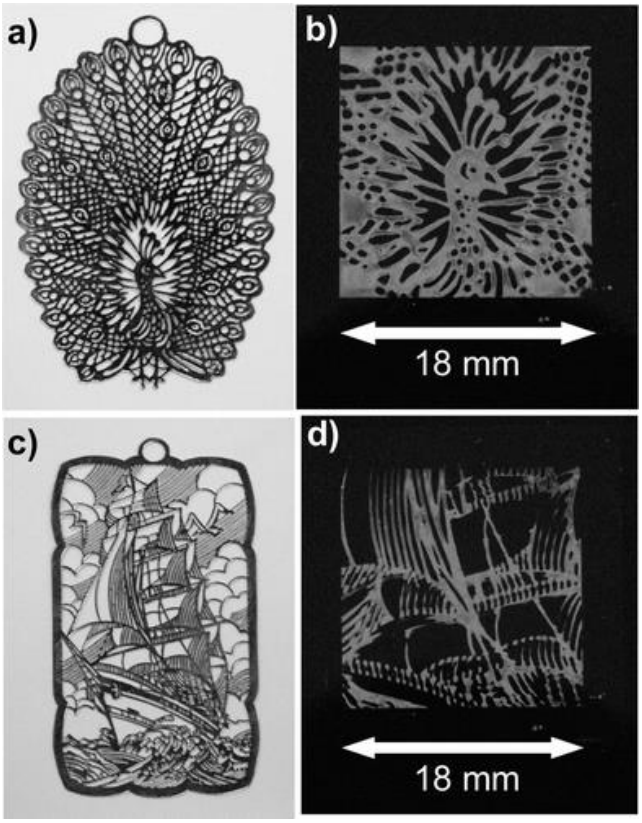


Figure 1.18 Examples of the photo-patterns that can be achieved using anthracene photodimerisation to crosslink polymers. The metal templates (left) are placed over the polymer sample, shielding areas from irradiation which therefore do not form photodimer species. Upon imaging with polarised light (right) the regions where photodimer has formed are easily distinguished from the crystalline monomer species. Reprinted with permission from Kihara *et al.*, *ACS Appl. Mater. Interfaces*, 2013, **5**, 2650-7. Copyright 2013 American Chemical Society.

Prior to these examples, anthracene had been incorporated into polymer systems as a fluorescent reporter. The groups of Pekcan, Winnik and Nuyken used the fluorescence of anthracene to monitor changes in polymer structure and swelling under particular conditions.[74–76] Wang *et al.* utilised the fluorescence output to monitor the structure of a gel which could be changed using anthracene photodimerisation.[77] Although in the system used, a gel also formed prior to irradiation, the group were able to show that photodimer formation resulted in a change to the polymer structure, and was accompanied by a decrease in fluorescence as the photodimer product formed.[77]

1.5.3 Uses of Anthracene within Biomolecules

Anthracene has also been incorporated into biomolecules to utilise its photochemical properties. It was first incorporated into phospholipids in 1984 to study the distribution of lipids in a bilayer. By using the photodimerisation properties of anthracene it could be established which phospholipids were in close proximity to one another in a liposome. The group also used anthracene fluorescence, to monitor the formation of the liposomes from the phospholipids.[78] They subsequently went on to demonstrate the use of the system in live bacteria.[79] More recently, anthracene has been utilised within an alginate system to create a photosensitiser, which the authors believe could find applications in sensitising organic reactions and in the development of sensors.[80]

1.5.3.1 Anthracene within DNA

Anthracene was first incorporated into DNA by Nakano and co-workers in 1995. The group attached an anthracene unit onto the 2 position of the sugar of a uridine phosphoramidite. Incorporation into an oligonucleotide allowed the group to show that duplex formation was still possible with the modification, and this was accompanied by an increase in anthracene fluorescence intensity.[81] Mortensen and co-workers reported on the synthesis of a unit in which the base of the phosphoramidite was replaced with anthracene, however, they did not go on to incorporate this into DNA.[82]

Jyo and colleagues took an alternative approach, attaching anthracene derivatives onto the ends of two short oligonucleotides. They then utilised the photodimerisation properties of anthracene, to link the oligonucleotides together. Upon binding to a template strand of DNA (comprised of complementary bases of the two oligonucleotides), the anthracene units were placed in close proximity to one another and UV irradiation afforded the linked product, Figure 1.19.[83,84] This strategy has also been used within locked nucleic acid (LNA) systems.[85]

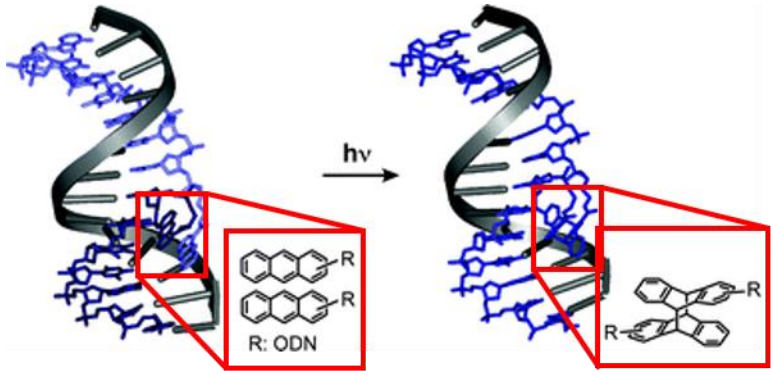


Figure 1.19 Work by Jyo and co-workers has focussed on linking two short oligonucleotides together using anthracene photodimerisation. The short oligonucleotides are preorganised on a template strand of DNA. Irradiation with UV light affords the crosslinked product. Adapted with permission from Mukae *et al.*, *Org. Biomol. Chem.*, 2009, **7**, 1349-54. Copyright 2009 Royal Society of Chemistry.

By incorporating the anthracene moieties into the same strand of DNA, Arslan *et al.* were able to "lock" DNA into a circular structure. The strand was first preorganised on a template so as to place the anthracene moieties into close proximity. This was done using a triplex forming sequence, Figure 1.20.[86]

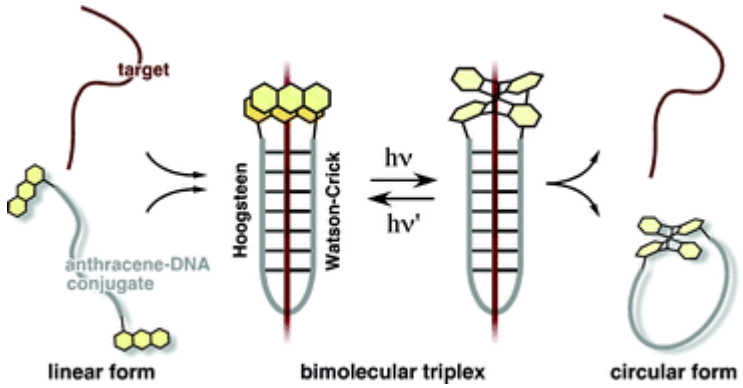


Figure 1.20 Schematic diagram of the formation of circular DNA using anthracene photodimerisation. The anthracene tagged strand is preorganised onto a template strand. Upon irradiation an anthracene photodimer is formed. Reprinted with permission from Arslan *et al.*, *Org. Biomol. Chem.*, 2010, **8**, 4843-8. Copyright 2010 Royal Society of Chemistry.

Tucker and co-workers utilised a similar approach to the early examples of anthracene incorporation, by replacing a base in the oligonucleotide sequence with an anthracene derivative. Anthracene was attached *via* a carbon chain linker to a serinol and phosphate backbone unit, which could be incorporated into DNA.[25] Later developments replaced serinol with threoninol to allow for control over the stereochemistry of the unit. These derivatives have been utilised as fluorescent sensors within DNA, as discussed in detail in Chapter 3. [26,87]

In addition to using anthracene fluorescence, the group have also utilised the photodimerisation properties of anthracene. Two anthracene moieties were incorporated into the same strand of DNA and, upon irradiation, the photodimer product was shown to bend the DNA and prevent binding to a complementary strand. However, duplex formation to the complementary strand prior to irradiation, prevented the photodimer formation and therefore a gated photochromic system was developed, Figure 1.21.[88]

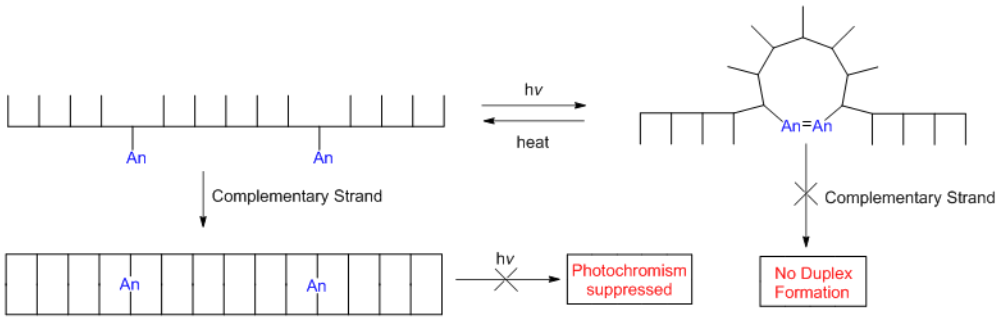


Figure 1.21 Anthracene units were incorporated into an oligonucleotide. Upon photodimer formation, a change in conformation of the DNA was induced which prevented duplex formation with a complementary strand of DNA. However, prior duplex formation prevented the reaction of the anthracene units, revealing a gated photochromic system. Adapted with permission from Manchester *et al.*, *J. Am. Chem. Soc.*, 2012, **134**, 10791-4. Copyright 2012 American Chemical Society.

Saito *et al.* have also worked to utilise anthracene fluorescence by attaching anthracene to uridine (Figure 1.22). On incorporation in DNA, they demonstrated the use of the fluorescence output to identify errors in a complementary sequence of DNA. This is discussed in more detail in Chapter 3.[89] Brown and co-workers adopted a similar approach using a thymine base.[90]

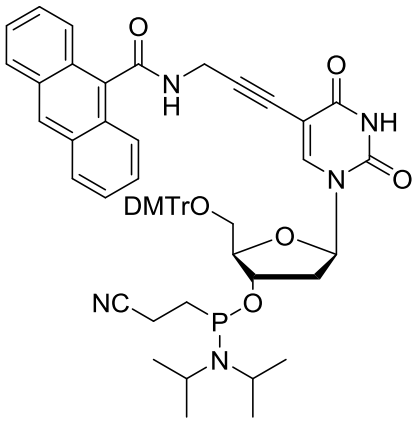


Figure 1.22 Chemical structure of the anthracene unit incorporated into DNA by Saito *et al.* for the detection of errors in DNA sequences.

1.5.3.2 Anthracene within Proteins

The use of anthracene within proteins and peptides has been very limited. There are a few examples that use the fluorescent output of small anthracene derivatives to study proteins.[13,91–94] However, to the best of my knowledge, there is only one example in the literature to attach an anthracene unit to a peptidic structure.[95] Within this example 9-carboxylic acid anthracene was coupled to the side chain of a lysine residue within a peptide expected to bind to DNA in an α helical conformation.[95] The group went on to assess the effect of adding PNA bases to the end of the peptide in the hope that this will increase the DNA binding affinity. The anthracene moieties acted as reporter groups, displaying changes in absorbance, fluorescence and CD signals upon binding to DNA.[95]

1.6 Project Aims

The work presented in this thesis aims to explore the use of anthracene within biomolecules, utilising both the fluorescence and photodimerisation properties it possess.

Anthracene fluorescence has previously been shown to allow for the detection of base identity at a SNP site.[25,87] This technique is to be further explored within a sequence associated with Alzheimer's disease. Several different anthracene probes will be synthesised and the fluorescence properties analysed. The optimal stereochemistry and linker length will then be used to assess the application of

the assay to quantify ratios of two possible bases at a SNP site. This assay would allow for quantitative determination of the bases that are present within a sample. In addition to this, the probes will be analysed in a different sequence to establish if they can be used to distinguish between guanine and the 8-oxoG modified base.

Anthracene will also be incorporated into a transcription factor that is known to bind to DNA as a homodimer. The incorporation will be made so as to replace the natural dimerisation region within the protein. The affect this incorporation has upon DNA binding will be assessed, utilising anthracene fluorescence amongst other techniques. Upon irradiation, the anthracene units are proposed to dimerise, but only when bound to the target DNA sequence and therefore brought into close proximity. The photodimerisation will in turn increase the DNA binding affinity. The reversibility of the system will also be analysed in order to develop a photo-switched peptide binding system.

Anthracene photodimerisation will be further explored within DNA structures. By incorporating anthracene units within a sequence capable of forming a hairpin structure, dimerisation is expected to induce the hairpin formation, and release the complementary strand of DNA. Several systems will be explored in order to optimise the amount of release achieved and the reversibility of the systems assessed. Incorporation of anthracene into a G-quadruplex is also

reported within the thrombin binding aptamer (TBA). Incorporations of two anthracene entities at various positions on the aptamer will, upon photodimerisation, stabilise or destabilise the structure depending upon where the incorporations are made. This will affect the folding of the structure and also the binding to the target protein. The reversibility of the process will be assessed and a photo-switched aptamer binding system could be developed.

1.7 References

1 J. D. Watson and F. H. C. Crick, *Cold Spring Harb. Symp. Quant. Biol.*, 1953, **18**, 123–31.
2 A. D. Hershey and M. Chase, *J. Gen. Physiol.*, 1952, **36**, 39–56.
3 E. T. Kool, *Chem. Rev.*, 1997, **97**, 1473–88.
4 S. Neidle, *Principles of Nucleic Acid Structure*, Elsevier, Oxford, 1st edn., 2008.
5 B. Norden, A. Rodger and T. Dafforn, *Linear Dichroism and Circular Dichroism: A Textbook on Polarized-light Spectroscopy*, The Royal Society of Chemistry, Cambridge, 2010.
6 K. Nakamoto and M. Tsuboi, *Drug-DNA Interactions: Structures and Spectra*, John Wiley & Sons, New Jersey, 2008.
7 P. Belmont, J. F. Constant and M. Demeunynck, *Chem. Soc. Rev.*, 2001, **30**, 70–81.
8 S. Doonan, *Nucleic Acids*, Royal Society of Chemistry, Cambridge, 2004.
9 G. M. Blackburn, M. J. Gait, D. Loakes and D. M. Williams, *Nucleic Acids in Chemistry and Biology*, Royal Society of Chemistry, Cambridge, 2006.
10 G. Walsh, *Proteins: Biochemistry and Biotechnology*, John Wiley & Sons, New Jersey, 2004.
11 W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai and K. C. Holmes, *Nature*, 1990, **347**, 37–44.
12 I. Johnson and M. T. Z. Spence, *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*, Life Technologies Corporation, 11th edn., 2010.
13 R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem. Rev.*, 2010, **110**, 2579–619.
14 B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem.*, 1996, **4**, 5–23.
15 L. Philip, in *DNA and RNA Nanobiotechnologies in Medicine: Diagnosis and Treatment of Diseases*, Springer Science & Business Media, 2013, pp. 225–41.
16 H. V. Nguyen, A. Sallustrau, L. Male, P. J. Thornton and J. H. R. Tucker, *Organometallics*, 2011, **30**, 5284–90.
17 H. V Nguyen, Z. Zhao, A. Sallustrau, S. L. Horswell, L. Male, A. Mulas and J. H. R. Tucker, *Chem. Commun.*, 2012, **48**, 12165–7.
18 J. N. Wilson and E. T. Kool, *Org. Biomol. Chem.*, 2006, **4**, 4265–74.
19 H. Asanuma, X. Liang, T. Yoshida and M. Komiyama, *Chembiochem*, 2001, **2**, 39–44.
20 K. Tanaka, A. Tengeiji, T. Kato, N. Toyama and M. Shionoya, *Science*, 2003, **299**, 1212–3.
21 A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari and F. E. Romesberg, *J. Am. Chem. Soc.*, 2008, **130**, 2336–43.
22 D. A. Malyshev, K. Dhami, H. T. Quach, T. Lavergne, P. Ordoukhanian, A. Torkamani and F. E. Romesberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 12005–10.

23 M. M. Georgiadis, I. Singh, W. F. Kellett, S. Hoshika, S. A. Benner and N. G. J. Richards, *J. Am. Chem. Soc.*, 2015, **137**, 6947–55.

24 L. Zhang, Z. Yang, K. Sefah, K. M. Bradley, S. Hoshika, M.-J. Kim, H.-J. Kim, G. Zhu, E. Jimenez, S. Cansiz, I.-T. Teng, C. Champanhac, C. McLendon, C. Liu, W. Zhang, D. L. Gerloff, Z. Huang, W. Tan and S. A. Benner, *J. Am. Chem. Soc.*, 2015, **137**, 6734–7.

25 N. Moran, D. M. Bassani, J.-P. Desvergne, S. Keiper, P. a S. Lowden, J. S. Vyle and J. H. R. Tucker, *Chem. Commun.*, 2006, 5003–5.

26 J.-L. H. A. Duprey, Z. Zhao, D. M. Bassani, J. Manchester, J. S. Vyle and J. H. R. Tucker, *Chem. Commun.*, 2011, **47**, 6629–31.

27 J.-L. H. A. Duprey, D. M. Bassani, E. I. Hyde, C. Ludwig, A. Rodger, J. S. Vyle, J. Wilkie, Z.-Y. Zhao and J. H. R. Tucker, *Supramol. Chem.*, 2011, **23**, 273–7.

28 J.-L. H. A. Duprey, PhD Thesis, University of Birmingham, 2010.

29 R. Saiki, S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich and N. Arnheim, *Science*, 1985, **230**, 1350–4.

30 J. M. S. Barlett and D. Stirling, *PCR Protocols*, Humana Press, New Jersey, 2nd edn., 2003.

31 Y. E. Khudyakov and H. A. Fields, *Artifical DNA*, CRC Press, Florida, 2003.

32 P. Saraswat, R. Soni, A. Bhandari and B. Nagori, *Indian J. Pharm. Sci.*, 2009, **71**, 488–98.

33 J. K. Watts and D. R. Corey, *J. Pathol.*, 2012, **226**, 365–79.

34 S. T. Crooke, *Annu. Rev. Med.*, 2004, **55**, 61–95.

35 J. Bath and A. J. Turberfield, *Nat. Nanotechnol.*, 2007, **2**, 275–84.

36 E. Winfree, F. Liu, L. A. Wenzler and N. C. Seeman, *Nature*, 1998, **394**, 539–44.

37 P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302.

38 A. Kuzuya and Y. Ohya, *Acc. Chem. Res.*, 2014, **47**, 1742–9.

39 B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605–8.

40 R. P. Goodman, M. Heilemann, S. Doose, C. M. Erben, A. N. Kapanidis and A. J. Turberfield, *Nat. Nanotechnol.*, 2008, **3**, 93–6.

41 P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield and J. H. Reif, *Angew. Chem. Int. Ed. Engl.*, 2004, **43**, 4906–11.

42 S. F. J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama and A. J. Turberfield, *Nat. Nanotechnol.*, 2012, **7**, 169–73.

43 M. Liber, T. E. Tomov, R. Tsukanov, Y. Berger and E. Nir, *Small*, 2015, **11**, 568–75.

44 Q. Zhang, Q. Jiang, N. Li, L. Dai, Q. Liu, L. Song, J. Wang, Y. Li, J. Tian, B. Ding and Y. Du, *ACS Nano*, 2014, **8**, 6633–43.

45 O. Boutureira and G. J. L. Bernardes, *Chem. Rev.*, 2015, **115**, 2174–95.

46 R. E. Feeney, B. A. Yamasaki and K. F. Geoghegan, *Modification of Proteins*, Americian Chemical Society, Washington, D.C., 1982, vol. 198.

47 G. T. Hermanson, *Bioconjugate Techniques*, Elseview Inc., London, 3rd edn., 2013.

48 D. Dougherty, *Curr. Opin. Chem. Biol.*, 2000, **4**, 645–52.
49 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–44.
50 W. Niu and J. Guo, *Mol. Biosyst.*, 2013, **9**, 2961–70.
51 A. J. de Graaf, M. Kooijman, W. E. Hennink and E. Mastrobattista, *Bioconjug. Chem.*, 2009, **20**, 1281–95.
52 N. Voloshchuk and J. K. Montclare, *Mol. Biosyst.*, 2010, **6**, 65–80.
53 Y. H. Lau, P. de Andrade, Y. Wu and D. R. Spring, *Chem. Soc. Rev.*, 2015, **44**, 91–102.
54 A. D. Bautista, C. J. Craig, E. A. Harker and A. Schepartz, *Curr. Opin. Chem. Biol.*, 2007, **11**, 685–92.
55 Z. E. Reinert and W. S. Horne, *Org. Biomol. Chem.*, 2014, **12**, 8796–802.
56 Z. E. Reinert, G. A. Lengyel and W. S. Horne, *J. Am. Chem. Soc.*, 2013, **135**, 12528–31.
57 W. S. Horne, M. K. Yadav, C. D. Stout and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2004, **126**, 15366–7.
58 M. S. T. Gonçalves, *Chem. Rev.*, 2009, **109**, 190–212.
59 H.-P. Josel and D. Heindl, *Labeling of Biomolecules with Fluorophores*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.
60 H. Bouas-Laurent, A. Castellan, J.-P. P. Desvergne and R. Lapouyade, *Chem. Soc. Rev.*, 2000, **29**, 43–55.
61 G. W. Breton and X. Vang, *J. Chem. Educ.*, 1998, **75**, 81–2.
62 H. Bouas-Laurent, A. Castellan, J.-P. P. Desvergne and R. Lapouyade, *Chem. Soc. Rev.*, 2001, **30**, 248–63.
63 A. Beyeler, P. Belser and L. De Cola, *Angew. Chemie Int. Ed. English*, 1997, **36**, 2779–81.
64 T. Jin, *Chem. Commun.*, 2000, 1379–80.
65 G. McSkimming, J. H. R. Tucker, H. Bouas-Laurent, J.-P. Desvergne, S. J. Coles, M. B. Hursthouse and M. E. Light, *Chemistry*, 2002, **8**, 3331–42.
66 G. McSkimming, J. Tucker, H. Bouas-Laurent and J. Desvergne, *Angew. Chem. Int. Ed. Engl.*, 2000, **39**, 2167–9.
67 Y. Molard, D. M. Bassani, J.-P. Desvergne, N. Moran and J. H. R. Tucker, *J. Org. Chem.*, 2006, **71**, 8523–31.
68 Y. Molard, D. M. Bassani, J.-P. Desvergne, P. N. Horton, M. B. Hursthouse and J. H. R. Tucker, *Angew. Chem. Int. Ed. Engl.*, 2005, **44**, 1072–5.
69 C.-K. Liang, G. V Dubacheva, T. Buffeteau, D. Cavagnat, P. Hapiot, B. Fabre, J. H. R. Tucker and D. M. Bassani, *Chemistry*, 2013, **19**, 12748–58.
70 A. Tron, P. J. Thornton, C. Lincheneau, J.-P. Desvergne, N. Spencer, J. H. R. Tucker and N. D. McClenaghan, *J. Org. Chem.*, 2015, **80**, 988–96.
71 J. T. Goldbach, T. P. Russell and J. Penelle, *Macromolecules*, 2002, **35**, 4271–6.
72 L. A. Wells, M. A. Brook and H. Sheardown, *Macromol. Biosci.*, 2011, **11**, 988–98.
73 H. Kihara and M. Yoshida, *ACS Appl. Mater. Interfaces*, 2013, **5**, 2650–7.
74 M. Erdoğan, G. Hizal, Ü. Tunca, D. Hayrabetyan and Ö. Pekcan, *Polymer*, 2002, **43**, 1925–31.

75 T. Kos, C. Strissel, Y. Yagci, T. Nugay and O. Nuyken, *Eur. Polym. J.*, 2005, **41**, 1265–71.
76 F. Caldérara, Z. Hruska, G. Hurtrez, T. Nugay, G. Riess and M. A. Winnik, *Die Makromol. Chemie*, 1993, **194**, 1411–20.
77 C. Wang, D. Zhang, J. Xiang and D. Zhu, *Langmuir*, 2007, **23**, 9195–200.
78 J. Bony and J.-F. Tocanne, *Eur. J. Biochem.*, 1984, **143**, 373–9.
79 M. Welby, Y. Poquet and J.-F. Tocanne, *FEBS Lett.*, 1996, **384**, 107–11.
80 D. R. Chejara, S. Kondaveeti and A. K. Siddhanta, *Polym. Bull.*, 2014, **72**, 35–48.
81 K. Yamana, R. Aota and H. Nakano, *Tetrahedron Lett.*, 1995, **36**, 8427–30.
82 R. S. Coleman and M. A. Mortensen, *Tetrahedron Lett.*, 2003, **44**, 1215–9.
83 T. Ihara, T. Fujii, M. Mukae, Y. Kitamura and A. Jyo, *J. Am. Chem. Soc.*, 2004, **126**, 8880–1.
84 M. Mukae, T. Ihara, M. Tabara and A. Jyo, *Org. Biomol. Chem.*, 2009, **7**, 1349–54.
85 K. Pasternak, A. Pasternak, P. Gupta, R. N. Veedu and J. Wengel, *Bioorg. Med. Chem.*, 2011, **19**, 7407–15.
86 P. Arslan, A. Jyo and T. Ihara, *Org. Biomol. Chem.*, 2010, **8**, 4843–8.
87 Z. Zhao, M. San, J.-L. H. A. Duprey, J. R. Arrand, J. S. Vyle and J. H. R. Tucker, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 129–32.
88 J. Manchester, D. M. Bassani, J.-L. H. A. Duprey, L. Giordano, J. S. Vyle, Z. Zhao and J. H. R. Tucker, *J. Am. Chem. Soc.*, 2012, **134**, 10791–4.
89 Y. Saito, K. Motegi, S. S. Bag and I. Saito, *Bioorg. Med. Chem.*, 2008, **16**, 107–13.
90 Q. Xiao, R. T. Ranasinghe, A. M. P. Tang and T. Brown, *Tetrahedron*, 2007, **63**, 3483–90.
91 A. J. Wilson, J. Hong, S. Fletcher and A. D. Hamilton, *Org. Biomol. Chem.*, 2007, **5**, 276–85.
92 M. Sisido and T. Hohsaka, *Appl. Microbiol. Biotechnol.*, 2001, **57**, 274–81.
93 T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 1999, **121**, 34–40.
94 H. Sasaki, M. Sisido and Y. Imanishi, *Langmuir*, 1991, **7**, 1944–8.
95 G. Balasundaram, T. Takahashi, A. Ueno and H. Mihara, *Bioorg. Med. Chem.*, 2001, **9**, 1115–21.

Chapter 2

Techniques

2.1 Introduction

Within this chapter the techniques used during the course of this work are presented. These will include solid-phase oligonucleotide synthesis, detailing anthracene incorporation and solid-phase peptide synthesis, followed by the purification and characterisation of these, as well as analytical techniques; ultraviolet-visible spectroscopy (UV-Vis), fluorescence spectroscopy, circular dichroism spectroscopy (CD), high performance liquid chromatography (HPLC), mass spectrometry (MS) and gel electrophoresis.

2.2 Oligonucleotide Synthesis

Over the past 60 years oligonucleotide synthesis has progressed from using standard organic chemistry techniques to using automated synthesisers that can produce long sequences of DNA.[1,2] The advancement in the methodology was pioneered by Caruthers and co-workers, utilising a reactive phosphoramidite to build the oligonucleotide chain in a series of steps; deprotecting the previous unit, coupling the new monomer, oxidising the phosphate group and repeating.[1,3,4] By ensuring all side chains are protected, and the main chain is protected until it needs to be coupled, the sequence can be made with a high specificity. The main chain is protected using an acid labile group, whilst the bases are protected with base labile groups. This ensures the removal of the correct groups and is known as orthogonal protection. The amine on the base is protected with benzoyl on adenine, an acetyl on cytosine and an isobutyryl on guanine (thymine is not protected as it does not have a primary amine group). The alcohol group on the 5′ of the sugar is protected with a dimethoxytrityl (DMT) group.

Oligonucleotides are synthesised in the 3′ to the 5′ direction. The synthesis is performed using a solid support which normally incorporates the initial base, typically using a controlled pore glass (CPG) bead. In the first step of the synthesis, the DMT protecting group is removed to expose the primary alcohol using trichloroacetic acid (Figure 2.1, step 1). The DMT cation liberated in this step is orange in colour and by measuring the absorbance at 498 nm, an

indication of the deprotection efficiency can be determined, which in turn can be an indication of the yield of the synthesis, allowing for step-by-step tracking of individual coupling yields. The next step is to attach the subsequent base in the sequence. The base is activated with 5-(ethylthio)-1H-tetrazole (ETT) and coupled to the available alcohol (Figure 2.1, steps 2 and 3). Any unreacted sequences are then capped with an acetyl group (Figure 2.1, step 4), to prevent sequences with deleted bases being synthesised, allowing for greater ease of subsequent purification. Finally, the phosphite is oxidised to a phosphate using iodine. The whole cycle is then repeated with the next base in the sequence and continues until the desired oligonucleotide sequence is complete. It can then be cleaved from the resin with aqueous ammonia and the protecting groups on the bases are then removed by heating in the same solution. The cleaved and deprotected oligonucleotides are purified by reversed-phase high performance liquid chromatography (RP-HPLC) which is described in Section 2.5.[5,6]

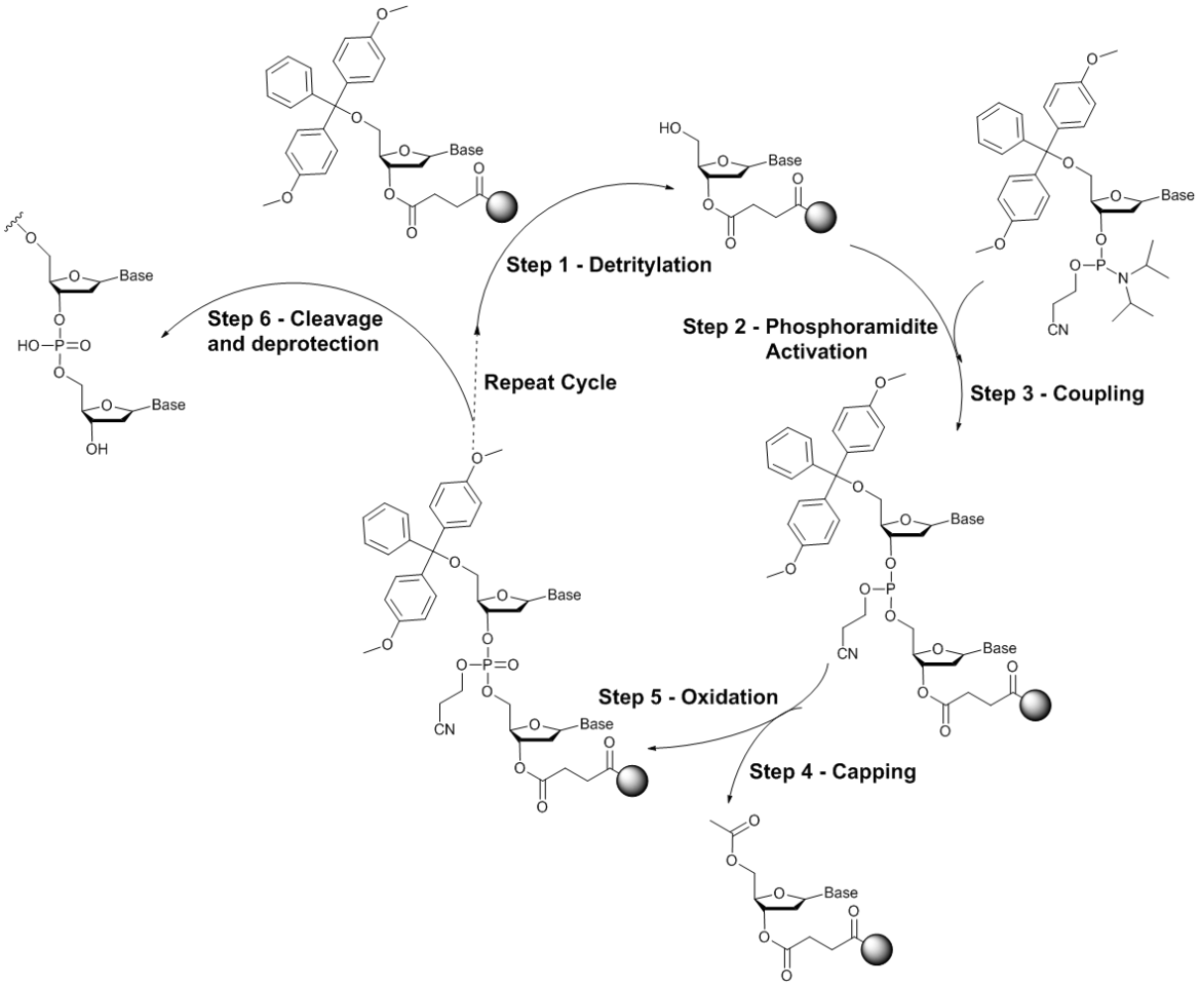


Figure 2.1 Solid-phase oligonucleotide synthesis cycle. Step 1 – removal of the DMT protecting group of the alcohol, Step 2 – Activation of the phosphoramidite for coupling, Step 3 – Coupling of the new phosphoramidite to the chain *via* the newly exposed alcohol, Step 4 – Capping of any unreacted alcohol groups, Step 5 – oxidation of the phosphite group. These 5 steps are repeated until the desired sequence has been made. The oligonucleotide can then be cleaved from the resin and the side chains deprotected (Step 6).

2.3 Anthracene Incorporation into Oligonucleotides

Phosphoramidites with an anthracene moiety in the place of the DNA base have been synthesised and successfully incorporated into oligonucleotides within the literature. A full discussion of the development of the synthetic procedures has

been previously reported;[7,8] a description of the currently used method is outlined below.

The anthracene is incorporated *via* the oligonucleotide backbone so as to replace one base (Figure 2.2). A variety of linker lengths to the anthracene, as well as L- and D-stereochemistries, have previously been made. Within these studies, linkers of 1 and 3-7 carbons were used with both L- and D-stereochemistries.

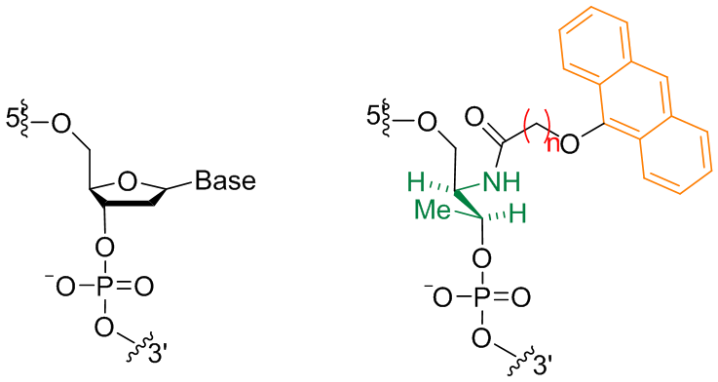


Figure 2.2 Comparison of the Tucker group probe (right) to natural DNA (left). Anthracene is incorporated into the oligonucleotide to replace one base. A threoninol linker (green) is used to control the stereochemistry and the length of the carbon linker (red) can be altered to control the position of the anthracene (orange).

The synthesis starts from anthrone, which is reacted with a bromo ethyl ester, this specifies the length of the carbon chain, and reacts with the anthrone *via* a base-induced nucleophilic substitution, Figure 2.3. Subsequent saponification yields the carboxylic acid, which can then be reacted with threoninol to form an amide bond with activation from N,N,N′,N′-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU) and N, N-diisopropylethylamine (DIPEA). The threoninol linkage is chosen with the stereochemistry desired for the final product. The primary alcohol is then protected with a DMT protecting

group and the secondary alcohol is phosphytilated ready for solid-phase oligonucleotide synthesis, as discussed in Section 2.1. Full experimental conditions and characterisation have been previously reported[7,8] and can also be found in the experimental section (Chapter 8.2).

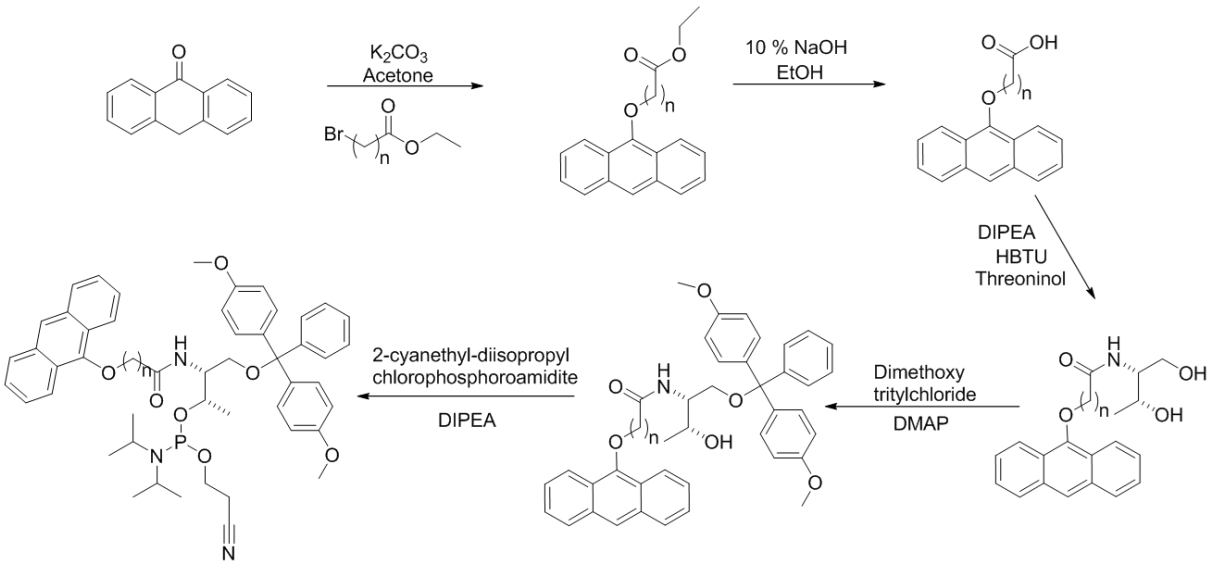


Figure 2.3 Synthesis of anthracene phosphoramidites for subsequent incorporation into DNA.

2.4 Peptide Synthesis

As with oligonucleotide synthesis, peptide synthesis is currently performed on a solid-support using chemistry initially developed by Merrifield.[9] Within these studies rink amide 4-methylbenzhydrylamine (MBHA) resin was used, although many alternatives are available. The peptide is synthesised from the C-terminus towards the N-terminus. The resin is first deprotected to remove the fluorenylmethyloxycarbonyl (Fmoc) group on the amine using piperidine (Figure 2.4, step 1). The first amino acid in the sequence can then be coupled. This amino

acid contains an Fmoc protected amine, as well as side chains protected with acid labile groups where appropriate, and is coupled upon activation from HBTU and DIPEA (Figure 2.4, step 2). The cycle can then be repeated by removing the Fmoc group from the coupled amino acid, followed by coupling the subsequent amino acid in the sequence. This cycle is repeated until the desired sequence has been completed. It is important to completely rinse the resin between each step so as to ensure that the previous reagents have been removed, for the step-by-step synthesis to work efficiently. Following Fmoc deprotection, the N-terminus is then capped with an acetyl group and the peptide cleaved from the resin with trifluoroacetic acid (TFA). The acid concomitantly deprotects the side chains of the amino acids. The peptides are then purified using RP-HPLC (Section 2.5).[10,11]

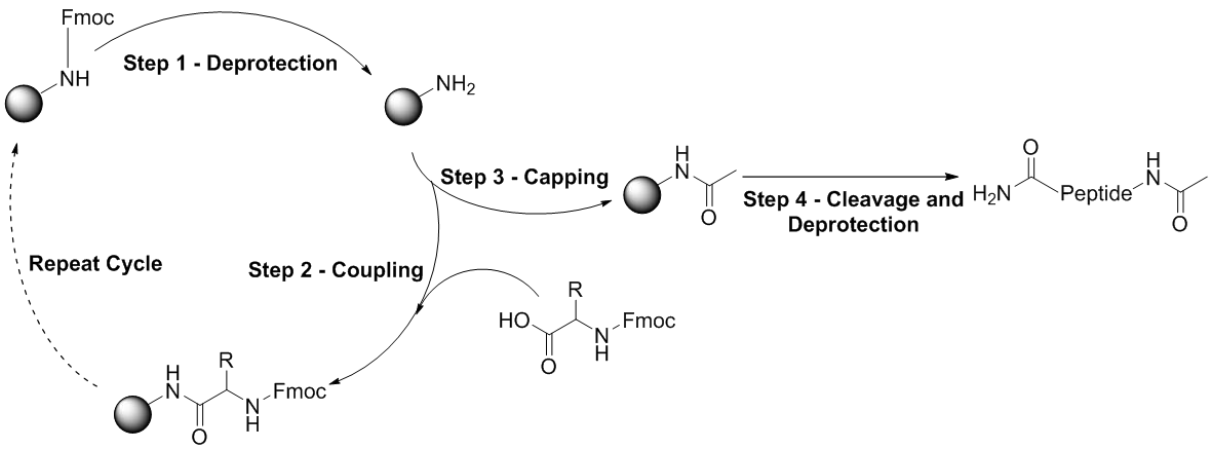


Figure 2.4 Solid-phase peptide synthesis (SPPS) cycle showing the deprotection of the main chain amine (step 1) followed by the coupling of the subsequent amino acid (step 2). These are repeated until the desired peptide has been made. The exposed amine is then capped (step 3) and the peptide cleaved from the resin (step 4). Side chain deprotection is also achieved in this step.

Within these studies anthracene is incorporated into the peptide *via* a lysine side chain. Details of this can be found in Chapter 4.4.1.

2.5 Reversed-Phase High Performance Liquid Chromatography

High performance liquid chromatography (HPLC) is a technique used to separate molecules according to their interaction with a solid-phase and a mobile phase. Traditionally a polar stationary phase (*i.e.* silica), which is packed into a column, is used with a non-polar mobile phase (*i.e.* chloroform). High pressure is applied to the column, forcing the mobile phase, and the dissolved compounds being analysed, to move through and the latter to separate according to the affinity they have with the stationary phase. Non-polar compounds interact with the stationary phase the least and hence are eluted first. The eluent is monitored by recording its optical properties (*i.e.* UV-Vis absorbance or fluorescence) and peaks can be seen when an analyte, which absorbs at the monitored wavelength, elutes. The technique can be used on a preparative or an analytical scale, dependent upon the size of the column used and the amount of material being loaded.[12,13]

With RP-HPLC the stationary phase is made to be non-polar by coating the silica particles with hydrocarbons. A polar mobile phase is then used (*i.e.* water). This method is well suited for more hydrophilic molecules such as oligonucleotides and peptides, and typically results in the most hydrophilic molecules being eluted first, although size can also play a factor. Since pH can alter the hydrophobic character of the compounds being analysed, the pH of the mobile phase is controlled. The work reported herein uses 0.1 M triethylammonium acetate (TEAA) at pH 7 to separate oligonucleotides and 0.05% trifluoroacetic acid at pH

2 to separate peptides, using a water and acetonitrile mobile phase. The separation can be performed using either an isocratic method (*i.e.* the eluting solvent is kept constant throughout the separation) or a gradient method in which the amount of polar solvent is increased during the run. Combinations of these two methods can also be used, to achieve optimum separation.[13,14]

2.6 Mass Spectrometry

Mass spectrometry is the technique used to determine the molecular weight of components within a sample and is the primary characterisation technique used for peptides and oligonucleotides. The technique works by ionising the sample and then detecting the ions formed, by separation according to their mass to charge ratio. All studies reported herein used a time of flight (TOF) detector. In this detection method, all ions are accelerated so that they all have the same kinetic energy. The time taken for the ions to reach the detector is then recorded, with lower mass species reaching the detector first.[15] A variety of techniques for ionising the compounds are available, however, within these studies electrospray (ESI) and matrix assisted laser desorption ionisation (MALDI) were used.

The ESI technique uses a fine nozzle with an electrical current supplied over it. The analyte is sprayed through the nozzle causing an aerosol to form and the compounds to be ionised. This technique is a soft ionisation method and typically gives multiply charged species for large biomolecules.[15] The fragmentation can be

controlled using the cone voltage, an increase in this results in more fragmentation.[16]

MALDI is an ionisation technique predominantly used for biomolecules. It works by firing a laser at the sample which is co-crystallised with a matrix. The matrix compounds are typically charged acids, which are strongly absorbing in the UV region to efficiently absorb the energy from the laser and transfer ionisation to the sample. The matrix is ionised by the laser and causes secondary ionisation of the sample. This technique gives rise to singly charged ions of the analyte, with very little fragmentation, making it an ideal technique for large biomolecules.[15,16]

2.7 Ultraviolet Visible Spectroscopy

Ultraviolet visible (UV-Vis) spectroscopy is an optical technique used to determine the absorbance profile of a compound. Upon light irradiation, electrons within a molecule can absorb the energy and become excited, moving from the highest occupied molecular orbital (HOMO) to a higher unoccupied molecular orbital. The extent of light absorbed at different wavelengths is recorded by the spectrophotometer to give an absorbance value as described in equation 2.1.

$$A = log\left(\frac{I_0}{I}\right) \qquad Eq\ 2.1$$

Where A is absorbance, I_0 is the intensity of incidence light and I is the intensity of light transmitted to the detector.

For a given compound, the absorbance is related to the concentration of the compound in the solution being measured, the pathlength the light needs to travel through the solution, and the extinction coefficient of the compound at a specific wavelength, as described by the Beer Lambert Law, shown in equation 2.2.

$$A = \varepsilon lc \qquad Eq\ 2.2$$

Where A is absorbance, ε is the extinction coefficient of the compound in a specific solvent, l is the sample path length and c is the concentration of the compound in solution.[17]

For peptides and oligonucleotides, the above relationship between concentration and absorbance is routinely used to determine an accurate concentration of solutions. Classical methods of weighing the material and dissolving in a known volume are not adopted herein, due to the small quantities of compound used and the tendency of the compounds to contain salts and water, making the theoretical molecular weight unreliable for use in concentration calculation. Within this study a variety of ways to determine the concentration optically were used and these are described below.

2.7.1 Oligonucleotide Concentration Determination

The concentration of oligonucleotides are routinely determined at 260 nm where the bases π → π* transition is observed. The most commonly accepted method for determining the extinction coefficient at this wavelength is the nearest-neighbour model, which takes into account the effect of the neighbouring bases upon one another.[18,19]

Anthracene is known to absorb at 260 nm and hence contributes to the absorbance value obtained. Duprey has previously calculated the extinction coefficient for the anthracene molecules used herein to be 51 444 M^{-1} cm^{-1} at 260 nm in aqueous solution at pH 7.[8] In order to accurately determine the extinction coefficient of anthracene containing sequences, the extinction coefficient of the flanking regions of the anthracene were calculated independently of one another and the total calculated by the sum of all three components.

2.7.2 Peptide Concentration Determination using Anthracene

Unmodified peptides are usually quantified using the absorbance of aromatic amino acids such as tryptophan, tyrosine and phenylalanine.[20] Free thiols on cysteine residues can also be quantified and the peptide concentration determined from this. The latter can be performed using an Ellman's assay.[21] A solution of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) reacts with free accessible thiols in a 1:1 ratio to produce 2-nitro-5-thiobenzoate, which is a yellow molecule

with known absorbance properties ($\varepsilon_{412\ nm}$ = 14 150 M^{-1} cm^{-1}).[22] The absorbance is measured and hence the concentration determined. Due to the exposed thiol on the peptide reacting with DTNB, this method is sacrificial.

The peptide used in these studies does not contain the standard amino acids used for quantification, however, it is possible to take advantage of the absorbance properties of the anthracene that has been attached. Anthracene has a strong absorbance at 253 nm and weaker absorbance's between 300 – 400 nm, as discussed in Chapter 1.5. Due to a slight absorbance of the peptide within the 253 nm region it was decided that the peptide concentration would be determined at 375 nm for more accuracy, using the extinction coefficient of 6 300 M^{-1} cm^{-1} for free anthracene in aqueous solutions pH 7.[23]

2.7.3 Thermal Melting of DNA

A routine method for the measurement of the relative stability of duplex DNA and folding of single stranded DNA is thermal melting analysis to deduce the melting temperature (T_m), typically tracked by measuring the change in UV-Vis absorbance as a function of temperature. DNA bases are known to display a stronger absorbance when in the single stranded form, compared to a double stranded system.[24,25] This is referred to as the hyperchromic effect and has been attributed to several causes, which have recently been summarised in a book by McLaren and Shugar.[26] By measuring the increase in absorbance as a sample is

heated, the transition from double stranded to single stranded DNA can be monitored, Figure 2.5. This transition occurs at a higher temperature for more stable duplexes i.e. those with more base pairs, and for more G-C rich sequences, due to the larger number of hydrogen bonds they contain. The T_m value is taken as the temperature at which the first derivative of the absorbance signal displays its maximum intensity.

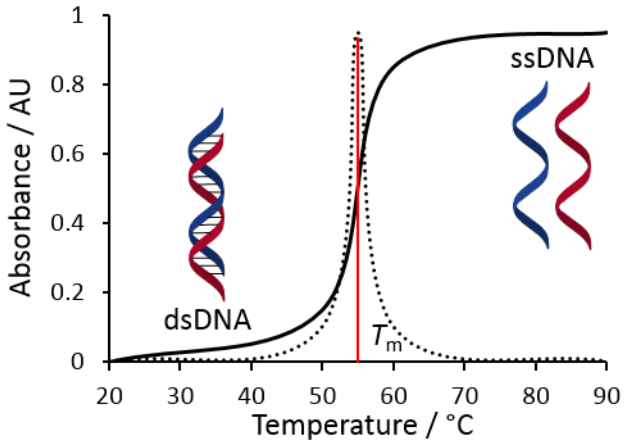


Figure 2.5 Schematic of a thermal melting curve showing the increase in absorbance as a function of temperature (—), the first derivative (····) and the point at which the T_m value would be taken.

Within the studies presented herein, 5 ramps are performed for each T_m measurement. Ramps 1, 3 and 5 are dissociation ramps, *i.e.* heating, and ramps 2 and 4 are annealing, *i.e.* cooling. The T_m value is only taken from the dissociation ramp data, while a temperature of association, T_a can be obtained from the association ramps. These values normally display slight differences due to the mechanisms of annealing and denaturing.

The T_m of an oligonucleotide is heavily dependent upon the sequence of the DNA, as this is directly correlated to the hydrogen bonds formed and the π-stacking of the system. Due to this, systems are also affected by mismatches in the DNA sequence and intercalation between the bases.[27,28]

Furthermore, thermal melting transitions are influenced significantly by the salts within the sample. Hard cations within the solution are known to stabilise the DNA duplex, by acting as a counter ion for the negative phosphate backbone. Altering the concentration or type of ions will alter the T_m value obtained. A change in the pH can also alter the T_m, as this will alter the hydrogen bonding propensity of the base pairs. For these reasons values are only directly comparable to samples performed under the same buffer and salt conditions.[29]

2.8 Fluorescence Spectroscopy

In fluorescence spectroscopy, light of a particular wavelength is used to excite electrons in the molecule into a higher energy electronic state. As these electrons relax back to the ground state, light can be emitted (Figure 2.6) and this emission of light is recorded within the technique. The term fluorescence only applies to light emitted resulting from an electron returning to the ground state from an excited singlet state. It is also possible for a similar process to occur from an excited triplet state, termed phosphorescence, which is not relevant to these studies.[23]

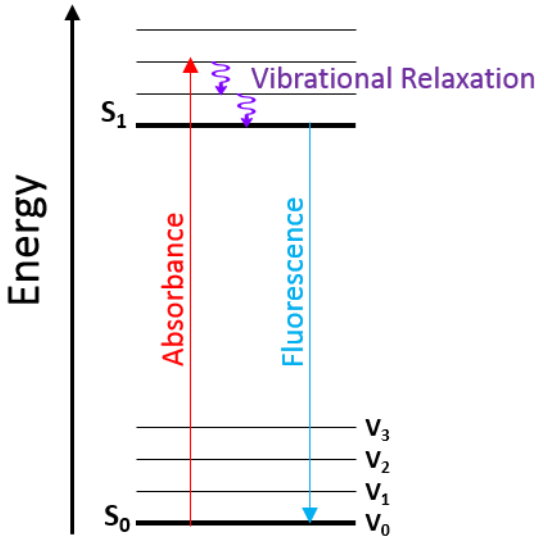


Figure 2.6 Jablonski diagram displaying the electronic excitation of an electron from the ground state (S_0) to an excited singlet state (S_1). As the electron returns to the ground state, light of a lower energy (higher wavelength) is emitted which is termed fluorescence.

Several values can be determined from fluorescence data, including quantum yields, a measure of how many of the excited molecules decay by fluorescence; and lifetimes of fluorescence, which can indicate quenching pathways of the excited state.[23]

2.9 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is a technique widely used to study the secondary structure of biomolecules. It works by measuring the difference in absorbance of left handed and right handed circularly polarised light, to give an ellipticity value. Chiral molecules, by their nature, will absorb the different helices of light by varying amounts, allowing them to be studied.[30] In addition to

chiral centres, DNA and proteins fold into organised structures which can also absorb a preferential direction of polarised light at various wavelengths.

It is known that B-DNA displays a characteristic CD signal of a positive band centred at ~275 nm and a negative band ~240 nm. A-DNA, a more compacted version of B-DNA in which the bases are more tilted, is shown by a positive signal centred at 260 nm and a negative band centred at 210 nm. Z-DNA, a left handed helix, is characterised by a weak negative band at 290 nm, a weak positive band at 280 nm and a large negative band between 220-280 nm.[30,31] Examples of these spectra are shown in Figure 2.7.

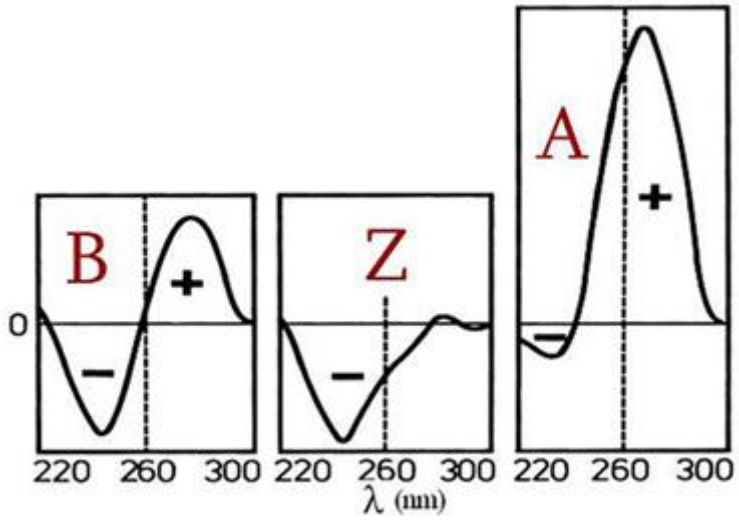


Figure 2.7 Schematic representation of the CD signals displayed by B-, Z- and A-DNA. Reprinted with permission from Dodero *et al.*, *Front. Biosci.*, 2011, **16**, 61-73. Copyright 2011 Frontiers in Bioscience.

Proteins and peptides can also be studied using this method, and the transitions displayed from them are much better understood than with DNA. As with DNA, characteristic bands are displayed for the secondary structures found within proteins, Figure 2.8.[30,32] The ellipticity values that are obtained are routinely converted to molar ellipticity values per residue (Equation 2.3) to allow for easier comparison between systems.[33]

$$\theta_m = \frac{\theta}{10 \times l \times c \times N_a} \qquad Eq\ 2.3$$

Where θ_m is the molar ellipticity value, θ is the measured ellipticity, l is the pathlength of the sample in cm, c is the concentration in M and N_a is the number of amino acids present.

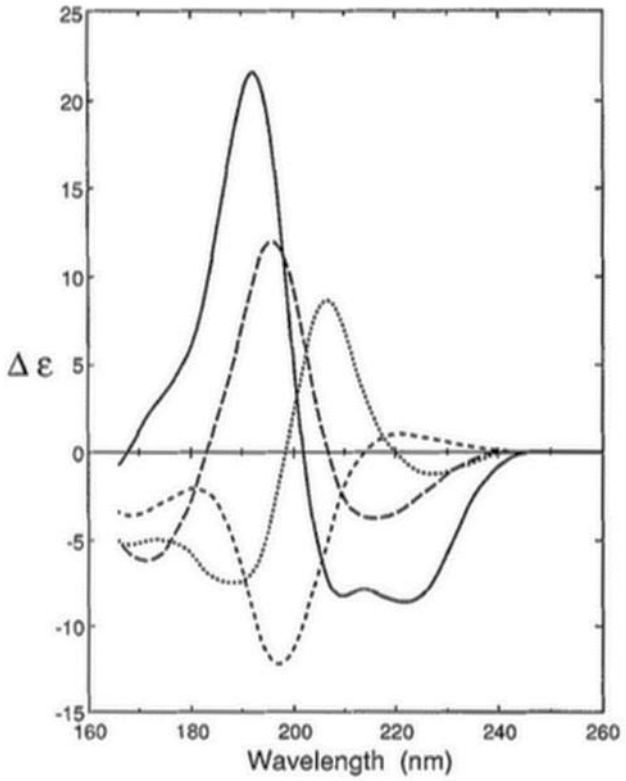


Figure 2.8 Schematic representation of the CD profiles obtained from various protein structures, α-helix (——), β-sheet (— — —), β-turn (·····) and random coil (-----). Reprinted with permission from Johnson *et al.*, *Proteins.*, 1990, **7**, 205-14. Copyright 1990 Wiley-Liss, Inc.

In addition to this qualitative evaluation, quantitative information can be obtained from CD measurements of proteins. For α-helical systems the percentage folding can be obtained using Equation 2.4.[34–36]

$$\%\, folding = \left(\frac{\theta_m - \theta_c}{\theta_h - \theta_c}\right) \times 100 \qquad Eq\ 2.4$$

Where θ_c is the mean residue ellipticity of a random coil (+640 deg cm^2 dmol^{-1}) and θ_h is the mean residue ellipticity of a complete helix as described by Equation 2.5.

$$\theta_h = -42500 \times \left(1 - \left(\frac{3}{n}\right)\right) \qquad Eq\ 2.5$$

Where n is the number of amino acid residues.

2.10 Gel Electrophoresis

Gel Electrophoresis is a technique used to study biomolecules, which separates species dependent upon their size, shape and charge. A gel is made with either polyacrylamide or agarose, and an electrical current is applied across the gel. Molecules move through the gel towards the positive electrode, separating depending upon their size, fold and charge. Highly charged molecules move faster through the gel due to the increased attraction to the electrode. Small molecules also move faster, as they are more freely able to move through the small pores in the gel. Once the run is complete the molecules can be visualised by various techniques and the gel imaged.[37,38]

A variety of techniques have been developed to study proteins, DNA and the interactions of molecules with DNA. Those used within these studies are discussed below.

2.10.1 DNA Gel Electrophoresis

DNA is a highly charged molecule, making it amenable to gel electrophoresis studies. Samples can be loaded onto the gel and, due to the highly negatively charged backbone of DNA, the species migrate towards the positive electrode. As oligonucleotides are considerably smaller than typically analysed DNA fragments, adaptions have to be made to the method to allow the smaller species to be successfully separated and analysed. This is achieved by increasing the percentage of acrylamide used in the gel and hence decreasing the pore size and restricting the flow of the molecules through the gel.

There are two types of gel used within these studies to analyse oligonucleotides; native and denaturing. A native gel maintains secondary structure and interactions between DNA strands, allowing for the interactions of oligonucleotides to be studied and for example the duplex DNA structure to be retained. Denaturing gels disrupt the secondary structure and interactions between strands, using formamide and urea, allowing the size of ssDNA to be more closely analysed and any covalently induced interactions to be assessed.[37]

DNA gels can be imaged *via* several techniques. The simplest is UV shadow. A thin layer chromatography (TLC) plate is placed under a UV lamp with the gel on top. The silica on the TLC plate fluoresces upon excitation at 254 nm, however, where the DNA bands are, the excitation light is absorbed and a shadow is seen on the TLC plate which can then be imaged. Although this is the cheapest and safest way to visualise gels, large quantities of sample need to be loaded on to the gel in order for the bands to be visualised. To overcome this problem many fluorescent dyes have been developed that fluoresce upon binding to DNA, typically through intercalation, such as ethidium bromide and SYBR Gold®. Although these dyes offer lower detection limits than UV shadow, they are toxic and care has to be taken in their use. They also limit the extraction of the DNA from the gel. In addition to these visualisation techniques, radio labelling of the DNA can also be used. This uses an enzyme to attach a ^{32}P to the DNA which can then be imaged.[14] The DNA can also be labelled with a fluorophore and subsequently imaged.

2.10.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is the standard technique used to analyse proteins. Unlike DNA, proteins and peptides are less highly charged. In addition to this they can be positive, negative or neutral making them harder to analyse *via* gel electrophoresis. In order to overcome this issue, all proteins are run using sodium dodecyl sulphate (SDS), a

chemical denaturant that is negatively charged. This results in all proteins having a similar charge and therefore separating solely dependent upon their size. As with DNA, due to the smaller nature of peptides compared with proteins, they are run using gels containing a higher percentage of acrylamide gel. Once run, protein gels can be imaged by a variety of techniques including silver staining, fluorescent dyes or Coomassie staining.[33,38]

2.10.3 Electrophoretic Mobility Shift Assay

In order to analyse binding of a molecule to DNA, a gel electrophoretic mobility shift assay (EMSA) is used. As before, the DNA is loaded onto the gel and migrates towards the positive electrode. In each well of the gel, an increasing amount of the binding molecule of interest is added, and as this is a native technique, the interactions with the DNA are maintained during the running of the gel. The binding of this molecule increases the effective size of the DNA, retarding its movement through the gel, resulting in a new band appearing. As the concentration of the molecule increases, the original DNA band in the gel disappears and eventually the only band that can be seen is that of the higher molecular weight species, corresponding to the bound DNA. The gel is imaged using the same techniques as discussed in Section 2.10.1.[39,40]

2.11 References

1 S. L. Beaucage and R. P. Iyer, *Tetrahedron*, 1992, **48**, 2223–2311.
2 C. B. Reese, *Org. Biomol. Chem.*, 2005, **3**, 3851–3868.
3 M. D. Matteucci and M. H. Caruthers, *J. Am. Chem. Soc.*, 1981, **103**, 3185–3191.
4 S. L. L. Beaucage and M. H. H. Caruthers, *Tetrahedron Lett.*, 1981, **22**, 1859–1862.
5 P. Herdewijn, *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, New Jersey, 2005.
6 S. Agrawal, *Protocols for Oligonucleotides and Analogs: Synthesis and Properties*, Humana Press, New Jersey, 1993.
7 J. Manchester, PhD Thesis, University of Birmingham, 2013.
8 J.-L. H. A. Duprey, PhD Thesis, University of Birmingham, 2010.
9 R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154.
10 W. C. White and P. D. Chan, *Fmoc solid phase peptide synthesis*, Oxford University Press: New York, New York, 2000.
11 M. Bodanszky, *Principles of Peptide Synthesis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.
12 T. Hanai, *HPLC: A Practical Guide*, Royal Society of Chemistry, Cambridge, 1999.
13 L. R. Snyder, J. L. Glajch and J. J. Kirkland, *Practical HPLC Method Development*, J. Wiley, 2nd edn., 1988.
14 V. A. Gault and N. H. McClenaghan, *Understanding Bioanalytical Chemistry: Principles and Applications*, John Wiley & Sons, Chichester, 2009.
15 E. De Hoffmann and V. Stroobant, *Mass Spectrometry: Principles and Applications*, John Wiley & Sons, 3rd edn., 2013.
16 J. Gross, *Mass Spectrometry: A Textbook*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
17 H.-H. Perkampus, *UV-Vis Spectroscopy and its Applications*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.
18 C. R. Cantor, M. M. Warshaw and H. Shapiro, *Biopolymers*, 1970, **9**, 1059–1077.
19 G. D. Fasman, *Handbook of Biochemistry and Molecular Biology, Volume 1: Nucleic Acids*, CRC Press, 1975.
20 S. C. Gill and P. H. von Hippel, *Anal. Biochem.*, 1989, **182**, 319–326.
21 G. L. Ellman, *Arch. Biochem. Biophys.*, 1959, **82**, 70–77.
22 C. K. Riener, G. Kada and H. J. Gruber, *Anal. Bioanal. Chem.*, 2002, **373**, 266–276.
23 J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer US, 3rd edn., 2006.
24 R. Thomas, *Gene*, 1993, **135**, 77–79.
25 J.-L. Mergny and L. Lacroix, *Oligonucleotides*, 2003, **13**, 515–37.

26 A. D. McLaren and D. Shugar, *Photochemistry of Proteins and Nucleic Acids: International Series of Monographs on Pure and Applied Biology: Modern Trends in Physiological Sciences*, Elsevier, 22nd edn., 2014.

27 K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi and E. T. Kool, *J. Am. Chem. Soc.*, 2000, **122**, 2213–2222.

28 P. Yakovchuk, E. Protozanova and M. D. Frank-Kamenetskii, *Nucleic Acids Res.*, 2006, **34**, 564–74.

29 C. Schildkraut and S. Lifson, *Biopolymers*, 1965, **3**, 195–208.

30 B. Norden, A. Rodger and T. Dafforn, *Linear Dichroism and Circular Dichroism: A Textbook on Polarized-light Spectroscopy*, The Royal Society of Chemistry, Cambridge, 2010.

31 V. I. Dodero, Z. B. Quirolo and M. A. Sequeira, *Front. Biosci.*, 2011, **16**, 61–73.

32 W. Curits Johnson, *Proteins*, 1990, **7**, 205–14.

33 T. E. Creighton, *Protein Structure: A Practical Approach*, Oxford University Press, Oxford, 2nd edn., 1997.

34 N. Choy, V. Raussens and V. Narayanaswami, *J. Mol. Biol.*, 2003, **334**, 527–539.

35 J. K. Myers, C. N. Pace and J. M. Scholtz, *Proc. Natl. Acad. Sci.*, 1997, **94**, 2833–2837.

36 C. A. Rohl, A. Chakrabartty and R. L. Baldwin, *Protein Sci.*, 1996, **5**, 2623–37.

37 P. Jones, *Gel Electrophoresis: Nucleic Acids*, Wiley-VCH Verlag GmbH & Co. KGaA, New York, 1995.

38 M. J. Dunn, *Gel Electrophoresis of Proteins*, IOP Publishing Ltd., Bristol, 1986.

39 T. Moss, *DNA-Protein Interactions: Principles and Protocols*, Humana Press, New Jersey, 2nd edn., 2001.

40 A. Travers and M. Buckle, *DNA-Protein Interactions*, Oxford University Press, New York, 2000.

Chapter 3

SNP Detection Using an Anthracene Probe

3.1 Introduction

Within this chapter the use of anthracene as a fluorescent single nucleotide polymorphism (SNP) reporter probe is explored. SNPs are the primary source of genetic diversity and it is thought that many may be indicators of a susceptibility to disease. As such they are currently the focus point of much research.

3.1.1 Single Nucleotide Polymorphisms

Over recent years the understanding of the human genome has increased significantly. This has led to the discovery of regions of DNA that are thought to be responsible for diseases. 90% of the differences in these regions are down to a single nucleotide polymorphism (SNP), where just one base is changed in the sequence, making them the most common form of genetic variation.[1–3] Within each coding region of a gene, there are an average of four SNPs; half of these result in mutations in the proteins the gene codes for.[3] This property arises from the DNA being read in codons (Chapter 1.2.2). If the SNP occurs at a particular position within the codon, the amino acid selected remains the same, resulting in no change to the protein, despite a SNP being present. These are referred to as synonymous or silent SNPs.[4] However, if the SNP does affect the amino acid coded for, a non-synonymous SNP, then the protein function can be altered. SNPs are known to be responsible for cystic fibrosis and sickle-cell anaemia.

3.1.2 Present Detection Methods

One of the most popular and established methods to identify the base at a SNP site, is the commercially available Taqman assay, which works by amplifying DNA using the polymerase chain reaction (PCR), Chapter 1.4.1.3.1. The probe consists of a DNA oligonucleotide with a fluorophore attached, typically to its 5′ end, and a quencher attached to its 3′ end. When the target DNA strand is present, a duplex is formed with the probe. As the PCR process replicates the DNA, *taq polymerase*

degrades the complementary strand, in this case the probe, from the 5′ to the 3′ end, releasing the fluorophore and associated quencher, thereby resulting in fluorescence being observed, Figure 3.1.[5] This technique relies upon the mismatched strand not binding, and hence not being degraded. Therefore, a difference in the binding affinities between the fully complementary and one base mismatch are required, in order for the system to work. As a consequence of this fine balance, the assay is performed at a slightly elevated temperature, typically 50-60 °C. If the temperature is too low, the probe will bind to both strands; if it is too high, neither will bind. Due to the elevated temperatures used and the need for PCR, this assay is unsuitable for use in live cellular environments.

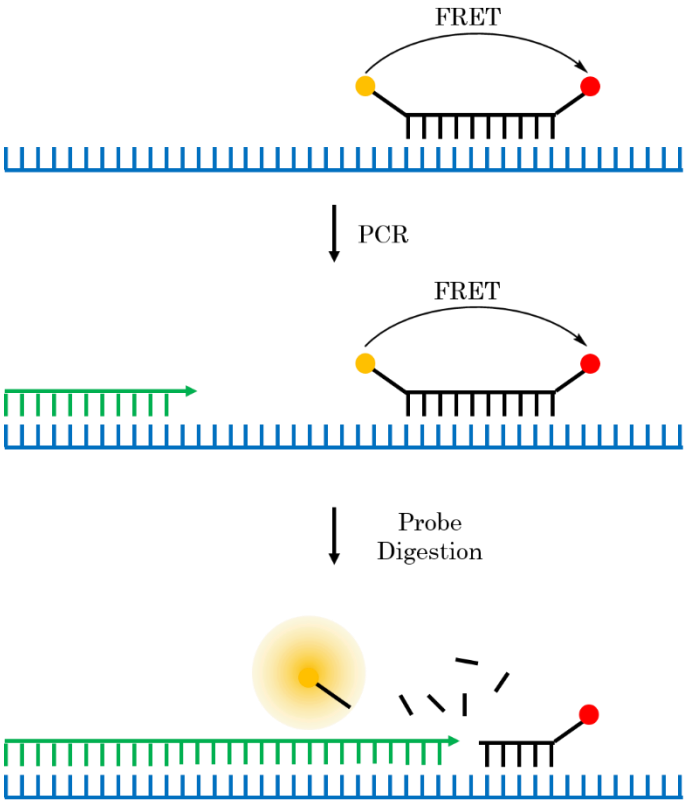


Figure 3.1 Schematic diagram illustrating the principle behind the Taqman assay. The probe containing a FRET pair of fluorophores binds to the target DNA (top). PCR is performed and the complementary strand is produced (middle). As the polymerase approaches the probe, it is digested and the fluorophores are released from close proximity, removing the FRET pathway and hence fluorescence occurs (bottom).

Another popular method is the use of molecular beacons.[6] These use a DNA hairpin structure in which the ends of the DNA probe sequence complement each other, Figure 3.2. Rather like the Taqman System, at one end of the probe is a fluorophore and at the other end is a quencher. In the hairpin form any fluorescence from the probe is quenched due to the close proximity of the quencher group and the fluorescent label. In the presence of target DNA, the probe unfolds to hybridise with the target, due to the presence of more complementary bases compared to the stem. Upon target binding and duplex formation, the fluorophore and quencher become separated and fluorescence is observed. As with the Taqman probe, this process is temperature dependent, since it relies upon a difference in binding affinity for target sequences with different SNP identities.

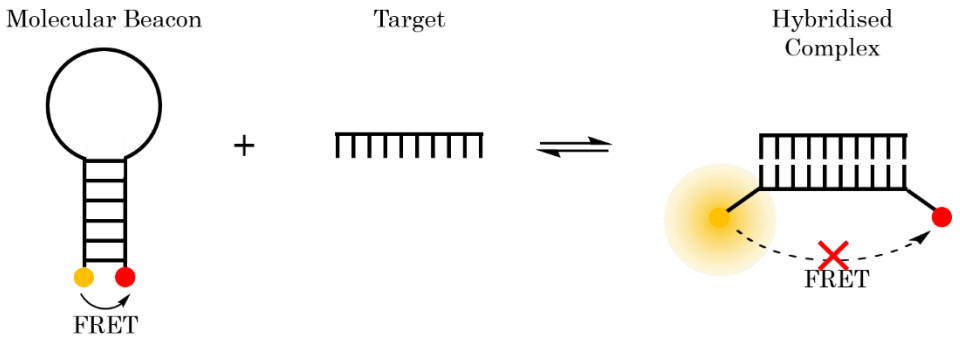


Figure 3.2 Schematic diagram illustrating the process of SNP sensing using a Molecular Beacon. In the stem loop form the fluorophore and quencher are in close proximity and FRET occurs (left). Upon hybridising with the target strand, the stem loop unfolds, separating the quencher from the fluorophore and fluorescence is observed (right).

Scorpion probes have recently been developed which combine the molecular beacon concept with a PCR reaction. These probes contain the primer for the PCR reaction joined to a molecular beacon, the loop of which is the same sequence as the target site (i.e. complementary to the PCR product being formed). Upon extension of the

primer, the molecular beacon unfolds with the loop region binding to the newly formed PCR product, leading to a structural rearrangement spatially, repositioning fluorophore and quencher, resulting in fluorescence, Figure 3.3. Due to the molecular beacon region being covalently attached to the PCR product, the annealing process is intramolecular and therefore faster and able to give more reliable results than the Taqman assay.[7]

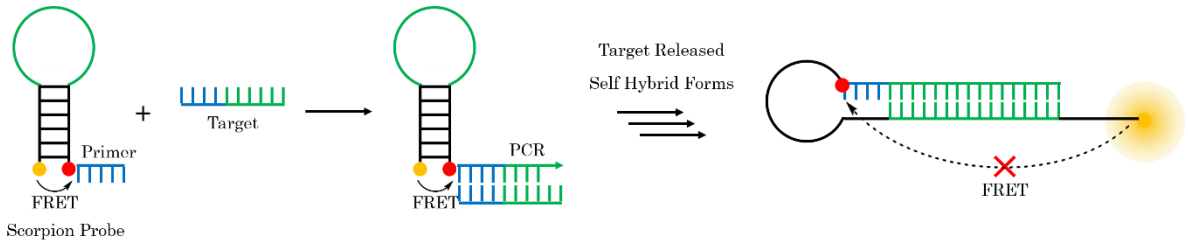


Figure 3.3 Schematic representation of the Scorpion Probe process. A primer is attached to a stem loop forming sequence in which a fluorophore and quencher are held in close proximity (left). PCR is performed, extending the primer, this extended sequence is complementary to that in the loop region of the stem loop (middle). Upon the next cycle of PCR, the target is released and the stem loop opens to hybridise with the newly formed sequence. This increases the distance between the fluorophore and quencher and hence fluorescence is observed (right).

3.1.3. Limitations of Present Methods

The Taqman, molecular beacon and scorpion probe methods all have the same limitation, they rely upon a difference in binding affinity of the probe to strands containing different SNP identities. As discussed above, this means that the assays have to be performed at an elevated temperature and more importantly, in a well-defined temperature range that is difficult to optimise and is never the same for any two SNPs. Another limitation is that the above methods are only able to identify the presence or absence of the base they are designed to. It would instead be more useful if a probe could be developed that could identify the base that is present, thus determining the sequence of DNA. This could enable biologists to

establish whether the SNP will have an effect on the amino acid that is coded for, and therefore an effect on phenotype.

3.1.4. Base-discriminating Fluorescent Nucleotides

By replacing a base within an oligonucleotide with a fluorescent molecule, the base that is opposite, or next to the probe can potentially be identified by the changes that are observed in fluorescence upon hybridisation. Seitz and co-workers have recently used a Peptide Nucleic Acid (PNA) with a thiazole orange fluorophore incorporated as a SNP sensor. The PNA binds to the DNA target, with the thiazole orange tag intercalating with the DNA bases. When a matched DNA base pair is present next to the tag, it is forced into a coplanar arrangement and fluorescence is observed. However, when a mismatch is present, the tag can move more freely and undergo torsional motions, depleting the fluorescence observed. This is shown schematically in Figure 3.4.[8,9]

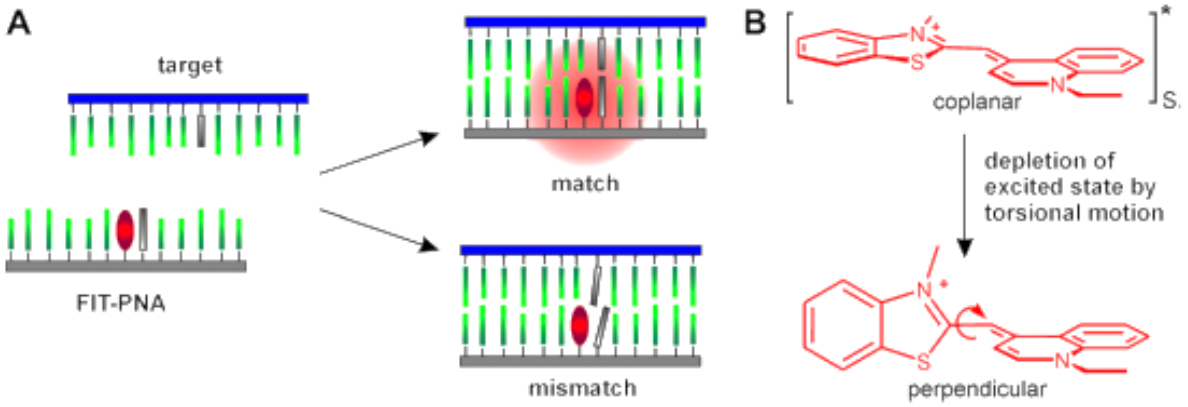


Figure 3.4 Schematic of the probe used by Seitz and co-workers to probe for SNPs within a DNA sequence. A) The FIT-PNA probe is hybridised with target DNA. If a fully matched sequence is formed, the thiazole orange tag is held in the coplanar conformation (B, top) and fluorescence is observed. However, when a mismatch is present in the sequence, the thiazole orange moiety can form the perpendicular arrangement (B, bottom) and fluorescence is depleted. Reprinted with permission from Socher *et al.*, Anal. *Biochem.*, 2008, **375**, 318-30. Copyright 2008 Elsevier, Inc.

The same group have also used a pyrene moiety attached to an adenine derivative to identify when a thymine base was opposite the probe, Figure 3.5.[10] When thymine is present the adenine forms a Watson-Crick base pair with it, forcing the pyrene into solution and fluorescence occurs. However, when thymine is not present (i.e. a SNP) there is no base pair formed and the pyrene moiety intercalates. In this arrangement the fluorescence is quenched by the surrounding DNA bases and hence no fluorescence is observed.[10,11]

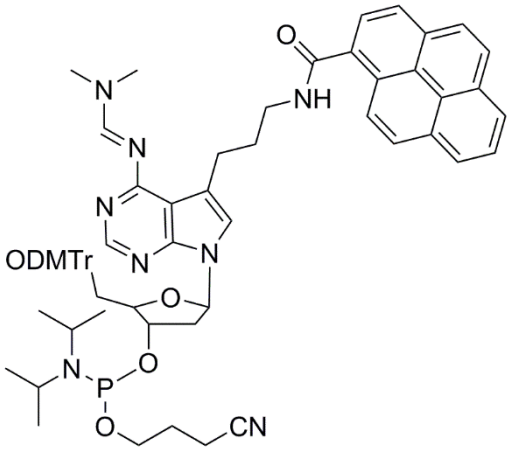


Figure 3.5 Pyrene phosphoramidite tag for incorporation into DNA probe. The pyrene moiety is attached to an adenine base to allow for the sensing of thymine on the complementary strand.

Another recent SNP detection method made use of a cyclodextrin tagged DNA sequence.[12] This binds to the complementary sequence of DNA up to the point of the SNP. A nucleobase specific fluorescent probe is then added. This molecule is made up of the nucleotide binding region, designed to only bind to a specific base, which is attached to a dansyl group. If the corresponding base is present, the probe binds, placing the dansyl group inside the cyclodextrin resulting in fluorescence,

Figure 3.6. Although the group has so far only made a probe to bind to guanine, they expect to be able to make probes for all four bases.

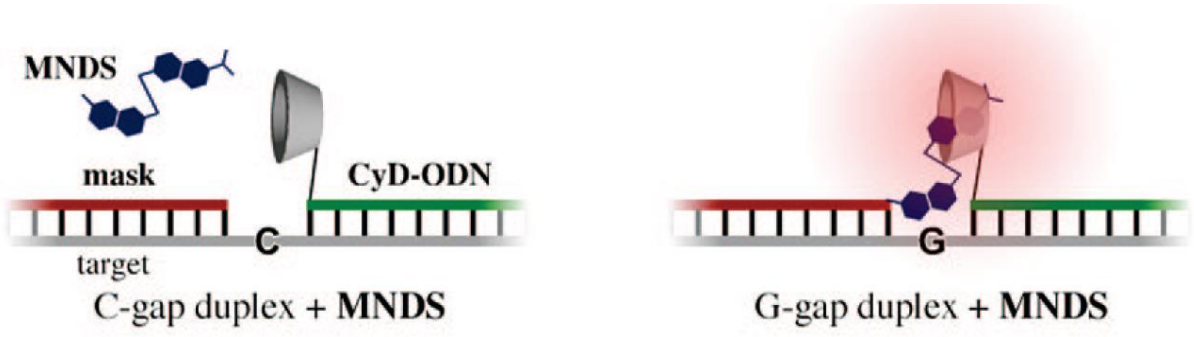


Figure 3.6 Schematic representation of cyclodextrin tagged DNA for SNP detection. An oligonucleotide with a cyclodextrin moiety attached, is bound up to the SNP site. A small molecule specific for the base of interest is introduced. If the complementary base is present, it binds and places a dansyl group within the cyclodextrin, resulting in fluorescence. Reprinted with permission from Ihara *et al.*, *J. Am. Chem. Soc.*, 2009, **131**, 1386-7. Copyright 2009 American Chemical Society.

Previous work within the Tucker group has focused on developing and utilising an anthracene tag, for incorporation into DNA as a fluorescent probe, briefly described in Chapter 2.3. The anthracene tag has been shown previously to display significantly different fluorescence intensities when placed opposite all four possible bases in a target DNA sequence.[13–15] The anthracene tag incorporates a carbon chain linker, the length of which can be varied, and a threoninol unit which allows for control of the stereochemistry, Figure 3.7. Both of these properties can be altered and provide a mechanism by which changes in fluorescence can be optimised.

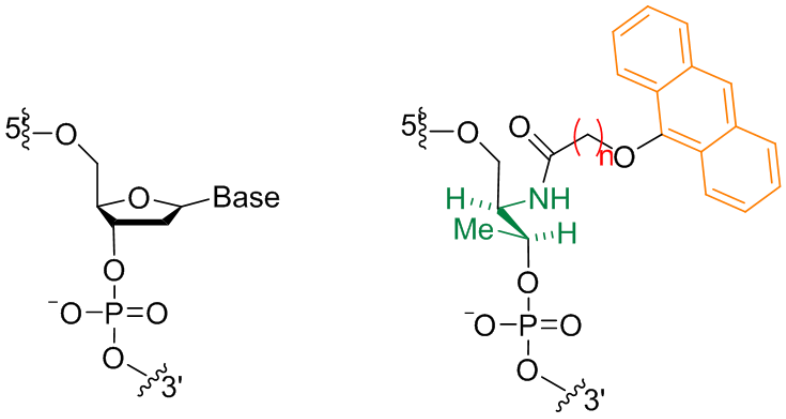


Figure 3.7 Comparison of the Tucker group probe (right) to natural DNA (left). The anthracene probe can be altered by varying the length of the carbon chain connecting the anthracene to the DNA backbone (n, in red) and the stereochemistry of the threoninol unit (green) to provide optimum differences between the various bases.

More recently Saito *et al.* have used anthracene incorporated into the DNA *via* uridine, which allows for the detection of adenine when located directly opposite the probe. In contrast, when other bases are located opposite, a quenching in the anthracene fluorescence is observed.[16]

The approach used within base discriminating fluorescent nucleotides (BDF's), removes the problem of working at a particular elevated temperature window, as needed with the popular Taqman assay, due to the probe sequence binding to the target strand irrespective of the SNP site base identity. Within the Tucker group probe, the change in fluorescence arises from the differences in intercalation of the anthracene, and any interactions it has with the adjacent and opposite bases. The research is now moving towards targeting relevant sequences known to be associated with diseases such as prostate cancer and Alzheimer's disease.[17]

3.2 Project Aims

Within this chapter the use of an anthracene SNP detection probe is assessed within a sequence of DNA thought to be associated with Alzheimer's disease. Various stereochemistry and linker length combinations are incorporated into a sequence of DNA complementary to the SNP site, with the anthracene incorporation being made directly opposite the base that is expected to change. The fluorescence of anthracene is then measured and tracked upon duplexing with the complementary DNA. Increases in fluorescence intensity are observed for some bases but decreases for other, Figure 3.8, allowing the base identity to be determined. Further to this, analysing a mixture of two bases is expected to give a fluorescence output proportional to the amount of each base present allowing for quantitative determination of the bases present.

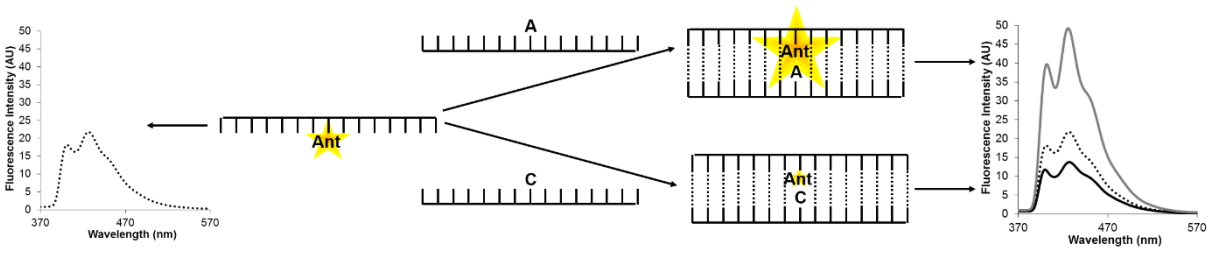


Figure 3.8 Anthracene is incorporated into a strand of DNA complementary to the strand of interest. Upon duplex formation a change in the intensity of fluorescence is observed depending on the base that is present.

In addition, the use of the anthracene probe to detect the base modification 8-oxoG will be assessed.

3.3. Anthracene SNP Sensing

3.3.1. Alzheimer's SNP Sensing Sequences

To further prove the applicability of the technique, it was decided that a sequence associated with Alzheimer's disease (rs1801270) should be investigated.[18,19] The target sequence is one of several SNP sites thought to be prevalent in Alzheimer's disease. The native sequence contains a C base which is altered to an A with a higher rate in Alzheimer's patients (Table 3.1). This causes a mutation of a serine to an alanine in the protein that the gene codes for, resulting in complications in the neuron cell cycle and fibril formation. An anthracene monomer was incorporated into a complementary probe strand directly opposite the site of the SNP. Anthracene probes of stereochemistries L and D and linker lengths of 1, 3-7 were synthesised as described in Chapter 2.3. A linker length of 2 was not used as it was previously established within the group that upon saponification to the carboxylic acid, a competing, faster elimination reaction yielded anthraquinone instead of the desired product.[20] Probe and target strands were synthesised using standard solid-phase oligonucleotide synthesis, purified by RP-HPLC and characterised by ESI-MS (Chapter 2). This was performed by Dr Zheng-Yun Zhao on some strands and these are indicated in the appendix.

Table 3.1 Sequences synthesised for SNP detection trials on Alzheimer's associated mutation. Probes labelled according to the linker length (n) and stereochemistry of the threoninol unit (L/D).

	Name	Sequence 5′ – 3′
Probe	Alz-Pr(nL/D)	AGT CGC G**X**C TCA GCT
Target	Alz-TarA/C	AGC TGA G**A/C**C GCG ACT

3.3.2. Fluorescence Spectroscopy Studies

Fluorescence spectra of the probe alone were recorded between 370 – 570 nm in the absence and presence of target strands, containing either SNP identity. After addition of each target strand, a change in the intensity of the fluorescence was observed, Figure 3.9.

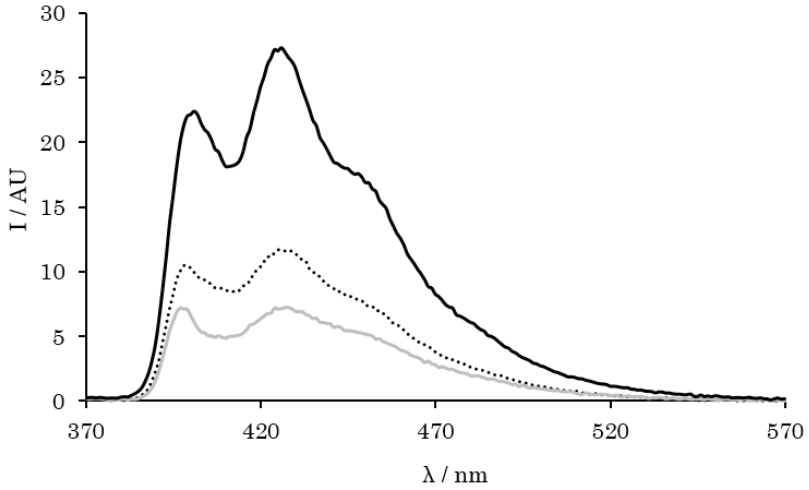


Figure 3.9 Fluorescence spectra of **Alz-Pr(5L)** (dot) in the presence of **Alz-TarC** (grey) and **Alz-TarA** (black), performed at 1 μM probe with 1.5 μM target, where applicable, in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, λ_{ex} 350 nm, 293 K.

In order to compare the fluorescence response between various probes, the percentage change in fluorescence at 426 nm was plotted as a function of increasing equivalence of target strand (Figure 3.10). It was found that after additions exceeded 1 molar equivalent of target to probe, there was little change in fluorescence, indicating that a 1:1 duplex had been formed.

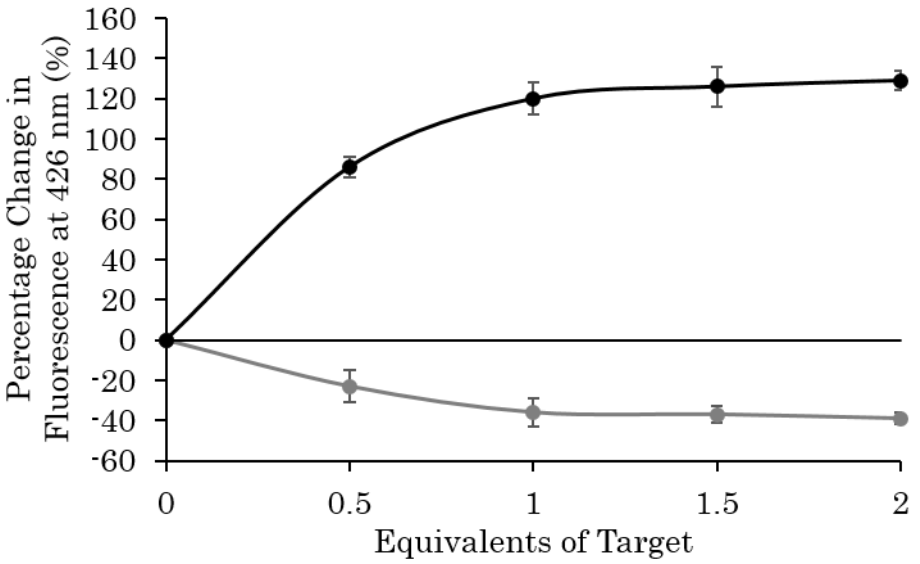


Figure 3.10 Fluorescence titration of **Alz-Pr(5L)** with **Alz-TarC** (grey) and **Alz-TarA** (black) performed at 1 μM probe in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, λ_{ex} 350 nm, 293 K. Error bars correspond to 1 standard deviation.

Titrations were performed in triplicate for all probe and target combinations, the results of which are summarised in Figure 3.11. It was found that in all the longer linker length probes (3-7) there is an increase in the fluorescence intensity upon addition of **Alz-TarA**, and a decrease with **Alz-TarC**. The **Alz-Pr(1L)** and **Alz-Pr(1D)** probes show little discrimination between the two targets. These results are in good agreement with those for a model system studied extensively by the group.[21] It is proposed that the 1L and 1D linkers are too short for the anthracene to fully interact within the duplex. However, the longer linkers allow for intercalation of the anthracene, and an interaction with the base opposite as well as those upstream of the incorporation site. It is thought that the change in fluorescence is caused by a combination of quenching from neighbouring bases, the amount of intercalation and photo-induced electron transfer (PET) to the bases. Interestingly, **Alz-Pr(5L)** shows a significantly larger increase in fluorescence when forming a

duplex with **Alz-TarA** than any of the other probes. The changes in fluorescence intensity observed are not as large as those found with other BDF's.[9,11] Despite this, they are significant and can easily be detected. In addition to this, the probes used within these studies can distinguish all bases,[21] a property that other BDF's do not possess.

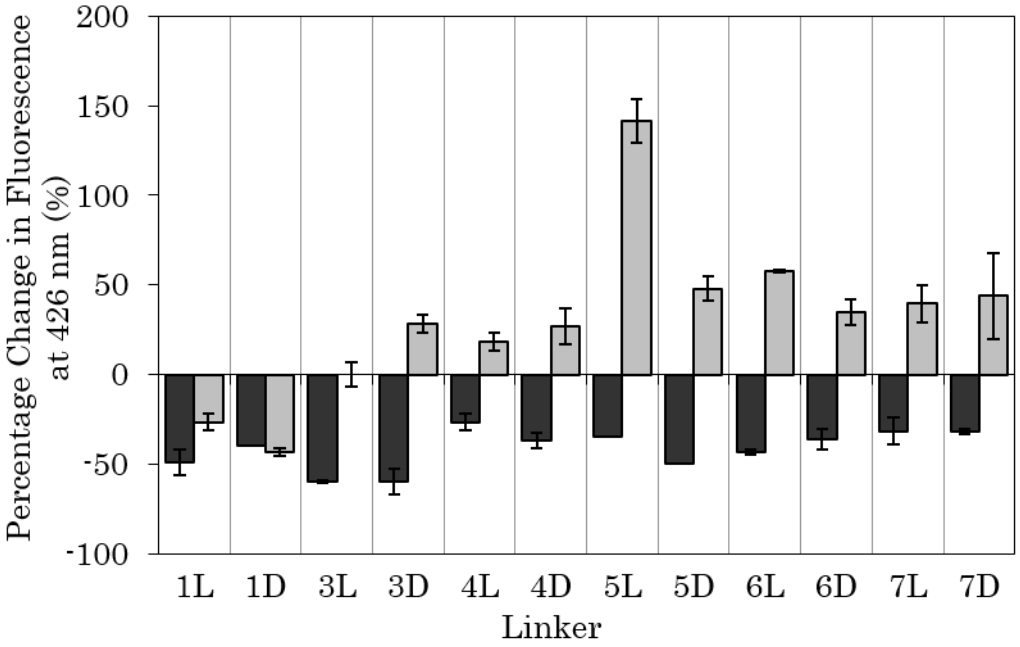


Figure 3.11 Fluorescence titration of all **Alz-Pr** probes with **Alz-TarC** (black) and **Alz-TarA** (grey) performed at 1 μM probe in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, λ_{ex} 350 nm, 293 K. Values taken after addition of 1.5 μM target and are the average of three repeat runs. Error bars correspond to 1 standard deviation.

3.3.3. Thermal Melting Studies

Thermal melting studies (as described in Chapter 2.7.3), were performed on the strands to ensure they were duplexes at room temperature, and to determine the extent of any differences seen in binding strength. It was found that all systems formed stable duplexes, with recorded T_m values being above both room and at the

biological relevant temperature 37.5°C, Table 3.2. This confirmed that the systems studied by fluorescence were fully hybridised at the temperature used (293 K, 20°C). In all cases there is a reduction in T_m upon replacement of a base with the anthracene probe. This is expected due to the removal of the stabilising hydrogen bonds provided by the base that has been removed. However, the anthracene is able to intercalate (depending upon the linker length and stereochemistry allowing the correct orientation) with the neighbouring bases, resulting in higher T_m values than would be expected for the complete removal of a base, based on work previously performed within the group.[21]

Table 3.2 T_m values (°C) for duplexes formed between **Alz-Pr** and **Alz-Tar**. Values are compared to unmodified target (UM) in which a G or T base replaces the anthracene probe when the C or A target is used, respectively. Studies are performed at 5 μM duplex oligonucleotide in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl. All values are an average of at least 3 repeat runs and are reported to the closest 0.5 °C due to the limits of instrument error.

	Probe **Alz-Pr**											
Target	UM	1L	1D	3L	3D	4L	4D	5L	5D	6L	6D	7L	7D
Alz-TarA	67.0	58.5	58.0	61.0	62.5	59.0	61.5	60.0	59.0	59.0	61.5	59.0	61.5
Alz-TarC	73.0	57.5	55.0	62.0	61.0	63.0	61.5	63.0	60.5	61.5	63.0	62.5	62.0

Notwithstanding the closeness of the values, in the majority of cases, the duplex formed with **Alz-TarC** has a higher T_m than that formed with **Alz-TarA**, indicating that a more stable duplex is formed in these cases. A potential explanation for this is the relative sizes of cytosine, a pyrimidine, and adenine, a purine. The smaller pyrimidine increases the cavity size in the duplex, allowing for the anthracene to fully intercalate, and the duplex to assemble into a more natural DNA

conformation. However, the larger purine could disrupt the local environment more, disrupting the bases surrounding the anthracene site, resulting in a reduction in the T_m. This observation is consistent with the fluorescence data for the longer linker length probes; it is possible that when the anthracene is able to intercalate more, it could interact with the upstream bases more, leading to fluorescence quenching as seen in Figure 3.10.

The increase in fluorescence observed in the case of the adenine target, could be due to the anthracene being removed from a solvent exposed environment upon duplex formation, restricting the quenching pathway that water offers. However, despite some degree of intercalation, in this system the anthracene cannot stack effectively with the bases upstream due to the limited space afforded by the adenine base, resulting in less quenching compared to the cytosine case and consequently a slightly lower T_m and a relative increase in the fluorescence. Molecular modelling of a control system has been performed previously and showed an intricate relationship between the anthracene unit, the base upstream of it and the bases directly opposite.[21]

The T_m values of the duplexes formed with **Alz-Pr(1L)** and **Alz-Pr(1D)** are considerably lower than those with the longer linker lengths. This is proposed to be due to the relative rigidity of these linkers not allowing for the intercalation of

the anthracene and therefore the systems do not benefit from the favourable stacking interactions formed between anthracene and the bases.

3.3.4. Circular Dichroism Studies

Circular dichroism (CD) studies were performed to evaluate the secondary structure of the duplexes formed between the anthracene probe and the DNA targets. Previous studies on a model system showed that all duplexes formed a B-DNA structure, with some showing an induced signal from the anthracene at 254 nm.[21] As expected, the duplexes studied herein all formed B-DNA structures with a positive band centred ~275 nm and a negative band centred ~240 nm, Figure 3.12 and 3.13.

Studies on the single stranded probes for linker lengths of 3 and above, showed them to be well folded, with an induced anthracene peak at 254 nm, the intensity of which varies between probes. It is expected that secondary structure is forming out of the single strand so as to remove anthracene from the polar water environment. CD spectra of single stranded probes showed the **Alz-Pr(1L)** and **Alz-Pr(1D)** to be similar to the expected signal for unmodified single stranded DNA, with no induced anthracene peak at 254 nm. This is expected to be due to the short linker being more rigid and restricting the amount of folding that can occur in order for the bases to stack with the anthracene. A comparison to the unmodified ssDNA and dsDNA strands can be found in the appendix (Chapter 9.1).

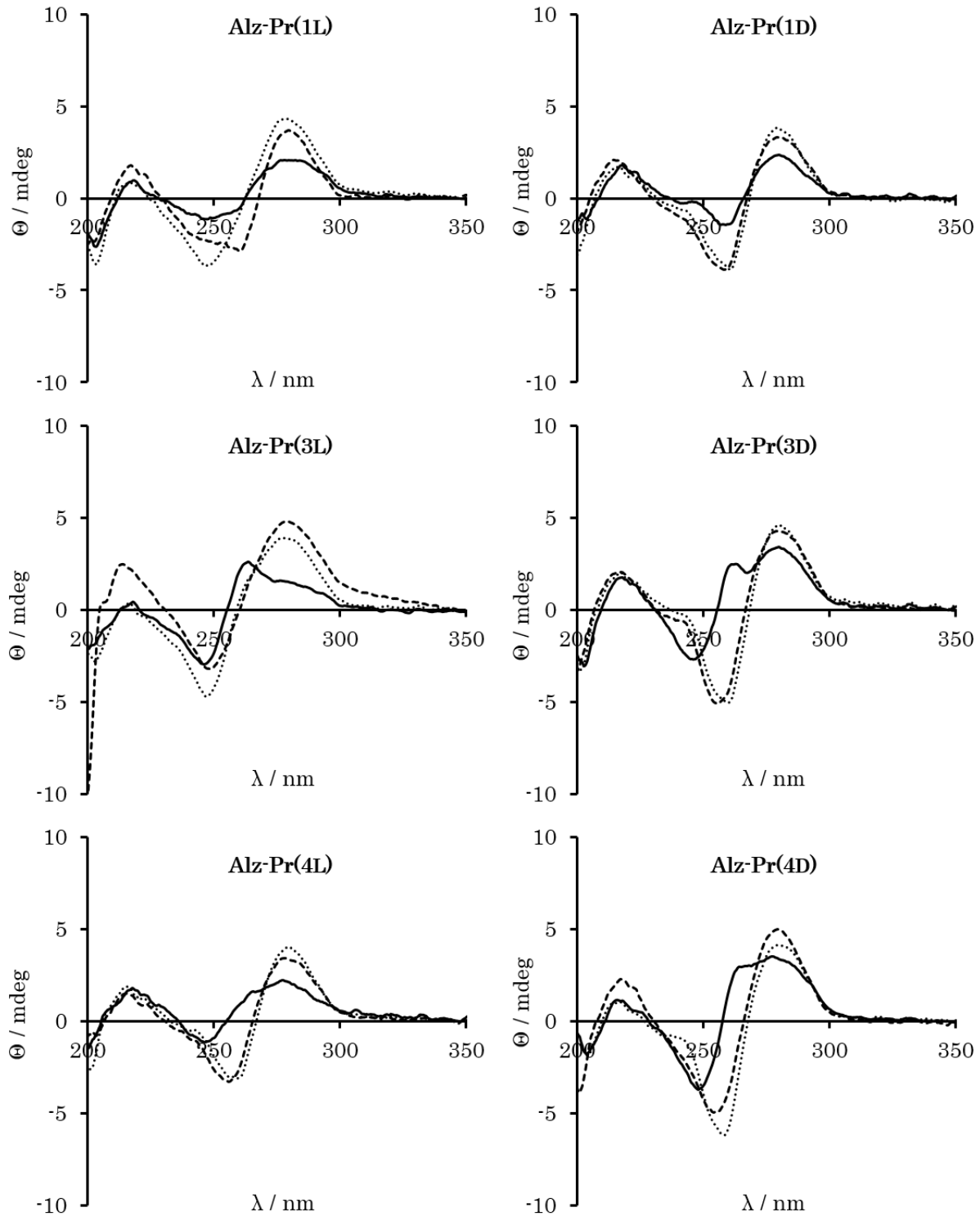


Figure 3.12 Circular Dichroism spectra of all **Alz-Pr** and **Alz-Tar** combinations, probe alone (—), with **Alz-TarC** (····) and with **Alz-TarA** (----). All spectra shown are recorded with 1 µM probe and target, where applicable, in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl at 293 K.

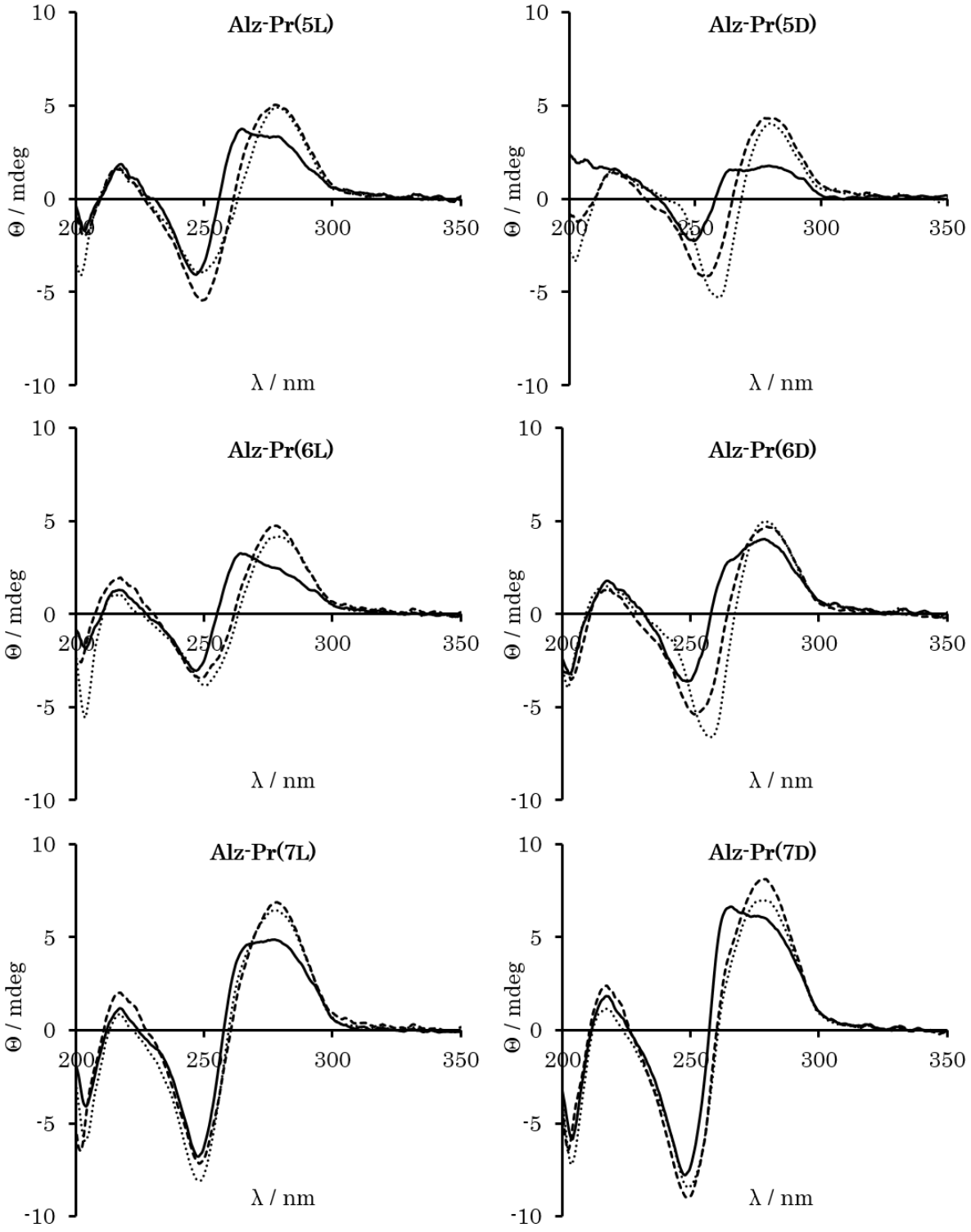


Figure 3.13 Circular Dichroism spectra of all **Alz-Pr** and **Alz-Tar** combinations, probe alone (—), with **Alz-TarC** (····) and with **Alz-TarA** (----). All spectra shown are recorded with 1 µM probe and target, where applicable, in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl at 293 K.

Upon addition of target DNA, an increase in the intensities of the positive band around 280 nm and the negative band around 245 nm was noted. This would be expected due to the formation of duplex B-DNA between the target and probe. It was also found that the anthracene induced signal present in the spectrum for the single stranded probes, was not visible upon duplex formation. This is possibly caused by the anthracene no longer being able to lay at the optimum angle (45° in relation to the base pair long axis) for an induced signal to occur, and therefore the signal is reduced.[21,22]

In general, the duplex formed with **Alz-TarA** shows a blue shift of the negative band at ~250 nm in comparison to the duplex formed with **Alz-TarC**. The positive band shows little change in position but some change in intensity with the **Alz-TarA** duplex displaying a higher intensity. Of particular interest is the case of **Alz-Pr(5L)** with **Alz-TarA**; this is the only duplex to show a marked increase in intensity of the negative band in relation to the duplex formed with **Alz-TarC** and a significant increase in the fluorescence intensity upon duplex formation with the **Alz-TarA** target. Within this region, signals from the B-DNA structure as well as induced signals from the anthracene are known to contribute, making assigning the change to a particular conformation of DNA, or orientation of the anthracene, quite difficult. [23,24]

3.4. Anthracene Heterozygous SNP sensing

3.4.1. Heterozygous SNP sensing

So far the studies presented herein have focussed on the identification on samples with homozygous SNP sites, *i.e.* with only one of the bases present. Of increasing interest to medical professionals is the study of heterozygous SNPs. These arise as a result of everyone having two copies of the same gene, one paternal and one maternal. These copies vary in their sequence and therefore different base identities may be present in one and not the other, Figure 3.14.

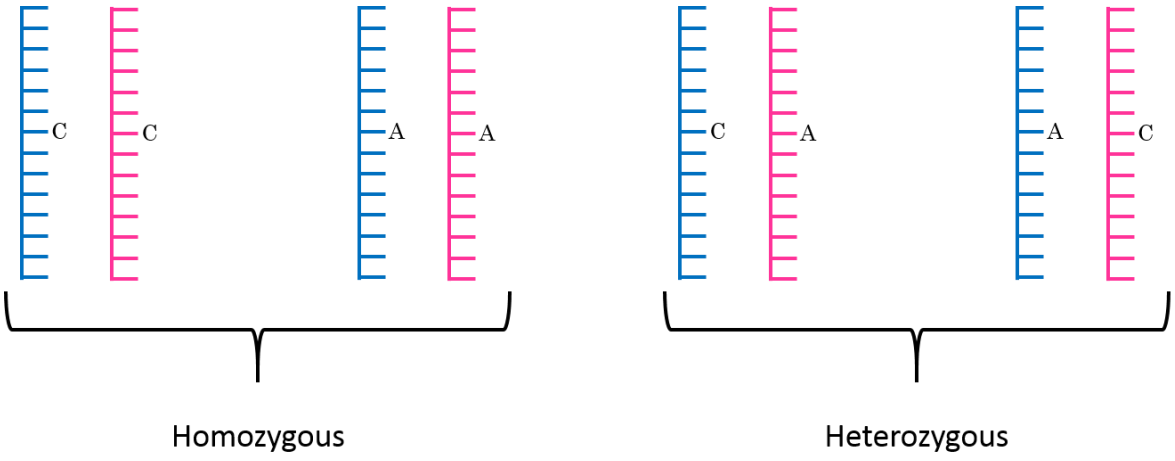


Figure 3.14 Schematic representation of the possible base combinations at a C/A SNP site with paternal DNA shown in blue and maternal DNA in pink. The two DNA strands can contain the same base, homozygous as seen on the left, or have different bases, heterozygous as seen on the right.

Previously the possibility of sensing the presence of heterozygous DNA using the anthracene probe has been explored with a prostate cancer SNP.[17] In this case a T/A SNP site was analysed with a mixture of 50% T base and 50% A base. It was found that the intensity of the anthracene fluorescence was in between the results obtained for the homozygous samples.

During the transcription process (Chapter 1.2.2), the DNA is converted into mRNA and, in heterozygous cases, this process may result in one allele being transcribed in preference to the other. This therefore alters the effect seen within a patient due to the SNP. Unusual allelic ratios of mRNA may be a marker for disease progression or susceptibility. Therefore the ability to assess how much one copy of a gene is transcribed compared to another, would be beneficial.

The SNP determination examples discussed in Section 3.1.2, the Taqman assay, molecular beacons and scorpion probes, can be made into systems that can study heterozygous cases. Two probes can be used, one for each allele, which contain different fluorophores. The probe that displays the highest binding (*i.e.* the one with the correct sequence) will bind and display fluorescence, the wavelength of which corresponds to the base that is present. The wavelength can then be read out and hence the base determined. These methods are limited by their need for a specific temperature window restricting their use in cellular environments.

In addition to this, there are a few methods currently available specifically designed for heterozygous sample determination. Firstly, PCR-based methods relying on complex design utilising SNPs that are present in restriction sites, followed by gel electrophoresis analysis of the PCR fragments.[25,26] An enzyme is used that cuts the DNA at a specific sequence, the restriction site. If this site is altered by the presence of an alternative base, then the enzyme does not recognise

the site and hence does not cut the DNA and it remains the full length. The fragments are then analysed by gel electrophoresis and according to the size of the DNA, either cut or full length, the SNP ratio can be deduced. This method cannot be applied in all cases due to the need for SNP sites to be in a restriction site. Furthermore, the quantification of the separated fragments can be inaccurate.

A second method is the commercially available SNaPshot system (Applied Biosystems Ltd.) which is based upon the dideoxy single base extension technique.[27–29] This method uses single-base extension of a primer. The primer is made to be complementary up to the SNP site, this primer is then extended by enzymes using fluorescently labelled bases. Upon incorporation into the sequence the fluorophore emission is read, with the wavelength of the fluorescence being specific to the base that has been incorporated, allowing for the read-out of the base present and the relative quantities of each base.[30]

Pyrosequencing has also been used for the quantification of allele expression.[31] In this technique single stranded DNA is tethered to a surface and forms a template for DNA polymerase to form the complementary strand. Upon incorporation of a base, pyrophosphate is released which is converted into adenosine triphosphate (ATP) by ATP sulfurylase. The amount of ATP is then quantitatively determined using the fluorescence produced by the firefly luciferase, upon consumption of ATP.

By repeating the process in cycles, using one base in turn, the sequence of DNA can be determined.[32]

All the techniques described above rely on either the PCR reaction, elevated temperatures or the DNA being analysed being tethered to a surface, meaning that they cannot be used in cellular environments. In many cases the technologies are also limited by the sequences they work on. If the assay developed by the Tucker group could be used to quantify the amounts of each base present at a SNP site, the technology could potentially be used in situations where the other assays fail or as an alternative. This could allow for real time tracking of allele specific expression within biological systems.

3.4.2. Heterozygous SNP Sensing using an Anthracene Probe

In order to evaluate whether the Tucker group probe could be used for this application, several target samples with varying ratios of wild type and SNP DNA present, were prepared. In order to maximise the sensitivity of the assay, the probe showing the largest difference between the C/A bases used in the Alzheimer's SNP sensing was used, **Alz-Pr(5L)**. The fluorescence intensity of the probe was monitored followed by addition of 1.5 equivalents of target mixture. The fluorescence intensity was then re-measured and the percentage change in

intensity at 426 nm calculated. It was determined that there was a linear trend between the ratio of base concentrations and fluorescence intensity, Figure 3.15.

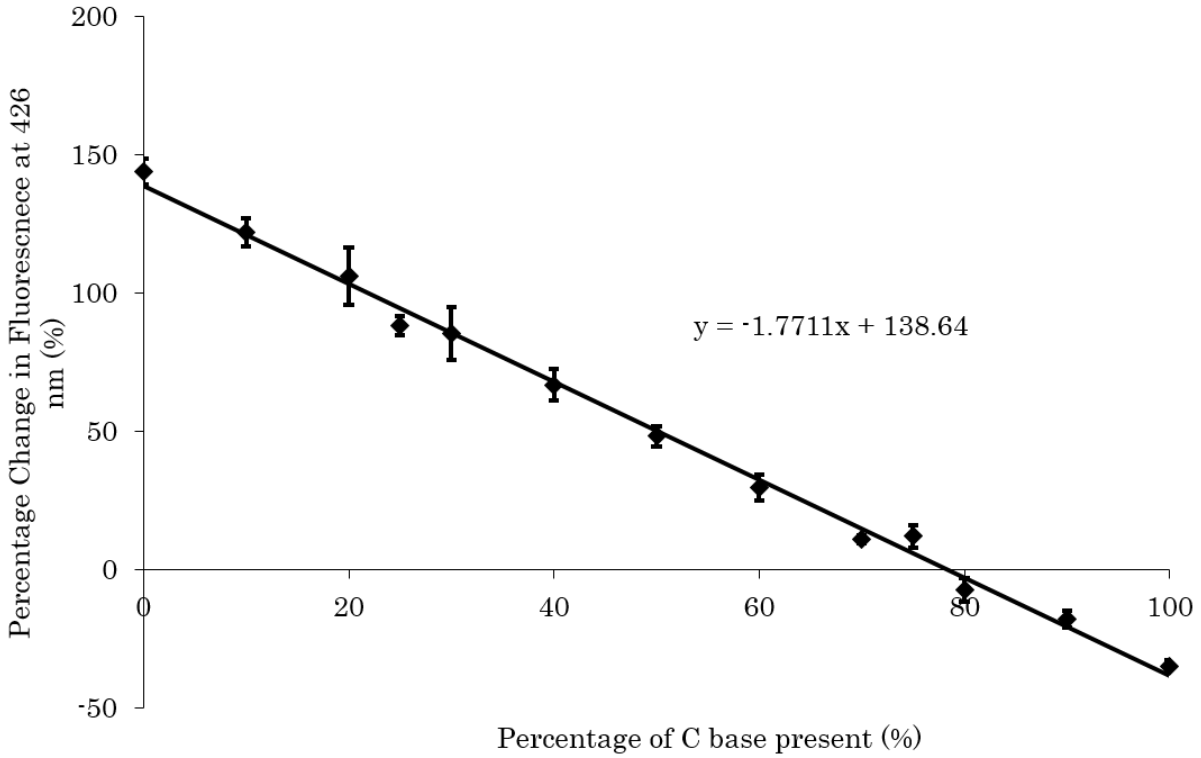


Figure 3.15 Calibration of heterozygous SNP quantification, where remaining percentage of base on x-axis is made up with A base. Performed on 1 mL samples with 1 µM probe and 1.5 µM target in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, λ_{ex} 350 nm, excitation and emission slit widths were 3 and 5 nm respectively, performed at 293 K. Data shown is the average of three repeat runs. Error bars correspond to 1 standard deviation.

Using the calibration in Figure 3.15, blind heterozygous samples of **Alz-TarA** and **Alz-TarC** were analysed to determine the percentage of C/A base present. The results were calculated using the equation of the linear trend line:

$$y = -1.7711x + 138.64$$

The y value was obtained by calculating the percentage change in fluorescence at 426 nm in relation to the probe alone fluorescence. The results of this trial can be seen in Table 3.3 and are shown to be very close to the actual values of the blind

samples. The results obtained when no target is present, highlights an important limitation of this assay. As the calibration crosses the x-axis, at this point you would be unable to tell if the assay was working or if there was ~80% of target with C base and 20% with A base present. This issue could be circumvented by running an identical assay with a different anthracene linker length, and/or stereochemistry, which would cross the x-axis at a different value. This would also give a double verification of the results obtained.

Table 3.3 Trial of anthracene probe heterozygous SNP sensing using the calibration shown in Figure 3.14. All samples run at 1 μM probe in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl. λ_{ex} 350 nm, 293 K.

Percentage Change in Fluorescence at 426 nm (%)	Calculated Percentage of C base present (%)	Actual Percentage of C base present (%)
+28	62	60
-37	99	100
+34	59	55
0	78	No target

Using the above calibration, quantities of allelic ratios in unknown samples would be able to be determined. However, in order for the assay to be biologically compatible, even with PCR amplification, the amount of target required would need to be reduced. Subsequently the volume was reduced to 50 μL in an ultramicro quartz cuvette and analysis was performed on a more sensitive fluorimeter (Edinburgh FLSP920 fluorescence spectrometer). A longer excitation wavelength of 370 nm was also used as the anthracene probe shows a higher absorbance at this wavelength and hence the fluorescence output should be increased. The slit widths used were also increased to allow more light to irradiate the sample and reach the detector. This results in a reduction in the resolution of the fluorescence

spectra obtained, but it was deemed more important to increase the intensity of the signals observed. As detailed previously, a calibration graph was produced which closely matches that shown in Figure 3.15. Due to the lower volume cuvette increasing the scatter within the instrument, the errors found in this case were much higher, Figure 3.16.

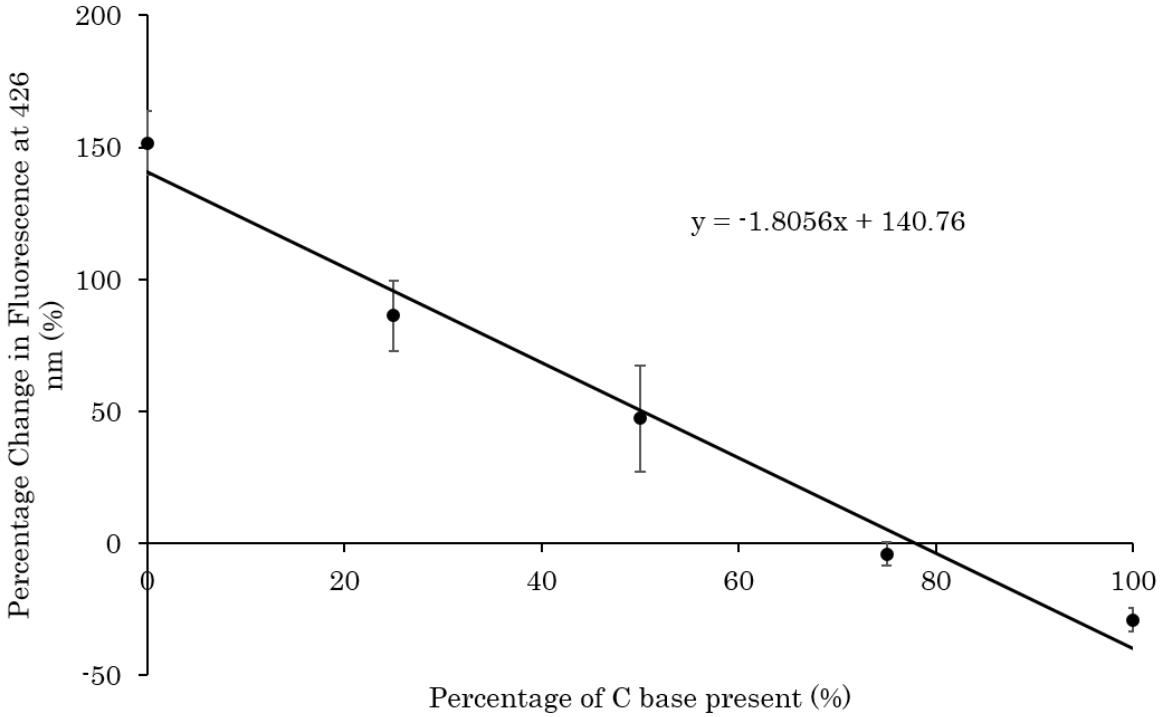


Figure 3.16 Calibration of heterozygous SNP quantification where remaining percentage of base on x-axis is made up with A base. Performed on 50 µL samples with 1 µM probe and 1.5 µM target in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, λ_{ex} 370 nm fitted with 375 nm long pass filter, excitation and emission slit widths were 8 nm, performed at 293 K. Data shown is the average of three repeat runs. Error bars correspond to 1 standard deviation.

Using the calibration in Figure 3.16, unknown heterozygous samples of **Alz-TarA** and **Alz-TarC** were analysed as done previously, using the new equation of the linear trend line:

$$y = -1.8056x + 140.76$$

The results of this trial can be seen in Table 3.4 and, although reasonably accurate, do show a greater error than that found when studying larger volumes. This is an inherent problem of using a smaller volume cuvette that increases the scatter of light. By optimising the fluorescence output of the probe, larger volumes at lower concentrations could be used. This would maintain the amount of material needed but decrease the scatter seen and hence decrease the error.

Table 3.4 Trial of anthracene probe heterozygous SNP sensing using the calibration shown in Figure 3.15. All samples run at 1 μM probe in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, 50 μL. λ_{ex} 370 nm fitted with 375 nm long pass filter, 293 K.

Percentage Change in Fluorescence at 426 nm (%)	Calculated Percentage of C base present (%)	Actual Percentage of C base present (%)
+34	59	50
+76	36	25
+14	70	60
-20	89	90
+13	71	75
+9	73	80
+59	45	50
+60	45	40

These results are extremely promising and demonstrate for the first time that a BDF can be used to quantitatively determine the ratio of bases at a SNP site. This could offer a solution to study DNA sequences in situations where the currently used methods are unable to be used, for example within cells or in G-C rich sequences which cannot be easily amplified.

Further optimisation of the assay is continuing in the group, with the aim of testing true biological samples and reducing the amount of target that is needed to make

the assay more applicable to the standard biological concentrations. Potential routes being explored use asymmetric PCR or Förster Resonance Energy Transfer (FRET), to display the relative anthracene fluorescence *via* a different fluorophore.[33]

3.5 Preliminary Studies: Identifying Base Modifications using an Anthracene Probe

3.5.1 Base Modifications

DNA is also known to alter by chemical changes to the base rather than a complete base change. This is known as an epigenetic change. In these situations the sequence of DNA is unchanged, rather a chemical alteration to the base occurs which could potentially lead to a change in gene expression. Examples of this are 8-oxoG and 5-MeC, which occur as a result of metabolic processes, Figure 3.17A.[34] The occurrence of 8-oxoG, a product of oxidative stress, results in the guanine being able to pair with either a thymine or adenine *via* Hoogsteen base-pairing, due to the additional hydrogen bond that can be formed. This confuses the usual pairing system, shown in Figure 3.17B.[35,36]

With 5-MeC a methyl group is added to the 5-position on the cytosine. This is of interest as there is evidence that unusual methylation patterns are associated with cancer development.[37] Therefore identification of these modifications would also be beneficial and is being addressed using Tucker group probes. We have previously

reported its use in identifying 5Me-C modifications,[14] within this thesis the possible sensing of 8-oxoG is explored.

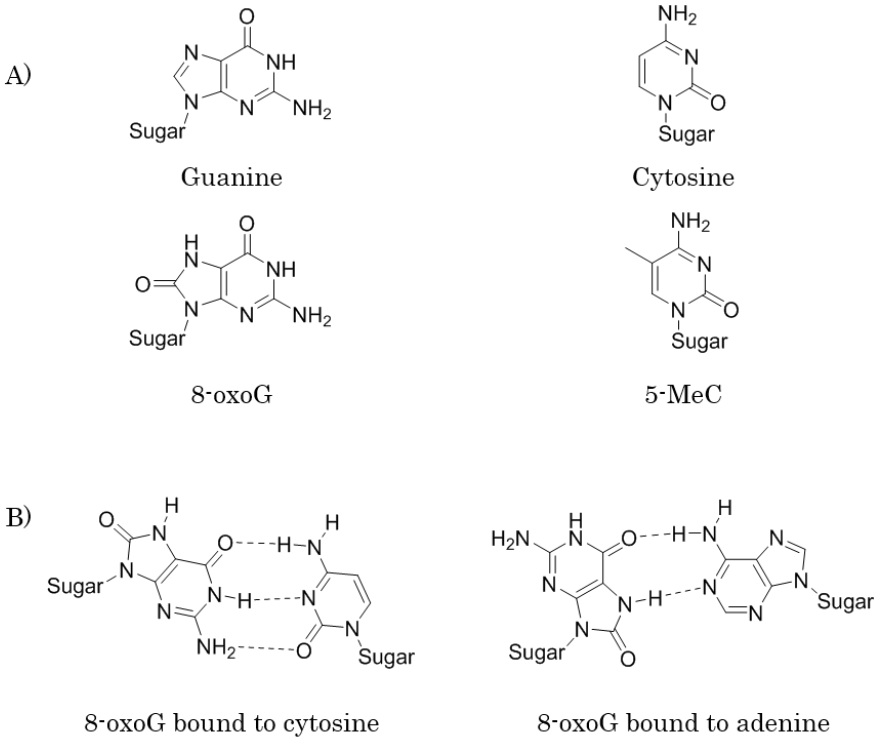


Figure 3.17 A) The structures of the base modifications 8-oxoG compared to guanine (left) and 5-MeC compared to cytosine (right). B) The possible binding modes of 8-oxoG to guanine's usual hydrogen bonding partner, cytosine (left) and through a Hoogsteen base-pair with an adenine (right).

3.5.2 Current Detection Methods

3.5.2.1. Global Analysis

The level of 8-oxoG in cells is a direct measure of the oxidative damage experienced and is hence extremely interesting to medics.[38] This has resulted in several assays being developed that examine the overall level of 8-oxoG within a system.

A common method used that is applicable to many base modifications, is enzymatic digestion followed by HPLC separation of the products. The DNA is broken down into the monomer constituents and subsequently analysed to determine the bases present. For 8-oxoG, electrochemistry can be used for its detection along with MS. This is a sensitive method that can detect oxidation of one base in every 10 000, however relatively large amounts of material are required.[39,40] Alternatively GC-MS can be utilised in a similar manner.[40,41]

Recently, fluorescent aptamers have been synthesised that allow for the detection of 8-oxoG, which could be utilised after an enzymatic digestion. It is envisioned that this will allow for quicker and cheaper determination compared to HPLC and MS.[42,43] Other less established methods for determination have included monitoring the electrochemical properties of strands containing 8-oxoG[44,45] and ion mobility through a nanopore.[46]

Alternative approaches use antibodies to bind 8-oxoG sites followed by a competitive enzyme-linked immunosorbent assay (ELISA) to determine the modification,[47] or use of an enzyme that binds to the modification which can subsequently be detected by gel electrophoresis.[48] One of the main problems with these detection methods is that during the extraction of samples, guanine can be oxidised giving false positive results.[49] In order to circumvent this problem and allow for the cell by cell determination of oxidation, imaging techniques have

subsequently been developed. Persinger *et al.* have summarised this work in a comprehensive review.[50]

3.5.2.2 Site Specific Detection

The above examples focus on determining the overall amount of oxidation within a sample. Assays that can deduce a specific site of oxidation could help in the research of this field to evaluate the affect oxidation has on the patient and also the factors affecting where the oxidation occurs.

The use of fluorescent DNA probes, in a similar manner to the approach already described in this work, is currently the only reported successful technique to perform site specific identification of 8-oxoG. Tor and co-workers have developed a fluorescently labelled cytosine adduct that can successfully distinguish between G, 8-oxoG and T bases. By incorporating the furan labelled cytosine opposite the site of interest in a complementary strand of DNA, the fluorescence intensity could be used to determine the identity of the base present.[51]

Sasaki and colleagues have also developed a fluorescent assay using 1, 3-diazaphenoxazine labelled nucleotides. Initially the group made a probe that would detect the presence of 8-oxoG without distinguishing its position within the DNA sequence.[52–54] They then further developed this assay by incorporating the group

onto adenine and, when placed opposite a site for a potential 8-oxoG ligand, this fluorescent nucleobase analogue showed a dramatic decrease in fluorescence when 8-oxoG was present.[55]

3.5.3. Strand Design – 8-oxoG Sensing

In order to prove the applicability of the Tucker probe system in sensing 8-oxo-G, a sequence was analysed that was recommended by Nik Hodges, University of Birmingham. The sequence is used within the studies of the Hodges group and the probe could have applications within their research. In addition to this, strands of the control system routinely used within the Tucker group were also assessed. This sequence has been extensively researched and will act to help understand any trends seen. The control system strands had previously been synthesised, purified and characterised by fellow members of the group.[21] The sequences used are shown within Table 3.4. Probes were synthesised with the same scope of linker length and stereochemistry possibilities as were used in Section 3.3. The strands were subsequently purified and characterised as described in the experimental section. This was performed by Dr Zheng-Yun Zhao on some strands and these are indicated in the appendix.

Table 3.4 Sequences synthesised for base modification detection trials. Probes labelled according to the linker length (n) and stereochemistry of the threoninol until (L/D)

	Name	Sequence 5′ – 3′
Control Probe	ConBM-Pr(nL/D)	TGG ACT C**X**C TCA ATG
Control Target	ConBM-TarG/oxo-G	CAT TGA G**G/oxoG**G AGT CCA
Probe	BM-Pr(nL/D)	GGG GAT **X**AC TAG TTC
Target	BM-TarG/oxo-G	GAA CTA GT**G/oxoG** ATC CCC

3.5.4. Base Modification Fluorescence Studies

As with the SNP sensing in Section 3.3.5, fluorescence spectra of all probes (**ConBM-Pr**'s and **BM-Pr**'s) were recorded in the absence and presence of both targets, **ConBM-TarG** and **ConBM-TaroxoG**, and **BM-TarG** and **BM-TaroxoG**, respectively. Percentage changes in the intensity of the fluorescence in relation to the probe alone were calculated and are shown in Figure 3.18.

Within the model system, the longer linker length probes (**ConBM-Pr(3-7)**) can all distinguish between the 8-oxo-G and G bases, whilst **ConBM-Pr(1L)** and **ConBM-Pr(1D)** fail to discriminate between the two bases. As with the SNP identification in Section 3.3.2, this can be ascribed to the lack of flexibility resulting from the shorter linker length arm, restricting the intercalation of the anthracene moiety. With the longer linker lengths, anthracene can intercalate and stack with the neighbouring bases. The resulting decrease in fluorescence seen when 8-oxo-G is present, could be explained one of two ways. Firstly, due to the increase in size of

8-oxoG compared to G because of the extra oxygen atom. This could restrict the intercalation of the anthracene, resulting in a significant amount of quenching by water. Secondly, 8-oxo-G is known to have a lower oxidation potential compared to G,[56–58] which would allow for more effective PET to the base and therefore fluorescence quenching as Tor and co-workers reported with their work.[51] In order to establish which of these routes is responsible, or if indeed both play a role, more in depth studies analysing the lifetimes of fluorescence and computational modelling of the systems would need to be performed.

In comparison, the **BM-Pr** system displayed different effects, with all the longer linker lengths exhibiting a decrease in fluorescence upon duplex formation, independent of the base opposite. In all cases, hybridisation with **BM-TaroxoG** resulted in less of a reduction than that with **BM-TarG**, indicating that the above hypothesis to explain the trends could still hold true.

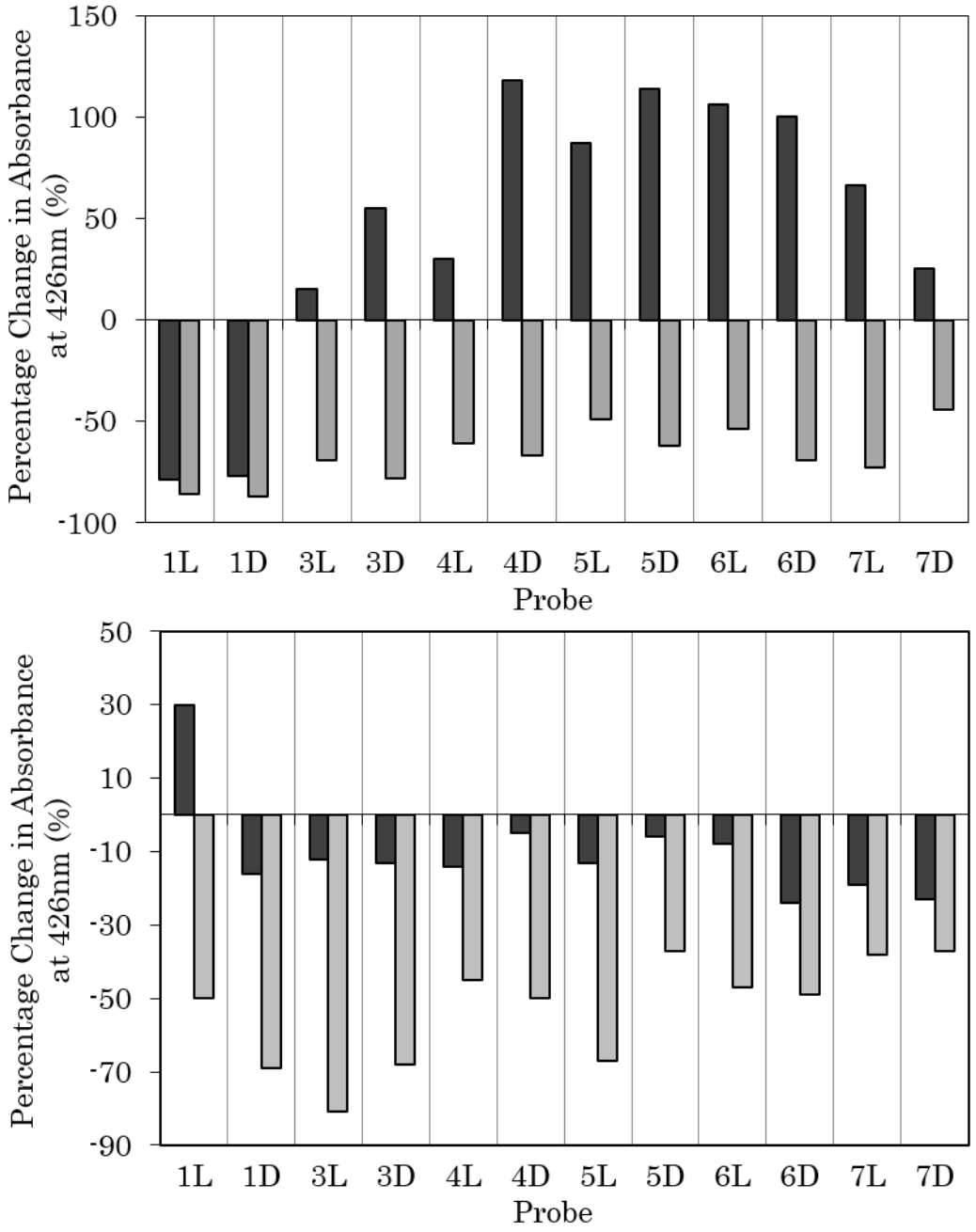


Figure 3.18 Percentage changes in fluorescence intensity upon addition of **ConBM-TaroxoG** (light grey) and **ConBM-TarG** (dark grey) with all **ConBM-Pr** configurations, top and **BM-TaroxoG** (light grey) and **BM-TarG** (dark grey) with all **BM-Pr** configurations, bottom. All performed at 1 µM probe in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl at 293 K. λ_{ex} 350 nm. Values taken after addition of 1.5 µM target.

The overall difference in the trends seen is thought to be as a consequence of the bases flanking the anthracene. Within **ConBM-Pr** systems, a C is upstream of the anthracene and within **BM-Pr** systems a T is in this position. It has been noted

previously that the anthracene is affected by the upstream base and that, upon duplex formation, the identity of this base can alter the fluorescence.[21] This is expected to arise as a result of the PET to the upstream base with which it stacks. It has been previously calculated that C would be a more efficient quencher of anthracene fluorescence than T.[21] It can therefore be hypothesised that in the **ConBM-Pr** systems, the initial quenching by the upstream C is reduced upon duplex formation with **ConBM-TarG**, resulting in a relative increase in the fluorescence. However, when duplexed with **ConBM-TaroxoG**, the 8-oxoG can quench by PET and a decrease in fluorescence is observed. In the case of the **BM-Pr** system, the initial quenching is not as significant due to the inefficient quenching of the T upstream. Upon duplex formation the change in fluorescence is therefore smaller and is seen as a slight decrease in both cases. Measurements of quantum yields of the single stranded probes would be able to confirm this.

An alternative explanation is that the 8-oxoG base lies in its preferred *syn* configuration (Figure 3.19), which does not allow the anthracene to intercalate as readily as a G base does. This would result in the anthracene being outside of the DNA duplex and hence exposed to the water environment, which is known to quench anthracene fluorescence.[15] These hypotheses could be investigated using CD, T_m and fluorescence quantum yield studies. Obtaining crystal structures and performing molecular dynamic simulations on the systems, could offer greater insight.

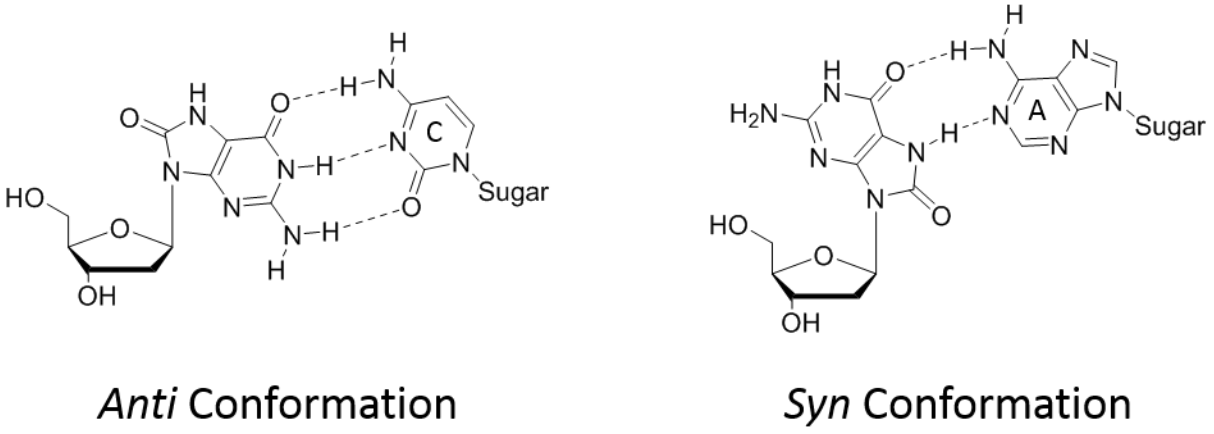


Figure 3.19 *Anti* and *Syn* conformations of 8-oxoG. In the preferred anti conformation, 8-oxoG forms hydrogen bonds with cytosine (left). When in the *syn* conformation it can form a Hoogsteen base pair with adenine (right).

The data shows that the best probe for further optimisation in the sequence of interest would be **BM-Pr(1L)**, as this probe shows the largest fluorescence differences between the two targets, **BM-TaroxoG** and **BM-TarG**. This is expected to be due to the constrained nature of the shorter linker length, restricting the optimal intercalation of the anthracene. This would affect the fluorescence when in the single stranded form, as well as when duplexed. In order to evaluate the causes of this, CD and T_m analysis should be performed, along with lifetime and fluorescence quantum yield studies. These would give an insight into the stability of the duplex structures and give an insight into the nature of the anthracene environment, enabling the system to be further understood.

In due course, it would be expected that this base modification identification technique could be developed into a quantitative method for 8-oxoG determination, in a similar way to that detailed with the SNP sensing in Section 3.4.2.

3.6 Conclusions and Future Work

The use of anthracene as a fluorescent SNP and base modification reporter molecule, has been presented. Within an Alzheimer's SNP sequence the fluorescence intensity of the anthracene has been shown to be sensitive to the nature of the base located directly opposite in the complementary strand. The change in this intensity is shown to display a linear relationship with respect to the ratios of base present at a SNP site, allowing for the technique to be used for quantitative determination of the DNA present. Future work will focus on the detection of RNA as this would be the target molecule for cellular studies. The RNA work has been performed by Rosemary Bamford and will be published at a later date.[33] In addition to this, samples extracted from patients are also to be analysed to establish the reliability of the assay when used on much longer sequences of DNA. Further studies on the origin of the trends seen, should be performed by assessing the lifetimes of the fluorescence and molecular modelling of systems.

The anthracene probe has also been shown to be an effective 8-oxoG sensor, although the influence that the surrounding bases have on the output has been identified to be a significant issue. Further work in this area to understand the systems further, could be performed using CD, T_m analysis, lifetime measurements and molecular modelling. It would also be of interest to assess whether sensing the base modification, displays the same linear trend to that seen in the related SNP system.

3.7 References

1 Y. Suh and J. Vijg, *Mutat. Res.*, 2005, **573**, 41–53.
2 X. Duan, L. Liu, F. Feng and S. Wang, *Acc. Chem. Res.*, 2009, **43**, 260–70.
3 Z. Wang and J. Moult, *Hum. Mutat.*, 2001, **17**, 263–70.
4 R. Hunt, Z. E. Sauna, S. V Ambudkar, M. M. Gottesman and C. Kimchi-Sarfaty, *Methods Mol. Biol.*, 2009, **578**, 23–39.
5 F. E. A. McGuigan and S. H. Ralston, *Psychiatr. Genet.*, 2002, **12**, 133–6.
6 S. A. E. Marras, F. R. Kramer and S. Tyagi, *Genet. Anal. Eng.*, 1999, **14**, 151–6.
7 N. Thelwell, S. Millington, A. Solinas, J. Booth and T. Brown, *Nucleic Acids Res.*, 2000, **28**, 3752–61.
8 L. Bethge, D. V. Jarikote and O. Seitz, *Bioorg. Med. Chem.*, 2008, **16**, 114–25.
9 E. Socher, D. V Jarikote, A. Knoll, L. Roeglin, J. Burmeister, O. Seitz and L. Röglin, *Anal. Biochem.*, 2008, **375**, 318–30.
10 Y. Saito, Y. Miyauchi, A. Okamoto and I. Saito, *Chem. Commun.*, 2004, 1704–5.
11 A. Okamoto, K. Kanatani and I. Saito, *J. Am. Chem. Soc.*, 2004, **126**, 4820–7.
12 T. Ihara, A. Uemura, A. Futamura, M. Shimizu, N. Baba, S. Nishizawa, N. Teramae and A. Jyo, *J. Am. Chem. Soc.*, 2009, **131**, 1386–7.
13 J.-L. H. A. Duprey, D. M. Bassani, E. I. Hyde, C. Ludwig, A. Rodger, J. S. Vyle, J. Wilkie, Z.-Y. Zhao and J. H. R. Tucker, *Supramol. Chem.*, 2011, **23**, 273–7.
14 J.-L. H. A. Duprey, Z. Zhao, D. M. Bassani, J. Manchester, J. S. Vyle and J. H. R. Tucker, *Chem. Commun.*, 2011, **47**, 6629–31.
15 N. Moran, D. M. Bassani, J.-P. Desvergne, S. Keiper, P. a S. Lowden, J. S. Vyle and J. H. R. Tucker, *Chem. Commun.*, 2006, 5003–5.
16 Y. Saito, K. Motegi, S. S. Bag and I. Saito, *Bioorg. Med. Chem.*, 2008, **16**, 107–13.
17 Z. Zhao, M. San, J.-L. H. A. Duprey, J. R. Arrand, J. S. Vyle and J. H. R. Tucker, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 129–32.
18 R. Scacchi, G. Gambina, G. Moretto and R. M. Corbo, *Dement. Geriatr. Cogn. Disord.*, 2013, **35**, 51–7.
19 S. Gravina, F. Lescai, G. Hurteau, G. J. Brock, A. Saramaki, S. Salvioli, C. Franceschi and I. B. Roninson, *Aging*, 2009, **1**, 470–80.
20 N. Moran, PhD Thesis, University of Exeter, 2006.
21 J.-L. H. A. Duprey, PhD Thesis, University of Birmingham, 2010.
22 B. Norden, A. Rodger and T. Dafforn, *Linear Dichroism and Circular Dichroism: A Textbook on Polarized-light Spectroscopy*, The Royal Society of Chemistry, Cambridge, 2010.
23 A. Rodger, I. S. Blagbrough, G. Adlam and M. L. Carpenter, *Biopolymers*, 1994, **34**, 1583–93.
24 A. Rodger, S. Taylor, G. Adlam, I. S. Blagbrough and I. S. Haworth, *Bioorg. Med. Chem.*, 1995, **3**, 861–72.

25 C. G. Simpson, G. Clark, D. Davidson, P. Smith and J. W. S. Brown, *Plant J.*, 1996, **9**, 369–80.
26 C.-P. Witte, E. Isidore, S. A. Tiller, H. V. Davies and M. A. Taylor, *Plant Mol. Biol.*, 2001, **45**, 169–79.
27 B. P. Sokolov, *Nucleic Acids Res.*, 1990, **18**, 3671.
28 M. N. Kuppuswamy, J. W. Hoffmann, C. K. Kasper, S. G. Spitzer, S. L. Groce and S. P. Bajaj, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 1143–7.
29 A. C. Syvänen, A. Sajantila and M. Lukka, *Am. J. Hum. Genet.*, 1993, **52**, 46–59.
30 C. R. Cowles, J. N. Hirschhorn, D. Altshuler and E. S. Lander, *Nat. Genet.*, 2002, **32**, 432–7.
31 J. G. Schaart, L. Mehli and H. J. Schouten, *Plant J.*, 2005, **41**, 493–500.
32 M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén and P. Nyrén, *Anal. Biochem.*, 1996, **242**, 84–9.
33 R. Bamford, PhD Thesis, University of Birmingham, 2015.
34 S. S. Wallace, *Free Radic. Biol. Med.*, 2002, **33**, 1–14.
35 S. Shibutani and A. P. Grollman, *Cancer Lett.*, 1994, **83**, 315–22.
36 G. A. Leonard, A. Guy, T. Brown, R. Teoule and W. N. Hunter, *Biochemistry*, 1992, **31**, 8415–20.
37 L. S. Kristensen, H. M. Nielsen and L. L. Hansen, *Eur. J. Pharmacol.*, 2009, **625**, 131–42.
38 S. Kanvah, J. Joseph, G. B. Schuster, R. N. Barnett, C. L. Cleveland and U. Landman, *Acc. Chem. Res.*, 2010, **43**, 280–7.
39 J.-L. Ravanat, E. Gremaud, J. Markovic and R. J. Turesky, *Anal. Biochem.*, 1998, **260**, 30–7.
40 J. Cadet and M. Weinfeld, *Anal. Chem.*, 1993, **65**, 675A–82A.
41 M. Dizdaroglu, *FEBS Lett.*, 1993, **315**, 1–6.
42 J. Roy, P. Chirania, S. Ganguly and H. Huang, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 863–7.
43 Q. Zhang, Y. Wang, X. Meng, R. Dhar and H. Huang, *Anal. Chem.*, 2013, **85**, 201–7.
44 B. Zhang, L.-H. Guo and M. M. Greenberg, *Anal. Chem.*, 2012, **84**, 6048–53.
45 A. Mugweru, B. Wang and J. Rusling, *Anal. Chem.*, 2004, **76**, 5557–63.
46 A. E. P. Schibel, N. An, Q. Jin, A. M. Fleming, C. J. Burrows and H. S. White, *J. Am. Chem. Soc.*, 2010, **132**, 17992–5.
47 B. Yin, R. M. Whyatt, F. P. Perera, M. C. Randall, T. B. Cooper and R. M. Santella, *Free Radic. Biol. Med.*, 1995, **18**, 1023–32.
48 A. R. Collins, S. J. Duthie and V. L. Dobson, *Carcinogenesis*, 1993, **14**, 1733–5.
49 J.-L. Ravanat, T. Douki, P. Duez, E. Gremaud, K. Herbert, T. Hofer, L. Lasserre, C. Saint-Pierre, A. Favier and J. Cadet, *Carcinogenesis*, 2002, **23**, 1911–8.
50 R. L. Persinger, R. Melamede, I. Bespalov, S. Wallace, D. J. Taatjes and Y. Janssen-Heininger, *Exp. Gerontol.*, 2001, **36**, 1483–94.
51 N. J. Greco, R. W. Sinkeldam and Y. Tor, *Org. Lett.*, 2009, **11**, 1115–8.

52 O. Nakagawa, S. Ono, Z. Li, A. Tsujimoto and S. Sasaki, *Angew. Chem. Int. Ed. Engl.*, 2007, **46**, 4500–3.
53 Y. Koga, Y. Fuchi, O. Nakagawa and S. Sasaki, *Tetrahedron*, 2011, **67**, 6746–52.
54 T. Nasr, Z. Li, O. Nakagawa, Y. Taniguchi, S. Ono and S. Sasaki, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 727–30.
55 Y. Taniguchi, R. Kawaguchi and S. Sasaki, *J. Am. Chem. Soc.*, 2011, **133**, 7272–5.
56 C. Sheu and C. S. Foote, *J. Am. Chem. Soc.*, 1995, **117**, 6439–42.
57 S. M. Gasper and G. B. Schuster, *J. Am. Chem. Soc.*, 1997, **119**, 12762–71.
58 S. Steenken, S. V. Jovanovic, M. Bietti and K. Bernhard, *J. Am. Chem. Soc.*, 2000, **122**, 2373–4.

Chapter 4

Controlling DNA Binding using an Anthracene Tagged Peptide

Parts of this chapter have recently been published: Exploiting Anthracene Photodimerization within peptides: Light Induced, Sequence-Selective DNA Binding, G. A. Bullen, J. H. R. Tucker and A. F. A. Peacock, *Chem. Comm.,* **51**, 8130-8133.

4.1 Introduction

The vast majority of biological processes rely upon several separate entities associating together. For instance DNA transcription relies upon the double stranded nature of the DNA, an enzyme binding to the DNA and the enzyme itself often comprises of several domains or even separate proteins.[1] Therefore, controlling oligomerisation could prove beneficial for a variety of applications. Within this chapter the first reported use of anthracene to control the oligomerisation of a peptidic structure is explored. A summary of this work has recently been published.[2]

4.1.1 Protein Crosslinking

Within nature there are known to be several ways in which proteins interact with one another, including hydrophobic interactions, hydrogen bonds, salt bridges and electrostatic interactions.[3] In addition to these they can also be covalently crosslinked, the most well-known being disulphide bond formation between two cysteine residues, but recently many more amino acids have been found to be capable of crosslinking such as tyrosines.[4] Chemists and biologists have spent many years working on ways to mimic and control the crosslinking of proteins to enable them to study the localisation of proteins, allowing for deduction of protein relationships and elucidation of the proteins function. In addition to this, crosslinking will allow the evaluation of the processes that oligomerisation is key for.[5–8]

The traditional way to induce oligomerisation, is to chemically link the entities. This is typically done using crosslinking reagents, many of which are commercially available and summarised in The Molecular Probes Handbook by Invitrogen.[9] The majority of these utilise an exposed thiol, typically from a cysteine residue, or an amine and can be used with either hetero- or homo- bifunctionality to link multiples of the same protein or attach different entities. This is done utilising reactive groups such as maleimides, succinimidyl esters or iodoacetamides (Figure 1.11), as well as many other derivatives which have been used. The vast array available are explored in the comprehensive Bioconjugate Techniques text book by Hermanson.[10]

4.1.2 Controlling Protein Crosslinking using Light

The use of light to control oligomerisation would have several advantages over the chemical alternatives. Notably it would offer increased control of the reaction by adjustments in the intensity and wavelength of the irradiation light used, small focussed areas of irradiation, and the advantage of no chemical activators being required. A recent review by Preston and Wilson covers techniques in this area.[11]

In recent years the Photo Induced Crosslinking of Unmodified Proteins (PICUP) technique has been developed which allows for the crosslinking of proteins *via* tyrosine residues.[12,13] This technique, although photo activated, requires the addition of a ruthenium complex and ammonium persulphate (APS) in order to

proceed. The tris-bipyridylruthenium (II) complex undergoes brief photolysis upon excitation with visible light when in the presence of APS, to generate Ru(III) and a sulphate radical. It is thought that the reaction proceeds *via* oxidation of a tyrosine residue by Ru(III), the resultant radical can then react either with another tyrosine residue or a nucleophilic cysteine or lysine side chain. Kodadek and co-workers have published proposed mechanisms for these two pathways, Figure 4.1,[12] and the technique has been applied to a variety of systems within the literature.[14–20]

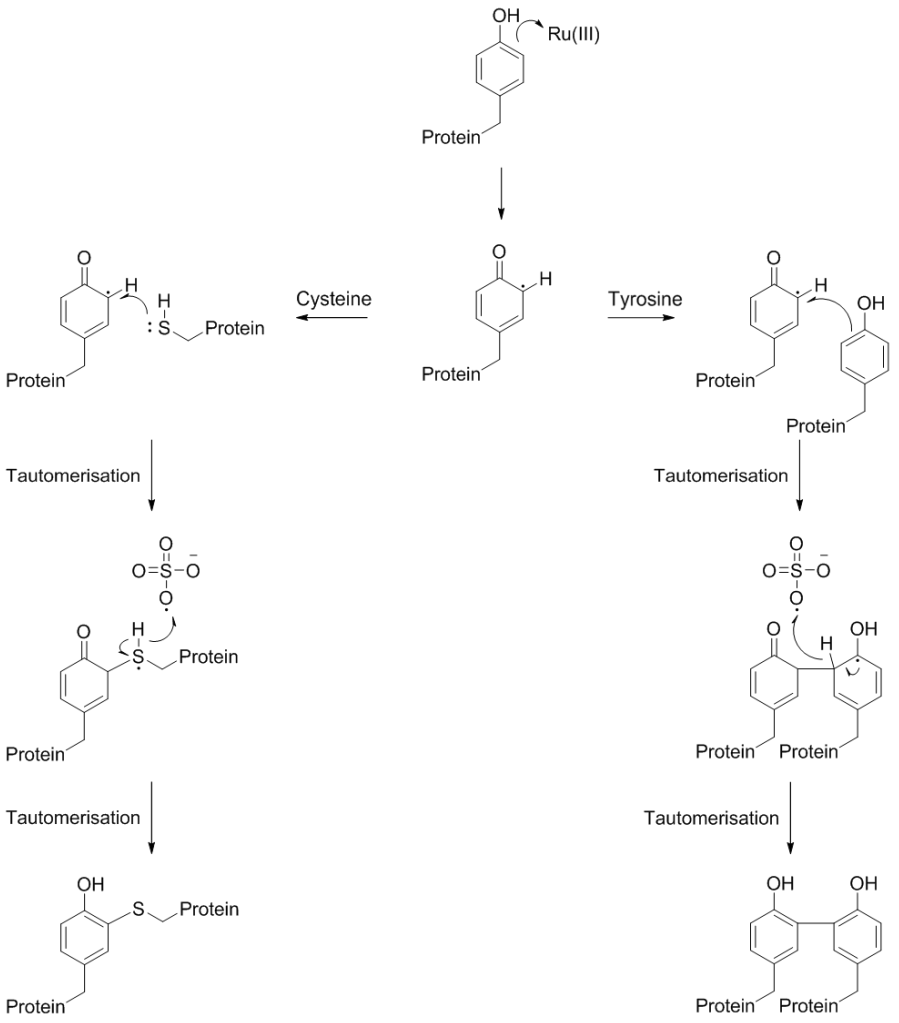


Figure 4.1 Proposed mechanism of the PICUP technique showing the two possible reaction routes with tyrosine (right) and cysteine (left).[12]

A limitation of the PICUP technique, along with the requirement for the use of APS and the ruthenium complex, is that the crosslinking is not site specific, *i.e.* you cannot direct the crosslinking to a particular protein or amino acid. This can be seen as an advantage to some, as it allows for the determination of protein oligomerisation within cells due to the ability to crosslink proteins in close proximity. However, for some applications there is a need to specifically control the site at which crosslinking occurs. In order to overcome this issue, several photo reactive alternatives have been developed, some of which are now commercially available, which allow for a more targeted crosslinking approach.[9]

Firstly, aryl azides have been utilised that, upon irradiation with UV-light, react with nucleophilic groups in the vicinity. The aryl azide can be coupled to a variety of chemical crosslinking reagents (as described in Section 4.1.1) to allow for coupling to the protein of interest. The crosslinker is first coupled to the protein and then, upon irradiation, the aryl azide will react with a nucleophilic group in close proximity to form the crosslinking. This technique has been applied to the photocrosslinking of proteins with both DNA as well as other proteins.[21,22]

Photo reactive amino acids have recently been reported which contain diazirine rings which, upon UV-light activation, form crosslinks with nearby amino acid side chains and protein backbones.[23] Methionine and leucine derivatives have been made commercially available, Figure 4.2, and it has been shown that these

modified amino acids can be incorporated endogenously into protein sequences, allowing for easy addition within cells.[23]

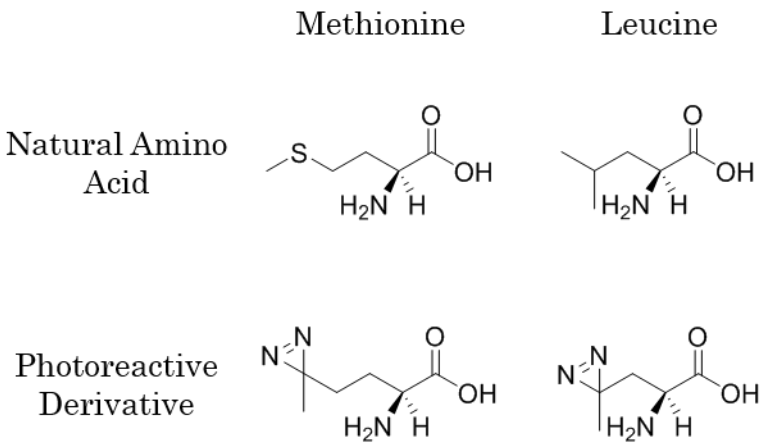


Figure 4.2 Structure of commercially available photoreactive amino acids (bottom) compared to the naturally occurring amino acids (top), methionine (left) and leucine (right).

The currently available methods allow for site specific incorporation of a photoreactive group into one protein, however, upon light activation, the modification reacts indiscriminately with several different amino acid side chains, as well as the peptide backbone, and the specific site of crosslinking cannot be controlled. However, those that crosslink with thiols are often used in a moderately selective manner due to the rarity of the cysteine residue, especially on the protein surface.[10]

4.1.3 Anthracene for Photocrosslinking

Anthracene is known to form photodimers upon UV-light irradiation (Chapter 1.5).[24–26] Incorporation of anthracene moieties at specific points in protein sequences could therefore allow for site specific crosslinking on both proteins.

Many anthracene derivatives are commercially available and could potentially be coupled to a variety of amino acid side chains. In order to evaluate the feasibility of photo-induced anthracene dimerisation within proteins, a model system based on the GCN4 transcription factor was investigated. This was chosen due to the fact that strong DNA binding is associated with protein dimerisation and its extensive study in the literature. This represents, to the best of my knowledge, the first time that anthracene photodimerisation has been used within a peptide or protein structure to control oligomerisation.

4.2 GCN4 Transcription Factor

The yeast transcription factor GCN4 is known to comprise of a homodimer protein,[27] the monomer of which displays weak binding to target site DNA.[28] The proteins are held together by a leucine zipper region towards the C-terminus of the protein,[27] which has been studied extensively in the literature and found to be essential for dimerisation and subsequent DNA binding. The leucine zipper is made of two α-helical regions held together by hydrophobic interactions between the two chains. The DNA binding residues are found in the basic region of the GCN4 structure, Figure 4.3.[29]

The native GCN4 transcription factor is known to bind to the AP1 site of DNA (**ATGA**C**TCAT**), however, it is also found to bind to the similar CRE site (**ATGA**CG**TCAT**), which differs in only a single base pair, and has been used

extensively for binding studies.[30] Both of these sequences contain two binding sites, one for each strand of GCN4 within the homodimer, separated by one base in the case of AP1 and two bases in the case of CRE. These sites are shown in the above sequences in bold.

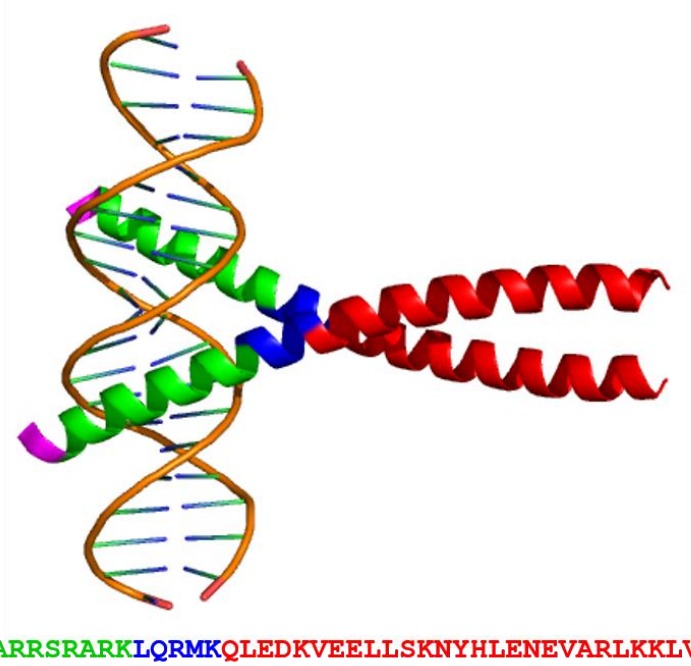


Figure 4.3 Pymol representation of the C-terminus of the GCN4 homodimer bound to target DNA (based on PDB 2DGC).[29] The coiled coil region (red) holds the protein together as a homodimer, which is joined to the basic region (green) *via* a linker region (blue). Towards the N-terminus is a capping region (purple).

4.2.1 Controlling DNA Binding using the Coiled Coil Region

The DNA binding of GCN4 is known to be a cooperative process with the dimer species displaying a higher DNA binding affinity.[28] Due to this, an interesting approach to control DNA binding, is to control dimerisation of the GCN4 units. This can be done using the natural dimerisation domain, the coiled coil. Controlling

the secondary structure of the peptide in this region has the potential to disrupt the coiled coil, controlling the association of the GCN4 units and hence the DNA binding.

Futaki and co-workers reported a GCN4-bZIP modified peptide with iminodiacetic acid groups attached on the exterior of the coiled coil region. Upon chelation of Co(II) by iminodiacetic acid residues, the α-helical structure is disrupted resulting in disruption to the coiled coil, Figure 4.4. The monomer peptides that are formed are shown to display a much weaker DNA binding affinity consistent with other GCN4 monomer reports.[31]

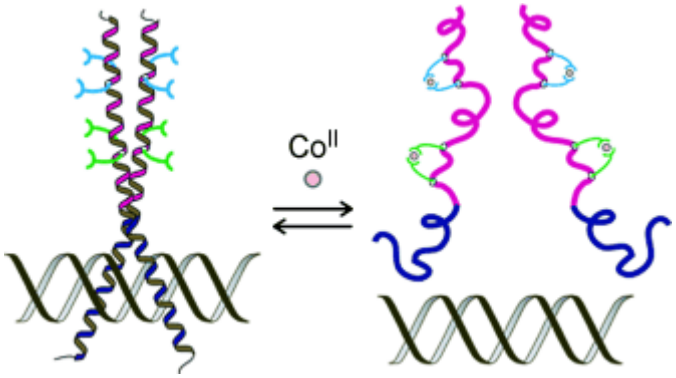


Figure 4.4 The iminodiacetic acid tagged GCN4-bZIP derivative shows the expected DNA binding (left) in the absence of a metal ion. However, upon addition of Co(II), the iminodiacetic acid groups chelate the Co(II), disrupting the α-helix and hence the coiled coil, resulting in monomer formation and reduced DNA binding (right). Reprinted with permission from Azuma *et al.*, *Angew. Int. Ed. Engl.*, 2009, **48**, 6853-6. Copyright 2009 WILEY-VCH GmbH & Co.

Light has also been used to disrupt the folding of the protein, and hence the DNA binding. In work by Woolley and colleagues, the DNA binding residues of GCN4 were grafted onto the sequence of the photo-activated yellow protein (PYP). In the native form of PYP, the DNA binding residues were not in the right arrangement

to form interactions with the DNA and hence bind. However, upon irradiation, the protein conformation changes, exposing the DNA binding residues, and binding to target DNA is achieved, Figure 4.5.[32]

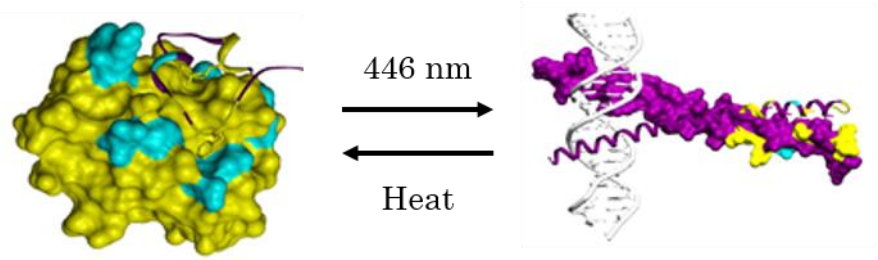


Figure 4.5 The DNA binding residues of the GCN4 transcription factor are grafted onto the photoactive yellow protein. In its native form the DNA binding residues are not exposed (left), but upon irradiation with 446 nm light the protein changes conformation, exposing the DNA binding residues and allowing for DNA binding. Adapted with permission from Morgan *et al.*, *J. Mol. Biol.*, 2010, **399**, 94-112. Copyright 2010 Elsevier Ltd.

Zhang *et al.* utilised the *trans-cis* isomerisation properties that azobenzene displays upon UV-light irradiation to control coiled coil formation within the leucine zipper region of GCN4. They attached azobenzene to amino acid residues seven units apart within the α-helices of the zipper region. In the *trans* form a bend is induced disrupting the α-helix and coiled coil formation. Upon light irradiation the *cis* form of azobenzene is formed, which allows for the coiled coil to form and hence DNA binding to be achieved, Figure 4.6.[33]

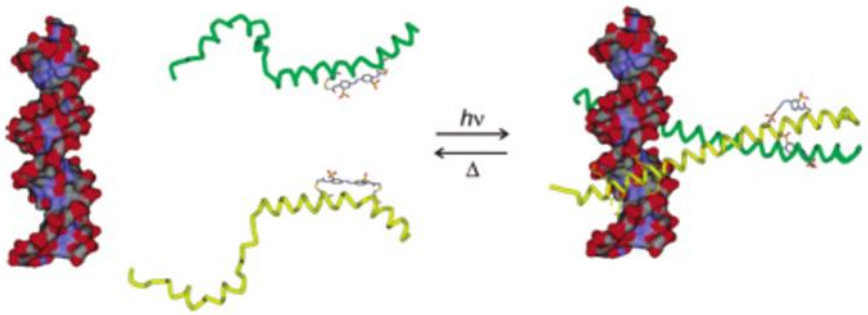


Figure 4.6 GCN4-bZIP with azobenzene attached in the coiled coil domain. In the *trans* configuration the protein cannot fold into an α-helix and DNA binding is disrupted. Upon light excitation the *cis* conformer is produced which allows for the α-helix and coiled coil formation, and hence DNA binding Reprinted with permission from Woolley *et al.*, *Biochemistry*, 2006, **45**, 6075-84. Copyright 2006 American Chemical Society.

4.2.2 Controlling DNA Binding using the Dimerisation of Short Peptides

An alternative approach is to use short readily synthesised peptides and control the dimerisation of these. Kim and co-workers were the first to demonstrate this approach within the GCN4 system. They replaced the coiled coil region with a single cysteine which could subsequently form a disulphide linkage with another chain. It was shown that the monomer version did not retain strong DNA binding, however, upon dimer formation, DNA binding was recovered.[34]

A few years later, a more elegant approach utilising non-covalent interactions was demonstrated by Ueno *et al.* utilising a cyclodextrin and guest compound to control dimerisation.[35] To one unit of the GCN4 binding domain, a cyclodextrin moiety was attached, with a 1-adamantaneacetic acid unit attached to the other. Upon addition of these two modified strands to target DNA, increased affinity for the DNA strand is observed in relation to the monomer peptide species when studied individually.[35] The group went on to study whether non-covalent interactions, such as that of the 1-adamantaneacetic acid with cyclodextrin, displayed the same DNA binding properties as covalent dimerisation techniques. They concluded that, although the covalently linked peptide displayed a slightly higher affinity for DNA, the non-covalent versions displayed enhanced selectivity.[36] It is thought that this is due to the monomer GCN4 units binding weakly to the DNA and pre-organising the dimerisation entities.

4.2.3 Controlling DNA Binding with Basic Domain Orientation

In order to display strong DNA binding, the basic domain of the protein not only needs to be dimerised, but also has to be held in the correct orientation. This property has also been utilised to control DNA binding.

Schepartz and co-workers attached the binding residues to iron complexes to allow for the control of the orientation and distance between the binding regions.[37] Peacock and Oheix took this further, developing a reversible metal based system to allow for control of the orientation and hence DNA binding. A terpyridine linkage between the two peptidic binding domains displayed weak DNA binding in the *apo* form, however, upon addition of Cu(II) or Zn(II), the terpyridine chelates to the metal and the resultant *cis* orientation of the peptide chains promotes DNA binding.[38]

Mascareñas and colleagues appended the DNA binding region of GCN4 to an azobenzene linker with the aim of using the *cis-trans* isomerisation properties of azobenzene to control DNA binding. They established that when the azobenzene linker was in the *trans* conformation, poor DNA binding was observed. Upon light excitation the azobenzene unit converted to the *cis* conformation, shortening the distance between the two DNA binding domains and resulting in an increase in DNA binding, Figure 4.7.[39]

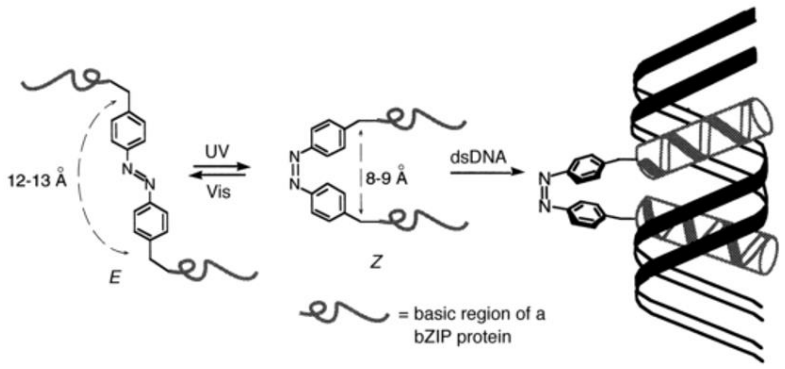


Figure 4.7 Schematic representation of azobenzene linked basic regions of GCN4. In the *trans* form there is a large distance between the chains (left) and no DNA binding is observed. Upon isomerisation, the *cis* form binds to the target site of DNA. Reprinted with permission from Caamaño *et al.*, *Angew. Int. Ed. Engl.*, 2000, **39**, 3104-7. Copyright 2000 WILEY-VCH GmbH & Co.

These approaches offer the advantage of using relatively short peptides, which are more readily synthesised than those that include the leucine zipper region of GCN4. Work has also shown that both dimerisation and orientation of the strands, can be used to control the sequence to which the peptide structure binds. By linking the peptide chains at the N-terminus as oppose to the natural C-terminus, the target site can be reversed.[40]

Most recently, Mascareñas and co-workers published work that utilised the different DNA sequence selectivity when dimerised at the C- or N- terminus. The C-terminus of the GCN4 DNA binding domain contained a Ni(II) dependent dimerisation unit. However, the N- terminus could be dimerised *via* a cysteine residue introduced for disulphide formation. Dimerisation through the natural C-terminus was shown to maintain selectivity for the natural DNA target site, whilst dimerisation through the N- terminus displayed selective binding for a reversed target site.[41]

4.3 Project Aims

Within this project anthracene will be incorporated on the terminus of the GCN4 transcription factor, in place of the natural dimerisation unit, the coiled coil. It is thought that this incorporation will not disrupt the ability of the peptide to bind to a specific sequence of DNA and the anthracene unit can be utilised to assess this by monitoring the fluorescence output in a similar fashion to Chapter 3. When the peptide binds to target site DNA, the anthracene units will be arranged in a close proximity to one another and, upon irradiation, they will form a photodimer which will result in an increased binding affinity for the target site, Figure 4.8. In the absence of the specific sequence, the anthracene units will not be arranged in close enough proximity to form photodimers. The reversibility of the system will also be assessed.

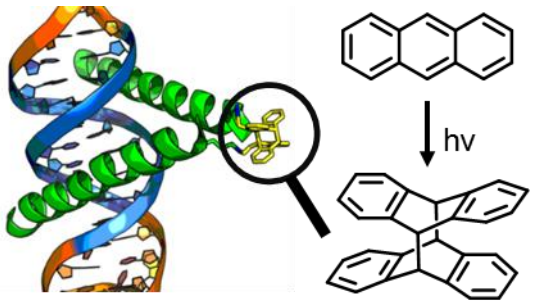


Figure 4.8 Schematic representation of the anthracene tagged peptide binding to target site DNA. Upon binding, the anthracene units are placed in close proximity and will form the photodimer upon irradiation.

4.4 Anthracene Tagged Peptide for Dimerisation Control

4.4.1 Anthracene Tagged Peptide Design

All essential residues for specific DNA binding[29,42] were maintained within the design (shown in bold), along with the linker region which is known to enhance the thermal stability of the protein-DNA complex (underlined).[43] A lysine was incorporated towards the C- terminus with a labile protecting group on the side chain, to allow for the preferential deprotection of this group, for the subsequent incorporation of the anthracene moiety (highlighted in black). Three additional residues, not essential to DNA binding, that form part of the natural GCN4 sequence were retained at the N- terminus to help maintain the α-helical nature of the peptide within the DNA binding region. A glycine was incorporated at the C- terminus to aid in the synthesis of the peptide. The final sequence is shown below:

Ac -A L K **R** A **R N T** E **A A R R S R** A **R K L** Q R M K Q **K** G - NH_2

A methyltrityl (Mtt) protecting group was chosen to protect the lysine at which the anthracene was to be incorporated and was selectively removed in mild acid, maintaining other side chain protecting groups and the bond between the resin and the peptide. The removal of the Mtt group was tracked using the Kaiser test.[44] It was found that three cycles of 3% TFA in DCM for 10 minutes were required so as to ensure complete deprotection. A carboxylic acid derivative of anthracene was

subsequently coupled *via* the side chain amine using standard coupling conditions for 15 hours.[45] The peptide sequence was then completed and the peptide cleaved, deprotected and purified using standard conditions.[46]

The initial anthracene derivative chosen for coupling was the carboxylic acid intermediate, formed during the synthesis of anthracene phosphoramidites for incorporation into DNA (Chapter 2.3), Figure 4.9. However, it was identified by MS that this derivative did not survive the peptide synthesis conditions. A mass which could be attributed to cleavage of the anthracene at the ester bond was observed and hence this derivative could not be used. Instead anthracene 9-carboxylic acid was successfully incorporated as desired.

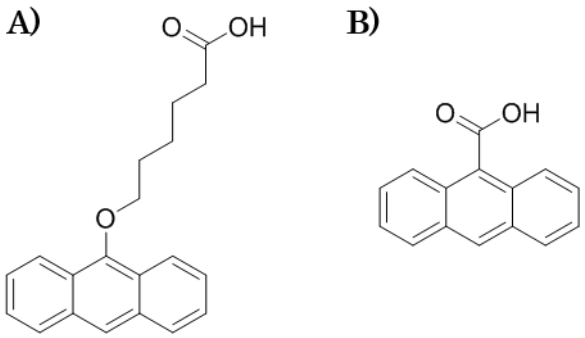


Figure 4.9 Structure of anthracene derivatives used in this study. A) 6-(anthracene-9-yloxy) hexanoic acid incorporation was unsuccessful. B) Anthracene 9-carboxylic acid was successfully incorporated.

4.4.2 DNA Sequences used within this Study

In order to study the DNA sequence selectivity, three strands of DNA were assessed, target site (CRE), a strand containing half the target site (half CRE) and a non-specific (NS) sequence.

Table 4.1 DNA sequences used within the study with target sites shown in bold.

Name	Sequence
CRE 1	5′-ACG AG**A TGA** CG**T CAT** CTC CA-3′
CRE 2	3′-TGC TC**T ACT** GC**A GTA** GAG GT-5′
Half CRE 1	5′-TGG AG**A TGA** CGT TGT CTC GT-3′
Half CRE 2	3′-ACC TC**T ACT** GCA ACA GAG CA-5′
NS 1	5′-TGG AGT ATG CGT CGA TTC GT-3′
NS 2	3′-ACC TCA TAC GCA GCT AAG CA-5′

The NS sequence contains the same number and type of bases as CRE and half CRE, but in a scrambled order. It is expected that the peptide will display some affinity for NS DNA due to favourable electrostatics between the positive peptide and the negative DNA. In addition to this, the hydrophobic, planar anthracene is expected to intercalate into the DNA so as to be shielded from the water environment. With the half CRE system, there is a binding site for only one peptide strand. It is expected that one peptide may fold into an α-helix and bind weakly to the DNA at the target site, whilst the other peptide will bind in a non-specific manner. Finally, with the CRE site, two peptides are proposed to be able to bind in a specific nature, forming contacts with the bases in the DNA sequence. This conformation preorganises the anthracene moieties to allow them to photodimerise upon irradiation. It is proposed that in the other systems, due to the anthracene

groups not being in close proximity, no photodimerisation will be observed, Figure 4.10.

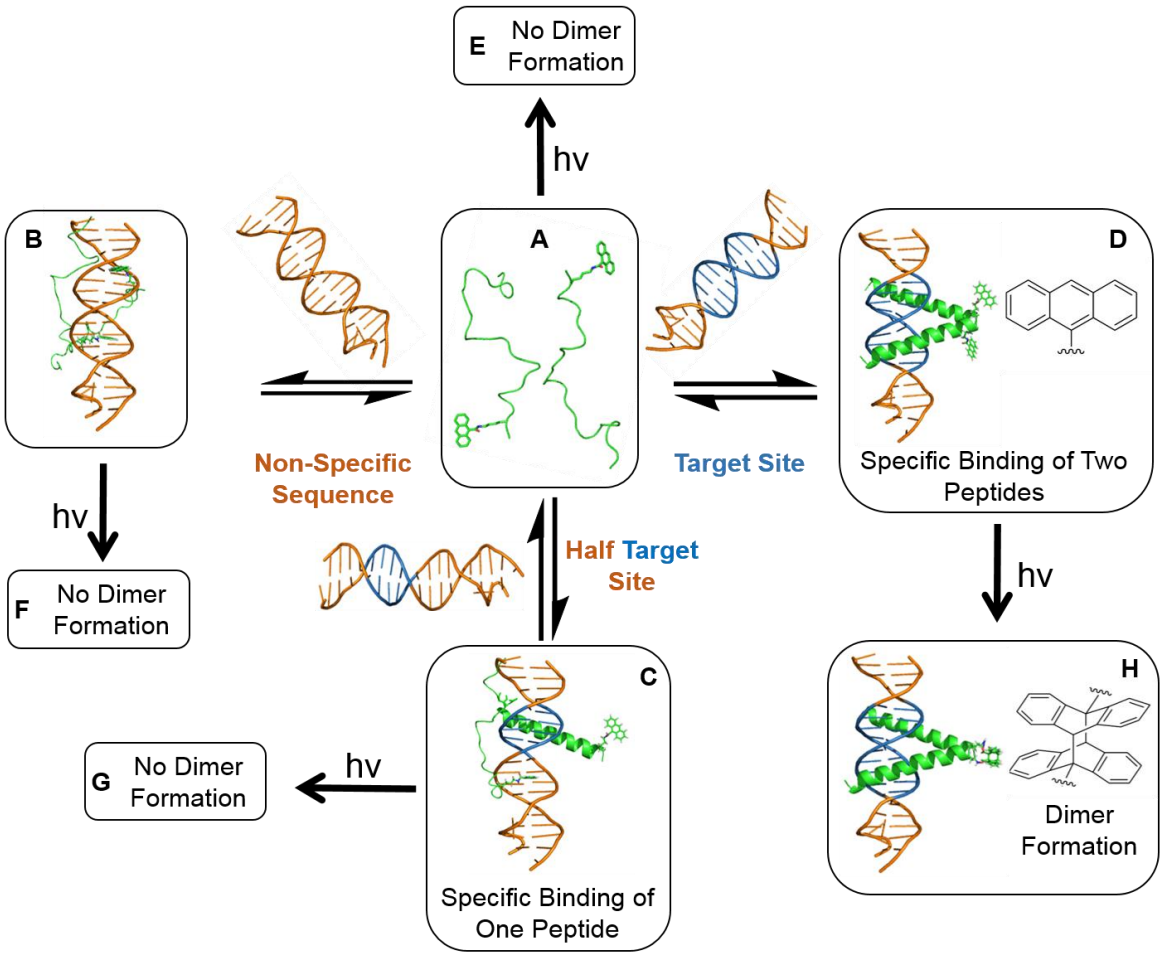


Figure 4.10 The peptide is expected to be unfolded in solution (A), upon addition of NS DNA the peptide will associate in a non-specific manner (B); with half CRE DNA one peptide can fold and form specific interactions (C); whilst with CRE DNA two peptides can bind (D); upon irradiation no photodimerisation is expected with no DNA present (E); when associated with NS DNA (F) or with half CRE DNA (G); when irradiated whilst preorganised with CRE DNA, photodimerisation is expected to occur (H).

4.5 DNA Binding of Anthracene Tagged Peptide

4.5.1 Fluorescence and UV-Vis Tracked Binding

The fluorescence of anthracene is known to be quenched by DNA bases and hence a reduction in fluorescence is expected upon DNA binding.[47,48] It was observed that, upon addition of NS DNA, there was a slight reduction in fluorescence intensity, this is accompanied by a small (2 nm) bathochromic shift in the λ_{max} and could be attributed to intercalation of the anthracene.[49] With the addition of half CRE a greater reduction is seen and a further decrease is seen with the addition of CRE (Figure 4.11). Both of these systems show no bathochromic shift possibly indicating a different binding mode in these systems.[48,50]

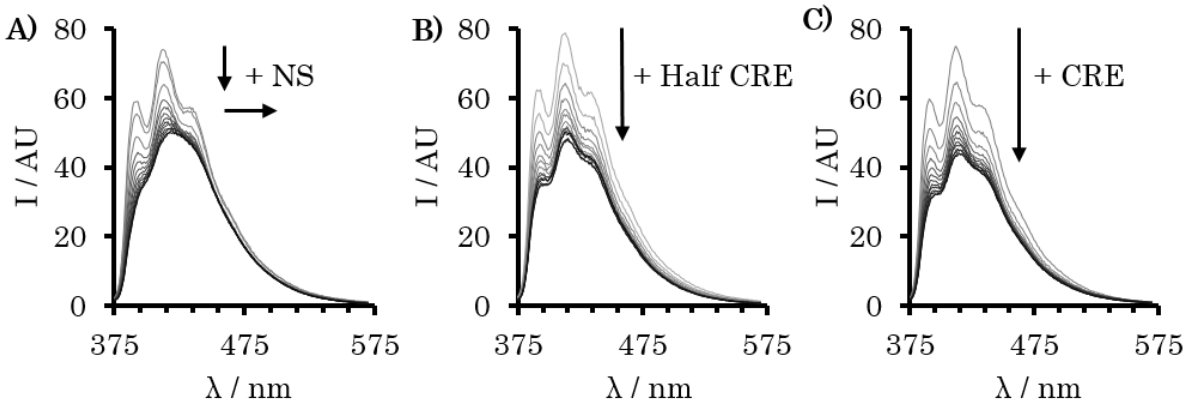


Figure 4.11 Fluorescence spectra of anthracene tagged peptide (2 μM) in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl, in the presence of increasing concentrations (0-3 μM) of duplex A) NS, B) Half CRE or C) CRE DNA at 293 K. λ_{ex} = 325 nm.

In an attempt to quantify the binding affinity and stoichiometry, binding fits were performed on the above fluorescence titrations for CRE and NS DNA, Figure 4.12, using Dynafit software (Biokin Ltd., Massachusetts, USA).[51] This software uses the Newton-Raphson algorithm for computing the composition of the reactants of the equilibrium process. The binding curves are then obtained using the Livermore

solver and a Levenberg-Marquardt least squares regression algorithm.[51] In both cases a 1:1 stoichiometry of peptide:DNA showed poor fitting. However, 2:1 and 3:1 peptide:DNA fits were comparable and could not be used to unambiguously assign DNA binding ratios. This is expected to be as a result of multiple binding modes occurring, such as electrostatics and intercalation of anthracene. Using a stoichiometry of 4:1 and 5:1 displayed poor fits. A similar study was performed using UV-Vis spectroscopy to track DNA binding, Figure 4.13, which displayed similar trends to above.

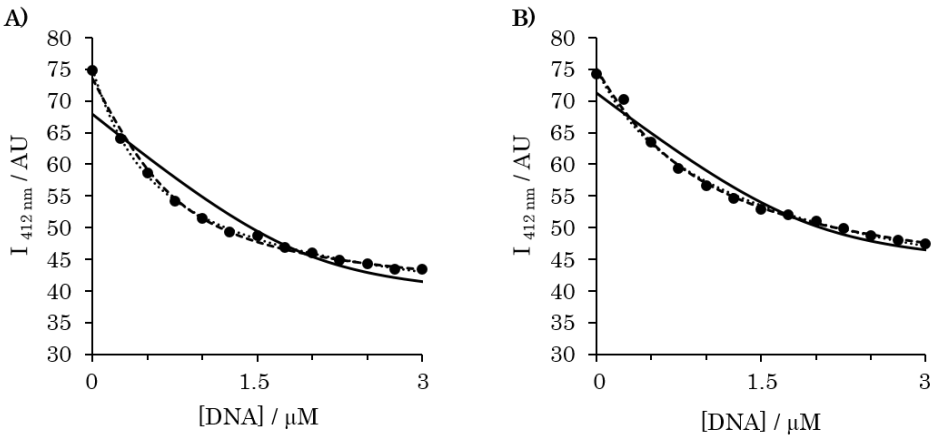


Figure 4.12 Fittings of fluorescence titrations of CRE (A) and NS DNA (B) into 2 μM anthracene tagged peptide using a 1:1 (solid), 2:1 (dash) and 3:1 (dotted) peptide:DNA binding model. Performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K.

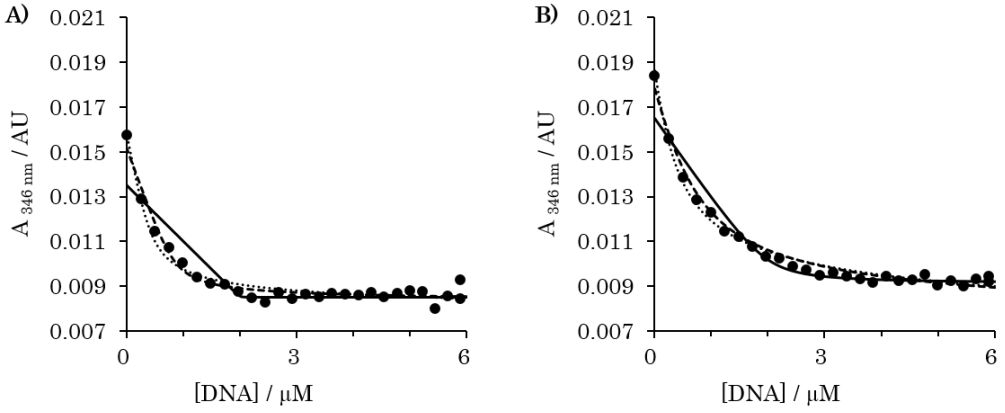


Figure 4.13 Fittings of UV-Vis titrations of CRE (A) and NS (B) DNA into 2 μM anthracene tagged peptide using a 1:1 (solid), 2:1 (dash) and 3:1 (dotted) peptide:DNA binding model. Performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K.

Dissociation constants were calculated using the fits shown in Figures 4.12 and 4.13, Table 4.2. The binding constants obtained are all within the micromolar range, which is comparable to binding of dimer systems reported in the literature.[52–54] Monomer systems have been reported and are typically an order of magnitude weaker.[28] Binding to NS DNA is not commonly reported due to the weaker binding affinity. Within our system, the similarity between the two DNA systems is proposed to be an indication of the relatively high binding affinity the anthracene moiety contributes due to intercalation, in combination with favourable electrostatic interactions between the positive peptide and negative DNA backbone.

Table 4.2 Dissociation constants calculated from fitting fluorescence and UV-Vis titrations of CRE and NS DNA into peptide. All performed using Dynafit software (Biokin Ltd., Massachusetts, USA).

System	Fluorescence $\log K_D$	UV-Vis $\log K_D$
CRE 2:1	-6.1	-6.7
CRE 3:1	-5.8	-6.2
NS 2:1	-5.9	-6.0
NS 3:1	-5.6	-5.9

Job plots were also attempted on the system (Appendix, Figure 8.4). Upon changing the concentration of the peptide, the anthracene signals change in intensity due to the concentration of the chromophore also changing. This was accounted for by completing an identical "Job plot" in the absence of DNA. However, when data was corrected for this, there was a large amount of variation in the points and the stoichiometry could not be deduced.

4.5.2 Circular Dichroism Tracked Binding

Circular dichroism (CD) has also been used to assess if the miniature peptide still recognises the natural DNA binding site. Anthracene tagged peptide monomer was shown to be relatively unfolded (38% ± 1) in the absence of DNA. Upon addition of 1:2 NS DNA to anthracene tagged peptide monomer, very little change to the folding of the peptide (40% ± 2) is observed, Figure 4.14A. This would be expected due to the lack of a binding site, and the formation of only non-specific interactions between the peptide and the DNA, Figure 4.10B. However, the analogous experiment performed with half CRE DNA, lead to an increase in the α-helical content of the peptide (57% ± 2). A further increase is seen upon addition of CRE DNA (75% ± 1), Figures 4.14B and C, respectively. In order to bind, the peptide needs to form specific bonds to the DNA bases and in order to do this the peptide needs to be folded into an α-helix. Upon addition of CRE DNA, two peptides can fold to bind the two target sites present (Figure 4.10D) and this is shown by an increase in the α-helical content from 38 → 75%. In the presence of half CRE, only one peptide can bind as an α-helix (Figure 4.10C) and hence results in half the increase of that seen with CRE DNA.

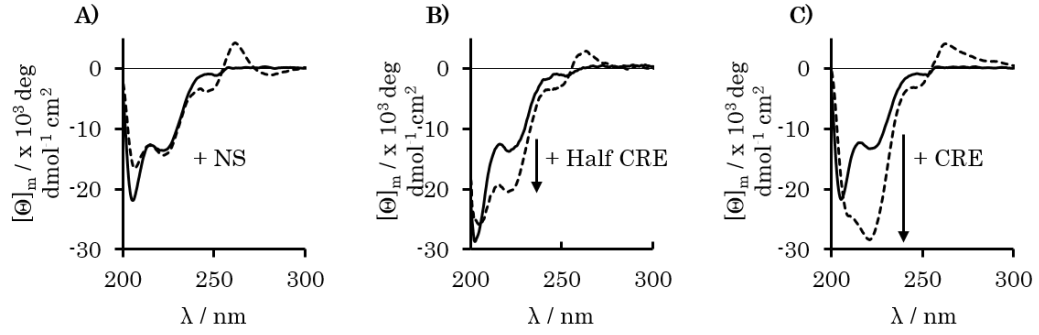


Figure 4.14 CD tracked addition of DNA (1 μM) to peptide (2 μM) in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Peptide (——) and with DNA (-----) for NS (A), half CRE (B) and CRE DNA (C). DNA signals have been subtracted from all spectra.

In all cases changes in the 250-300 nm region of the spectra are observed. These have been attributed to a combination of effects from the anthracene induced signal and changes to the DNA backbone.[55,56]

4.5.3 Thermal Melting Analysis of DNA

It is known that when molecules bind to DNA, by intercalation or *via* groove binding, that the T_m of the DNA can change.[57] As such the T_m values of CRE and NS DNA in the absence and presence of anthracene tagged peptide were measured, Table 4.3. Values are reported to the nearest 0.5°C, although the repeats show error lower than this value, the instrumental error is 0.5°C and therefore the values cannot be reported to more accuracy. It can be seen that with both DNA sequences there is an increase upon addition of anthracene tagged peptide. This agrees with the observations from earlier studies, as both intercalation (in the case of NS DNA),[50,58] but also with major groove binding (in the case of CRE DNA) is expected to be associated with an increase.[57]

Table 4.3 T_m values obtained for CRE and NS DNA (1 μM) in the absence and presence of anthracene tagged peptide (2 μM). All performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7. All values are average of 2 repeat runs and are reported to the closest 0.5 °C due to the limits of experimental error.

DNA Strand	T_m without peptide (°C)	T_m with peptide (°C)
CRE	65.5	68.0
NS	65.5	69.0

4.6 Formation of Anthracene Tagged Peptide Photodimer – Long Irradiation

4.6.1 UV-Vis and Fluorescence Tracking

Initial studies into the photodimerisation of the anthracene tagged peptide were performed by irradiating at 365 nm until the anthracene band in fluorescence and UV-Vis had fully disappeared, as would be expected upon dimer formation. It was found that this occurred over 90 minutes and that with CRE and NS DNA strands, a similar rate of reduction was seen, Figures 4.15 and 4.16.

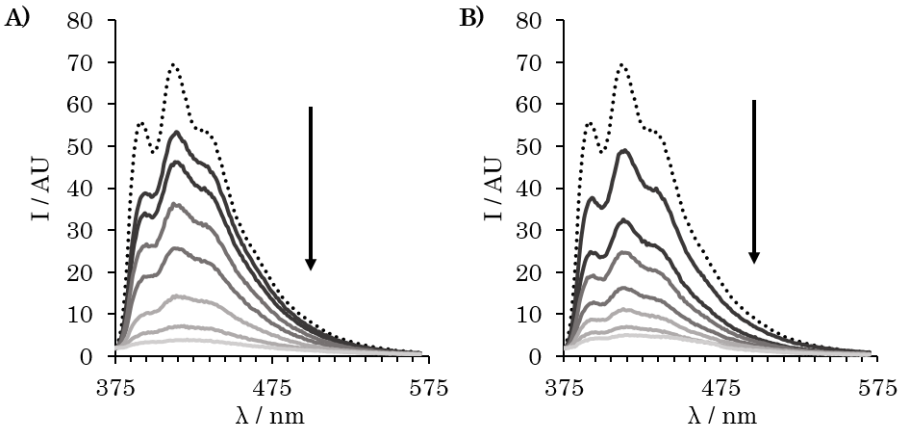


Figure 4.15 Fluorescence tracked irradiation of anthracene tagged peptide with NS (A) and CRE (B) DNA. Peptide (2 μM) shown prior to addition of DNA (·····) and after addition of 1 μM DNA (——) followed by irradiation with 365 nm light for 15, 30, 45, 60, 75 and 90 minutes. Performed in 100 mM NaCl, 10 mM sodium phosphate buffer pH 7 at 293 K. λ_{ex} = 325 nm.

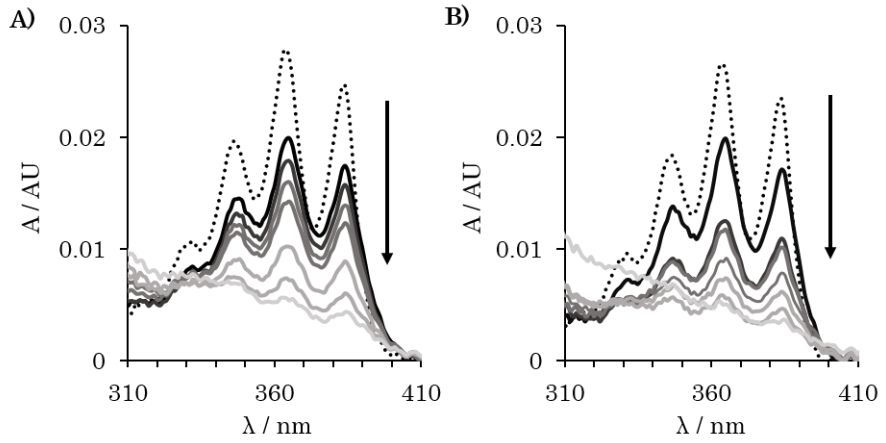


Figure 4.16 UV-Vis tracked irradiation of anthracene tagged peptide with NS (A) and CRE (B) DNA. Peptide (2 μM) shown prior to addition of DNA (·····) and after addition of 1 μM DNA (——) followed by irradiation with 365 nm light for 15, 30, 45, 60, 75 and 90 minutes. Performed in 100 mM NaCl, 10 mM sodium phosphate buffer pH 7 at 293 K.

This would not have been predicted as it was expected that only the CRE DNA would allow for the preorganisation of the anthracene moieties, which is proposed to be a requirement for photodimerisation to occur under these experimental conditions, Figure 4.10D.

4.6.2 Monitoring Photodimer Formation using SDS-PAGE

In order to evaluate if these events are associated with photodimer formation, the irradiated samples were subsequently analysed via SDS-PAGE, Figure 4.17. It can be seen that in the samples that were irradiated in the presence of NS DNA, there is only a very weak band at higher molecular weight, indicating that very little photodimerised anthracene tagged peptide is present. In the samples irradiated in the presence of CRE DNA, a higher molecular weight species is identified, however over the course of the extended irradiation there appears to be no further increase in the amount of this present.

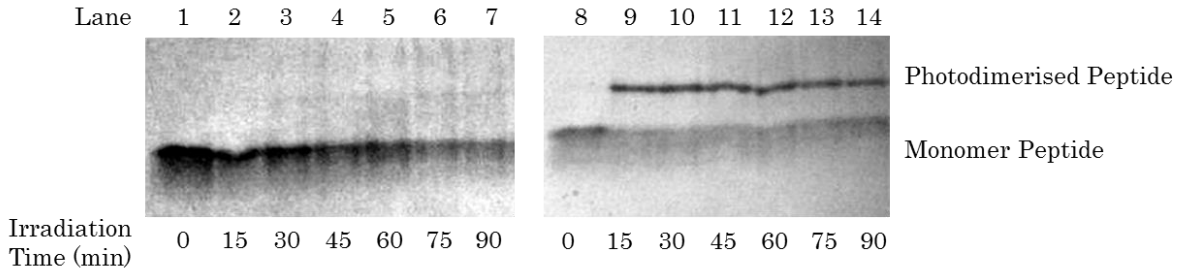


Figure 4.17 SDS-PAGE of 2 μM peptide in the presence of either 1 μM NS (lanes 1-7) or 1 μM CRE (lanes 8-14) DNA in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl irradiated with 365 nm light at 293 K. Aliquots taken prior to irradiation and at 15, 30, 45, 60, 75 and 90 minutes. Visualised using Coomassie brilliant blue R-250 protein stain.

4.6.3 Monitoring Photodimer Formation with Mass Spectrometry

Irradiated samples were analysed by mass spectrometry, to assess whether the new higher molecular weight species is a dimer product and to provide insight into the origin of the decrease in the anthracene bands (not associated with photodimer formation) in the presence of NS DNA. It was found to be problematic to obtain mass spectra for samples which had been irradiated in buffer (and in the presence of salt) and subsequently desalted. In order to overcome this, samples were irradiated in water and analysed by Dr Andrew Jones. The results indicated the presence of several species with the most prevalent corresponding to masses of 3137.8, 3341.9 and 6683.8 Da. Peptide prior to irradiation was run using the same method and gave the expected mass of 3325.9 Da.

A mass of 3341.9 Da species is consistent with the addition of an oxygen atom to the peptide. The corresponding dimer of this would have a mass of 6683.8 Da which is also detected. If we assume the mass of 3137.8 Da is also from an oxidised species, the reduction in relation to 3341.9 Da of 204.1 Da could correspond to the loss of the anthracene and the carbonyl group used to attach it, with the addition of one proton. However, there is no further experimental evidence to support this hypothesis, although the loss of anthracene could account for the loss in fluorescence and UV-Vis signals.

4.6.4 Peptide Irradiated Alone – UV-Vis and MS Analysis

In order to understand the system further, anthracene tagged peptide was irradiated in the absence of DNA and the anthracene band tracked by UV-Vis spectrometry, Figure 4.18. This also indicated a significant reduction in absorbance over the course of two hours.

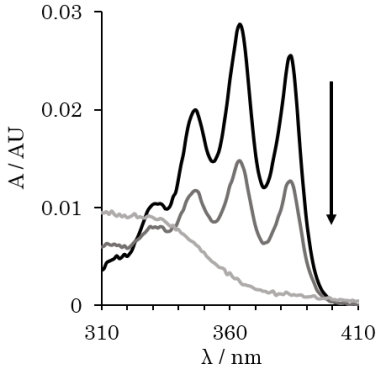


Figure 4.18 UV-Vis tracked irradiation of anthracene tagged peptide (2 μM) shown prior to irradiation (black) and after 1 hour (dark grey) and 2 hour (light grey) irradiation with 365 nm light. Performed in 100 mM NaCl, 10 mM sodium phosphate buffer pH 7 at 293 K.

Analysis of the sample by MS indicated the major species to be of mass 3340.89 Da, a +16 Da species, along with a small amount of 3162.81 Da, corresponding to loss of 162 Da, and no dimer formation. It is proposed that the +16 Da species is the result of oxidation. Subsequent MS fragmentation studies, performed by Dr Cleidiane Zampronio, indicated this to be the case with oxidation of methionine identified, Figure 4.19. Oxidation of methionine is known to occur during light exposure over extended periods of time.[59,60] Interestingly, irradiation of a control peptide (Ac-ALKRARNTEAARRSRARKLQRMKQCG-NH_2) prepared previously in the group by Xun Zhong,[61] which lacks our anthracene unit, but which contains the same methionine residue, showed no oxidation. It is therefore proposed that the anthracene moiety is enhancing the oxidation of the methionine residue.

However, this observation does not account for the reduction in the anthracene signals seen in the UV-Vis and fluorescence spectra recorded after irradiation. If oxidation is occurring on the methionine, then it is assumed that the anthracene it is not being chemically altered, and a decrease in the optical spectra would not be expected.

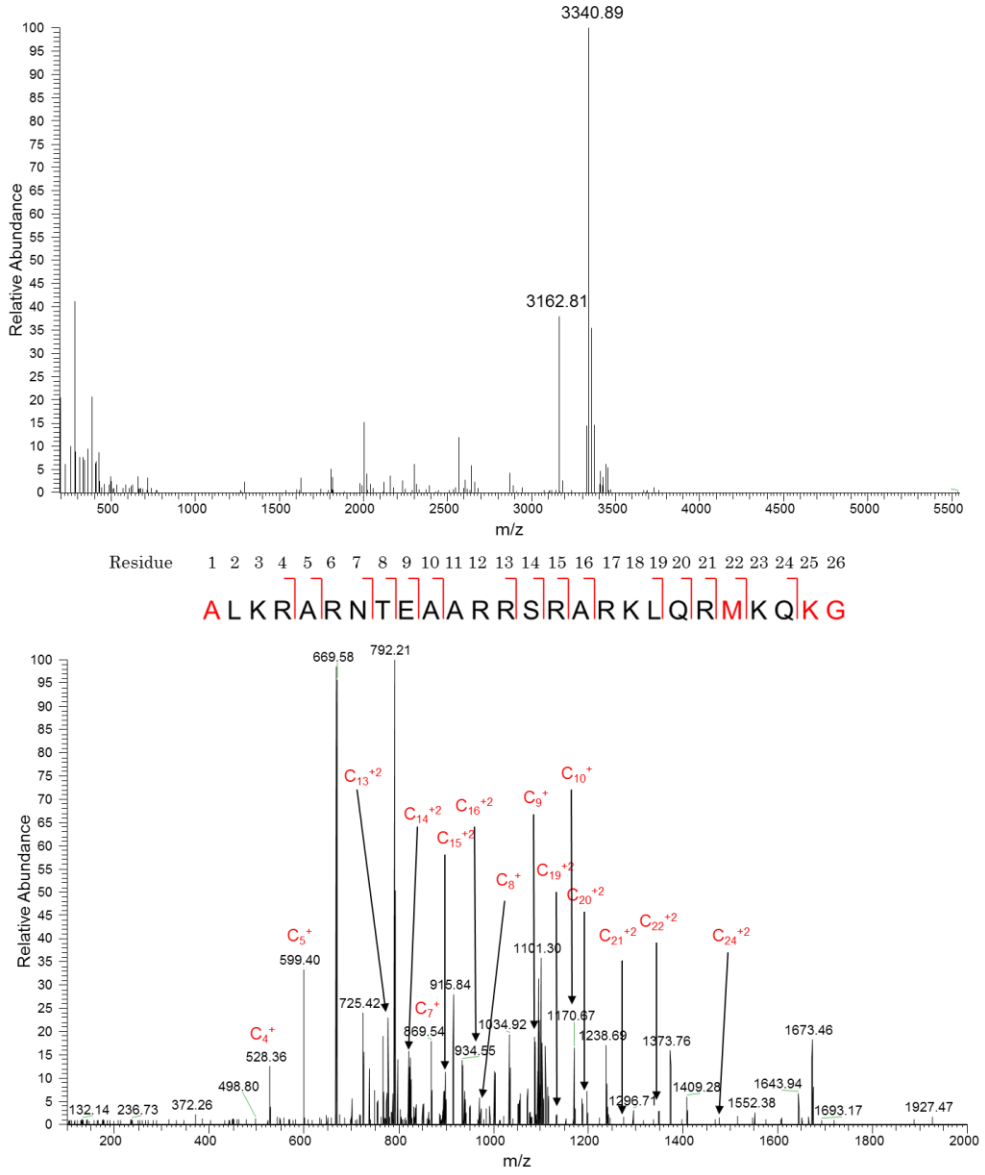


Figure 4.19 ESI-MS of irradiated peptide (top) with fragmentation of the 3340.89 Da peak showing oxidation of methionine (bottom). Experiments performed by Cleidiane Zampronio. Residues shown in red are found to have modifications, alanine is acetylated and glycine amidated as part of the peptide synthesis, the lysine is modified with the anthracene moiety and methionine with the oxygen species being investigated.

It should be noted that mass spectrometry probes the molecules in the gaseous state and it is therefore results are not always indicative of the molecules that are in solution. This could account for the unexplained behaviour observed.

4.6.5 Small Molecule Irradiation

In an attempt to determine the cause of the reduction in UV-Vis and fluorescence signals, small molecule anthracene derivatives were subsequently irradiated and investigated. These studies were performed at low micromolar concentrations to mimic the conditions used in the peptide system, and where possible, in aqueous solvent. This proved difficult and experiments were hindered by precipitation over time. However, MS and NMR of dilute samples of anthracene-9-carboxylic acid were obtained prior to and following, irradiation. The spectra taken after irradiation indicated the complete conversion to anthraquinone, a known product of anthracene irradiation in an oxygen rich environment (Figure 4.20).[62,63] If we apply the same reaction scheme for the formation of anthraquinone proposed by Mallakin and co-workers to our system, it would result in detachment of the anthracene from the peptide. Within the mass spectrum presented in Section 4.6.4, a species of mass 3162.82 Da was detected which could correspond to loss of the anthracene and an additional proton if the methionine is also oxidised. The fragmentation studies reported would also be consistent with this scenario.

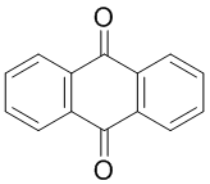


Figure 4.20 Chemical structure of anthraquinone, a known oxidation product of anthracene.

The formation of anthraquinone would be consistent with the reduction in the fluorescence and UV-Vis spectra, as was observed upon irradiation of the anthracene tagged peptide. However, the MS studies on the samples indicate that not all of the peptide is degraded, despite the spectral studies indicating no remaining characteristic anthracene bands.

4.6.6 Degassing Trials

It was known from previous studies[64] that the anthracene moiety used in the DNA work did not display the same degradation problems. A trial was performed irradiating an anthracene tagged oligonucleotide system with and without degassing the sample to monitor the effect, Figure 4.21. Degassing the sample will removed oxygen and therefore decrease the oxidation products formed. This showed a stark difference between the two systems, with no reduction in fluorescence signal when irradiation was performed after degassing.

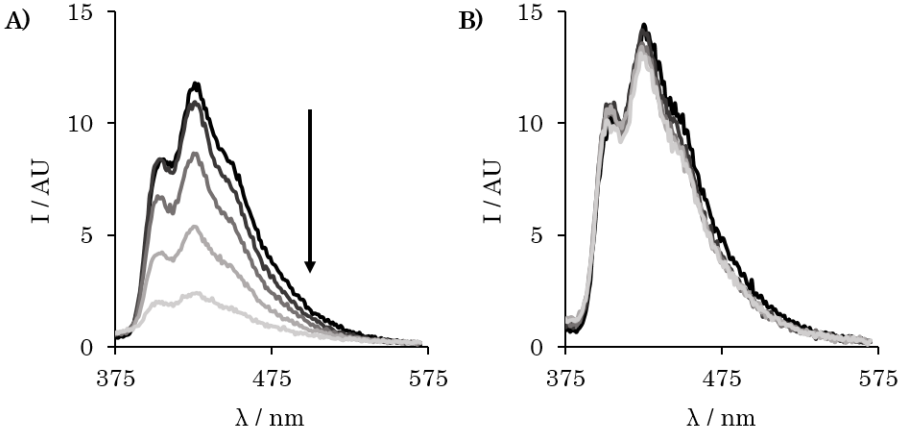


Figure 4.21 Fluorescence tracked irradiation of anthracene tagged oligonucleotide **BM-Pr(3L)** in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Scans taken at 0, 30, 60, 90 and 120 minutes with no prior degassing (A) and with degassing for 10 minutes prior to irradiation and then continued degassing during (B).

In order to minimise the amount of oxidative damage caused with the anthracene tagged peptide, samples were degassed to remove the oxygen as above. Comparison of the degassed sample to one which was subject to no degassing, Figure 4.22, showed the opposite trend to above. The degassed sample showed a greater reduction in the anthracene fluorescence over the course of 30 minutes irradiation.

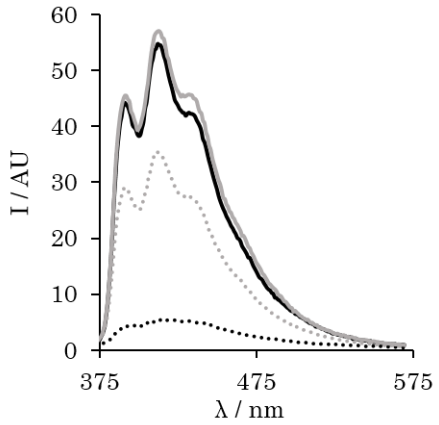


Figure 4.22 Fluorescence tracked irradiation of anthracene tagged peptide without degassing (grey) and with argon degassing (black), prior to irradiation (solid) and after 30 minutes irradiation with 365 nm light (dot). All performed on 2 μM peptide in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K.

In order to conclude that the degassing was not causing this effect due to mixing of the sample rather than the oxygen displacement, a similar experiment was performed with a sample with air being bubbled through, Figure 4.23. This sample displayed the same reduction in fluorescence as that which had not been degassed, confirming the difference is due to the reduction in dissolved gases and not a result of agitation of the sample.

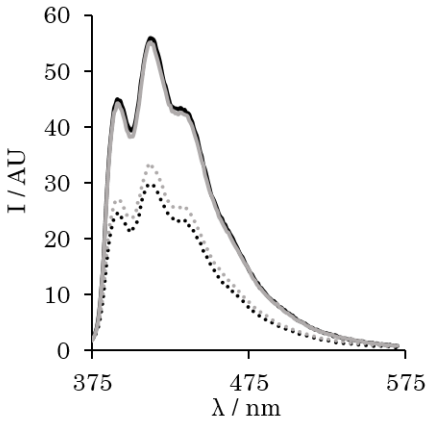


Figure 4.23 Fluorescence tracked irradiation of anthracene tagged peptide without agitation (grey) and with air bubbling (black), prior to irradiation (solid) and after 30 minutes irradiation with 365 nm light (dot). All performed on 2 μM peptide in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K.

These observations are contradictory to the proposed anthraquinone formation, (Section 4.6.3) which would be reduced upon removal of oxygen from the irradiation environment. Further studies need to be performed to analyse the systems further. Possible studies could look to alter the sequence by removing the methionine, and monitoring the influence on irradiation. This could indicate whether this particular residue located close to the anthracene, is problematic. In addition to this it could be helpful to perform NMR studies on the irradiated peptide and compare any changes. Due to the large amounts of material required to do this, it could not be performed at this time. However, this could provide insight in to the site of oxidation and the reduction in the anthracene signals.

4.7 Shorter Irradiations

Subsequent irradiations were performed over shorter time frames (10 minutes). It was hypothesised that this would reduce the amount of photo side products formed.

Analysis of photodimer formation over a period of 30 minutes showed that within 10 minutes photodimer species had formed and the yield (qualitatively by SDS-PAGE) did not increase upon further irradiation, Figure 4.24. The gel was performed with a control covalently linked dimer **(GCN4bd1)$_2$Pyr** which had previously been synthesised within the group.[65] This displayed similar mobility through the gel as the higher molecular weight species present, when the anthracene tagged peptide is irradiated in the presence of CRE DNA, strongly suggesting that this species is the dimer product. Due to these observations subsequent irradiations were performed for a duration of 10 minutes.

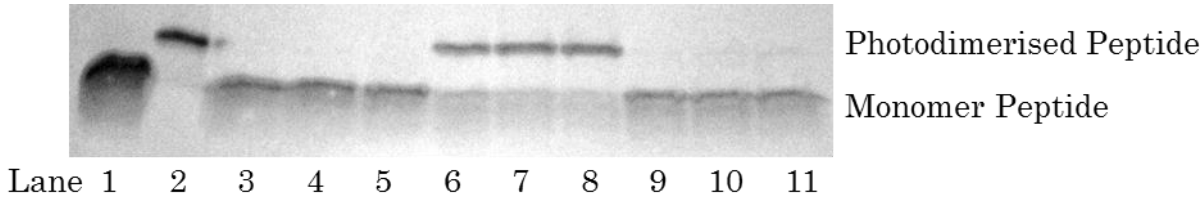


Figure 4.24 SDS-PAGE tracking photodimer formation of 2 μM anthracene tagged peptide with 1 μM DNA where applicable. Lane 1 – anthracene tagged peptide, 2 – **(GCN4bd1)$_2$Pyr** covalent dimer control, 3 – peptide alone irradiated for 10 minutes, 4 – peptide alone irradiated for 20 minutes, 5 – peptide alone irradiated for 30 minutes, 6 – peptide with CRE DNA irradiated for 10 minutes, 7 – peptide with CRE DNA irradiated for 20 minutes, 8 – peptide with CRE DNA irradiated for 30 minutes, 9 – peptide with NS DNA irradiated for 10 minutes, 10 – peptide with NS DNA irradiated for 20 minutes and 11 – peptide with NS DNA irradiated for 30 minutes. Gel visualised using Coomassie brilliant blue R-250 protein stain.

4.7.1 Fluorescence, UV-Vis and CD Tracked Irradiation

Formation of the photodimer species was performed with irradiation with 365 nm light for 10 minutes and monitored by UV-Vis and fluorescence spectroscopy, Figures 4.25 and 4.26. Using this shorter irradiation time, a clear difference can be seen when the anthracene tagged peptide is irradiated with CRE DNA. In both

the fluorescence and UV-Vis spectra, a greater reduction is seen in this sample than any of the other systems.

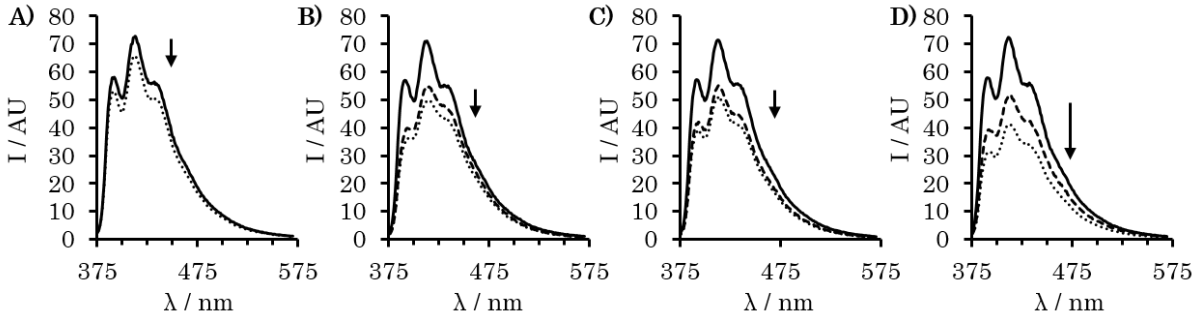


Figure 4.25 Fluorescence tracked irradiation of anthracene tagged peptide (2 μM). Peptide alone (——), with DNA where applicable (----) and after 10 minute irradiation with 365 nm light (·····). A) Peptide in the absence of DNA, B) with NS, C) with half CRE and D) with CRE DNA. All performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. λ_{ex} = 325 nm.

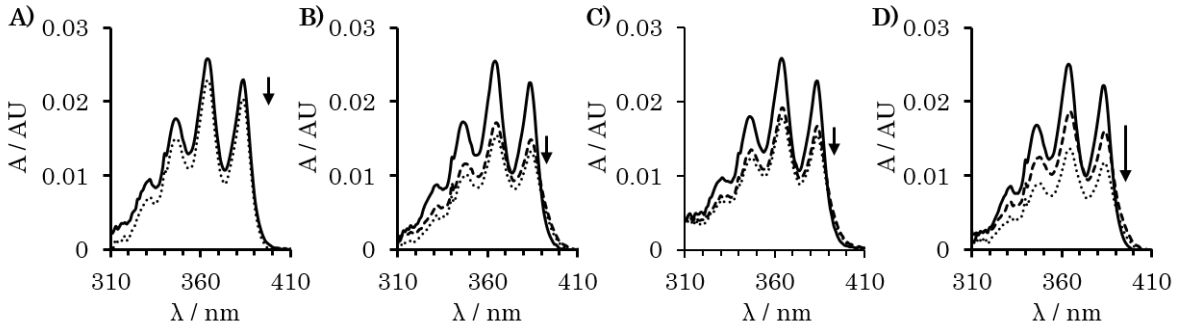


Figure 4.26 UV-Vis tracked irradiation of anthracene tagged peptide (2 μM). Peptide alone (——), with DNA where applicable (----) and after 10 minute irradiation with 365 nm light (·····). A) Peptide in the absence of DNA, B) with NS, C) with half CRE and D) with CRE DNA. All performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K.

Tracking of the irradiation *via* CD also indicates a difference in the four cases, Figure 4.27. The only system which displays a change upon irradiation is the sample irradiated with CRE DNA present, which displays a further increase in the α-helical content.

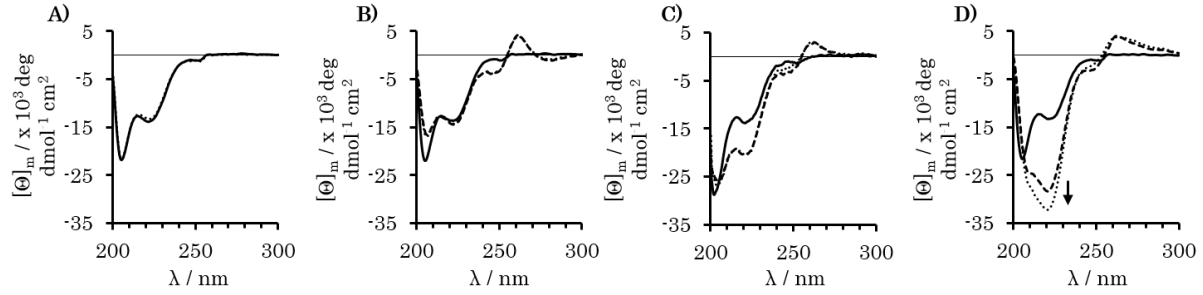


Figure 4.27 CD tracked irradiation of anthracene tagged peptide (2 μM). Peptide alone (——), with DNA where applicable (----) and after 10 minute irradiation with 365 nm light (·····). A) Peptide in the absence of DNA, B) with NS, C) with half CRE and D) with CRE DNA. All performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K.

4.7.2 Summary of Photodimer Formation

In order to have a more quantitative evaluation of the results, the percentage changes in the signals, shown in the above results, were calculated. All values are relative to those of the peptide recorded in the presence of DNA, prior to irradiation (signals are known to change upon DNA addition as discussed at the beginning of this chapter, Figures 4.12, 4.13 and 4.14), Table 4.4.

Table 4.4 Percentage changes in relative absorbance at 364 nm, fluorescence intensity at 426 nm and percentage folding values from ellipticity at 222 nm after 10 minute irradiation with 365 nm with error calculated as standard deviation. All relative to initial value taken after DNA addition and average of 3 repeats.

Sample	% change in absorbance at 364 nm (%)	% change in fluorescence at 426 nm (%)	% folded before irradiation (%)	% folded after irradiation (%)
No DNA	-8 ± 3	-7 ± 2	38 ± 1	36 ± 1
NS DNA	-9 ± 2	-8 ± 3	40 ± 2	40 ± 1
Half CRE DNA	-6 ± 2	-4 ± 3	57 ± 2	56 ± 2
CRE DNA	-27 ± 1	-20 ± 1	75 ± 1	84 ± 1

Reductions in the fluorescence intensity and absorbance, of the anthracene tagged peptide, when irradiated alone or in the presence of half CRE or NS DNA, are comparable. However, when irradiated in the presence of CRE DNA a much larger reduction is seen *via* these two techniques. This extra reduction is proposed to be due to photodimer formation, and illustrates that photodimer formation only occurs when the anthracene moieties are pre-organised in close proximity to one another. In addition to this, the CRE system is the only system that displays a change in α-helical content upon irradiation. This increase in the folding is proposed to be due to an increase in binding affinity associated with the dimer peptide.

4.7.3 Thermal Melting Analysis After Irradiation

Following from the T_m analysis reported in Section 4.5.3, the values were record after irradiation with 365 nm light for 10 minutes, Table 4.5. Results of the irradiated samples were within error of those determined for samples prior to irradiation. This indicates the thermal stability of the DNA duplex has not been effected by the formation of the photodimer peptide species or by the irradiation process.

Table 4.5 T_m values obtained for CRE and NS DNA (1 μM) in the absence and presence of anthracene tagged peptide (2 μM) before and after irradiation for 10 minutes with 365 nm light. All performed in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7. All values are average of 2 repeat runs and are reported to the closest 0.5 °C due to the limits of experimental error.

Sample	T_m before irradiation (°C)	T_m after irradiation (°C)
CRE	65.5	66.0
CRE + Peptide	68.0	68.5
NS	65.5	66.0
NS + Peptide	69.0	68.5

4.8 Confirmation of Photodimer formation

4.8.1 SDS-PAGE

In order to confirm that the reduction seen in the anthracene signals (Figures 4.25 and 4.26) is accompanied by photodimer formation, the samples were analysed by SDS-PAGE, Figure 4.28. As described previously, a control of covalently linked peptide **(GCN4bd1)$_2$Pyr** was also run to allow for comparison. It can be seen that the only sample to display a higher molecular weight species, is that for the peptide irradiated in the presence of CRE DNA. The new species formed shows the same mobility through the gel as that of the covalently linked dimer control, indicating that the species have similar mass and confirming that photodimer has been formed. The presence of the photodimer band is accompanied by a reduction in the

monomer species, as would be expected due to it being converted to the photodimer species. The other irradiated samples show no new bands.

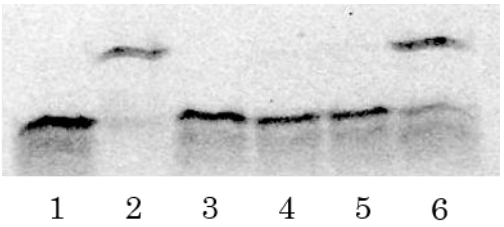


Figure 4.28 SDS-PAGE showing photodimer formation of 2 μM anthracene tagged peptide with 1 μM DNA where applicable. Lane 1 – anthracene tagged peptide, 2-**(GCN4bd1)$_2$Pyr** covalent dimer control, 3 – peptide irradiated alone, 4 – with NS DNA, 5 – with half CRE DNA and 6 – with CRE DNA. Gel visualised using Coomassie brilliant blue R-250 protein stain.

It is known that SDS-PAGE with Coomassie brilliant blue R-250 protein staining cannot be used as a way to quantify the amount of protein present,[66] therefore HPLC analysis was performed on the samples in an effort to determine a yield of dimer formation. However, the resultant chromatogram displayed poor separation and the technique was not pursued any further.

4.8.2 Mass Spectrometry Studies on Short Irradiated Samples

Several attempts were made to obtain mass spectra of peptide irradiated in the absence and presence of DNA strands after short irradiations, for comparison with those obtained in Section 4.6.3. It was proposed that the shorter irradiation times would limit the amount of degradation and oxidation occurring. Despite utilising several desalting methods and performing irradiations in water, suitable spectra could not be obtained.

4.9 Effect of Photodimer Formation on DNA Binding

4.9.1 Gel EMSA Tracked DNA Binding

Although photodimer formation has been proved, the increased binding affinity that this offers had yet to be assessed. Gel EMSA studies were performed which indicate that anthracene tagged peptide shows weak binding affinity to NS, half CRE and CRE DNA strands, Figure 4.29A. This is shown in the gel by smudging of the bands and would be expected due to the various interactions the peptide could have with the DNA, such as electrostatics and anthracene intercalation. After irradiation, gel EMSA images of NS and half CRE DNA samples are similar to those run prior to irradiation. Once again this is as would be hypothesised as there has been no evidence of dimer formation within these systems. In stark difference, with irradiated in the presence of CRE DNA (Figure 4.29B, lanes 13-18), a new band of higher molecular weight corresponding to the photodimer peptide bound to the DNA strand can be observed. Due to the relative strength of this binding, this complex remains intact during the running of the gel.

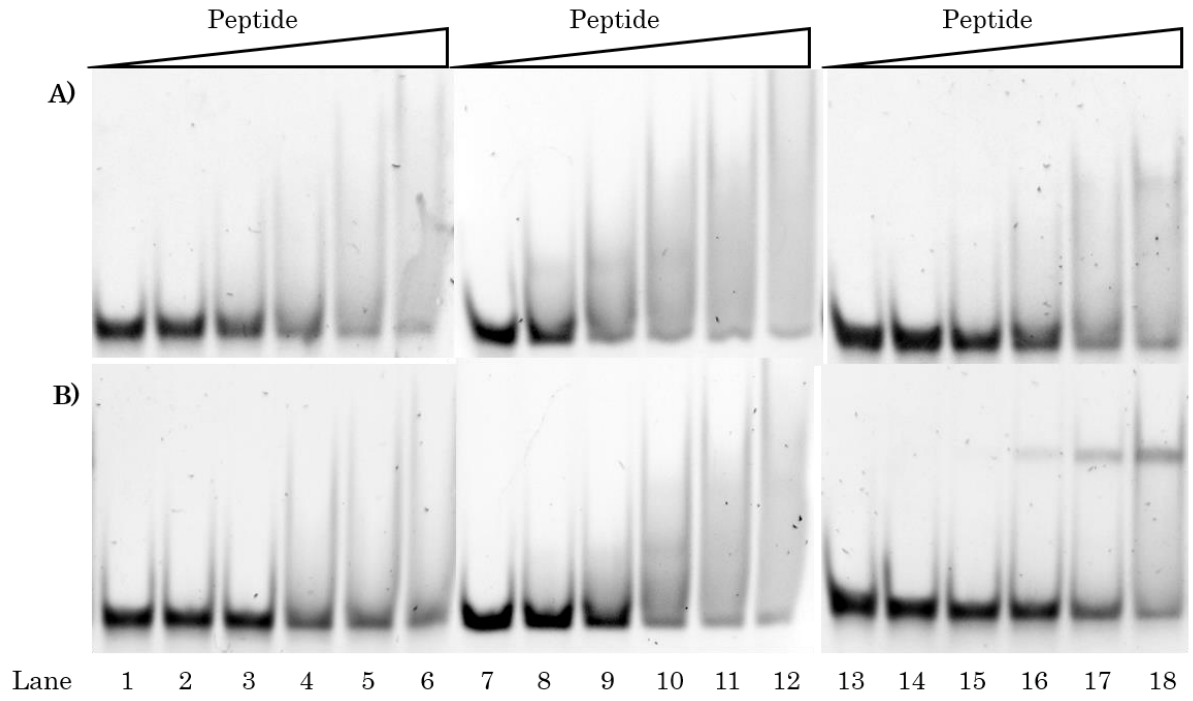


Figure 4.29 Gel EMSA of DNA (100 nM) with increasing anthracene tagged peptide concentrations (0, 100, 200, 400, 600 and 1000 nM) performed prior to irradiation (A) and after 10 minute irradiation with 365 nm light (B) for NS (lanes 1-6), half CRE (7-12) and CRE (13-18) DNA. Gel visualised using SYBR Gold® nucleic acid stain.

4.10 Thermal Reversion Studies

Anthracene dimerisation is known to be a reversible process upon heating and irradiation with lower wavelength light (Chapter 1.5). Samples of anthracene tagged peptide irradiated in the absence of DNA, and with NS and CRE DNA, were heated to 90°C for 10 hours. They were subsequently analysed *via* SDS-PAGE, Figure 4.30. It can be seen that there is no change in the peptide irradiated in the absence of DNA or the peptide irradiated with NS DNA samples, as would be expected. However, there is also no evidence of reversion within the peptide irradiated with CRE DNA samples, shown by retention of the higher molecular weight species (Figure 4.30).

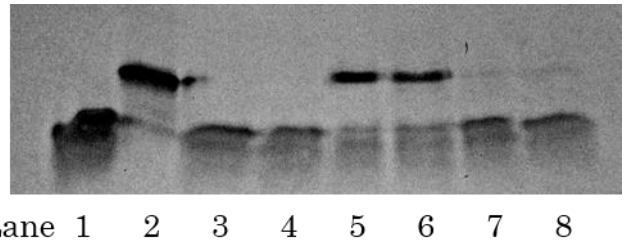


Figure 4.30 SDS-PAGE tracking photodimer reversion. Lane 1 – peptide, 2 – **(GCN4bd1)$_2$Pyr** control, lanes 3-8 show irradiated samples (30 minutes with 365 nm light) and heated samples (10 hours 90 °C) 3-4 – peptide in the absence of DNA, 5-6 – peptide with CRE DNA and 7-8 – peptide with NS DNA. Visualised using Coomassie brilliant blue R-250 protein stain.

It is unknown why this species does not revert upon heating. Previous systems within DNA have been shown to readily revert, and are accompanied by similar trends presented in Chapter 5.[64] It would be interesting to analyse if reversion was achieved with irradiation of a higher energy light. Reversion would increase the useability of this system as it would allow for the development of a fully controllable system.

4.11 Conclusions and Future Work

Within this work the photodimerisation of anthracene has been investigated within a peptide structure for the first time. This has been used to demonstrate a light-triggered, highly specific dimerisation event which occurs at low, biologically relevant micromolar concentrations. It relies upon the preorganisation of the anthracene moieties, adding to the specificity of dimer formation. In addition, photodimers only form between anthracene sub units, allowing for site specific photodimerisation, a property which is not common in current commercially available methods. Studies previously performed on the GCN4 motif have used

photoisomerisation to control binding to DNA. Within the system used in this study, the change in binding affinities between the monomer and photodimer species, is expected to be much larger, although this is yet to be assessed. In the future it is hoped that this technology could be used to control a variety of protein processes, such as the transcription process.

Although shorter irradiation times have been shown to produce the desired photodimer, the products of longer irradiation times are yet to be fully characterised. Further studies on model systems should be performed to allow for NMR and MS characterisation of the products. Deduction of the degradation pathway, could allow for a system to be developed that could prevent such degradation and increase the photodimerisation efficiency. Furthermore, incorporation into different systems would allow the analysis of the dependence this process has on the amino acid sequence, as well as evaluating the photodimerisation process using a different process.

4.12 References

1 D. S. Goodsell and A. J. Olson, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 105–53.
2 G. A. Bullen, J. H. R. Tucker and A. F. A. Peacock, *Chem. Commun.*, 2015, **51**, 8130–3.
3 M. H. Ali and B. Imperiali, *Bioorg. Med. Chem.*, 2005, **13**, 5013–20.
4 A. J. Bailey, *Amino Acids*, 1991, **1**, 293–306.
5 C. T. Rollins, V. M. Rivera, D. N. Woolfson, T. Keenan, M. Hatada, S. E. Adams, L. J. Andrade, D. Yaeger, M. R. van Schravendijk, D. A. Holt, M. Gilman and T. Clackson, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7096–101.
6 P. J. Belshaw, S. N. Ho, G. R. Crabtree and S. L. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 4604–7.
7 E. N. Salgado, R. A. Lewis, S. Mossin, A. L. Rheingold and F. A. Tezcan, *Inorg. Chem.*, 2009, **48**, 2726–8.
8 L.-G. Milroy, T. N. Grossmann, S. Hennig, L. Brunsveld and C. Ottmann, *Chem. Rev.*, 2014, **114**, 4695–748.
9 I. Johnson and M. T. Z. Spence, *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*, Life Technologies Corporation, 11th edn., 2010.
10 G. T. Hermanson, *Bioconjugate Techniques*, Elseview Inc., London, 3rd edn., 2013.
11 G. W. Preston and A. J. Wilson, *Chem. Soc. Rev.*, 2013, **42**, 3289–301.
12 D. A. Fancy and T. Kodadek, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6020–4.
13 D. A. Fancy, C. Denison, K. Kim, Y. Xie, T. Holdeman, F. Amini and T. Kodadek, *Chem. Biol.*, 2000, **7**, 697–708.
14 F. Rahimi, P. Maiti and G. Bitan, *J. Vis. Exp.*, 2009.
15 K. Deville, V. A. M. Gold, A. Robson, S. Whitehouse, R. B. Sessions, S. A. Baldwin, S. E. Radford and I. Collinson, *J. Biol. Chem.*, 2011, **286**, 4659–69.
16 G. Bitan and D. B. Teplow, *Acc. Chem. Res.*, 2004, **37**, 357–64.
17 R. Kodali and R. Wetzel, *Curr. Opin. Struct. Biol.*, 2007, **17**, 48–57.
18 K. Ono, T. Ikeda, J. Takasaki and M. Yamada, *Neurobiol. Dis.*, 2011, **43**, 715–24.
19 G. Bitan, A. Lomakin and D. B. Teplow, *J. Biol. Chem.*, 2001, **276**, 35176–84.
20 G. Bitan, M. D. Kirkitadze, A. Lomakin, S. S. Vollers, G. B. Benedek and D. B. Teplow, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 330–5.
21 U. C. Krieg, P. Walter and A. E. Johnson, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8604–8.
22 C. Wood and M. O'Dorisio, *J. Biol. Chem.*, 1985, **260**, 1243–1247.
23 M. Suchanek, A. Radzikowska and C. Thiele, *Nat. Methods*, 2005, **2**, 261–7.
24 H. Bouas-Laurent, A. Castellan, J.-P. P. Desvergne and R. Lapouyade, *Chem. Soc. Rev.*, 2001, **30**, 248–63.

25 H. Bouas-Laurent, A. Castellan, J.-P. P. Desvergne and R. Lapouyade, *Chem. Soc. Rev.*, 2000, **29**, 43–55.
26 G. W. Breton and X. Vang, *J. Chem. Educ.*, 1998, **75**, 81–2.
27 T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223–37.
28 W. Cao, L. Liu, D. E. Klein, L. Wei and L. Lai, *Thermochim. Acta*, 2000, **360**, 47–56.
29 W. Keller, P. König and T. J. Richmond, *J. Mol. Biol.*, 1995, **254**, 657–67.
30 J. W. Sellers, A. C. Vincent and K. Struhl, *Mol. Cell. Biol.*, 1990, **10**, 5077–86.
31 Y. Azuma, M. Imanishi, T. Yoshimura, T. Kawabata and S. Futaki, *Angew. Chem. Int. Ed. Engl.*, 2009, **48**, 6853–6.
32 S.-A. A. Morgan, S. Al-Abdul-Wahid and G. A. Woolley, *J. Mol. Biol.*, 2010, **399**, 94–112.
33 G. A. Woolley, A. S. I. Jaikaran, M. Berezovski, J. P. Calarco, S. N. Krylov, O. S. Smart and J. R. Kumita, *Biochemistry*, 2006, **45**, 6075–84.
34 R. V Talanian, C. J. McKnight and P. S. Kim, *Science*, 1990, **249**, 769–71.
35 M. Ueno, A. Murakami, K. Makino and T. Morii, *J. Am. Chem. Soc.*, 1993, **115**, 12575–76.
36 Y. Aizawa, Y. Sugiura and T. Morii, *Biochemistry*, 1999, **38**, 1626–32.
37 B. Cuenoud and A. Schepartz, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 1154–9.
38 E. Oheix and A. F. A. Peacock, *Chem. Eur. J.*, 2014, **20**, 2829–39.
39 A. Caamaño, M. Vázquez, J. Martínez-Costas, L. Castedo and J. Mascareñas, *Angew. Chem. Int. Ed. Engl.*, 2000, **39**, 3104–7.
40 T. Morii, Y. Saimei, M. Okagami, K. Makino and Y. Sugiura, *J. Am. Chem. Soc.*, 1997, **119**, 3649–55.
41 J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vázquez and J. L. Mascareñas, *Nat. Commun.*, 2013, **4**, 1874.
42 R. V. Talanian, C. J. McKnight, R. Rutkowski and P. S. Kim, *Biochemistry*, 1992, **31**, 6871–5.
43 W. T. Pu and K. Struhl, *Proc. Natl. Acad. Sci.*, 1991, **88**, 6901–5.
44 E. Kaiser, R. L. Colescott, C. D. Bossinger and P. I. Cook, *Anal. Biochem.*, 1970, **34**, 595–598.
45 G. Balasundaram, T. Takahashi, A. Ueno and H. Mihara, *Bioorg. Med. Chem.*, 2001, **9**, 1115–21.
46 W. C. White and P. D. Chan, *Fmoc solid phase peptide synthesis*, Oxford University Press: New York, New York, 2000.
47 A. Rodger, S. Taylor, G. Adlam, I. S. Blagbrough and I. S. Haworth, *Bioorg. Med. Chem.*, 1995, **3**, 861–72.
48 N. K. Modukuru, K. J. Snow, B. S. Perrin, J. Thota and C. V Kumar, *J. Phys. Chem. B*, 2005, **109**, 11810–8.
49 N. Moran, D. M. Bassani, J.-P. Desvergne, S. Keiper, P. a S. Lowden, J. S. Vyle and J. H. R. Tucker, *Chem. Commun.*, 2006, 5003–5.

50 M. Kozurková, D. Sabolová, H. Paulíková, L. Janovec, P. Kristian, M. Bajdichová, J. Busa, D. Podhradský and J. Imrich, *Int. J. Biol. Macromol.*, 2007, **41**, 415–22.

51 P. Kuzmic, *Anal. Biochem.*, 1996, **237**, 260–73.

52 X. Wang, W. Cao, A. Cao and L. Lai, *Biophys. J.*, 2003, **84**, 1867–75.

53 C. Berger, L. Piubelli, U. Haditsch and H. Rudolf Bosshard, *FEBS Lett.*, 1998, **425**, 14–18.

54 S. Cranz, C. Berger, A. Baici, I. Jelesarov and H. R. Bosshard, *Biochemistry*, 2004, **43**, 718–27.

55 M. R. Duff, V. K. Mudhivarthi and C. V Kumar, *J. Phys. Chem. B*, 2009, **113**, 1710–21.

56 M. A. Weiss, T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison and K. Struhl, *Nature*, 1990, **347**, 575–8.

57 K. R. Fox, *Drug-DNA Interaction Protocols*, Springer, New York, 2 nd., 2010.

58 W. B. Tan, A. Bhambhani, M. R. Duff, A. Rodger and C. V Kumar, *Photochem. Photobiol.*, **82**, 20–30.

59 X. M. Lam, J. Y. Yang and J. L. Cleland, *J. Pharm. Sci.*, 1997, **86**, 1250–5.

60 H. Liu, G. Gaza-Bulseco and L. Zhou, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 525–8.

61 X. Zhong, PhD Thesis, University of Birmingham, 2014.

62 A. Mallakin, D. George Dixon and B. M. Greenberg, *Chemosphere*, 2000, **40**, 1435–41.

63 H. Ihmels, D. Leusser, M. Pfeiffer and D. Stalke, *Tetrahedron*, 2000, **56**, 6867–75.

64 J. Manchester, PhD Thesis, University of Birmingham, 2013.

65 E. Oheix, University of Birmingham, 2014.

66 V. Neuhoff, R. Stamm, I. Pardowitz, N. Arold, W. Ehrhardt and D. Taube, *Electrophoresis*, 1990, **11**, 101–17.

Chapter 5

Photo-Triggered Release of DNA

5.1 Introduction

This chapter focuses on efforts to achieve the photo-triggered release of single stranded DNA, from an assembly that contains a region of duplex DNA. The strategy adopted will use anthracene photodimerisation to "lock" DNA into a particular conformation, and this is hypothesised to trigger the release of a complementary strand.

5.1.1 DNA as a Therapeutic

In recent years the use of nucleic acids as therapeutic agents has been explored.[1,2] These drugs have the potential to silence genes and insert new genetic material into a person's DNA, and hence supress or completely remove symptoms of particular diseases.[1,2] The nature of DNA makes these therapeutics extremely specific, but due to the body's natural ability to destroy foreign DNA, hard to deliver.[3] Several approaches to overcome this have been developed using a variety of different delivery and targeting mechanisms.[1]

Of particular note is the field of antisense RNA research, an area of nucleic acid therapeutics which has progressed more rapidly than others. This technique takes advantage of the propensity of nucleic acids to form duplexes with complementary strands. During the gene expression process, messenger RNA (mRNA) is formed (transcription) and it is this which is "read" to create proteins in the translation process (Chapter 1.2.2). This process only occurs when the mRNA is single stranded and hence binding a complementary strand to the region of interest prevents its expression. Fomivirsen was the first such drug to make it to market in 1998, although it has since been removed due to poor demand.[4] A recent article details over 20 drugs utilising the antisense strategy which are in late stage clinical trials.[5]

5.1.2 DNA Origami and Machines

In addition to its obvious biological applications, DNA has also proved to be a useful tool in the development of molecular machines. Due to the specific and predictable nature of its binding and folding, elaborate structures can be produced from DNA. In 2006 Rothemund reported on the use of a relative simple system formed from combining a single stranded DNA scaffold (obtained from the M13 bacteriophage) with several short oligonucleotide staples. The staples link the scaffold at particular locations *via* H-bonded base-pairing, which folds the DNA into a variety of 2-dimensional structures, Figure 5.1.[6] More complex 3D structures which can be manipulated using external stimuli, have since been developed and are summarised in a recent review by Saccà and Niemeyer.[7]

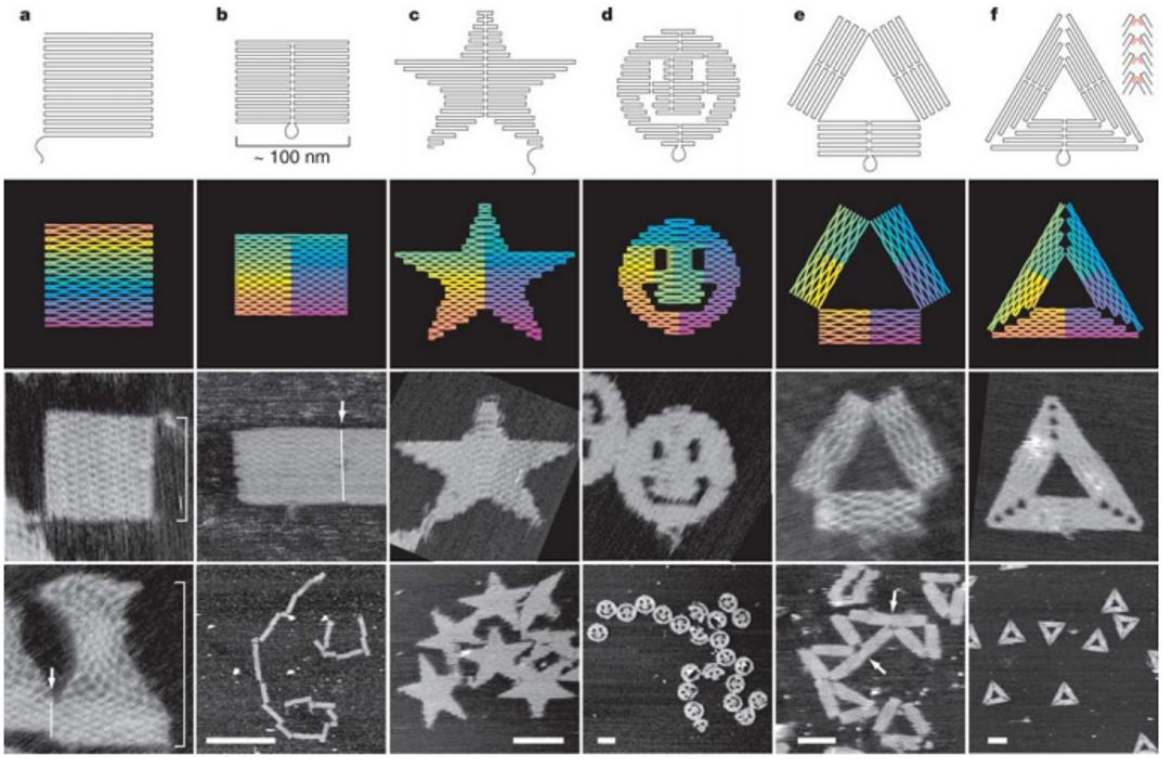


Figure 5.1 2D structures made from DNA strands in a technique known as DNA origami. Top two rows show the designs of the DNA strands; the second row shows this with the sequence colour coded from red to purple to visualise the folding. Bottom two rows are AFM images of the annealed complexes. Reprinted with permission from Rothemund *et al.*, *Nature*, 2006, **440**, 297-302. Copyright 2006 Nature Publishing Group.

It is anticipated that coupling this predictable folding and structure with the ability of DNA to be programmed to perform various tasks upon addition of an external stimulus, would allow molecular machines to be made. These properties have led to extensive research in the area of DNA-based devices, research which is vast and beyond the scope of this thesis. In 2011 Krishnan and Simmel published a comprehensive review into work that had been completed so far.[8] Of particular note is the work by Turberfield and co-workers who have developed molecular machines fuelled by DNA strands.[9] These include DNA walkers that can be programmed to follow a set route,[10,11] and reversible, shape changing DNA nanostructures.[12]

5.1.3 Use of Photo-Triggered Release

Much research has focussed on releasing a single stranded oligonucleotide (ssDNA) from a duplex, which could then be used as an antisense strand or as a stimulus to control a molecular machine. Aside from the traditional ways to denature the DNA duplex and release ssDNA, such as increasing temperature or altering the pH, release can be controlled electrochemically.[13] This work incorporated a daunomysin group into the DNA structure, the oxidised form of which intercalates in the DNA structure giving a stable duplex. However, the reduced form significantly lowers the T_m of the duplex and, at particular temperatures, ssDNA is formed.

Of interest to this work, is the ability to trigger the release of an oligonucleotide *in situ* utilising light. Groups have made efforts towards this goal utilising photo-cross-linked hydrogels[14] and photoresponsive catanionic vesicle complexes, which contain the ssDNA of interest.[15]

Many other groups have focused on releasing a complementary strand from duplex DNA using light, and work in this area has been summarised in a recent review, below follows highlights from this.[16] Asanuma and co-workers incorporated azobenzene derivatives into DNA to replace one base. Upon irradiation with light, azobenzene undergoes a *cis-trans* isomerisation. In the *trans* form the derivative can intercalate with surrounding bases and a duplex is formed. However, upon isomerisation the duplex is disrupted, releasing the complementary strand.[17] This technique has since been utilised in many molecular devices.[18–22] This method, as with the electrochemical technique described above, relies upon the change in T_m afforded, by the reduced ability of the modification to intercalate into the duplex of DNA. In order for complete release to be achieved, several azobenzene modifications are required. A recent study investigated various duplexes of the *cis* and *trans* isomers and showed that both incorporations distorted the B-DNA fold.[23] The authors commented that they were analysing the melting of DNA-azobenzene derivatives, but this work is yet to be published.

Spiropyran[24–26] and diarylethene[27] derivatives have also been incorporated into DNA. Although both these groups display photoisomerisation properties that are maintained once incorporated into the DNA structure, they induce minimal effects on the stability of DNA duplexes.

In addition to the work performed using azobenzene, Asanuma and colleagues have also reported on the irreversibe photocrosslinking of duplex DNA. The group incorporated *p*-stilbazole derivatives into each of the DNA strands and showed that the duplex structure could still form with these modifications. Upon irradiation with 340 nm light, an irreversible crosslink was formed, Figure 5.2.[28]

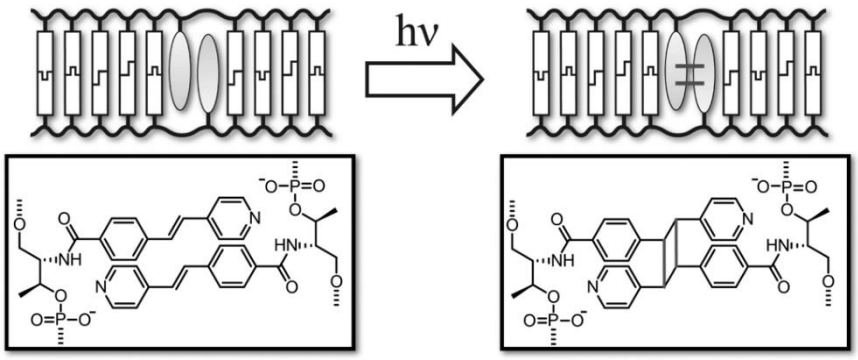


Figure 5.2 Schematic representation of the photocrosslinking of DNA using *p*-stilbazole derivatives. Incorporation to replace one base on each of the complementary strands results in stacking of the *p*-stilbazole units. Subsequent irradiation with 340 nm light crosslinks the moieties and hence the DNA. Reprinted with permission from Kashida *et al.*, *J. Am. Chem. Soc.*, 2013, **135**, 7960-6. Copyright 2013 American Chemical Society.

In 2012, Tucker and co-workers reported on the photocontrolled binding of DNA strands using anthracene photodimerisation. It was demonstrated that when two anthracene moieties within the same sequence of DNA were photodimerised, a

duplex with the complementary strand could not be formed. However, duplex formation followed by irradiation prevented the anthracene photodimer from forming, Figure 5.3.[29]

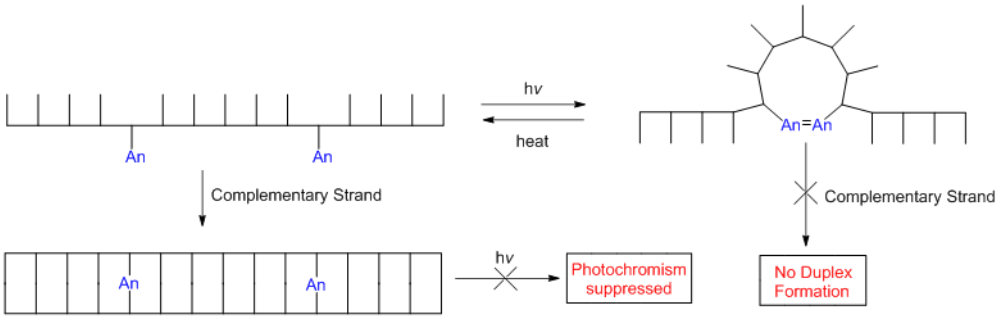


Figure 5.3 Anthracene moieties have been incorporated into an oligonucleotide to replace one base. Upon irradiation with 365 nm light, the anthracene photodimer product can be formed, folding the DNA and preventing duplex formation. However, if the anthracene modified oligonucleotide forms a duplex prior to irradiation, no photodimer product is observed. Adapted with permission from Manchester *et al.*, *J. Am. Chem. Soc.*, 2012, **134**, 10791-4. Copyright 2012 American Chemical Society.

It was anticipated that adding to the existing tools, a photo-activated reagent that could crosslink a specific duplex of DNA and hence release a complementary strand, would allow for more complex DNA machines to be made.

5.1.4 Previous Efforts Towards Photo-Triggered Release

Previously within the Tucker group, research was directed towards achieving the photo-triggered release of a short oligonucleotide strand by building on the work described above.[30] Two anthracene moieties, as described in Chapter 2.3, were incorporated into one DNA sequence to allow for photodimerisation, with the hypothesis that the oligonucleotides could be "locked" into a conformation, and release the complementary strand. The initial system consisted of an

oligonucleotide with anthracene attached to either end. Upon irradiation it was hypothesised that the tagged oligonucleotide would form a loop and the complementary strand would be displaced (Figure 5.4 A). However, very little or no photodimer formation was observed within this system, presumably due to the base pairing preventing the anthracene units from coming into close proximity. In an effort to overcome this, the anthracene tagged strand was extended to incorporate extra bases between the anthracene tags and the duplex region, using so-called "floppy ends" (Figure 5.4 B). However, upon irradiation, if anything this modification made the situation worse, with now no photodimer formed at all.

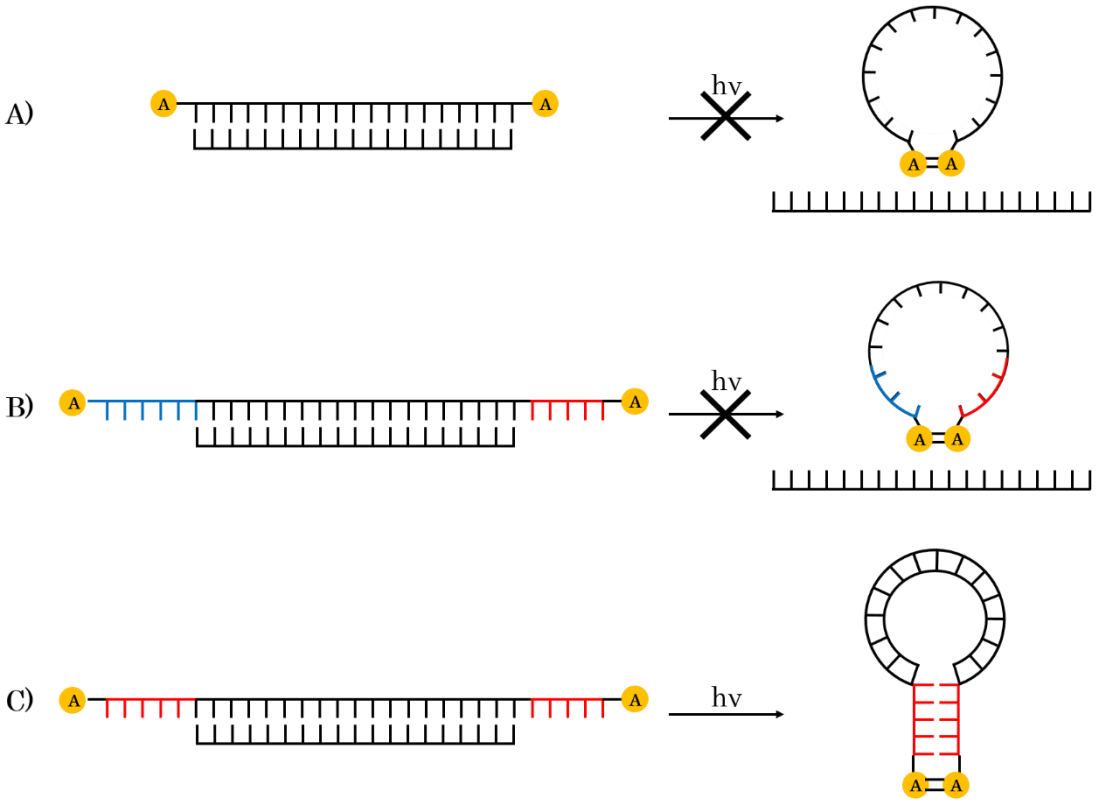


Figure 5.4 Designs previously studied within the Tucker group for photo-triggered release of an oligonucleotide. A) A duplex system with anthracene incorporated at the ends. It was hypothesised that, upon irradiation this would form a loop, displacing the complementary sequence. B) "Floppy ends" modification with extra bases between the anthracene and the duplex DNA (red and blue) was shown to not form photodimers. C) A stem loop system with complementary overhangs (red), to help force the system into a folded conformation, to aid photodimer formation. It was found that, although photodimer did form in this case, the complementary strand remained bound in the loop.

A third generation system utilised a stem loop system. Within this the "floppy ends" were designed to be complementary to one another, with the aim to push the equilibrium towards the loop folded system and increase the extent of photodimer formation (Figure 5.4 C). This modification proved successful and yielded photodimer species. However, further analysis utilising fluorescence tagged systems indicated that, although the photodimer stem loop had been formed, the complementary strand was not displaced by this process, and in fact remained bound to the stem loop, Figure 5.4 C.

5.2 Project Aims

Within this project, the stem loop system will be taken and extra bases will be added to the complementary strand to form an overlapping region. It is hypothesised that this will allow for the photodimer stem loop to still form, but due to the overlapping bases, the complementary strand will now be displaced. This will be assessed using optical techniques along with gel electrophoresis. The reversibility of the system will also be assessed.

5.3 Stem Loop System

5.3.1 Design of Stem Loop Release System

It had been previously suggested that incorporating additional bases into the complementary strand, which overlapped with the stem loop, may allow for the release of the strand.[30] Therefore a sequence was designed which incorporated two

extra bases, overhanging on each end of the release strand (**OH1**), which were complementary to the stem sequence, Figure 5.5. No changes were made to the stem loop sequence (**SL**), which utilised anthracene units with D stereochemistry and a carbon linker length of 7, which had previously been shown to form photodimer products. The sequences synthesised are shown in Table 5.1.

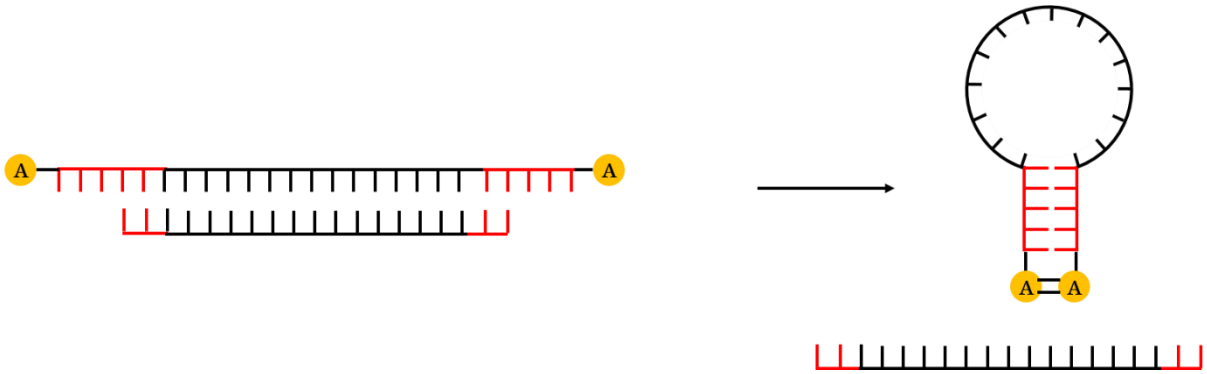


Figure 5.5 Design of extended complementary system, which incorporates complementary bases to the stem loop (red). It is hypothesised that the stem loop will form and become locked in this conformation by photodimerisation, which is expected to displace the complementary strand.

Table 5.1 Oligonucleotide sequences used within this study. The stem loop sequence (**SL**) has anthracene moieties identified as "X" (carbon linker of 7, D stereochemistry) incorporated on the ends. The extended ends of the **SL** sequence (underlined, bold) are complementary to one another to promote stem loop formation. Additional bases (underlined, bold) were added to the complementary strand, **OH1**.

Name	Sequence
SL	5'- X**GCGAC**TGGACTCTCTCAATG**GTCGC**X -3'
OH1	3'- **TG**ACCTGAGAGAGTTAC**CA** -5'

5.3.2 Thermal Melting Studies

The T_m values of the sequences were measured as described in Chapter 2.7.3. A T_m of 66 °C was recorded for the **SL** sequence alone, and 65 °C for **SL** duplex with **OH1**. The high T_m for the hairpin **SL**, despite having only five complementary

bases pairs, is thought to be due to the two anthracene units stacking with one another and hence increasing its stability. The T_m for the duplex of **SL** and **OH1** is higher than the previous target (the duplex without the extended bases, as used in previous studies, gave a T_m of 54 °C).[30] This would be expected due to the extra base pairs now available and theoretical calculations predicted a T_m of 62 °C. The increase in relation to this prediction can be rationalised by stabilising interactions provided by the anthracene modifications. This was also observed in the original system which had a predicted T_m of 53 °C and has also been seen in other duplex systems.[30,31] The two systems, duplex (**SL:OH1**) and stem loop (**SL**), have T_m values within error of one another, suggesting they have very similar stabilities. This should allow for a dynamic equilibrium between the two conformations and therefore photodimer formation is expected to be possible.

5.3.3 Photodimerisation Studies

Initial irradiation studies were performed on **SL** in the absence of target to ensure photodimer formation was possible, and to confirm the results obtained previously on this strand, Figure 5.6.[30]

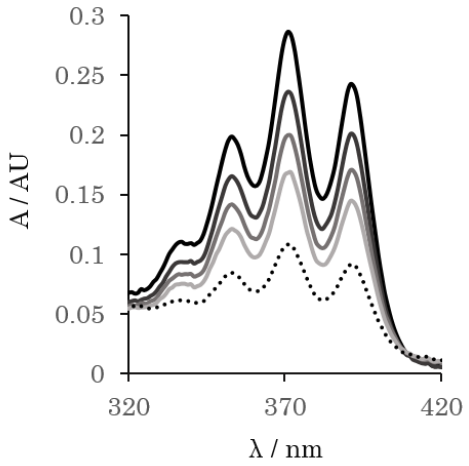


Figure 5.6 UV-Vis spectra tracking photodimerisation of **SL**. Performed with 20 µM **SL** in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Sample irradiated with 365 nm light after 10 minutes degassing with argon. Scans recorded every 10 minutes for 30 minutes (black to grey) followed by a further 30 minute irradiation (dotted).

Over the course of an hour, the majority of the anthracene signal had disappeared at a rate that was comparable to that seen previously, and attributed to the conversion to photodimerised species, **SL-D**.

Subsequently, the **SL:OH1** duplex (1:1 ratio) was irradiated and the anthracene signal tracked as before, Figure 5.7.

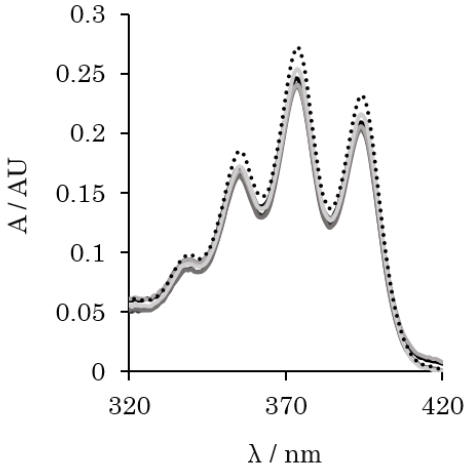


Figure 5.7 UV-Vis spectra tracking photodimerisation of **SL:OH1** duplex. Performed with 20 µM **SL** and 20 µM **OH1** in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Sample irradiated with 365 nm light after 10 minutes degassing with argon. Scans recorded every 10 minutes for 60 minutes (black to grey) followed by a further 30 minute irradiation (dotted).

However, despite an extended irradiation period of 90 minutes, there was no decrease in the anthracene signal, indicating that no photodimer had been formed. This was consistent with duplex formation dominating the equilibrium and restricting the folding of the stem loop, despite the T_m values being similar. In this conformation, the anthracene moieties would be held far apart (approximately 60 Å), preventing the photodimer forming.

5.3.4 New Reduced Length Targets

In an attempt to overcome this issue, further target strands were synthesised, with fewer complementary bases within the stem region of the stem loop sequence, Table 5.2. It was hypothesised that these sequences would still allow for photodimer formation as seen in previous work (Figure 5.4),[30] whilst also resulting in the displacement of the complementary sequence by virtue of the photodimer not binding the dimerised stem loop strand.

Table 5.2 New oligonucleotide sequences compared to those previously used (Table 5.1). Bases complementary to the stem region of **SL** (underlined, bold) are varied between sequences.

Name	Sequence (5′ → 3′)
SL	5′- X**GCGAC**TGGACTCTCTCAATG**GTCGC**X -3′
OH1	3′- **TG**ACCTGAGAGAGTTAC**CA** -5′
OH2	3′- **TG**ACCTGAGAGAGTTAC -5′
OH3	3′- ACCTGAGAGAGTTAC**CA** -5′

5.3.5 Photodimerisation with Reduced Length Targets

SL:OH2 and **SL:OH3** duplexes were irradiated with 365 nm light and UV-Vis spectra recorded at set time intervals, Figure 5.8.

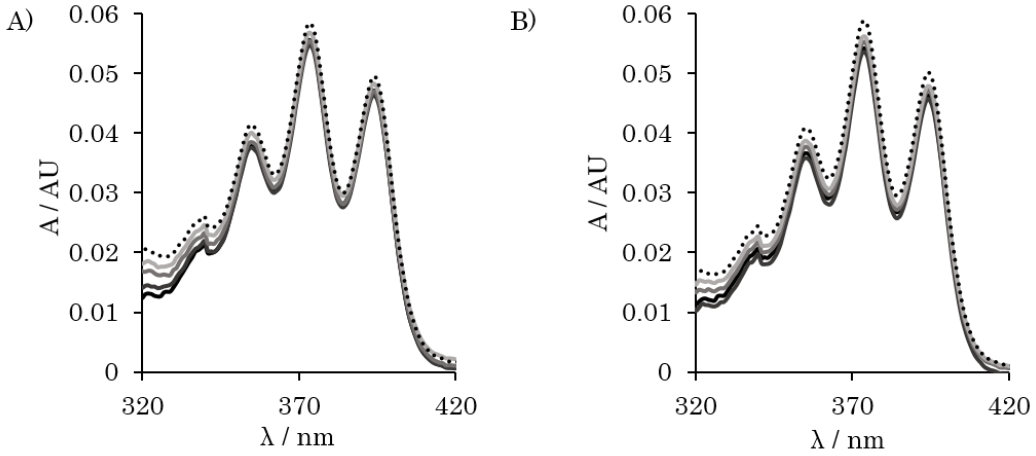


Figure 5.8 UV-Vis spectra tracking photodimer formation of 5 µM A) **SL:OH2** duplex and B) **SL:OH3** in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Sample irradiated with 365 nm light after 10 minutes degassing with argon. Scans recorded every 10 minutes for 30 minutes (black to grey) followed by a further 60 minute irradiation (dotted).

However, once again it was found that, despite the reduced length of the target strands, no reduction in the anthracene band was observed, indicating a lack of photodimerisation under these conditions. Subsequent analysis of the samples by native gel electrophoresis confirmed this, Figure 5.9. Upon irradiation of **SL** alone, a slight shift in the mobility of the band in the gel is observed, as **SL-D** is formed (Figure 5.9 Lanes 1 and 2). Upon addition of target sequence (**OH1**, **OH2** and **OH3**), a duplex is formed which shows a slower mobility through the gel due its increased molecular mass. However, upon photoirradiation, there is no change in the intensity of the band, and no development of bands corresponding to those of free **OHX** sequences or **SL-D**. This is further confirmation that the photodimer species cannot form once the stem loop is duplexed with a target strand.

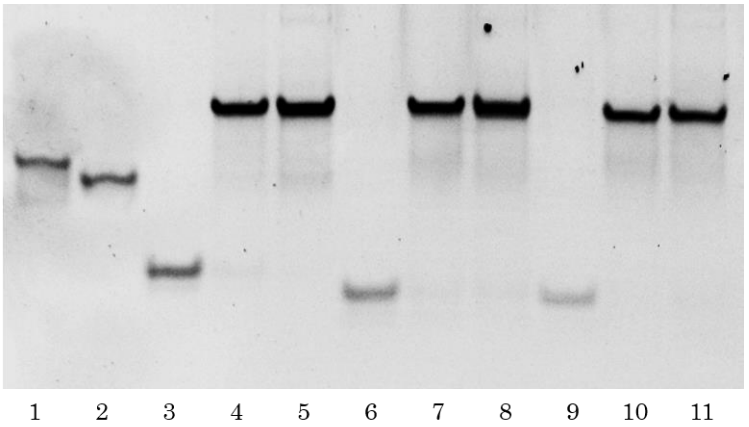


Figure 5.9 Native polyacrylamide gel of **SL** and target strands performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA. Lane: 1 **SL**, 2 **SL-D**, 3 **OH1**, 4 **SL:OH1**, 5 **SL:OH1** irradiated, 6 **OH2**, 7 **SL**:**OH2**, 8 **SL**:**OH2** irradiated, 9 **OH3**, 10 **SL**:**OH3**, 11 **SL**:**OH3** irradiated. Irradiated samples are taken after final UV-Vis spectra recorded (see Figures 5.7 and 5.8). Visualised with SYBR Gold® nucleic acid stain.

5.3.6 Binding of Overhang Strands to Locked Stem Loop

One question that remained was whether, if the stem loop (**SL**) was photodimerised in the absence of a complementary strand, would the target be able to bind the photoproduct (**SL-D**). In previous studies it was found that a shorter length target could still bind, despite the system being folded and locked in the stem loop conformation (Figure 5.4C).[30] In order to assess this, samples of **SL-D** in the presence of complementary **OHX** sequences, were analysed by native gel electrophoresis, Figure 5.10.

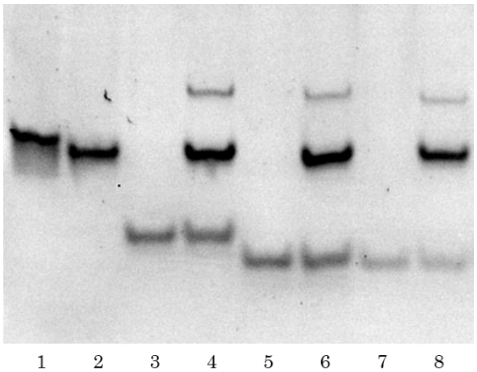


Figure 5.10 Native polyacrylamide gel of **SL-D** and target strand strands performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA at 293 K. Lane: 1 **SL**, 2 **SL-D**, 3 **OH1**, 4 **SL-D:OH1**, 5 **OH2**, 6 **SL-D:OH2**, 7 **OH3**, 8 **SL-D:OH3**. Visualised with SYBR Gold® nucleic acid stain.

Upon addition of the complementary **OHX** sequences, a new band of higher molecular weight was present, which corresponds to the **OHX** sequences bound to **SL-D**. However, the intensity of these bands were weak, and there was little change in the intensity of the **SL-D** and **OHX** bands in the gel. This suggests that, although there is evidence of binding, only a small proportion was now bound, and far less than previously observed with the earlier system containing no overlap with the stem, Figure 5.4C.[30]

Although this work showed some improvements to the previous systems, it was clear that, without developing a more complex system, for example through incorporating mismatches within sequences or competing strands, the intended photorelease mechanism was not going to work and a new design was needed.

5.4 Toe-Hold System

5.4.1 Current Uses of Toe-Hold Systems

In recent years, complex systems of oligonucleotides have been developed, that allow for the displacement of complementary strands by the addition of a competing strand (Figure 5.11). The displacement relies upon a section of a template strand (shown in black) being single stranded, to allow the competing strand (shown in blue) to be able to get a "toe-hold" (Figure 5.11, Step 2). The competing toe-hold strand can then "zip up" with the target strand, displacing the original duplexed strand (shown in red).

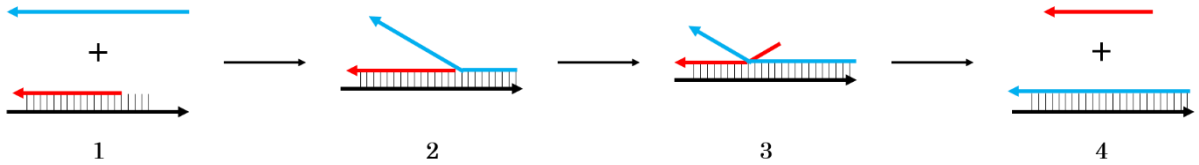


Figure 5.11 Schematic of the toe-hold system. 1) A duplex of DNA is taken which contains a region of single stranded DNA on the template strand (black). 2) Upon addition of a strand which has a region complementary to the single stranded section (blue), the new strand binds to the template. 3) If this new sequence displays a higher binding affinity for the template strand than that which is currently duplexed, the original strand is displaced (red). 4) The new duplex is formed between the template strand and the toe-hold strand.

In the last few years this approach has been used for several applications in both medical assays and DNA machines.[32,33] Of particular interest to the projects presented in this thesis is the use of toe-hold strands to deduce the genotype of a subject[32] and detect base mutations within DNA.[33] Within the genotyping assay, a simple system using a FRET pair is utilised, Figure 5.12. The PCR product, containing a toe-hold region and a fluorophore, is mixed with a strand complementary to the target and labelled with the FRET partner. If the target is present, this strand can bind in the toe-hold region, displacing the other strand and quenching the fluorescence. The authors also trialled the use of this assay to detect SNP sites, however, they concluded that the efficiency was not sufficient and required additional investigations.

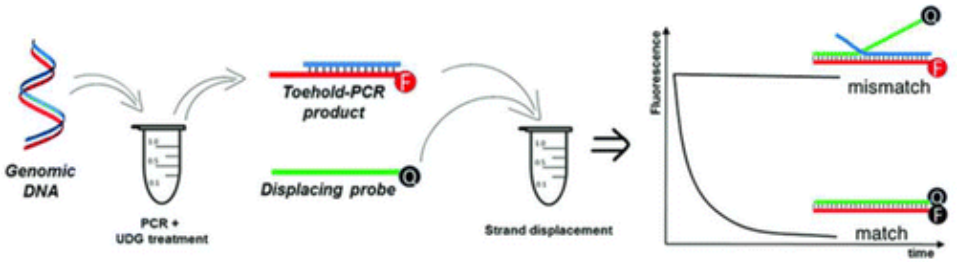


Figure 5.12 Genotyping assay developed using the toe-hold method. PCR is performed on the DNA sample, the primer used has a fluorophore attached to it. A competing toe-hold strand is added to the reaction which is tagged with a quencher molecule. If the fully complementary sequence from the PCR is present, the template strand is displaced and the quencher is brought into close proximity of the fluorophore, reducing the fluorescence signal. If there is a mismatch in the sequence, displacement does not take place and fluorescence is observed. Adapted with permission from Khodakov *et al.*, *J. Am. Chem. Soc.*, 2013, **135**, 5612-9. Copyright 2013 American Chemical Society.

Work published by Jiang and co-workers to detect base modifications utilises a much more complex system in which the toe-hold method is used in combination with DNA splicing. Upon fully complementary DNA binding to the toe-hold region, the DNA is extended to incorporate a splicing site, which is subsequently "cut" to release a strand that can fold into a G-quadruplex. A fluorescent dye that shows selectivity for G-quadruplexes is then added to the system and fluorescence is observed.[33]

The toe-hold system has also been utilised to make DNA machine components. Winfree and colleagues developed a variety of complex systems that work as logic gates and other circuit components, all relying on the recognition and strand displacement of oligonucleotides.[34] Deng *et al.* also reported on the development of AND and OR logic gates which rely upon the presence of the toe-hold strand, as well as a metal which enhances the toe-hold strand binding through dative bond formation.[35]

5.4.2 Design of a Toe-Hold System for Photo-Triggered Release

It was proposed that a toe-hold system could be exploited to achieve photo-triggered release of an oligonucleotide. Rather than utilising strands with different affinities for the release, if strands were designed with similar affinities, an equilibrium between the two duplexes would be expected to be achieved (if all were present in solution at the same concentration). Tagging both the template strand

and one of the complementary strands with anthracene moieties could then allow for the photodimerisation of the anthracene as a crosslink, locking the duplex together and preventing the other strand from binding (an effect that has been demonstrated in work previously performed in the group[30]). This would result in an effective "release" of the second strand, Figure 5.13.

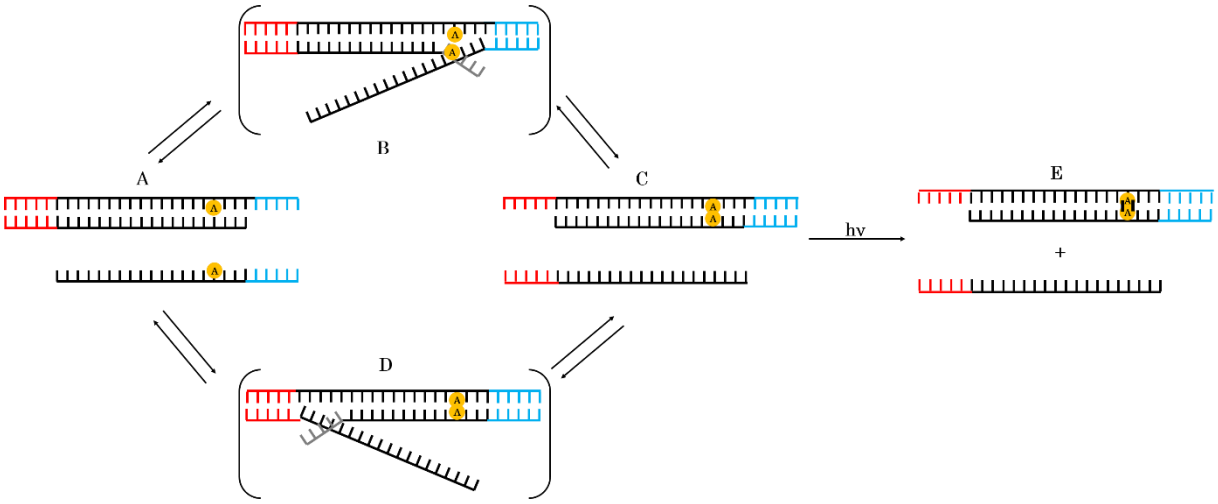


Figure 5.13 A) A duplex is formed between the template strand and a complementary sequence. B) A competing strand is able to get a "toe-hold" (through the bases shown in blue) on the target strand and displaces the original strand. C) The new duplex is formed. D) The new duplex can readily revert to the original duplex if the first complementary strand gets a "toe-hold" (through the bases shown in red). E) Irradiation of the system will photodimerise anthracene moieties in close proximity, locking the duplex of the two anthracene containing strands and releasing the unmodified strand.

Sequences for oligonucleotides used within this study were derived from work by Winfree and Zhang;[36] a schematic of these is shown in Figure 5.14. The toe-hold systems are made of three strands, a target (T), a photo-trigger (PT) and a complementary (C). The target and photo-trigger strands contain anthracene moieties placed opposite one another in the duplex; these were expected to photodimerise upon irradiation and "lock" the duplex, preventing the complementary strand binding. Four systems have been developed, **T1:PT1:C1-S**, **T1:PT2:C2**, **T2:PT3:C1-S** and **T2:PT4:C2**, the **C1-S** strand can be interchanged

with the **C1-M** and **C1-L** sequences. These systems were expected to produce the equilibrium mixture depicted in Figure 5.13.

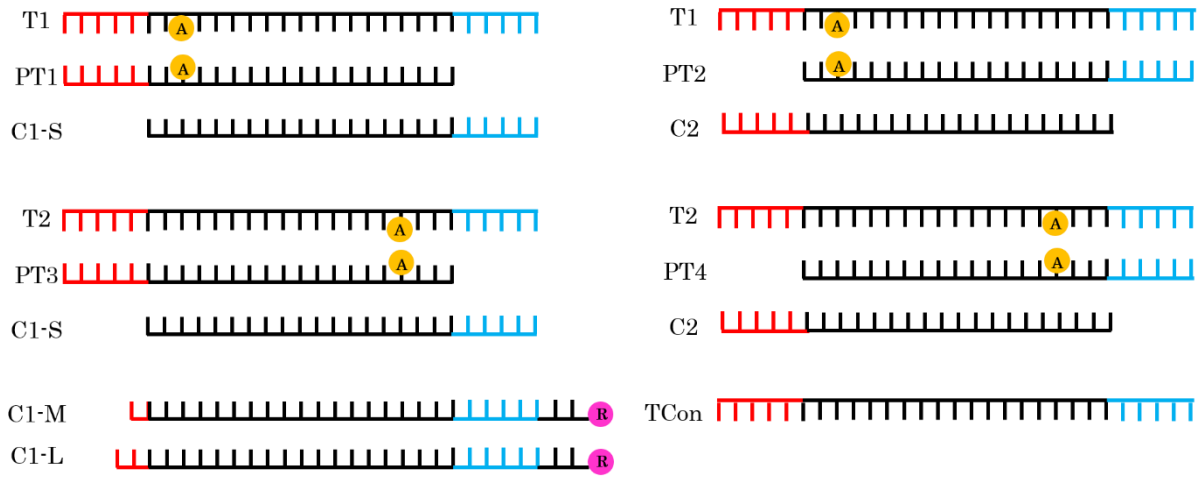


Figure 5.14 Sequences developed for photo-triggered toe-hold release. Photo-trigger (PT) and Template (T) strands contain anthracene (A) opposite one another in the duplex to allow for the photodimerisation and "locking" of the duplex. Complementary (C) strands contain a region of bases that are identical to those in the PT strands (shown in black) and compete for binding to the T strand. Each system will be comprised of three strands in which the PT and C strands form duplexes with the T stand (**T1:PT1:C1**, **T1:PT2:C2**, **T2:PT3:C1** and **T2:PT4:C2**). The **C1** strand is made with various lengths (**C1-S**, **C1-M** and **C1-L**) to allow for the analysis of the change in complementary bases. In addition to this, the longer strands (**C1-M** and **C1-L**) are labelled with a fluorescent dye, TAMRA (R).

The parent duplexes (**T1:PT1**, **T1:PT2**, **T2:PT3**, **T2:PT4**, **T1:C1-S**, **T1:C2**, **T2:C1-S** and **T2:C2**) all had predicted T_m values of 63 °C, and all contained toe-hold regions (shown in blue or red in Figure 5.14) of 6 bases in length. Anthracene moieties were placed towards the ends of the complementary section (shown in black in Figure 5.14) since it was considered that a run of unmodified bases would allow for displacement by the competing strand. However, on the other hand substitution on the end base could result in a disruption in the duplex folding, or the photodimerisation being affected by the ends of the DNA "fraying". They were therefore moved to a site a few bases from the end of the sequence. The incorporation was made at an A-T base pair site, as this was expected to give the least change to the T_m.

In order to assess the impact that the length of the competing strand had on the release, **C1-S** was extended by 1 (**C1-M**) and 2 (**C1-L**) base pairs. In addition to this, the two longer strands were labelled with a fluorescent dye (TAMRA), to allow for easier differentiation in gels. The dye was separated from the complementary region of the sequence by two bases, to prevent any interference with binding. The sequences of all the strands are shown in Table 5.3.

Table 5.3 Oligonucleotide sequences used for toe-hold photo-triggered release. Anthracene units (shown as X) were used with D stereochemistry and a 6 carbon linker as this has been shown previously to form photodimer products in similar systems. Extra bases on the longer **C1** strands (**C1-M** and **C1-L**) are shown underlined.

Name	Sequence
TCon	5′- TGG AGA CGT AGG GTA TTG AAT GAG GG -3′
C1-S	3′- GCA TCC CAT AAC TTA CTC CC -5′
C1-M	3′- <u>T</u> GCA TCC CAT AAC TTA CTC CCT T(TAMRA) -5′
C1-L	3′- <u>CT</u> GCA TCC CAT AAC TTA CTC CCT T(TAMRA) -5′
C2	3′- ACC TCT GCA TCC CAT AAC TT -5′
T1	5′- TGG AGA CG**X** AGG GTA TTG AAT GAG GG -3′
T2	5′- TGG AGA CGT AGG GTA T**X**G AAT GAG GG -3′
PT1	3′- ACC TCT GC**X** TCC CAT AAC TT -5′
PT2	3′- GC**X** TCC CAT AAC TTA CTC CC -5′
PT3	3′- ACC TCT GCA TCC CAT A**X**C TT -5′
PT4	3′- GCA TCC CAT A**X**C TTA CTC CC -5′

5.4.3 Thermal Melting Studies

All possible duplex combinations were analysed by variable temperature UV-Vis spectroscopy to obtain T_m values, as a measure of the duplex stability, Table 5.4.

Table 5.4 T_m data for all duplex combinations. Recorded at 5 μM duplex DNA in 100 mM NaCl, 10 mM sodium phosphate buffer pH 7. All values are average of at least 3 repeat runs and are reported to the closest 0.5 °C due to the limits of instrument error.

Duplex	T_m (°C)	Duplex	T_m (°C)
TCon:C1-S	64.0	**TCon:C2**	62.5
T1:PT1	63.0	**T1:PT2**	66.0
T2:PT3	65.5	**T2:PT4**	65.0
T1:C1-S	63.0	**T1:C2**	59.0
T2:C1-S	61.5	**T2:C2**	61.5
T1:C1-M	64.0	**T2:C1-M**	63.0
T1:C1-L	64.5	**T2:C1-L**	64.5

The control systems, **TCon:C1** and **TCon:C2,** have T_m values close to the predicted T_m of 63 °C (predictions are performed as described in Chapter 2.7.3). However, substitution of the anthracene moieties was found to make varying contributions to the T_m. In all cases, the duplexes containing two anthracene moieties (duplexes formed between T and PT strands) display higher T_m values than those with one modification (duplexes formed between T and C strands). In particular **T1:PT2** and **T2:PT3** displayed a significant increased stability. It is likely that these sequences allow for the anthracene moieties to stack, resulting in a relative increase in T_m. This could be aided by the incorporation of anthracene close to the end of the duplex

in these systems. Due to the space required by stacking, it would be expected that bases surrounding the substitution site would be disrupted. However, being closer to the end of the duplex would allow for greater flexibility and therefore a relative increase in stability would be observed.

In contrast, a marked destabilisation is seen for duplex **T1:C2**. A possible explanation for this, is that the anthracene moiety is incorporated into a C-G rich region. The anthracene is expected to intercalate in the DNA duplex and could disrupt the bases surrounding the substitution site. As the C-G base pairing offers the greatest affinity (containing three hydrogen bonds compared to two in A-T base pairs), a disruption to these would have a greater destabilisation effect on the duplex. Within the **T1:PT2** system, this effect could be countered by the stacking of the anthracene moieties, as hypothesised above.

A requirement for the toe-hold design to work is that the duplexes within the system have similar affinities, so that the strands can displace one another. The **T1:PT2:C2** system displays the highest variation in T_m at 7 °C. The only system in which both duplexes have similar stabilities is within the **T1:PT1:C1** system. Upon increasing the length of **C1-S** in sequences **C1-M** and **C1-L**, an increase in the T_m values was observed. In the case of the **T1:PT1:C1-M** and **T1:PT1:C1-L** systems, the T_m values with the **CX** strands exceeded that of the duplex between **T1:PT1**. However it was considered that this should allow for more of the complementary

strand to bind, and therefore leaves the potential for a higher amount of photo-triggered release. Within the **T2:PT3:C1-M** and **T2:PT3:C1-L** systems, the T_m values with the **CX** strands were higher compared to that recorded for **T2:C1-S**, however, the melting temperatures were still below that of **T2:PT3**.

5.4.4 Photodimerisation Studies

Prior to analysing whether the systems could be used for photo-triggered release, it needed to be confirmed that the anthracene containing duplexes could form crosslinked photodimers. All four combinations, **T1:PT1**, **T1:PT2**, **T2:PT3** and **T2:PT4**, were irradiated with 365 nm light and changes to the anthracene signal monitored using UV-Vis spectroscopy, Figure 5.15.

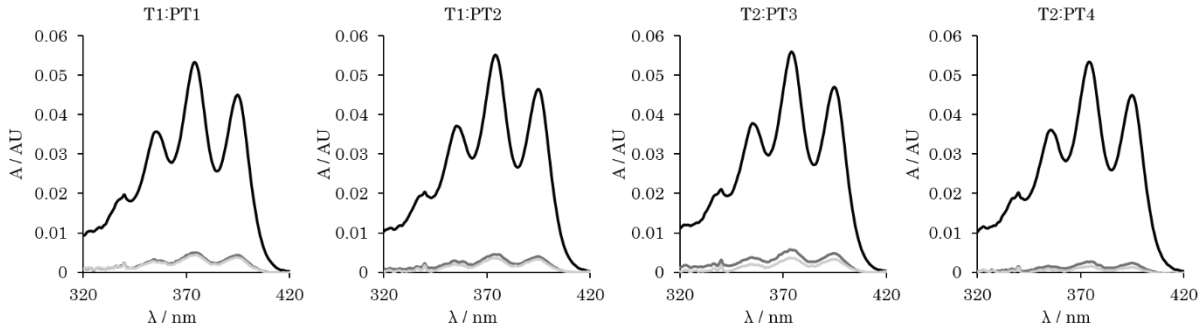


Figure 5.15 UV-Vis spectra of anthracene containing duplexes. All performed at 5 μM duplex in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Samples were recorded at 0 minutes (black) following degassing with argon for 10 minutes. They were then irradiated with 365 nm light for 10 (dark grey) and 20 (light grey) minutes.

All combinations showed a significant decrease within 10 minutes of irradiation, with little change thereafter. This is a comparable rate of reaction as that seen with similar systems in the group,[29,30] and is also comparative to the work by Asanuma and co-workers utilising a *p*-stilbazole photocrosslink.[28] As expected,

control systems with **T1** and **T2** duplexed with **C1-S** and **C2**, as well as the **PTX** strands duplexed with **TCon**, showed no change in absorbance upon irradiation (see appendix 8.3).

The photodimerised samples were subsequently analysed using a denaturing gel (this type of gel disrupts interactions between DNA strands and allows for covalently linked species to be analysed, Chapter 2.10.1), to confirm the presence of photodimer species, Figure 5.16. All samples contained a band consistent with a higher molecular weight than the ssDNA controls, indicating that photodimerisation had been successful.

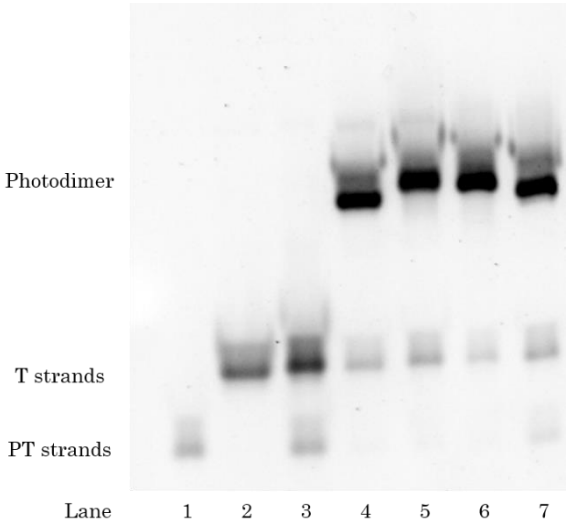


Figure 5.16 Denaturing polyacrylamide gel of toe-hold duplex dimers performed in 1 × tris borate EDTA running buffer with 500 nM DNA in 50% formamide. Lane: 1 **PT2**, 2 **T1**, 3 **PT2:T1**, lanes 4-7 duplexes were degassed and irradiated for 20 minutes lane: 4 **T1:PT2**, 5 **T1:PT1**, 6 **T2:PT4**, 7 **T2:PT3**. Visualised using SYBR Gold® nucleic acid stain.

UV-Vis analysis performed on the photodimer species, showed that no T_m value could be obtained in all cases, it being higher than the bounds of the experiment (85 °C). The stabilisation afforded by the photodimerisation of anthracene to form

a crosslink has been explored previously, and this result correlates well with the trends observed.[30]

5.4.5 Photodimer Binding Studies

A prerequisite of this toe-hold system is that anthracene photodimerisation blocks the binding of the **CX** strands, as it has previously been shown to do.[30] Native gel electrophoresis was performed on the crosslinked photodimer species (formed in Figure 5.15) and the complementary strand (**CX**), which was to be used in the toe-hold system (Figure 5.14), to ensure that photodimerisation inhibits binding to the template strand, Figures 5.17 (**C1-S** and **C2**) and 5.18 (**C1-M** and **C1-L**). It would be expected that samples of photodimer would have a band at high molecular weight from the photodimer linked duplex and **CX** strands remain unbound and be seen as a low molecular weight band.

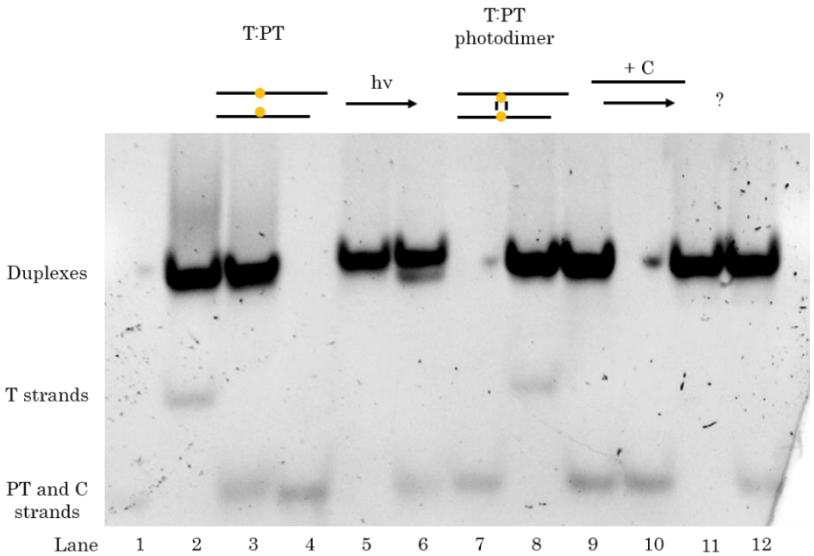


Figure 5.17 Top: Schematic representation of photodimer formation with subsequent binding to competing complementary strand assessed. Bottom: Native polyacrylamide gel of toe-hold control strands performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA, 293 K. Lane: 1 **C1-S**, 2 **T1:PT1 photodimer**, 3 **T1:PT1 photodimer** with **C1-S**, 4 **C1-S**, 5 **T2:PT3 photodimer**, 6 **T2:PT3 photodimer** with **C1-S**, 7 **C2**, 8 **T1:PT2 photodimer**, 9 **T1:PT2 photodimer** with **C2**, 10 **C2**, 11 **T2:PT4 photodimer**, 12 **T2:PT4 photodimer** with **C2.** Visualised using SYBR Gold® nucleic acid stain.

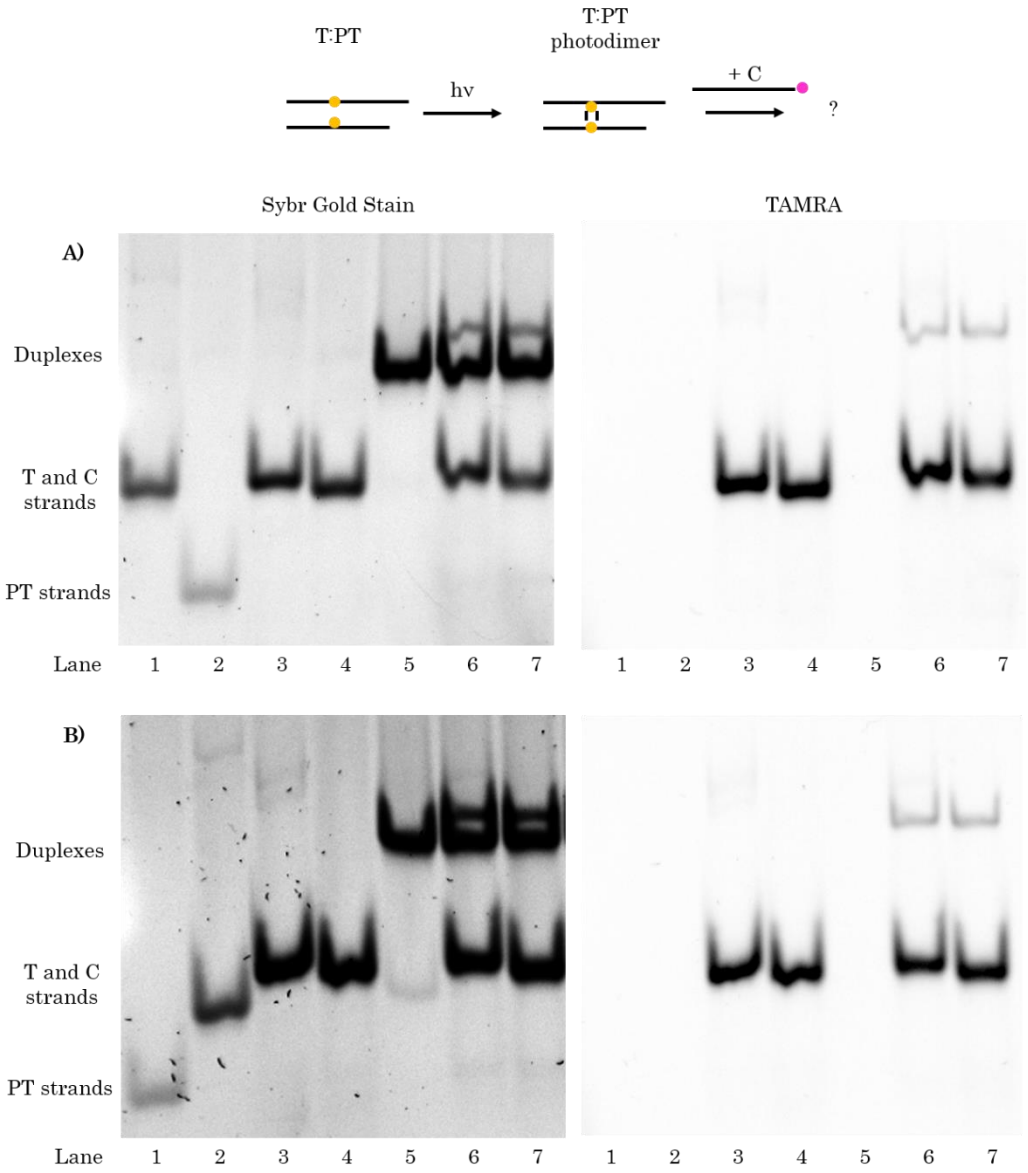


Figure 5.18 Top: Schematic representation of photodimer formation with subsequent binding to competing complementary strand assessed. Bottom: Native polyacrylamide gel of new complementary strands binding to photodimer performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA, 293 K. A) Lane: 1 **T1**, 2 **PT1**, 3 **C1-L**, 4 **C1-M**, 5-7 with **T1:PT1 photodimer** alone and with **C1-L** and **C1-M** respectively. B) Lane: 1 **PT3**, 2 **T2**, 3 **C1-L**, 4 **C1-M**, 5-7 with **T2:PT3 photodimer** alone and with **C1-L** and **C1-M** respectively. Visualised using SYBR Gold® nucleic acid stain (left) and TAMRA fluorescence (right).

In all cases no significantly higher molecular weight species were found which would be seen at a higher position in the gel, above the duplexes. This confirms that no three stranded systems could be detected. In lane 1 of Figure 5.17, it appears that the control **C1-S** has not been loaded or prepared correctly. An

"identical" sample is also in lane 4 which does display the expected band. When using the TAMRA tagged sequences, Figure 5.18, a new duplex can be seen which is also visible in the TAMRA image of the gel. This indicates that some duplex is forming with both the **C1-M** and **C1-L** strands. This is thought to be due to incomplete photodimerisation of the **TX** and **PTX** systems, allowing for the undimerised T strand to form duplexes with the **CX** strands. This is supported by the weak band corresponding to the **TX** strand, in the **TX:PTX** photodimer samples (Figure 5.18, Lane 5), which disappears on addition of the **CX** strands (Lanes 6 and 7). It is therefore concluded, that the photodimerisation has successfully blocked duplex formation between the **TX** and **CX** strands under these conditions.

5.4.6 Binding Studies of Toe-Hold Systems

Initial binding studies were performed using the control sequences, Figure 5.19, to assess if both competing strands (**CX**) could bind to the target (**T**) strand. If this is the case, when all three strands are mixed, a duplex band containing duplexes of both **TCon:C1-S** and **TCon:C2** is expected along with a band at lower molecular weight from the excess **C1-S** and **C2** strands. There should be no excess **TCon** and therefore no band for this when all three strands are present. The same is true for the strands containing anthracene, although the trends will be more easily distinguished utilising the anthracene fluorescence to assign which bands contain anthracene tagged strands. Likewise, the TAMRA tagged systems can also be identified by using the TAMRA fluorescence.

It was observed that both **TCon:C1-S** and **TCon:C2** (Figure 5.19, Lanes 4 and 5, respectively) formed duplexes. The control toe-hold system containing equimolar amounts of all three strands, **TCon:C1-S:C2**, contained two bands, one corresponding to duplex species and one to single stranded DNA (this band is expected to contain both single stranded **C1-S** and **C2**). It should be noted that the SYBR Gold® nucleic acid stain used, exhibits weaker bind to ssDNA than dsDNA, as well as a reduced fluorescence output with ssDNA.[37] The intensity between the bands can therefore not be taken as an indication of quantities present. It can be seen that the ssDNA bands are very weak and it is therefore hard to distinguish them. It was decided that for future gels, analysing strands which do not contain a TAMRA label, more material would be loaded to allow for the ssDNA bands to be more clearly seen. In addition to this, anthracene fluorescence was to be utilised to image the gels, allowing for the identification of any ssDNA. However, this approach was found to result in some overloading of the dsDNA band.

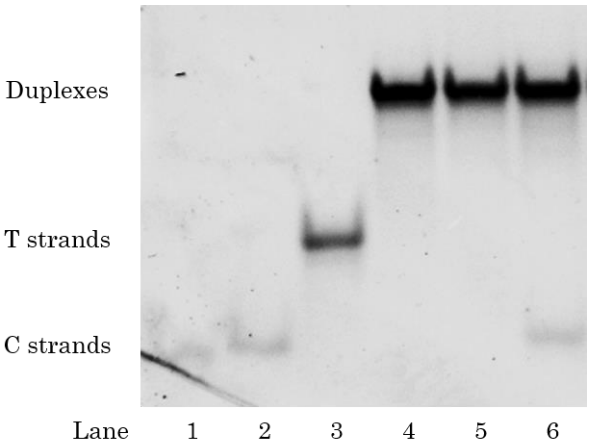


Figure 5.19 Native polyacrylamide gel of toe-hold control strands performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA at 293 K. Lane: 1 **C1-S**, 2 **C2**, 3 **TCon**, 4 **TCon:C1-S**, 5 **TCon:C2**, 6 **TCon:C1-S:C2**. Visualised using SYBR Gold® nucleic acid stain.

Analysis of all toe-hold combinations (Figure 5.14) was performed using native gel electrophoresis, Figures 5.20 and 5.21. It was expected that all **TX:CX** and **TX:PTX** combinations would form duplexes and be identified as a higher molecular weight band. In systems containing all three strands (**TX:CX:PTX**) an equilibrium of duplexes would be expected and as such a mixture of **PTX** and **CX** would be in single stranded form and observed as a lower molecular weight species in the gel. The bands containing anthracene (both duplexes and single stranded) will be identified using the anthracene fluorescence. It would be expected that in the **TX:CX:PTX** mixture, anthracene will be present in the duplex and single stranded bands.

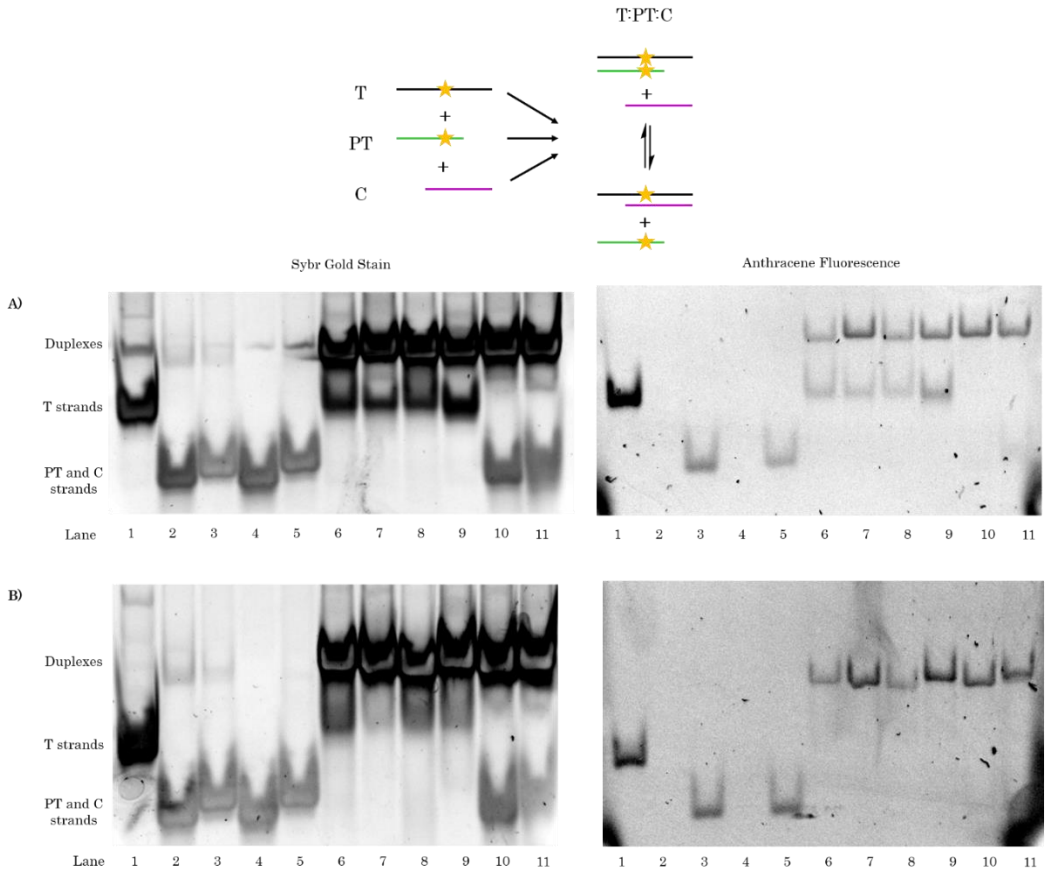


Figure 5.20 Top: Schematic representation of toe-hold system. Bottom: Native polyacrylamide gel of toe-hold systems performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 10 μM DNA at 293 K. A) Lane: 1 **T1**, 2 **C2**, 3 **PT2**, 4 **C1-S**, 5 **PT1**, 6 **T1**:**C2**, 7 **T1**:**PT2**, 8 **T1**:**C1-S**, 9 **T1**:**PT1**, 10 **T1**:**C2**:**PT1**, 11 **T1**:**C1-S**:**PT1**. B) Lane: 1 **T2**, 2 **C2**, 3 **PT4**, 4 **C1-S**, 5 **PT3**, 6 **T2**:**C2**, 7 **T2**:**PT4**, 8 **T2**:**C1-S**, 9 **T2**:**PT3**, 10 **T2**:**C2**:**PT4**, 11 **T2**:**C1-S**:**PT3**. Visualised using SYBR Gold® nucleic acid stain (left) and anthracene fluorescence (right).

Using the anthracene fluorescence (Figure 5.20, right), the strands containing anthracene are definitively identified. It can be seen that a significant amount of **T1** remains unbound, when expected to be in a 1:1 duplex with **PTX** and **CX** strands (Figure 5.20 Lanes 6-9). This will obviously hinder the toe-hold ability, as an excess of this strand will allow for additional binding to the **PTX** and **CX** strands.

Within the SYBR Gold® nucleic acid stain images of the gels, it can be seen that when all three strands are present (Figure 5.20, Gel A and B, lanes 10 and 11), there is dsDNA and ssDNA within the samples. However, correlating this with the anthracene fluorescence image shows anthracene to mainly be present in the dsDNA band, indicating that all of the **PTX** strands are bound and that the ssDNA must be due to the **CX** strands. This was not the planned scenario, since the **CX** strand had been released before any irradiation. It is thought that this is due to the higher T_m values associated with the **TX:PTX** duplexes, compared to those formed with **C1-S** and **C2**. However, the **T1:PT1:C1** system displayed no differentiation between the two strands *via* T_m and is accompanied by a very weak band corresponding to **PT1** in the anthracene fluorescence image of the gel (Figure 5.20, Gel A, Lane 11). This was therefore deemed the most promising system.

Using the TAMRA fluorescence (Figure 5.21, right) also allows for definitive identification of the **CX** strands. However, due to the increased length of **C1-M** and **C1-L**, their mobility through the gel has also been slowed, allowing for more

conclusive assignment of these bands, as they no longer overlap with those from the **PTX** strands. This trend is also seen in the duplexes formed with them, resulting in two bands being seen for the toe-hold system. Lanes 8 and 9 (Figure 5.21) show that duplexes are formed with both stands, and that some of each strand (**PTX** and **CX**) are present in their single stranded forms. The TAMRA image identifies that the slowest moving duplexes, are those which contains the **C1-M** and **C1-L** strands.

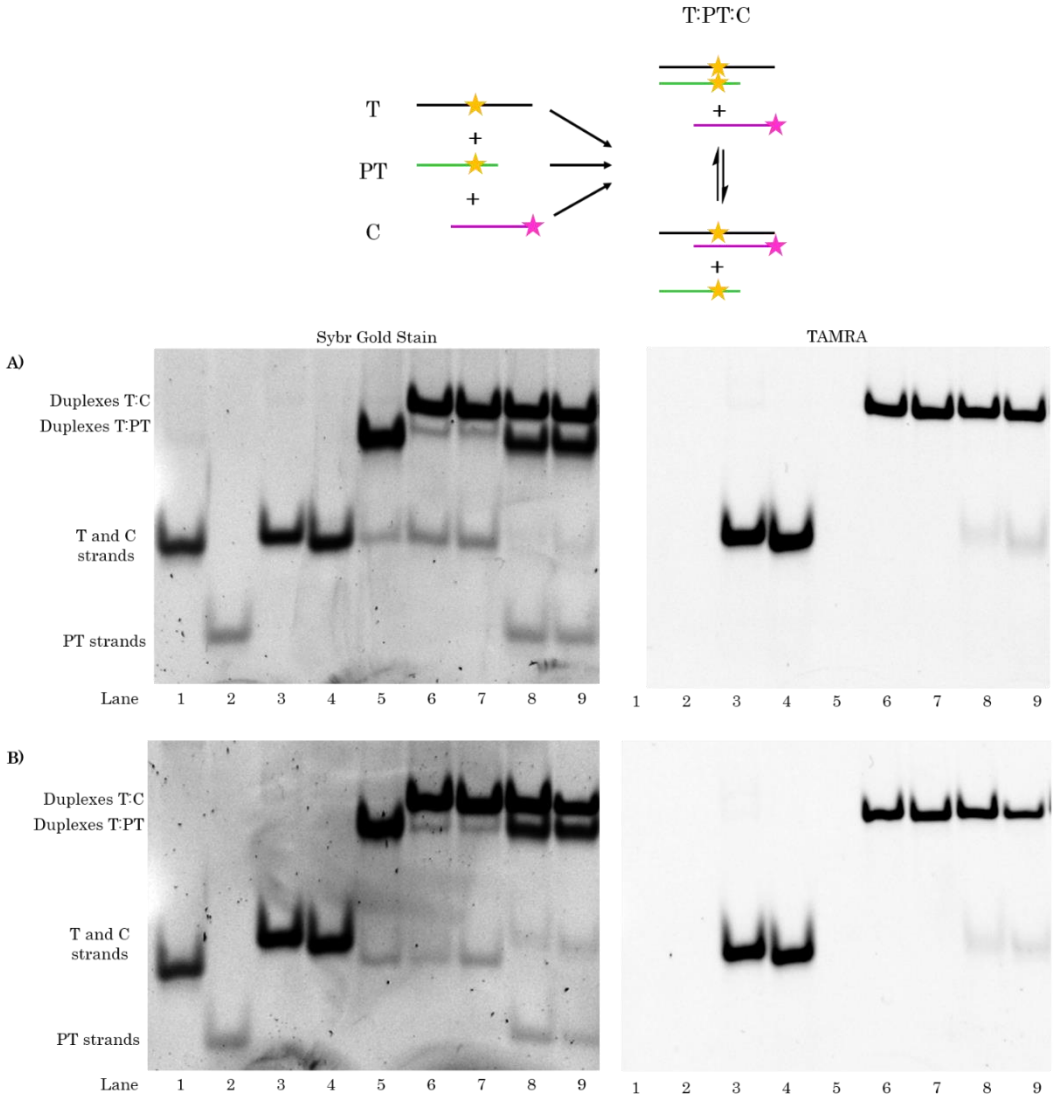


Figure 5.21 Top: Schematic representation of toe-hold system. Bottom: Native polyacrylamide gel of toe-hold system performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA at 293 K. A) Lane: 1 **T1**, 2 **PT1**, 3 **C1-L**, 4 **C1-M**, 5 **T1:PT1**, 6 **T1:C1-L**, 7 **T1:C1-M**, 8 **T1:PT1:C1-L**, 9 **T1:PT1:C1-M**. B) Lane: 1 **T2**, 2 **PT3**, 3 **C1-L**, 4 **C1-M**, 5 **T2:PT3**, 6 **T2:C1-L**, 7 **T2:C1-M**, 8 **T2:PT3:C1-L**, 9 **T2:PT3:C1-M**. Visualised using SYBR Gold® nucleic acid stain (left) and TAMRA fluorescence (right).

The system now displays the desired balance between binding to the **PTX** and **CX** strands, which has the potential to be manipulated using the photodimerisation of the **TX** and **PTX** strands. Therefore, only the systems using **C1-M** and **C1-L** were used for further studies.

5.4.7 Release by Addition of Competing Strand

Aside from the release by photodimerisation, as is being studied herein, the release by addition of the **PTX** strand itself needed to be addressed. Due to the two duplexes having similar T_m values and both binding to the template strand, an equilibrium is formed when all three are in solution and hence some of the **CX** strands are displaced by the **PTX** strands. In order to evaluate this, native gel electrophoresis studies were performed in the presence of increasing concentrations of **PTX** strands, Figure 5.22. It would be expected that as the amount of **PTX** strand is increased, the amount of single stranded **CX** strand would increase. This would be seen as a lower molecular weight band and will increase in intensity as the concentration of **PTX** increases. This will be identified using the TAMRA fluorescence associated with the **CX** strand. In addition to this, the amount of **TX:PTX** duplex will also increase as the **PTX** concentration is increased. This will be observed as a duplex band of slightly lower molecular weight (and therefore lower on the gel) than the **TX:CX** duplex.

The release of the **CX** strand was shown to be dependent on the concentration of **PTX** strand and can be clearly seen in the TAMRA fluorescence image of the gels (Figure 5.22, right), as the intensity of the single stranded **CX** oligonucleotides increases across the gels (lanes 5-8 and 9-12). The **T2:PT3** system is shown to be more susceptible to release in this manner with the single stranded **C1-M** and **C1-L** being visible at lower concentrations of **PT3** (Figure 5.22, gel B, lanes 5-8 and 9-12). This is due to the **T2:PT3** duplex having a higher T_m than either the **T2:C1-L** and **T2:C1-M** duplexes.

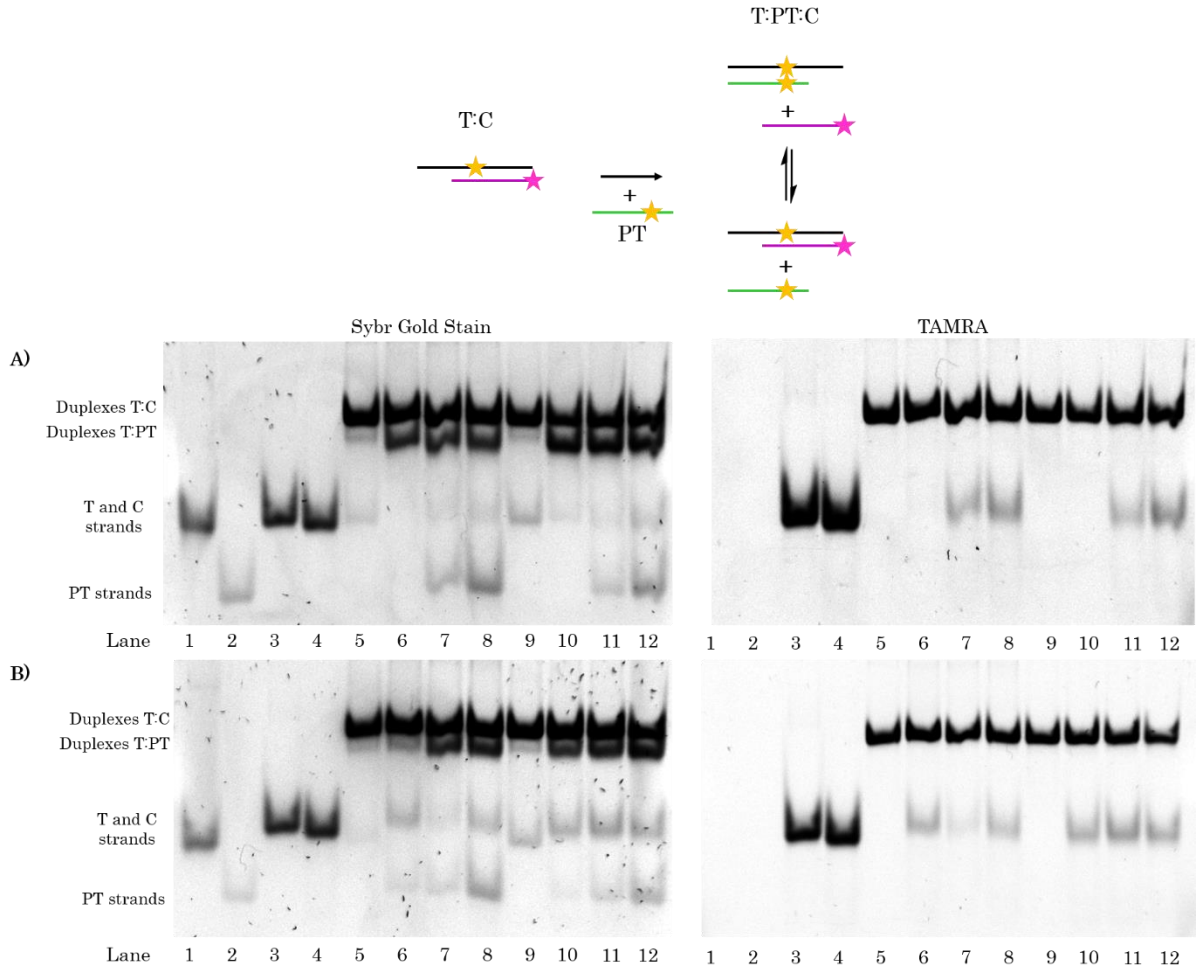


Figure 5.22 Top: Schematic representation of toe-hold system with displacement by competing strand. Bottom: Native gel electrophoresis performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM at 293 K. DNA visualised using SYBR Gold® nucleic acid stain (left) and TAMRA fluorescence (right). Gel A lane: 1 **T1**, 2 **PT1**, 3 **C1-L**, 4 **C1-M**, 5-8 **T1:C1-L** with 0, 250, 500 and 1000 nM **PT1** respectively, 9-12 **T1:C1-M** with 0, 250, 500 and 1000 nM **PT1** respectively. Gel B lane: 1 **T2**, 2 **PT3**, 3 **C1-L**, 4 **C1-M**, 5-8 **T2:C1-L** with 0, 250, 500 and 1000 nM **PT3** respectively, 9-12 **T2:C1-M** with 0, 250, 500 and 1000 nM **PT3** respectively.

5.4.8 Photodimerisation with TAMRA Release Strands

These new systems, containing all three components, were subsequently exposed to irradiation and changes to the anthracene region of the UV-Vis spectrum tracked, Figure 5.23. Upon irradiation, all samples showed a decrease in absorbance between 320-420 nm, which could be attributed to anthracene photodimerisation. The reductions in signal are lower than would have been expected, with approximately 20% reduction in the cases with **C1-L** and 40% reduction in the samples with **C1-M**. If this results in a proportional amount of photodimer formation and therefore release of the complementary strand, then not even half of the complementary strands are being released upon irradiation. However, this is a marked improvement on previous efforts, which did not generate any release.

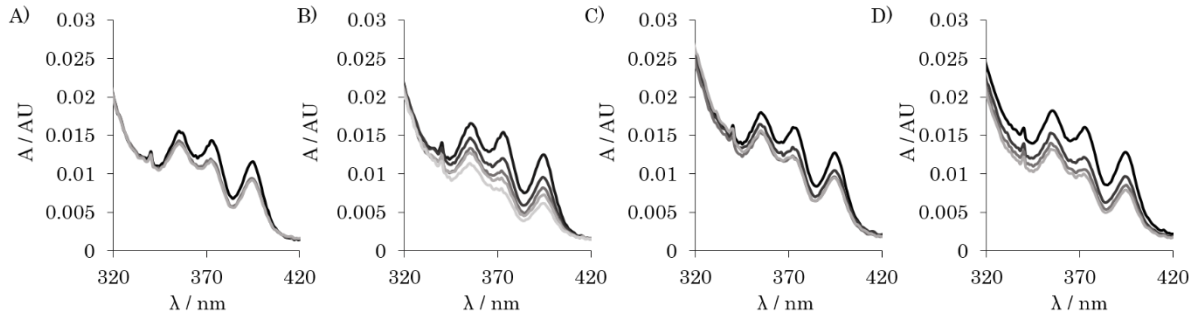


Figure 5.23 UV-Vis spectra of 1 μM of each oligonucleotide strand in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Sample irradiated with 365 nm light after 10 minutes degassing with argon at 293 K. Scans recorded every 30 minutes until no further change observed (black to grey). A) **T1:PT1:C1-L**, B) **T1:PT1:C1-M**, C) **T2:PT3:C1-L**, D) **T2:PT3:C1-M**.

The samples containing **C1-M** displayed a greater decrease, suggesting more photodimer species is formed in these cases than with **C1-L**. This would be expected due to the lower T_m of **C1-M** with the **TX** strands, due its shorter length. This would allow for a greater amount of **PTX:TX** duplex to form, and therefore a higher amount of photodimerisation.

5.4.9 Monitoring Photodimer Formation

Aliquots of sample, at the irradiation time steps used for the UV-Vis monitoring (Figure 5.23), were analysed using denaturing gel electrophoresis, to assess if photodimers were forming and to qualitatively analyse the change in amount over time, Figure 5.24. The photodimer would be shown as a higher molecular weight band, with any non-covalently linked species dissociating under the gel conditions and therefore being observed as low molecular weight, single stranded DNA.

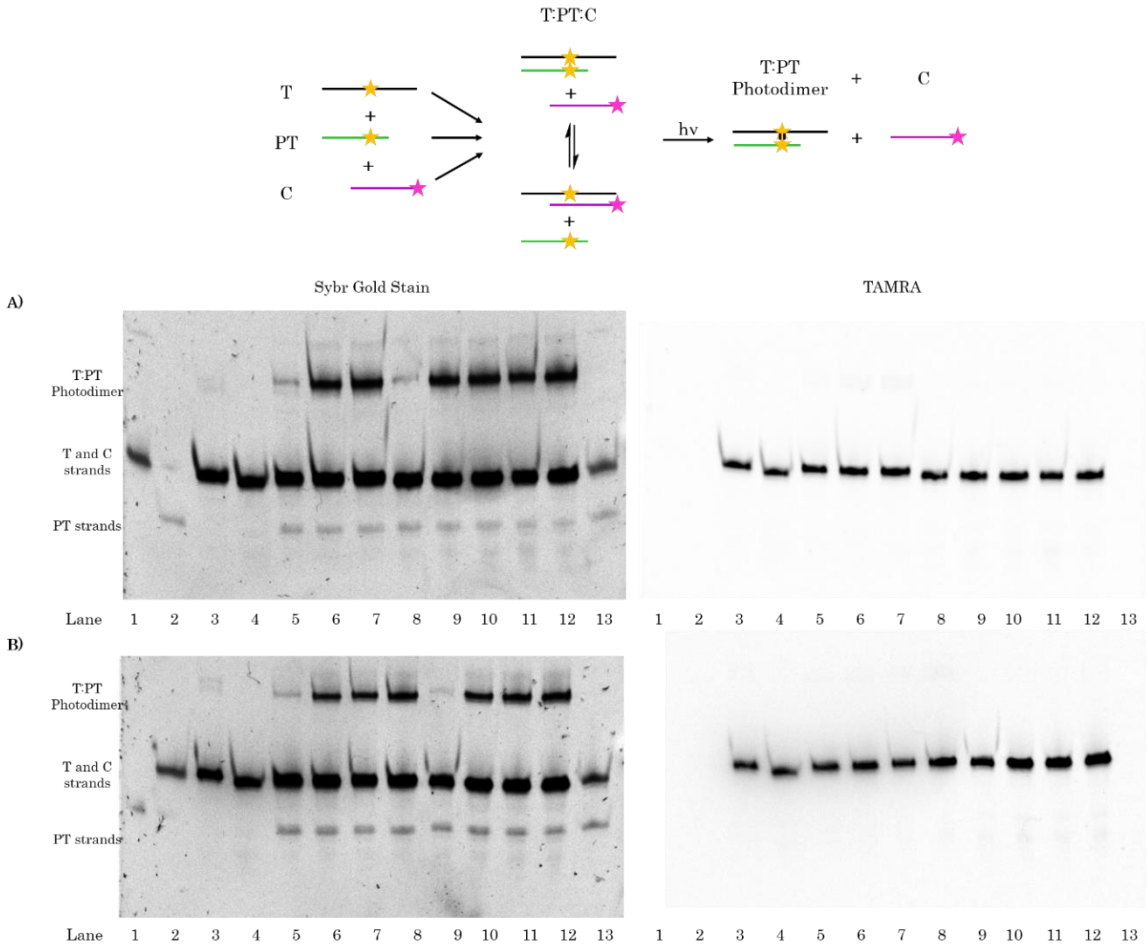


Figure 5.24 Top: Schematic representation of toe-hold system and expected photorelease. Bottom: Denaturing polyacrylamide gel of TAMRA release strand irradiated samples performed in 1 × tris borate EDTA running buffer with 500 nM DNA in 50% formamide at 293 K. Irradiation performed with 365 nm light after 10 minutes degassing with argon at 293 K. A) Lane: 1 **T1**, 2 **PT1**, 3 **C1-L**, 4 **C1-M**, 5 – 7 **T1:PT1:C1-L** at 0, 30 and 60 minutes irradiation respectively, 8-12 **T1:PT1:C1-M** at 0, 30, 60, 90 and 120 minutes irradiation respectively, 13 **T1:PT1** control. B) Lane: 1 **T2**, 2 **PT3**, 3 **C1-L**, 4 **C1-M**, 5 – 8 **T2:PT3:C1-L** at 0, 30, 60 and 90 minutes irradiation respectively, 9-12 **T1:PT1:C1-M** at 0, 30, 60 and 90 minutes irradiation respectively, 13 **T2:PT3** control. Visualised using SYBR Gold® nucleic acid stain (left) and TAMRA fluorescence (right).

Evidence of photodimer formation was observed within the first time frame of irradiation. The TAMRA image of the gels (Figure 5.24, right) confirms that the photodimer species does not contain **C1-L** or **C1-M** oligonucleotides. Interestingly, the amount of photodimer formed appears not to change over the course of further irradiation, as would have been expected from the UV-Vis spectra in Figure 5.23. This could be due to the camera being saturated at the first timeframe and therefore subsequent increases in the intensity of the band are not observed. Alternative techniques to quantify the amount of photodimer would need to be undertaken to confirm this, for example, HPLC separation followed by quantification, or the use of ^{31}P labelled oligonucleotides for quantitative detection in the gels.

5.4.10 Monitoring the Photo-Triggered Release

The aliquots used in the denaturing gel were also analysed using native gel conditions, in order to monitor the amount of photo-triggered release achieved, Figure 5.25. It would be expected that as the photodimer species forms, the **CX** strand will be released and therefore an increase in intensity of this band will be observed over the irradiation time. In addition to this the amount of **TX:PTX** duplex is also expected to increase.

Over the course of irradiation, the intensity of the duplex bands appears not to change, consistent with the observations made with the denaturing gel. However, the intensity of the **C1-L** and **C1-M** band does increase with irradiation time. This

is most prominent in the **T1:PT1:C1-M** sample (Figure 5.25, gel A, lanes 8-12) and is accompanied by a decrease in intensity of the single stranded **PT1** species. This result is very promising and confirms that photo-triggered release using the toe-hold system, is possible. Although changes in the **T2:PT3** system (Figure 5.25, gel B) are observed, these are not as prominent as the **T1:PT1** system. It is thought this could be due to the T_m of the **T2:PT3** duplex exceeding that of the **T2:C1-L** and **T2:C1-M** duplexes (Section 5.4.3). This limits the binding of **C1-L** and **C1-M** prior to irradiation and therefore limits the extent of release that can be achieved.

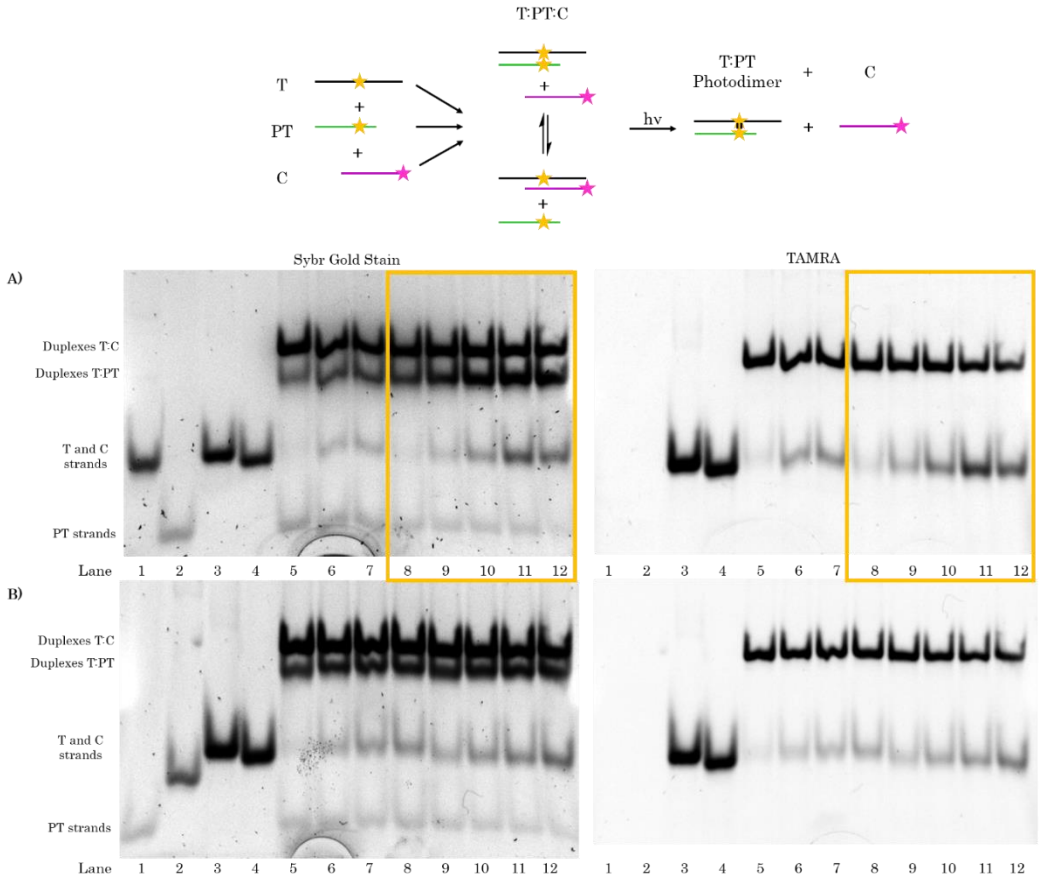


Figure 5.25 Top: Schematic representation of toe-hold system and expected photorelease. Bottom: Native polyacrylamide gel of TAMRA release strand irradiated samples performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA at 293 K. Irradiation performed with 365 nm light after 10 minutes degassing with argon at 293 K. A) Lane: 1 **T1**, 2 **PT1**, 3 **C1-L**, 4 **C1-M**, 5 – 7 **T1:PT1:C1-L** at 0, 30 and 60 minutes irradiation respectively, 8-12 **T1:PT1:C1-M** at 0, 30, 60, 90 and 120 minutes irradiation respectively. B) Lane: 1 **T2**, 2 **PT3**, 3 **C1-L**, 4 **C1-M**, 5 – 8 **T2:PT3:C1-L** at 0, 30, 60 and 90 minutes irradiation respectively, 9-12 **T1:PT1:C1-M** at 0, 30, 60 and 90 minutes irradiation respectively. Visualised using SYBR Gold® nucleic acid stain (left) and TAMRA fluorescence (right).

5.4.11 Efforts to Quantify Photo-Triggered Release

Due to the limitations associated with quantifying the amount of DNA within the gel, using the imaging techniques described above, alternative methods were needed. A potential technique to quantify the amount of **C1-L** and **C1-M**, bound to the duplex, was to utilise the fluorescence of the TAMRA tag. Within this thesis the changes in fluorescence output of a fluorophore upon incorporation into DNA, have been explored and this could also be utilised here. In order to evaluate the feasibility of this method, titrations of **T1** into **C1-L** and **C1-M** were performed, Figure 5.26. The intensity of the TAMRA fluorescence was shown to vary between **C1-L** and **C1-M,** which was expected to be a result of differences in their sequences. Upon addition of **T1**, a steady decrease in the intensity was observed in both cases, levelling off at a 1:1 ratio of **C1-M** to **T1**, consistent with the formation of a 1:1 duplex. The **C1-L:T1** system levels off earlier, suggesting errors in the concentration of the **C1-L** strand. However, the data presented is from one experiment and therefore repeats would be needed in order to confirm the results obtained.

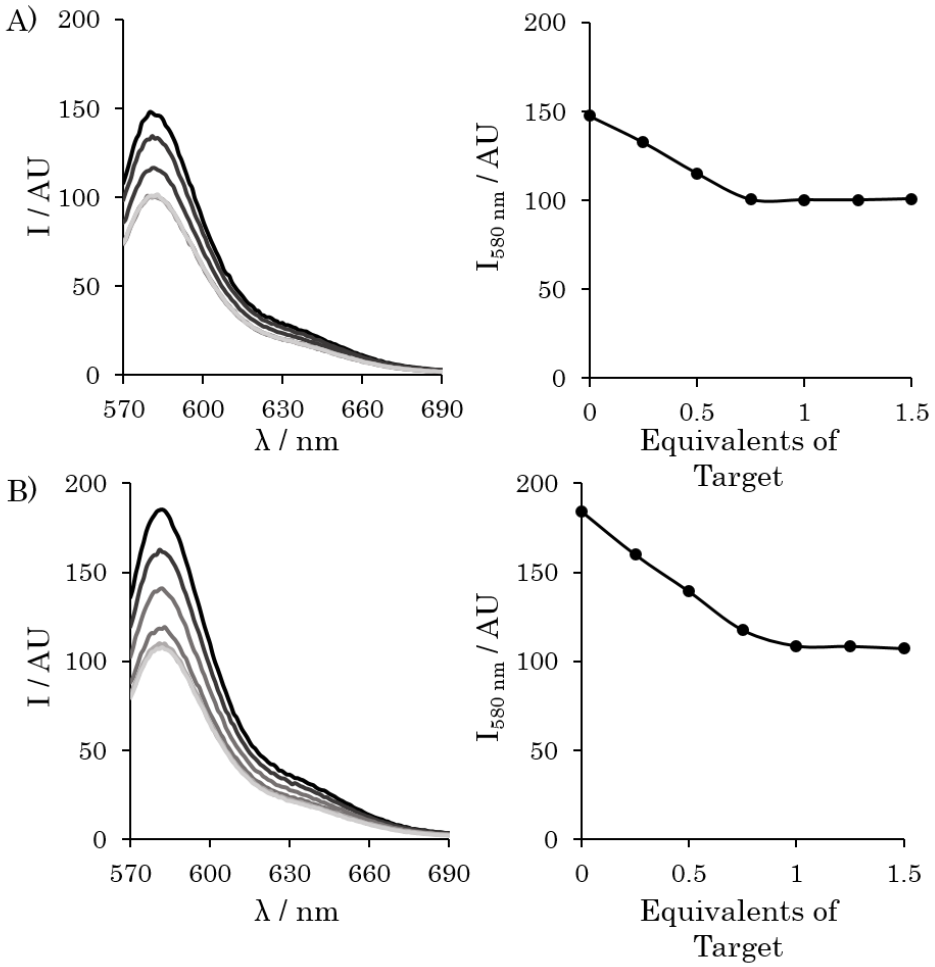


Figure 5.26 Fluorescence spectra (left) of titration of A) **T1** into **C1-L** and B) **T1** into **C1-M** with 1 µM **C1-L** or **C1-M** and plot of equivalents of target against change in fluorescence intensity (right). in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. **T1** added in 0.25 µM steps (black to grey). λ_{ex} = 556 nm.

Subsequent fluorescent analysis of the toe-hold system **T1:PT1:C1-L** was performed (Figure 5.27), to determine if fluorescence could be used to quantify the amount of TAMRA tagged strand in the duplex. The system containing all three oligonucleotides gave an intensity similar to that of the **C1-L:T1** duplex, suggesting that all **C1-L** was duplexed. However, this observation does not agree with the gel studies performed above, which suggested an equilibrium of binding between the **T1:PT1** and **T1:C1-L** duplexes.

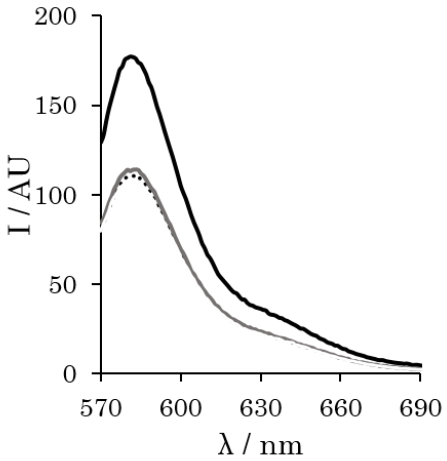


Figure 5.27 Fluorescence spectra of toe-hold system **C1-L** (black solid), **C1-L:T1** (black dot) and **C1-L:T1:PT1** (grey solid) all at 1:1 molar equivalents at 1 µM in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. λ_{ex} = 556 nm.

Further to this unexpected observation, a reduction in fluorescence intensity was observed upon irradiation of the **C1-L** oligonucleotide alone, Figure 5.28. Subsequent MS analysis of the sample showed no change in the molecular mass. It is not known what is causing this decrease, but it may be as a result of photobleaching, although the fluorescence signal did not reappear over time. Future attempts to quantify photo-triggered release utilising this method, were not explored further.

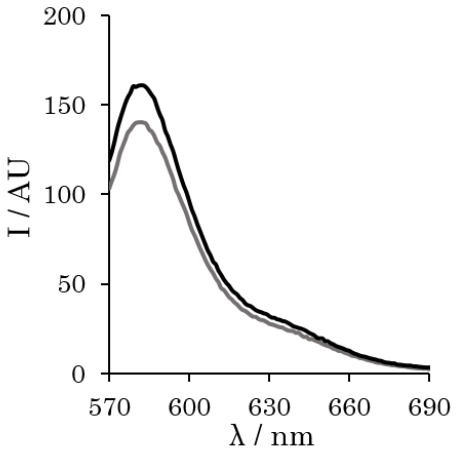


Figure 5.28 Fluorescence spectra of 1 µM **C1-L** prior to irradiation (black) and after 1 hour irradiation with 365 nm light (grey) in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. λ_{ex} = 556 nm

Qualitative analysis of the amount of photo-release achieved can be inferred from the native gels shown in Figure 5.25. It can be seen that with the **C1-L** complementary system, a modest release is achieved. This is more pronounced with the **C1-M** system, particularly **T1:PT1:C1-M** in which a significant amount is released. This level of release falls short of that observed in the azobenzene examples in the literature in which complete release is observed.[17] However, it should be noted that, at particular temperatures, the azobenzene system fails to work. This is due to binding to the complementary strand not being blocked, but only destabilised. Therefore, lowering the temperature would allow for the strand to still form a duplex. Within the system presented herein, the complementary strand would be unable to bind at these temperatures, due to the **PTX** strand being bound.

5.4.12 Thermal Reversion of Photodimerisation

It is known that anthracene photodimerisation can be reversed upon the application of heat. The reversibility of the system was assessed by taking the irradiated samples (Figure 5.23) and heating to 90 °C for 13 hours. A denaturing gel was then performed to compare before and after heat treatment, Figure 5.29. It would be expected that a high molecular weight species would be present from the photodimerised duplex (as in Figure 5.24). Upon heating, the photodimer would be expected to revert to the monomer species and therefore this high molecular weight band would reduce in intensity (or disappear completely) and the

amount of low molecular weight (single stranded DNA) would increase. It was observed that, although a small proportion of photodimer did remain, the majority had reverted to single stranded species as was expected. This reversibility is extremely promising and would allow the photo-triggered release to be reverted.

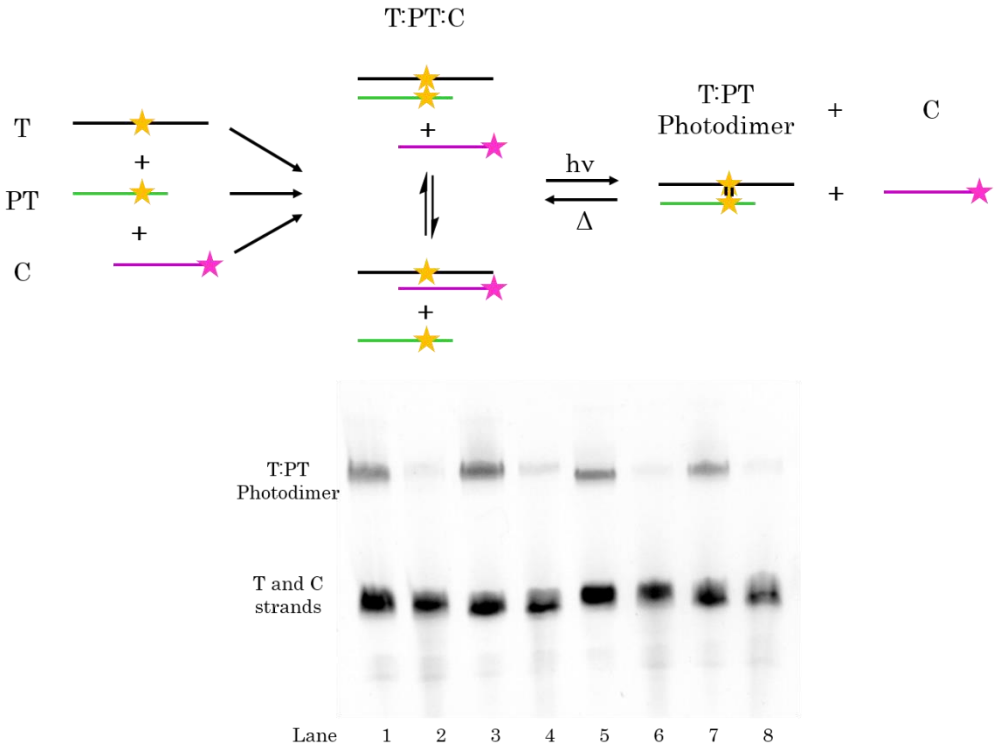


Figure 5.29 Top: Schematic representation of toe-hold system showing expected photorelease and reversion upon heating. Bottom: Denaturing polyacrylamide gel of toe-hold systems following irradiation with 365 nm light until no further change in absorbance was observed. Samples then heated to 90 °C for 13 hours and analysed. Gel performed in 1 × tris borate EDTA running buffer with 500 nM DNA in 50% formamide at 293 K. Lanes shows irradiated sample and heat treated sample respectively. Lane 1-2 **T1:PT1:C1-L**, 3-4 **T1:PT1:C1-M**, 5-6 **T2:PT3:C1-L** and 7-8 **T2:PT3:C1-M**. Visualised using SYBR Gold® nucleic acid stain.

5.4.13 Cycling of Photo-Triggered Release

The thermal reversion of the photodimers could allow for the system to be cycled and reused, as shown in the schematic in Figure 5.29. This was assessed using cycles of 90 minutes irradiation, followed by 13 hours heating at 90 °C with aliquots

taken at each point. The aliquots were then analysed by denaturing (Figure 5.30) and native (Figure 5.31) gel electrophoresis. In the denaturing gel it would be expected that the amount of photodimer (shown as a high molecular weight band) would alternate between being high concentration (an intense band at high molecular weight) after irradiation and lower (and therefore a weaker band) after heating. This will be accompanied by weak single stranded (low molecular weight) bands from the **PTX** and **TX** strands when photodimerised and strong single stranded bands when the anthracene photodimer is reversed to monomers. In all cases the bands from the **CX** strands is expected to remain constant as in no cases is this band covalently linked to another strand.

Within the native gel an increase in the intensity of the **CX** strand is expected after irradiation (and associated photorelease). Upon heating, the photorelease will be reversed and the intensity of the single stranded **CX** band will decrease as it forms duplexes with the **TX** strand. In accordance with this, the amount of **TX:CX** duplex will be higher after heating.

There is some photodimer present in the pre-irradiated systems, which have not been purposely exposed to light (lane 5). This suggests that the samples have been subject to photodimer inducing conditions prior to running the gel. Despite this slight amount of photodimer formation, the increase and decrease in the amount of photodimer present can clearly be seen in the denaturing gels (Figure 5.30)

suggesting that cycling of the system was indeed possible. Changes in the intensity of the **CX** strands can also be seen in the native gel (Figure 5.31). In the third cycle, fatigue of the system is more apparent. This level of fatigue is much higher that is seen in the azobenzene systems used by Asanuma and co-workers.[17] More detailed analysis on any side product formation which may be causing the fatigue is needed. This could allow for the degradation route to be blocked and therefore the cycling ability of the system improved.

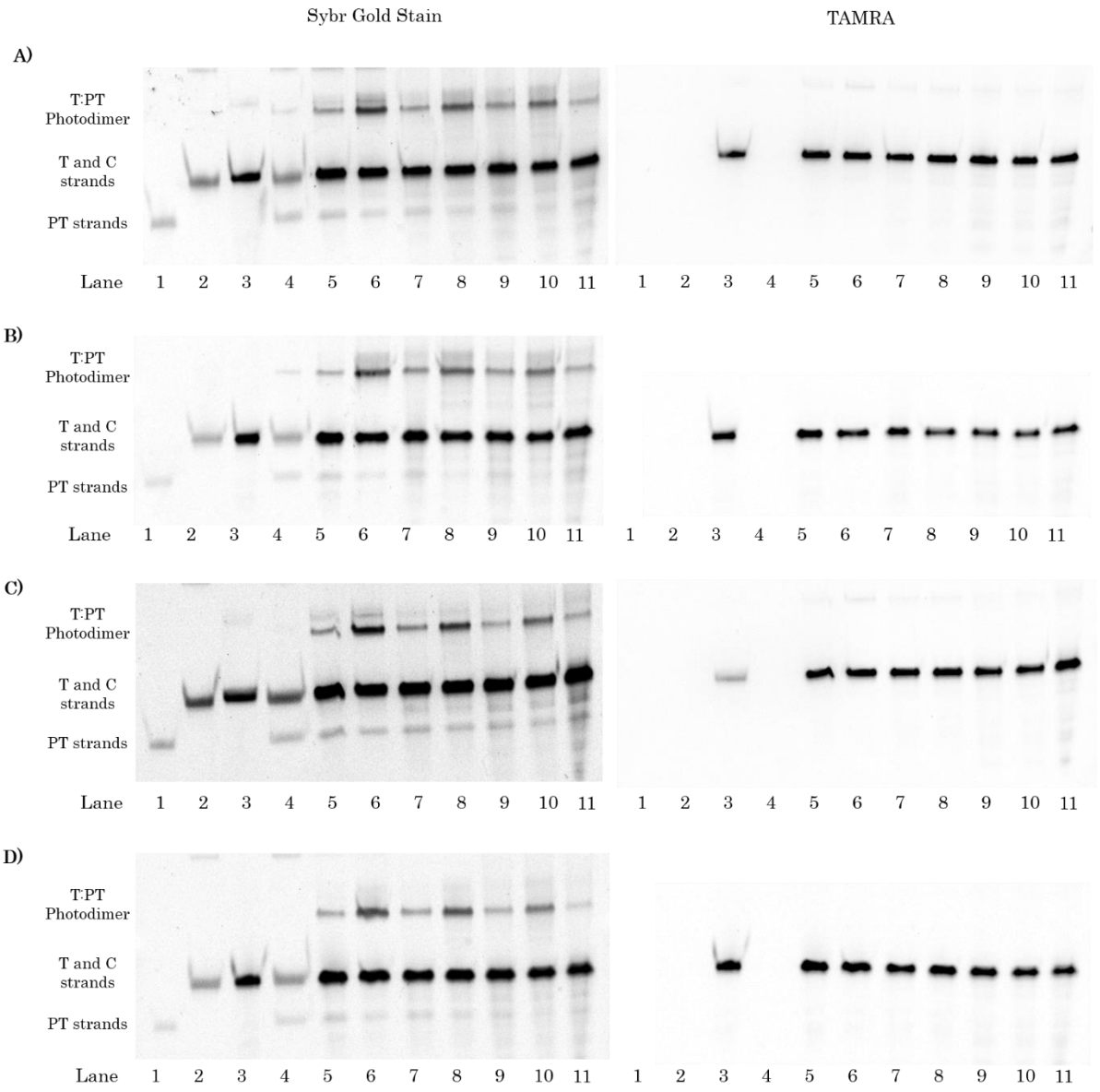


Figure 5.30 Denaturing gel electrophoresis to monitor cycling of photodimerisation with 90 minute irradiation with 365 nm light followed by heating to 90 °C for 13 hours. All performed in 1 × tris borate EDTA running buffer with 500 nM DNA in 50% formamide at 293 K and visualised using TAMRA fluorescence (right) and SYBR Gold® nucleic acid stain (left). Gel A Lane: 1 **PT1**, 2 **T1**, 3 **C1-L**, 4 **T1:PT1**, 5 **T1:PT1:C1-L**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T1:PT1:C1-L** over 3 cycles. Gel B Lane: 1 **PT1**, 2 **T1**, 3 **C1-M**, 4 **T1:PT1**, 5 **T1:PT1:C1-M**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T1:PT1:C1-M** over 3 cycles. Gel C Lane: 1 **PT3**, 2 **T2**, 3 **C1-L**, 4 **T2:PT3**, 5 **T2:PT3:C1-L**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T2:PT3:C1-L** over 3 cycles. Gel D Lane: 1 **PT3**, 2 **T2**, 3 **C1-M**, 4 **T2:PT3**, 5 **T2:PT3:C1-M**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T2:PT3:C1-M** over 3 cycles.

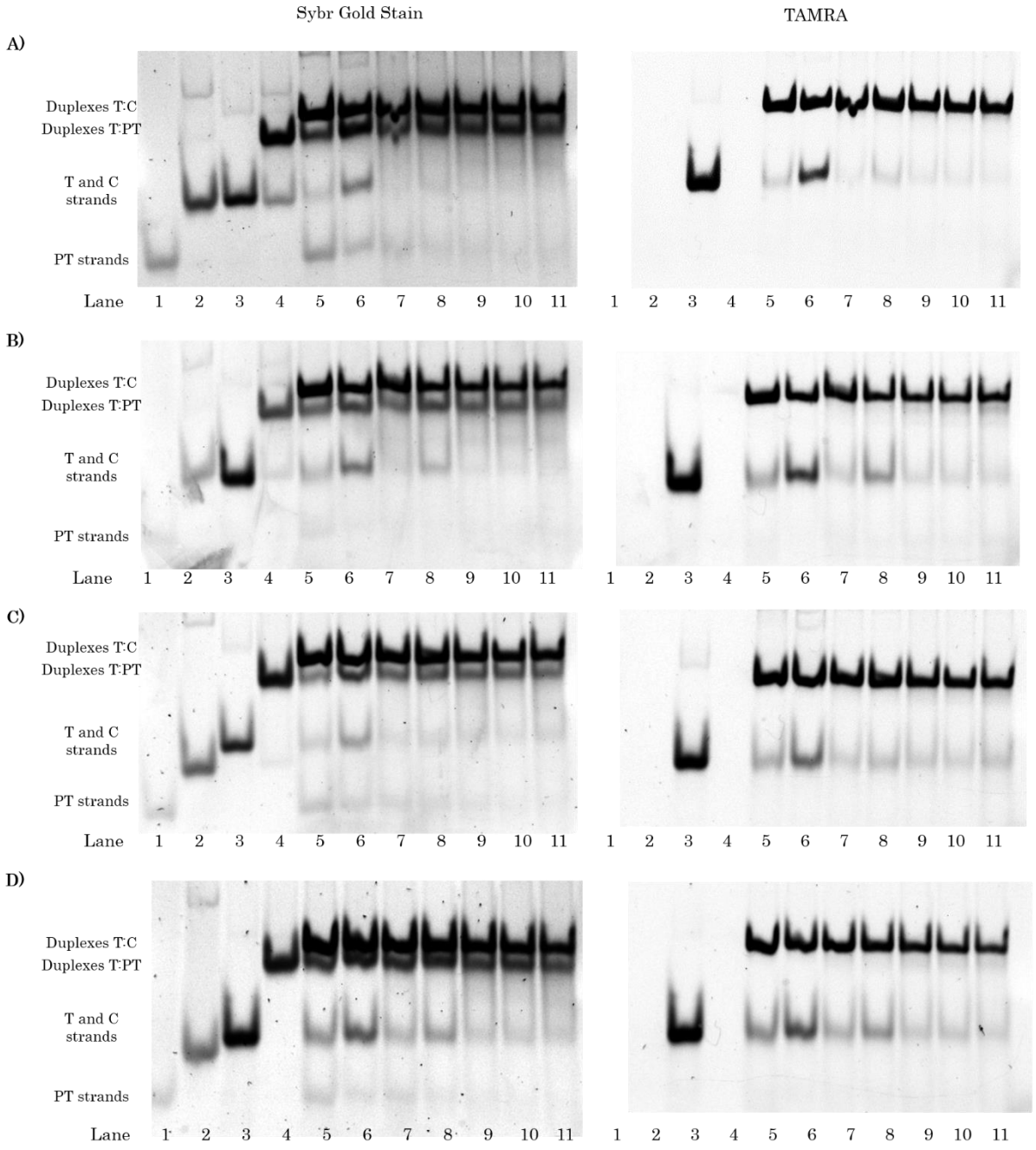


Figure 5.31 Native gel electrophoresis to monitor cycling of photodimerisation with 90 minute irradiation with 365 nm light followed by heating to 90 °C for 13 hours. All performed in 25 mM NaCl and 1 × tris borate EDTA running and sample buffer with 500 nM DNA at 293 K and visualised using TAMRA fluorescence (right) and SYBR Gold® nucleic acid stain (left). Gel A Lane: 1 **PT1**, 2 **T1**, 3 **C1-L**, 4 **T1:PT1**, 5 **T1:PT1:C1-L**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T1:PT1:C1-L** over 3 cycles. Gel B Lane: 1 **PT1**, 2 **T1**, 3 **C1-M**, 4 **T1:PT1**, 5 **T1:PT1:C1-M**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T1:PT1:C1-M** over 3 cycles. Gel C Lane: 1 **PT3**, 2 **T2**, 3 **C1-L**, 4 **T2:PT3**, 5 **T2:PT3:C1-L**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T2:PT3:C1-L** over 3 cycles. Gel D Lane: 1 **PT3**, 2 **T2**, 3 **C1-M**, 4 **T2:PT3**, 5 **T2:PT3:C1-M**, 6-11 cycles of irradiated (even lanes) and heated (odd lanes) samples of **T2:PT3:C1-M** over 3 cycles.

5.5 Conclusions and Future Work

Within this chapter a photo-triggered oligonucleotide release system has been designed and evaluated. The initial stem-loop system was shown to either not photodimerise with long "release" sequences, or to photodimerise and still allow the binding of the "release" sequence with shorter versions. A new design was subsequently used, which made use of a toe-hold system and was shown to display promising traits. This system was able to form a photodimerised duplex upon irradiation, which blocked the binding to the unmodified strand and therefore resulted in an increase in the free single stranded complementary oligonucleotide. In addition to this, the system was shown to be reversible and reusable, although the cycling ability of the system needs to be improved. This is the first example of the use of anthracene photodimerisation to trigger the release of DNA strands.

Future studies should focus on optimising the toe-hold system. Utilising a FRET pair of fluorophores could allow for the quantification of release, assuming those fluorophores are not affected by the irradiation process as the TAMRA label appeared to be. The system could then be optimised by varying the ratios of each strand present to create conditions where the optimum amount of complementary strand can be released. It is expected that this would never be able to reach 100% release due to the equilibrium shifting as more strand is released. However, more significant release efficiencies would need to be achieved than those presented herein in order to make the system a useful tool. It is anticipated that this could

be achieved by altering the length of the **PTX** strands. This would increase the number of hydrogen bond pairs and therefore the T_m of the strand, which should displace the competing strand more effectively. However, this may result in the displacement of the strand prior to irradiation.

An alternative approach, would be to consider altering the ratios of the three strands used. Within this chapter all irradiation studies were performed on systems containing equimolar amounts of **PTX**, **TX** and **CX** strands. Figure 5.22 shows that by altering the ratios of this strands, differing ratios of duplexes can be formed. By performing irradiation on these samples, a more effective release may be achieved. However, it is expected that there will be a fine balance, between achieving the release upon irradiation and the release being a prerequisite of the system due to excesses of competing strands.

5.6 References

1 S. Pushpendra, P. Arvind and B. Anil, *From Nucleic Acids Sequences to Molecular Medicine*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
2 J. C. Burnett and J. J. Rossi, *Chem. Biol.*, 2012, **19**, 60–71.
3 L. Xu and T. Anchordoquy, *J. Pharm. Sci.*, 2011, **100**, 38–52.
4 R. S. Geary, S. P. Henry and L. R. Grillone, *Clin. Pharmacokinet.*, 2002, **41**, 255–60.
5 J. K. Watts and D. R. Corey, *J. Pathol.*, 2012, **226**, 365–79.
6 P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302.
7 B. Saccà and C. M. Niemeyer, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 58–66.
8 Y. Krishnan and F. C. Simmel, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 3124–56.
9 B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605–8.
10 S. F. J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama and A. J. Turberfield, *Nat. Nanotechnol.*, 2012, **7**, 169–73.
11 P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield and J. H. Reif, *Angew. Chem. Int. Ed. Engl.*, 2004, **43**, 4906–11.
12 R. P. Goodman, M. Heilemann, S. Doose, C. M. Erben, A. N. Kapanidis and A. J. Turberfield, *Nat. Nanotechnol.*, 2008, **3**, 93–6.
13 S. N. Syed, H. Schulze, D. Macdonald, J. Crain, A. R. Mount and T. T. Bachmann, *J. Am. Chem. Soc.*, 2013, **135**, 5399–407.
14 K. W. Chun, J. B. Lee, S. H. Kim and T. G. Park, *Biomaterials*, 2005, **26**, 3319–26.
15 Y.-C. Liu, A.-L. M. Le Ny, J. Schmidt, Y. Talmon, B. F. Chmelka and C. T. Lee, *Langmuir*, 2009, **25**, 5713–24.
16 W. Szymański, J. M. Beierle, H. A. V Kistemaker, W. A. Velema and B. L. Feringa, *Chem. Rev.*, 2013, **113**, 6114–78.
17 H. Asanuma, X. Liang, T. Yoshida and M. Komiyama, *Chembiochem*, 2001, **2**, 39–44.
18 H. Nishioka, X. Liang, T. Kato and H. Asanuma, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 1165–8.
19 H. Asanuma, X. Liang, H. Nishioka, D. Matsunaga, M. Liu and M. Komiyama, *Nat. Protoc.*, 2007, **2**, 203–12.
20 Y. Kamiya and H. Asanuma, *Acc. Chem. Res.*, 2014, **47**, 1663–72.
21 C. Dohno and K. Nakatani, *Chem. Soc. Rev.*, 2011, **40**, 5718–29.
22 T. Goldau, K. Murayama, C. Brieke, S. Steinwand, P. Mondal, M. Biswas, I. Burghardt, J. Wachtveitl, H. Asanuma and A. Heckel, *Chemistry*, 2015, **21**, 2845–54.
23 M. Biswas and I. Burghardt, *Biophys. J.*, 2014, **107**, 932–40.
24 C. Brieke and A. Heckel, *Chemistry*, 2013, **19**, 15726–34.
25 J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.*, 2008, **130**, 11836–7.

26 H. Asanuma, K. Shirasuka, T. Yoshida, T. Takarada, X. Liang and M. Komiyama, *Chem. Lett.*, 2001, 108–9.
27 H. Cahová and A. Jäschke, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 3186–90.
28 H. Kashida, T. Doi, T. Sakakibara, T. Hayashi and H. Asanuma, *J. Am. Chem. Soc.*, 2013, **135**, 7960–6.
29 J. Manchester, D. M. Bassani, J.-L. H. A. Duprey, L. Giordano, J. S. Vyle, Z. Zhao and J. H. R. Tucker, *J. Am. Chem. Soc.*, 2012, **134**, 10791–4.
30 J. Manchester, PhD Thesis, University of Birmingham, 2013.
31 J.-L. H. A. Duprey, PhD Thesis, University of Birmingham, 2010.
32 D. A. Khodakov, A. S. Khodakova, A. Linacre and A. V Ellis, *J. Am. Chem. Soc.*, 2013, **135**, 5612–9.
33 J. Zhu, Y. Ding, X. Liu, L. Wang and W. Jiang, *Biosens. Bioelectron.*, 2014, **59**, 276–81.
34 G. Seelig, D. Soloveichik, D. Y. Zhang and E. Winfree, *Science*, 2006, **314**, 1585–8.
35 W. Deng, H. Xu, W. Ding and H. Liang, *PLoS One*, 2014, **9**, e111650.
36 D. Y. Zhang and E. Winfree, *J. Am. Chem. Soc.*, 2009, **131**, 17303–14.
37 R. S. Tuma, M. P. Beaudet, X. Jin, L. J. Jones, C. Y. Cheung, S. Yue and V. L. Singer, *Anal. Biochem.*, 1999, **268**, 278–88.

Chapter 6

Controlling DNA Folding using Anthracene Photodimerisation

6.1 Introduction

Within this chapter anthracene photodimerisation is used to control the folding of a G-quadruplex structure, and efforts towards evaluating the binding of this moiety to a natural enzyme, are assessed. This work is a continuation of the studies previously performed within the group.[1]

6.1.1 The G-Quadruplex Structure

The most common structure found within DNA is the B-DNA double helix as discussed in Chapter 1.2.1. However, several other folding conformations are now known to be naturally present, such as the i-motif and the G-quadruplex.[2] In recent years, the biological relevance of these structures has begun to be understood.[2] The G-quadruplex structure arises in guanine rich sequences. The DNA strand folds to allow four guanine bases to form Hoogsteen base pairs, with a monovalent cation in the centre of the cavity, Figure 6.1.[3] The fold that is adopted by the DNA strand is highly dependent upon the DNA sequence and the spacing between the guanine repeating units; this has been studied in detail and summarised in recent reviews.[4,5] Circular dichroism (CD) spectroscopy can be used to study the G-quadruplex structure, and the characteristic signals of the various conformations are now well documented.[6] The expected CD spectrum of the anti-parallel conformation is shown in Figure 6.2.

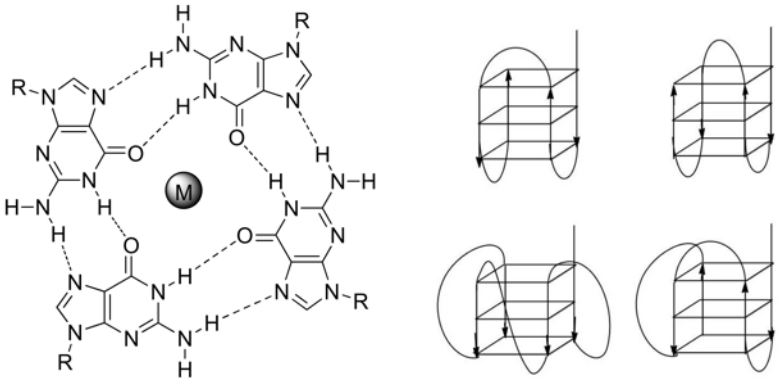


Figure 6.1 Hoogsteen bonded guanine bases form a quadruplex with a monovalent cation in the central cavity (left). The DNA strand can fold in various conformations to allow for this interaction, four of the most common are shown (right). Reprinted with permission from Ma *et al.*, *Oncogene and Cancer – From Bench to Clinic*, 2013, Chapter 6. Copyright 2013 InTech.

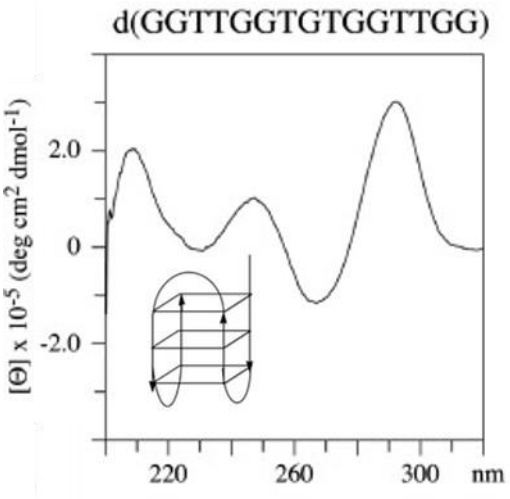


Figure 6.2 Circular Dichroism signal expected for the chair conformation of the G-quadruplex. Adapted with permission from Paramasivan *et al.*, *Methods*, 2007, **43**, 324-31. Copyright 2007 Elsevier Inc. and Ma *et al.*, *Oncogene and Cancer – From Bench to Clinic*, 2013, Chapter 6. Copyright 2013 InTech.

6.1.2 Biological Relevance

In recent years, the study of the G-quadruplex system has increased significantly as the biological relevance of the conformation becomes more understood.[7] It is now known that the structure is one of the main components of telomeres[8] and also plays a role, along with the i-motif, in the regulation of the transcription, translation and replication processes.[9–12] As such these structures have become interesting targets to understand and treat diseases. Research is now progressing to allow for the control of these stuctures.[13]

6.1.3 Controlling the G-Quadruplex

One potential way to control the formation of the G-quadruplex, is to use ligands that stabilise the structure. Many ligands have been developed and are now being explored as potential anti-cancer therapies.[14] The G4 ligand database

details the reported ligands that are known to bind to the G-quadruplex.[15] It also incorporates a tool to help design and target molecules to bind to specific G-quadruplex sequences and structures. The majority of G-quadruplex ligands exploit two modes of binding, a π-π stacking interaction with the guanine bases, and electrostatic interactions with the DNA backbone. The most promising of these molecules (the only one to make phase II clinical trials) utilises groove binding in addition to π-π stacking, Figure 6.3.[16,17]

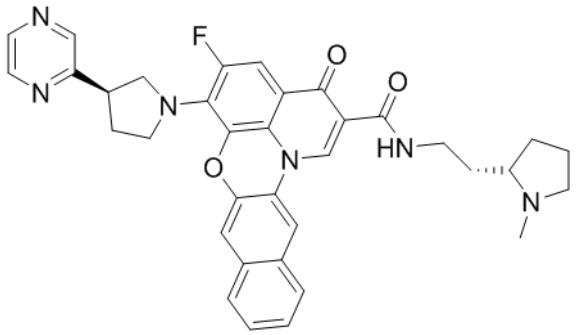


Figure 6.3 Structure of a G-quadruplex stabilising ligand, fluoroquinolone quarfloxin, comprised of an electron rich unit which forms π- π stacking interactions with the guanine bases, and side chains which form van der Waals interactions with the quadruplex grooves.

Alternative approaches aim to induce the quadruplex folding. This is known to be aided by metal ions (particularly potassium), but recently ligands that induce a particular conformation of G-quadruplex have been discovered. Figure 6.4 shows two of these structures. Binding to A has been shown to induce an antiparallel quadruplex, where the DNA strands align 5′-3′ with 3′-5′ (Figure 6.1, top right), whilst binding to B leads to the formation of a parallel quadruplex, where the DNA strands align 5′-3′ with 5′-3′ (Figure 6.1, bottom left). The authors do not suggest reasons for this differentiation, although they do observe that the

polyamine of B threads through the quadruplex core, a significantly different binding mode to molecule A.[18,19]

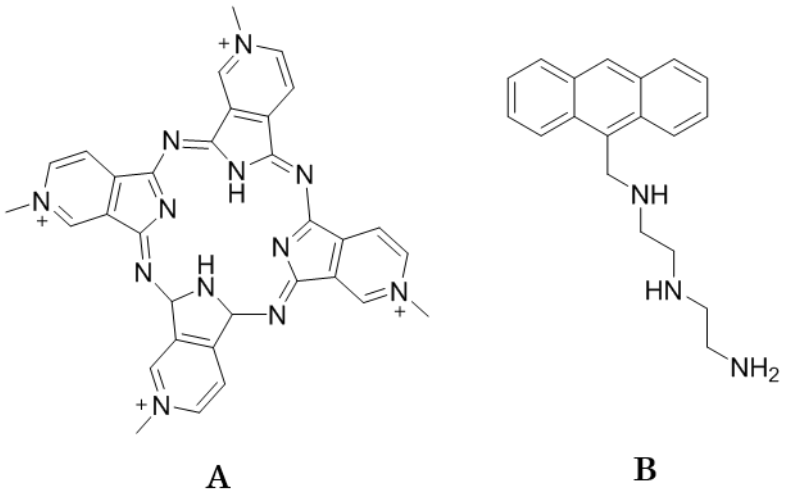


Figure 6.4 Structures of G-quadruplex inducing ligands. Binding of ligand A forms an antiparallel quadruplex, whilst binding of B induces the formation of a parallel quadruplex.

In addition to these, ligands that disrupt the quadruplex structure have also been developed.[20,21] These typically contain regions which are able to form π-π stacking interactions, coupled with large groups, which prevent the DNA folding into the quadruplex structure due to the formation of unfavourable steric interactions.[14] An example of this is the triarylpyridine shown in Figure 6.5 developed by Waller and co-workers.[20]

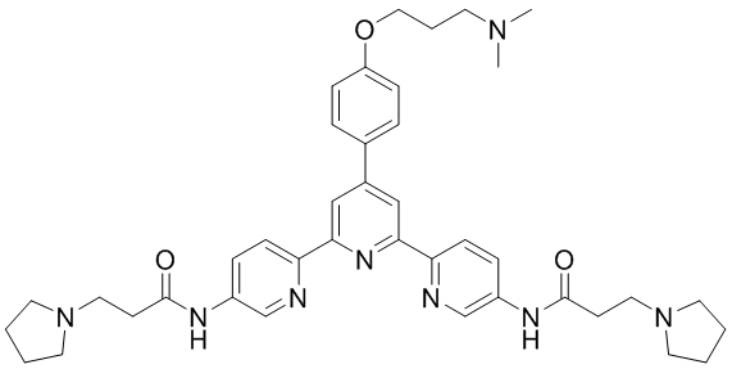


Figure 6.5 The triarylpyridine derivative developed by Waller and co-workers has been shown to disrupt G-quadruplex formation.

6.1.4 Using Light to Control the G-Quadruplex

Of particular interest to the studies presented herein, is using light to control G-quadruplex folding. Maeda and co-workers were the first to demonstrate this using a photo-regulated guanine nucleobase that underwent *cis-trans* isomerisation upon light irradiation. The researchers incorporated the nucleobase, at various positions within the thrombin binding aptamer sequence, a well-known G-quadruplex forming aptamer. Upon isomerisation to the *cis* form, it was shown that the quadruplex unfolded, leading to reduced binding to thrombin, Figure 6.6.[21]

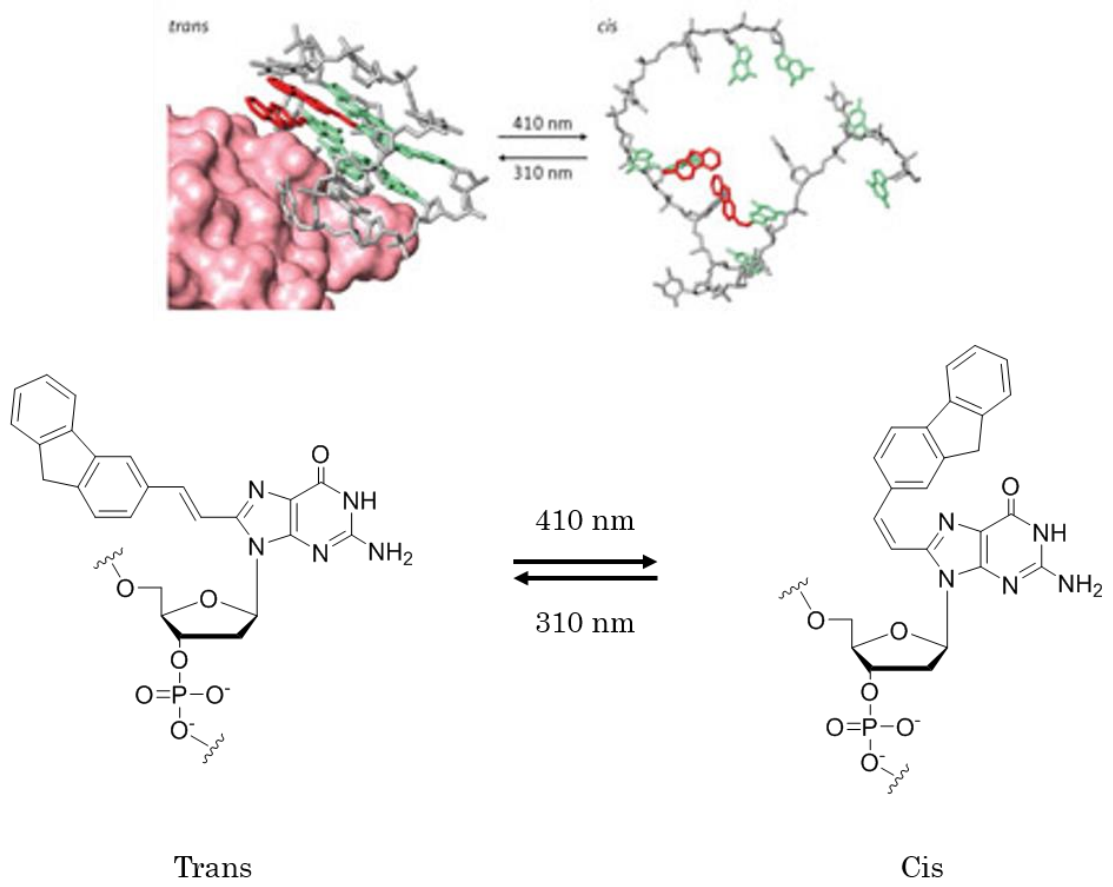


Figure 6.6 Photo-regulated nucleobase used to control G-quadruplex formation. In the *trans* form the quadruplex can form and binding to thrombin is observed. Isomerisation to the *cis* form disrupts the quadruplex and hence the thrombin binding. Adapted with permission from Ogasawara *et al.*, *Angew. Chem. Int. Ed. Engl.*, 2009, **48**, 6671-4. Copyright 2009 WILEY-WCH Verlag GmbH & Co.

Zhou and co-workers have utilised an azobenzene derivative to control the folding of a G-quadruplex sequence found within telomeres. They showed that the *trans* form of the molecule interacted with the G-quadruplex structure. Upon isomerisation, the *cis* form could not maintain the same stabilising interactions and the quadruplex unfolded, Figure 6.7.[22,23]

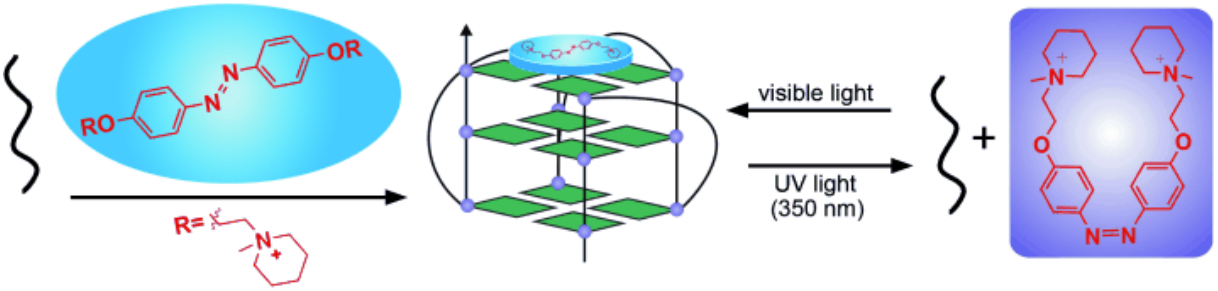


Figure 6.7 Schematic representation of the light controlled folding of a G-quadruplex structure utilising an azobenzene ligand. In the *cis* form the ligand interacts and stabilises the quadruplex structure (left). Upon isomerisation to the *trans* conformation (right), the ligand can no longer bind and the quadruplex does not form. Reprinted with permission from Wang *et al.*, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 5305-9. Copyright 2010 WILEY-WCH Verlag GmbH & Co.

Research into the photo crosslinking of G-quadruplex structures, in order to lock DNA into a particular conformation, has also been conducted. Verga *et al.* attached two photoactive molecules to bases within the quadruplex structure, one being an azide derivative the other a benzophenone. Upon irradiation, the molecules reacted to crosslink the DNA and lock it into the quadruplex structure. This technique was also shown to work within cells.[24] It is hoped that this type of

approach will allow for the trapping of quadruplexes within natural systems, so further studies can be conducted on the biological relevance of such structures.

6.1.5 Thrombin Binding Aptamer

The thrombin binding aptamer (TBA), 5′-GGTTGGTGTGGTTGG-3′, has been utilised to interrogate quadruplex binding in several studies, due to its relatively short sequence, classical anti-parallel folded G-quadruplex structure and well understood binding to thrombin. Within this study, the TBA sequence is adapted to evaluate the ability of anthracene photodimerisation, to control the G-quadruplex fold.

Thrombin is a protease enzyme essential to the coagulation cascade. Its primary function is to convert the soluble protein fibrinogen, into insoluble, clot-forming fibrin, so as to reduce blood loss. The TBA was first discovered in 1992, using the systematic evolution of ligands by exponential enrichment (SELEX) technique.[25] It has been shown to bind strongly to thrombin in exosite I, the natural binding site of the fibrinogen substrate, inhibiting the coagulation pathway. Other aptamers that bind to thrombin are known and vary considerably in their structure, conformation and binding site.[26] These aptamers are generally referred to by alternative names to avoid confusion with the original TBA sequence.

The TBA folds into a chair conformation (Figure 6.8) with a monovalent cation, typically potassium, in the centre. The sequence is comprised of repeated guanine bases, separated by loops containing thymine and a guanine base. In recent studies it has been identified that TBA actually binds to two thrombin entities (Figure 6.9), with one side binding in exosite I and the other to exosite II.[27,28] The binding is found to be highly dependent upon the cations that are present, as these are known to affect not only the folding of TBA, but also the folding of thrombin itself. It has been established that the presence of potassium enables the strongest binding between TBA and thrombin.[29,30]

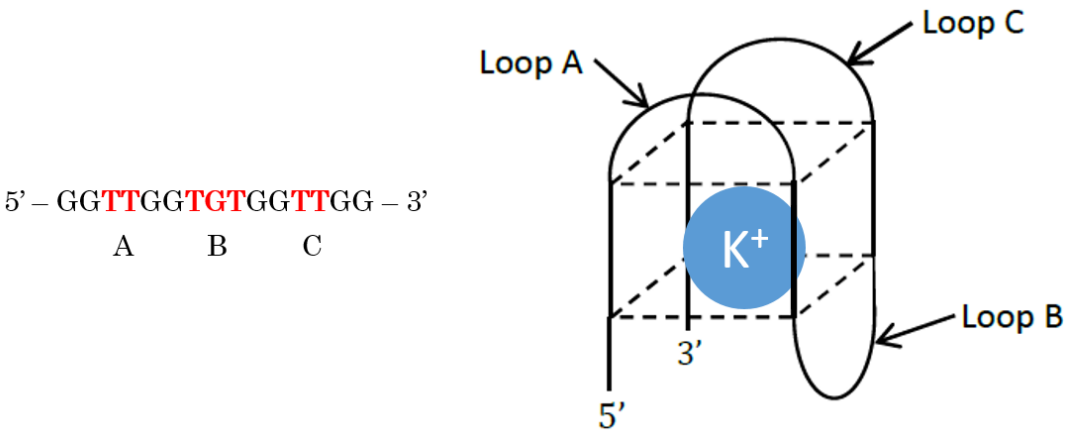


Figure 6.8 Thrombin binding aptamer sequence and folded conformation when bound to potassium. Loop regions are shown in red in the sequence. Adapted from Manchester, PhD Thesis, University of Birmingham, 2013.

Figure 6.9 Model structure of thrombin binding aptamer (yellow) bound to two thrombin proteins. One through exosite I (green) and the other through exosite II (blue). Reprinted with permission from Gonçalves *et al.*, *Chem. Commun.*, 2006, 4685-7. Copyright 2006 Royal Society of Chemistry.

The work presented herein utilises the TBA sequence, to assess the ability of anthracene photodimerisation to control the folding of the quadruplex structure.

6.1.6 Previous Work

Previously within the Tucker group, several systems incorporating anthracene at various points within the TBA sequence have been developed.[1] Subsequent photodimerisation of the anthracene moieties was found to progress at differing rates, depending upon the position of these and the stereochemistry and linker length used to attach the anthracene. Some combinations were found to display

no photodimer formation. Photodimers that did form were purified using RP-HPLC and variable temperature analysis identified that head-to-head (H-H) and head-to-tail (H-T) photodimers were present. Due to the instability of the H-H photodimers and their rapid reversion to monomer species, only the H-T photodimers were subjected to further analysis.[1]

It was established that, upon photodimerisation, the thermal melting (T_m) temperatures of the G-quadruplexes in the presence of K^+, were dramatically reduced, with some displaying no transition suggesting the complete disruption of the structure. This hypothesis was confirmed using CD analysis on two systems, but the full series of sequences was not completed. Thermal reversion of the photodimer species was performed by heating the sample to 80°C for 16 hours. Upon cooling to room temperature, it was observed by CD that the complex had reverted to the original quadruplex fold. Initial studies to assess the effect of anthracene incorporation and photodimerisation on binding to thrombin, using isothermal calorimetry and gel electrophoresis studies, proved inconclusive.[1] Therefore this area of the project is re-addressed herein.

6.2 Project Aims

Within this project, four anthracene tagged, G-quadruplex forming sequences are analysed to monitor their folding dependence on the monovalent cations present. The sequences are then subjected to irradiation conditions to form photodimer

species with the aim of locking, or disrupting the G-quadruplex formation. The structures are then re-assessed to monitor any changes and the binding of the aptamer to thrombin analysed.

6.3 Anthracene Tagged Thrombin Binding Aptamer

6.3.1 Anthracene Tagged Thrombin Binding Aptamer Sequences

Within this thesis the most promising sequences identified by previous work, are assessed, Table 6.1.

Table 6.1 Sequences used for thrombin binding aptamer where X denotes anthracene moiety with a 6 carbon linker length and D stereochemistry. Loop assignments are shown in Figure 6.8.

Name	Sequence (5′ – 3′)	Placement of Modification
TBA	GG TTG GTG TGG TTG G	N/A
M1-TBA	**X**TGG TTG GTG TGG TTG GT**X**	Ends
M2-TBA	GG TTG G**X**G **X**GG TTG G	Both in Loop B
M3-TBA	GG TTG GTG **X**GG **X**TG G	Loops B and C

These incorporate anthracene at various positions within the G-quadruplex, along with an unmodified TBA for direct comparison. All incorporations are made within the loop regions, or at the ends of the aptamer (Figure 6.8), so as to minimise the disruption to the hydrogen bonding regions and folding, prior to photodimerisation.

Previous work has shown that, upon photodimerisation in the presence of K^+, aptamer **M1-TBA** no longer displays a transition by T_m analysis, suggesting that it no longer has a quadruplex structure. **M2-TBA** and **M3-TBA** display reduced T_m values which would be consistent with a less stable quadruplex structure. CD studies had only been performed on the photodimer of **M3-TBA**, which displayed weak signals characteristic of a G-quadruplex structure.[1]

6.3.2 Circular Dichroism Studies

CD analysis was performed on the unmodified and anthracene tagged samples in various buffers, to assess the effect of cations on the folding. As discussed in Section 6.1.4, TBA folding is stabilised by the presence of potassium, however, it is also known that sodium can infer a degree of stability. As such the folding was analysed in Tris HCl buffer and sodium phosphate buffer, both in the presence and absence of potassium ions, Figure 6.10. The sequence used within this study is expected to fold into a chair like structure (Figure 6.8).

Samples of **TBA** and **M1-TBA** (Figure 6.10A and B, respectively) display similar properties. When performed in Tris HCl buffer, the aptamers are unfolded and in the presence of just sodium, signals characteristic of a weak chair G-quadruplex are observed which are slightly more pronounced for **M1-TBA**. It is only upon addition of potassium that strong anti-parallel G-quadruplex signals are observed, with little change between the Tris HCl and sodium phosphate

samples. These observations suggest that when the anthracene modifications are incorporated at the ends of the aptamer, there is little change to the folding behaviour of the sequence.

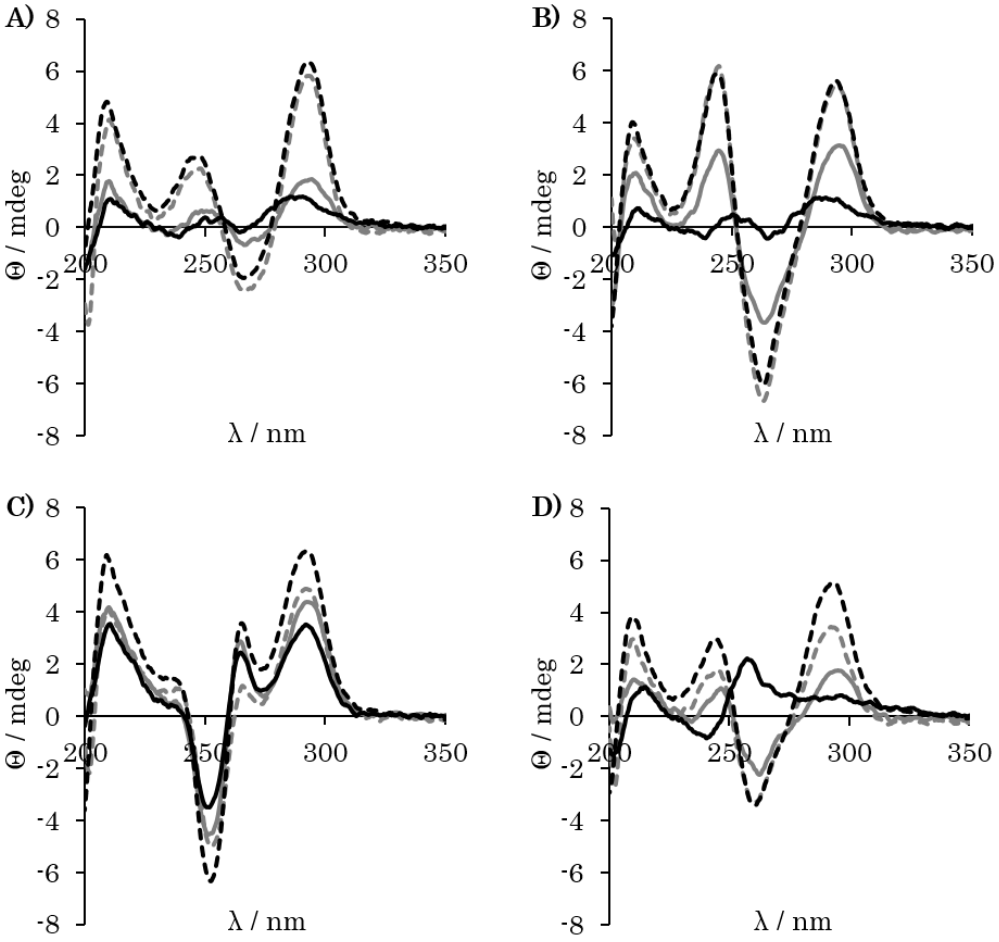


Figure 6.10 CD analysis of TBA derivatives. A) **TBA**, B) **M1-TBA**, C) **M2-TBA** and D) **M3-TBA**. Performed at 1 µM oligonucleotide in 10 mM Tris HCl pH 7 (black solid) and with 100 mM KCl added (black dash) and in 10 mM sodium phosphate buffer pH 7 (grey solid) and with 100 mM KCl added (grey dash). All performed in 1 cm pathlength cuvette at 293 K.

Aptamer **M2-TBA** is the only sample which does not display a folding dependence in the presence of cation (Figure 6.10 C). With this aptamer, all signals are very similar between the different buffer conditions used, with an anti-parallel conformation displayed. Additionally, all spectra display a strong anthracene

induced peak at 265 nm. This strongly suggests that the anthracene is intercalating, which could increase the stability of the folded conformation. Interestingly, the negative band centred around 250 nm is much stronger than would be expected from the chair G-quadruplex, and far more characteristic of a B-DNA structure. This could indicate that the anthracene incorporation at the centre of the sequence has completely disrupted the fold, and could account for the reduced folding dependence upon the presence of a cation.

Aptamer **M3-TBA** in Tris HCl buffer displays a positive band at around 255 nm, this could be due to an anthracene induced peak, or a parallel G-quadruplex structure as both display signals in this region. In the presence of sodium ions, weak signals characteristic of a quadruplex structure are observed with a positive band at ~290 nm and a negative band ~260 nm. The addition of potassium enhances the signals expected for an anti-parallel G-quadruplex.

The above results show that strong anti-parallel G-quadruplex signals are observed in the presence of potassium ions. Subsequently, analysis of the H-T photodimer species was performed in sodium phosphate buffer in the presence and absence of potassium, Figure 6.11, to establish the effect photodimerisation has on the folding.

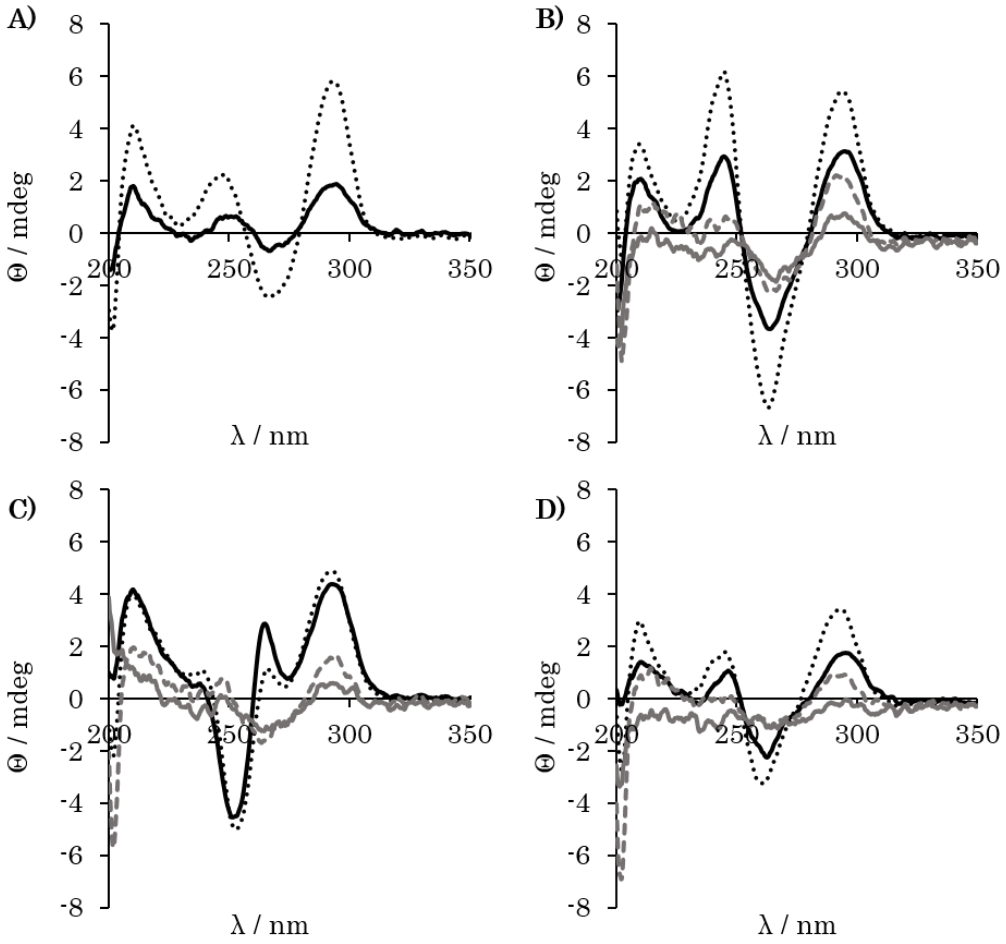


Figure 6.11 CD analysis of TBA derivatives. A) **TBA**, B) **M1-TBA**, C) **M2-TBA** and D) **M3-TBA** unirradiated samples (black) and H-T photodimers isolated from HPLC (grey). Performed at 1 µM oligonucleotide in 10 mM sodium phosphate buffer pH 7 (solid) and with 100 mM KCl added (dashed). All performed in 1 cm pathlength cuvette at 293 K.

In all cases the H-T photodimer species display a reduction in the signals observed, suggesting the structures are less well folded. In addition to this, addition of potassium has little effect on the photodimerised structure. Although samples **M1-TBA** and **M3-TBA** are observed to have transitions in the regions expected for a G-quadruplex fold, these are considerably weaker, suggesting that the secondary structure is less well formed, and resembles that of the unirradiated sample in the absence of potassium. Aptamer **M2-TBA** appears to display a shift in the negative band upon photodimerisation. It is possible that

this is due to the eradication of the induced positive anthracene band at 265 nm upon photodimerisation, giving a resultant spectrum that is similar to those observed for the other modified aptamers. These results confirm that anthracene photodimerisation can impart a degree of control over the G-quadruplex fold and the signals obtained for the photodimerised product are similar to those found in the work of Maeda and co-workers.[21]

6.3.3 Gel Electromobility Shift Assay

An important aspect of this work was to establish the effect of the anthracene moieties on the binding to thrombin. It was thought that this incorporation would disrupt the binding to thrombin to some extent due to the proximity of their attachments to the binding region of the aptamer. Upon photodimerisation, it was expected that the disruption to the folding would inhibit the thrombin binding further. Previous work using ITC to assess the binding was inconclusive.[1] Within this study it was initially decided that gel EMSA would be an ideal technique to qualitatively determine the binding affinities of the aptamers to thrombin. Initial studies were performed on **TBA**, Figure 6.12.

It can be observed that the intact protein-aptamer species is visible at 50 nM of thrombin (Figure 6.12, Lane 5), however, the **TBA** is not fully bound until 1000 nM (Figure 6.12, Lane 12), a much higher concentration than expected compared to previous studies.[31] This is potentially due to errors in the concentration of

thrombin present. The protein is supplied as a lyophilised powder that is quantified using the absorbance of the sample in solution. This method does not account for any impurities that could also absorb at the monitored wavelength (280 nm), such as other proteins, making the solution seem more concentrated than it is. In order to make this method more viable, the protein would need to be purified and re-quantified.

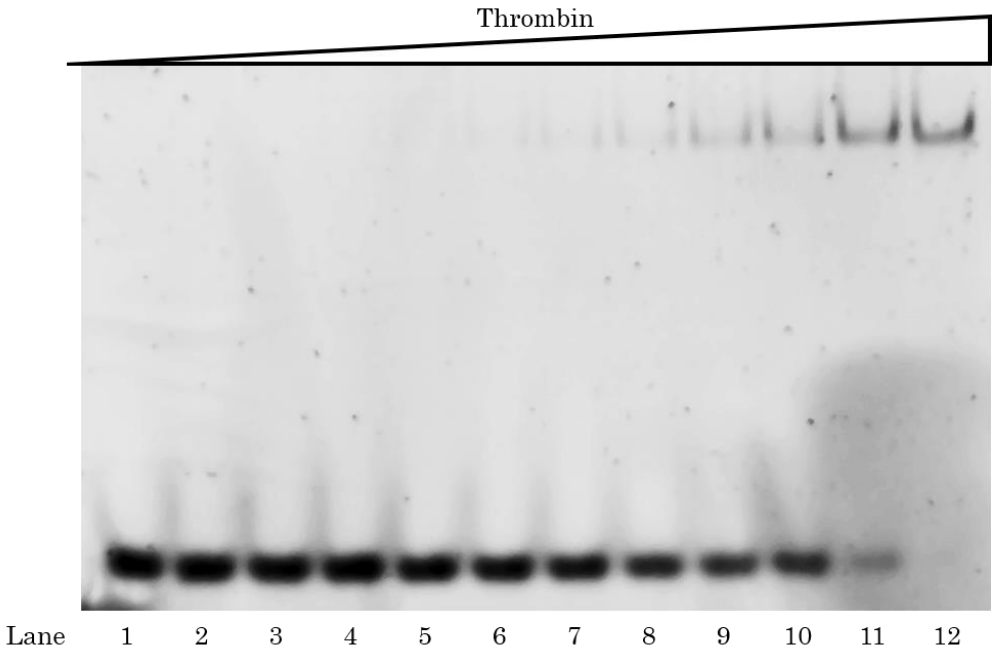


Figure 6.12 Gel EMSA of **TBA** (100 nM) with increasing thrombin concentrations (0, 12.5, 16.67, 25, 50, 75, 100, 150, 200, 250, 500 and 1000 nM). Gel visualised using SYBR Gold® nucleic acid stain.

Although this technique looked promising, it was not used in subsequent studies, as it was found that anthracene incorporation inhibited the fluorescence output of the SYBR Gold® nucleic acid stain. It was not clear whether this was a result of the anthracene inhibiting the binding of the dye to the DNA strands, or if the anthracene moiety was interfering with the fluorescence of the dye. It was found

that the amount of DNA needed to analyse the strands was five times the amount of unmodified TBA, i.e. 500 nM rather than 100 nM (see Figure 6.12). This increase would need to be matched by the same relative increase in protein concentration for the assay to work, an amount of protein that was deemed too excessive.

6.3.4 Binding assay

An alternative approach to monitor the binding is to exploit the natural clotting pathway that thrombin displays. It is known that fibrinogen is converted to fibrin by thrombin, Section 6.1.2. The fibrin then forms a clot, reducing the flow of the liquid and increasing the optical density of the sample. This clot formation is routinely used to measure fibrinogen concentrations within medical samples, and is referred to as the Clauss assay.[32]

Upon binding of TBA to thrombin, the exosite is blocked, and hence the time it takes for a clot to form is either slowed or halted completely. A measurement of the clot time, can therefore be used to infer the binding affinity of the inhibitor, as long as all other variables are kept constant. These type of experiment have been reported in the literature.[25,33–36] Thorough analysis of the clotting curves has also been reported and the elucidation of binding constants from the curves has been achieved.[37]

The aptamers used herein were analysed using the clotting time of fibrinogen when in the presence of thrombin, Figure 6.13. The thrombin, salts and aptamers were mixed prior to addition to the fibrinogen, to ensure they had reached a binding equilibrium.

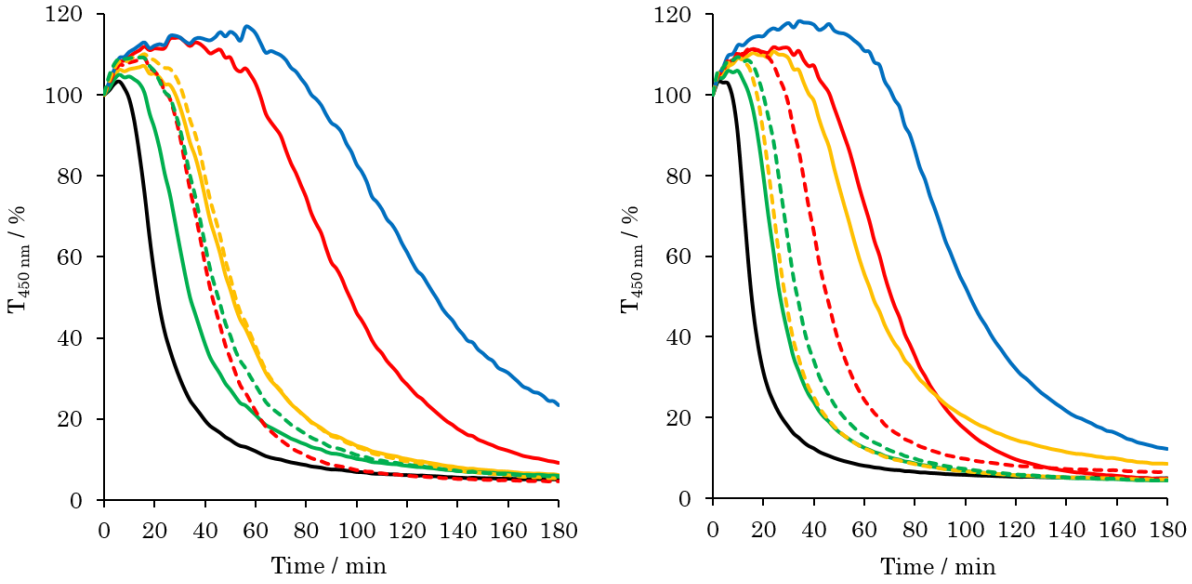


Figure 6.13 Two identical repeats of the clotting time assays performed using 3 μM fibrinogen with 0.1 NIH units of thrombin and 37 nM of aptamer in 50 mM Tris HCl pH 7 with 100 mM KCl, all performed at 25 °C. No aptamer (black), **TBA** (blue), **M1-TBA** (red), **M2-TBA** (yellow), **M3-TBA** (green) with isolated H-T photodimers of anthracene modified sequences shown as the corresponding dashed line.

Initial studies looked promising, with all aptamers displaying clotting times that fell in between the two expected extremes of no aptamer present and **TBA** present. From the CD results (Section 6.3.1) it would have been hypothesised that **M1-TBA** would display the closest resemblance to **TBA,** due to the extremely similar folding. The experiment is consistent with the hypothesis, with **M1-TBA** displaying the slowest clotting time of all the modified aptamers. Similarly, all the H-T photodimers, which were shown to have poor folding, display limited inhibition and have fast clotting times. Aptamer **M3-TBA** contains anthracene

modifications in both of the thrombin binding sites, and would hence be predicted to display the weakest binding of all the modified aptamers. Once again this trend is observed within the clotting times, with this sequence displaying the closest clotting curve to the sample containing no TBA.

Despite the trends being as one would have predicted, the results should not be over interpreted. This assay was repeated several times using several different stocks and suppliers of both fibrinogen and thrombin. Each time unmodified aptamer was used first, but unfortunately consistent results could not be obtained. In addition to this all buffers, cuvettes and preparation equipment was also changed, but the source of the true discrepancy could not be identified.

6.4 Conclusions and Future Work

This study has focussed on gaining a deeper understanding on the use of anthracene photodimerisation to control the folding of oligonucleotides. It has been shown by CD studies, that photodimerisation can reduce the folding of the TBA G-quadruplex. Different methods were then used in an attempt to assess the effect of photodimerisation on thrombin binding and activity. However, further studies would be required to verify initial results which indicate an effect on clotting time in the Clauss assay.

It is hoped that future work will allow for further experiments towards determining and quantifying the binding affinities of the aptamers. This could potentially be performed by optimising methods such as ITC, gel EMSA and the Clauss assay. Confirmation of the successful application of anthracene photodimerisation to control aptamer folding and therefore function, would eventually allow for other more advanced photocontrollable biological systems to be developed.

6.4 References

1 J. Manchester, PhD Thesis, University of Birmingham, 2013.
2 R. R. Sinden, *DNA Structure and Function*, Elseview Inc., California, 1994.
3 D.-L. Ma, V. Pui-Yan Ma, K.-H. Leung, H.-J. Zhong, H.-Z. He, D. Shiu-Hin Chan and C.-H. Leung, *Oncogene and Cancer - From Bench to Clinic*, InTech, 2013.
4 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–15.
5 J. L. Huppert, *FEBS J.*, 2010, **277**, 3452–8.
6 S. Paramasivan, I. Rujan and P. H. Bolton, *Methods*, 2007, **43**, 324–31.
7 Y. Wu and R. M. Brosh, *FEBS J.*, 2010, **277**, 3470–88.
8 S. Neidle and G. N. Parkinson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 275–83.
9 D. J. Patel, A. T. Phan and V. Kuryavyi, *Nucleic Acids Res.*, 2007, **35**, 7429–55.
10 C. Sissi, B. Gatto and M. Palumbo, *Biochimie*, 2011, **93**, 1219–30.
11 S. Kendrick, H.-J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2014, **136**, 4161–71.
12 H.-J. Kang, S. Kendrick, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2014, **136**, 4172–85.
13 Y. Xue, J. Liu, K. Zheng, Z. Kan, Y. Hao and Z. Tan, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 8046–50.
14 N. S. Ilyinsky, A. M. Varizhuk, A. D. Beniaminov, M. A. Puzanov, A. K. Shchyolkina and D. N. Kaluzhny, *Mol. Biol.*, 2014, **48**, 778–94.
15 Q. Li, J.-F. Xiang, Q.-F. Yang, H.-X. Sun, A.-J. Guan and Y.-L. Tang, *Nucleic Acids Res.*, 2013, **41**, D1115–23.
16 A. K. Mehta, Y. Shayo, H. Vankayalapati, L. H. Hurley and J. Schaefer, *Biochemistry*, 2004, **43**, 11953–8.
17 W. Duan, A. Rangan, H. Vankayalapati, M.-Y. Kim, Q. Zeng, D. Sun, H. Han, O. Y. Fedoroff, D. Nishioka, S. Y. Rha, E. Izbicka, D. D. Von Hoff and L. H. Hurley, *Mol. Cancer Ther.*, 2001, **1**, 103–20.
18 D. P. N. Gonçalves, R. Rodriguez, S. Balasubramanian and J. K. M. Sanders, *Chem. Commun.*, 2006, 4685–7.
19 R. Rodriguez, G. D. Pantoş, D. P. N. Gonçalves, J. K. M. Sanders and S. Balasubramanian, *Angew. Chem. Int. Ed. Engl.*, 2007, **46**, 5405–7.
20 Z. A. E. Waller, S. A. Sewitz, S.-T. D. Hsu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 12628–33.
21 S. Ogasawara and M. Maeda, *Angew. Chem. Int. Ed. Engl.*, 2009, **48**, 6671–4.
22 X. Wang, J. Huang, Y. Zhou, S. Yan, X. Weng, X. Wu, M. Deng and X. Zhou, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 5305–9.
23 X. Xing, X. Wang, L. Xu, Y. Tai, L. Dai, X. Zheng, W. Mao, X. Xu and X. Zhou, *Org. Biomol. Chem.*, 2011, **9**, 6639–45.

24 D. Verga, F. Hamon, F. Poyer, S. Bombard and M.-P. Teulade-Fichou, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 994–8.

25 L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564–6.

26 D. M. Tasset, M. F. Kubik and W. Steiner, *J. Mol. Biol.*, 1997, **272**, 688–98.

27 B. Pagano, L. Martino, A. Randazzo and C. Giancola, *Biophys. J.*, 2008, **94**, 562–9.

28 I. Russo Krauss, A. Merlino, C. Giancola, A. Randazzo, L. Mazzarella and F. Sica, *Nucleic Acids Res.*, 2011, **39**, 7858–67.

29 S. Nagatoishi, Y. Tanaka and K. Tsumoto, *Biochem. Biophys. Res. Commun.*, 2007, **352**, 812–7.

30 I. Russo Krauss, A. Merlino, A. Randazzo, E. Novellino, L. Mazzarella and F. Sica, *Nucleic Acids Res.*, 2012, **40**, 8119–28.

31 A. Sosic, A. Meneghello, E. Cretaio and B. Gatto, *Sensors (Basel).*, 2011, **11**, 9426–41.

32 I. J. Mackie, S. Kitchen, S. J. Machin and G. D. O. Lowe, *Br. J. Haematol.*, 2003, **121**, 396–404.

33 A. D. Gunjal, M. Fernandes, N. Erande, P. R. Rajamohanan and V. a Kumar, *Chem. Commun.*, 2014, **50**, 605–7.

34 G. Lupidi, M. Angeletti, A. M. Eleuteri, L. Tacconi, M. Coletta and E. Fioretti, *FEBS Lett.*, 1999, **462**, 236–40.

35 S. Monaco, M. Gioia, J. Rodriguez, G. F. Fasciglione, D. Di Pierro, G. Lupidi, L. Krippahl, S. Marini and M. Coletta, *Biochem. J.*, 2007, **402**, 503–13.

36 L. Martino, A. Virno, A. Randazzo, A. Virgilio, V. Esposito, C. Giancola, M. Bucci, G. Cirino and L. Mayol, *Nucleic Acids Res.*, 2006, **34**, 6653–62.

37 R. De Cristofaro and E. Di Cera, *J. Protein Chem.*, 1991, **10**, 455–68.

Chapter 7

Conclusions and Future Work

7.1 Introduction

Within this thesis the use of anthracene incorporation into biomolecules has been assessed, utilising the photodimerisation and fluorescence properties that anthracene possess to monitor, and control, the biomolecules.

7.1.1 SNP Detection using an Anthracene Probe

In the first project presented herein, anthracene was used as a fluorescent reporter molecule to establish the base present on a complementary strand of DNA. Several linker lengths and stereochemistry combinations of the anthracene were used and the fluorescent properties of each was assessed in the probe system. A linker length of 5 carbons and L stereochemistry was shown to give the greatest differentiation between the A and C bases that needed to be distinguished in the complementary sequence. Further to this, the probe was shown to be able to quantitatively determine the ratio of the two bases on the complementary strand by the relative intensity of fluorescence. This is an advancement on the currently available methods as this assay is able to be performed at room and biological temperatures.

Future work should focus on reducing the amounts of material needed to perform the assay, making it more biologically relevant. This could be done by reducing the concentrations and volumes required, although studies presented herein have shown that this increases the error of the assay. Alternatively, an additional fluorophore could be incorporated which could form a FRET partner with the anthracene reporter.

In addition to SNP sensing, the base modification 8-oxoG was also shown to be distinguished by the anthracene probe. This was assessed in two sequences of DNA and the fluorescent output was shown to be highly dependent on the bases surrounding the fluorophore. Future work should address this with more detailed studies on the origin of this effect. The probe could also be developed into a quantitative system as with the SNP sensing.

7.1.2 Controlling DNA Binding using an Anthracene Tagged Peptide

A DNA binding, homodimer peptide was taken and the coiled coil dimerisation region removed and replaced with anthracene. Studies showed that this incorporation did not disrupt the DNA binding or selectivity of the peptide. The anthracene was used as a fluorescent output to assess the DNA binding of the system. Subsequent irradiation to form photodimer species showed that the photodimer only forms when target DNA is present. This is due to the peptide units binding in a specific fashion and pre-organising the anthracene moieties into close proximity. With the other DNA sequences, the anthracene moieties are too far apart to form photodimers. The formation of a covalently linked homodimer peptide was shown to increase the binding affinity for target DNA. This project is the first example of the use of anthracene photodimerisation within a protein structure.

Despite the positive outcomes of these project, several questions remain unanswered, such as why the system is not reversible and the products that are formed from extensive irradiation. Future work should focus on these areas potentially using a different amino acid sequence to monitor the effect of the individual amino acids on the process, such as methionine oxidation.

7.1.3 Photo-Triggered Release of DNA

The use of anthracene to force the release of a segment of single stranded DNA was also assessed. This started by focussing on a stem loop system with anthracene moieties incorporated on the ends of the sequence. This system was shown to be difficult to optimise and a new toe-hold system was used as an alternative. Within this system three strands of DNA were used that were in an equilibrium mixture. Upon irradiation, two of the strands which incorporated an anthracene unit, were "locked" together, shifting the equilibrium and releasing the third strand. Initial results from this project are promising with a degree of release shown by gel electrophoresis analysis.

Further optimisation of the toe-hold system is needed to quantify the release of the single stranded oligonucleotide. This information could then be used to help optimise the system and increase the amount of release.

7.1.4 Controlling DNA Folding using Anthracene Photodimerisation

The final project presented herein places an anthracene monomer at various locations on a G-quadruplex aptamer. Photodimerisation was shown to alter the folding of the systems, dependent upon where the anthracene incorporations were made. The systems were then studied to monitor the effect of the incorporation and subsequent dimerisation, on the binding of the aptamer to its target, thrombin. These studies were hindered by error and, although initially promising, the results cannot be used to draw any conclusions on the effects this incorporation has had on the binding.

Further studies should focus on assessing the binding of the aptamer utilising gel electromobility shift assays. Alternative DNA folds could also be analysed in order to develop the use of anthracene dimerisation to control DNA tertiary structure.

Chapter 8

Experimental

8.1 Materials

Potassium phosphate dibasic, potassium phosphate monobasic, sodium chloride, acrylamide, tetramethylethylenediamine (TEMED), ammonium persulphate (APS), tris base, hydrochloric acid 32%, sodium dodecyl sulphate (SDS), sodium hydroxide, magnesium chloride, potassium chloride, potassium carbonate, sodium carbonate, trifluoroacetic acid (TFA), glacial acetic acid, ethyl acetate, ethylenediaminetetraacetic acid (EDTA), methanol, acetone, dichloromethane, ethanol and triethyl amine (TEA) were all purchased from Fisher Scientific Ltd, Loughborough UK. Tricine, anthrone, and 4,4'-dimethyoxytrityl chloride was

obtained from Alfa Aesar, Massachusetts, USA; anisole, thioanisole and 1,2-ethanedithiol were purchased from Merck Millipore, Darmstadt, Germany. Sigma Aldrich Ltd., Missouri, USA were used to supply glycerol, Igpeal CA 630, bovine serum albumin, ethyl bromoacetate, Ethyl-4-bromobutyrate, Ethyl-5-bromovalerate, Ethyl-6-bromohexanoate, Ethyl-7-bromoheptanoate, magnesium sulphate, silica, L-threoninol, D-threoninol, dimethylformamide (DMF), alumina, solid supported benzyl alcohol, HBTU, DIPEA, 4-dimethylaminopyridine and 9-anthracenecarboxylic acid. SYBR Gold was sourced from Life Technologies Ltd., Paisley, UK; hexane from VWR, Lutterworth, UK and Coomassie brilliant blue stain R-250 protein stain from AppliChem GmbH, Darmstadt, Germany.

Fmoc-Arg(Pbf)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, HBTU, DIPEA and rink amide MBHA resin were purchased from AGTC Bioproducts Ltd. Fmoc-Ala-OH, Fmoc-Asn(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Lys(Boc)-OH, 20% piperidine in DMF and peptide synthesis grade DMF were purchased from Pepceuticals Ltd, Leicester, UK. Fmoc-Lys(Mtt)-OH was obtained from Merck Millipore, Darmstadt, Germany. All oligonucleotide synthesis reagents were purchased from Link Technologies Ltd., Bellshill, UK, including 2-cyanethyl-diisopropylchlorophosphoramidite, potassium carbonate in anhydrous methanol and TAMRA-dT-CE phosphoramidite.

Thrombin and Fibrinogen were obtained from Sigma Aldrich Ltd., Missouri, USA and Merck Millipore Darmstadt, Germany. An assay kit "Clauss Fibrinogen Assay" from Helena Laboratories (UK) Ltd. was also used.

Unless otherwise stated, all reagents were used as per the manufacturer's instructions without further purification. All water was ultrapure grade obtained from a Millipore Elix-Gradient A10 system (resistivity > 18 μΩ cm toc ≤ 5 ppb, Millipore, France).

^{1}H and ^{31}P NMR were performed on a Bruker AV300 spectrometer, ^{13}C NMR was performed on a Bruker AV400 spectrometer. MALDI-TOF mass spectrometry was performed on a Water Synapt G2-Si Mass Spectrometer. Electrospray-TOF mass spectrometry was performed on a Waters LCT Premier or a Water Synapt G2-Si Mass Spectrometer.

8.2 Anthracene Monomer Synthesis

The synthesis of anthracene phosphoramidites to allow for incorporation into oligonucleotides was performed as previously reported.[1] Herein follows the characterisation of anthracene linkers 1 and 3-6 with both L and D stereochemistries. 7L and 7D versions were synthesised by Jack Manchester.[2] A reaction scheme for these can be found in Chapter 2.2.1.

8.2.1 General Synthesis Procedure

Anthrone was refluxed in degassed acetone in the dark overnight with 1 molar equivalent of potassium carbonate and 1 molar equivalent of the bromoethyl ester of the desired length:

n=1	Ethyl bromoacetate
n=3	Ethyl-4-bromobutyrate
n=4	Ethyl-5-bromovalerate
n=5	Ethyl-6-bromohexanoate
n=6	Ethyl-7-bromoheptanoate

The solution was then filtered and acetone removed *in vacuo* to afford an orange oil. The solid was re-suspended in dichloromethane (DCM), washed with water, dried with magnesium sulphate and the solvent removed *in vacuo*. The crude product was then purified on silica using column chromatography in 10% hexane in DCM.

The ester product was then hydrolysed by refluxing overnight in the dark in a 1:1 degassed solution of 10% sodium hydroxide and ethanol to produce the carboxylic acid. The ethanol was then removed *in vacuo* and the solution diluted to 3 times the volume with water. Concentrated hydrochloric acid was added dropwise to produce a cream precipitate which was collected by suction filtration.

The stereochemistry was introduced using L- or D-threoninol. The carboxylic acid was dissolved in degassed dimethylforamide (DMF) and 1 molar equivalent of O-

(benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HBTU) was added and the solution stirred for 15 minutes. 1 molar equivalent of threoninol with 1 molar equivalent of N,N-diisopropylethylamine (DIPEA) was then added and the reaction stirred at 40ºC for 40 hours. The solution was diluted with methanol and DCM and washed with water before the solvent was removed under vacuum. The crude product was column purified on silica using a gradient of 1-5% methanol in DCM to afford an orange oil.

The primary alcohol of the threoninol was then protected with a dimethyloxytrityl (DMT) group. The threoninol derivatised anthracene was dissolved in pyridine and 1 molar equivalent of dimethoxytrityl chloride was added with 0.15 molar equivalents of 4-dimethylaminopyridine. The solution was stirred in the dark overnight under argon. Water was added to dilute the solution and the product was extracted with DCM and dried over magnesium sulphate. The crude product was purified using a silica column with 40% hexane, 59% ethyl acetate and 1% triethylamine as an eluent.

Finally the phosphoramidite group was added by dissolving the DMT protected alcohol in dry DCM with 1 molar equivalent DIPEA and 1 molar equivalent 2-cyanethyl-diisopropylchlorophosphoramidite and stirring for 2 hours. 1 molar equivalent of solid supported benzyl alcohol was then added and stirred for a further hour. The solution was filtered and rinsed with ethyl acetate, washed with

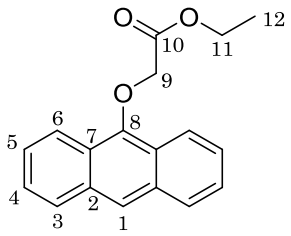
sodium carbonate, water and brine before being dried with magnesium sulphate. The solvent was removed *in vacuo* and the crude product purified on activated alumina using an eluent of 40% hexane, 59% ethyl acetate and 1% triethylamine.

The below compounds have been previously synthesised.[1,2]

8.2.2. Characterisation of n=1L and 1D

8.2.2.1. n=1 Ester

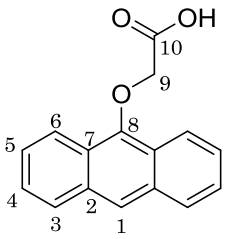
Ethyl 2-(anthracen-9-yloxy)acetate



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.36 (d, *J* = 7.7 Hz, 2H, H_6), 8.27 (s, 1H, H_1), 8.01 (d, *J* = 7.5 Hz, 2H, H_3), 7.53 – 7.46 (m, 4H, H_4 and H_5), 4.81 (s, 2H, H_9), 4.38 (q, *J* = 7.1 Hz, 2H, H_{11}), 1.38 (t, *J* = 7.2 Hz, 3H, H_{12}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 169.1 (C_{10}), 149.5 (C_8), 132.3 (C_2), 128.4 (C_3), 125.6 (C_5), 125.6 (C_4), 124.3 (C_7), 123.1 (C_1), 122.2 (C_6), 72.1 (C_9), 61.4 (C_{11}), 14.3 (C_{12}); m/z (ES+) calculated for $C_{18}H_{16}O_3Na$ 303, found 303. Yield: 28%, R_f = 0.59 in CH_2Cl_2.

8.2.2.2. n=1 Carboxylic Acid

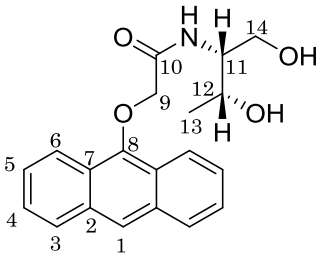
2-(Anthracen-9-yloxy)acetic acid



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.30 (s, 1H, H_1), 8.26 (d, *J* = 8.1 Hz, 2H, H_6), 8.02 (d, *J* = 8.2, 2H, H_3), 7.56 – 7.45 (m, 4H, H_4 and H_5), 4.85 (s, 1H, H_9); ^{13}C (100 MHz, Chloroform-*d*) δ 172.4 (C_{10}), 149.5 (C_8), 132.3 (C_2), 128.6 (C_3), 126.0 (C_5), 125.7 (C_4), 124.2 (C_7), 123.5 (C_1), 121.7 (C_6), 71.3 (C_9); m/z (ES+) calculated for $C_{16}H_{12}O_3Na$ 275, found 275. Yield: 90%.

8.2.2.3. n=1L Diol

2-(Anthracen-9-yloxy)-N-((2R,3R)-1,3-dihydroxybutan-2-yl)acetamide

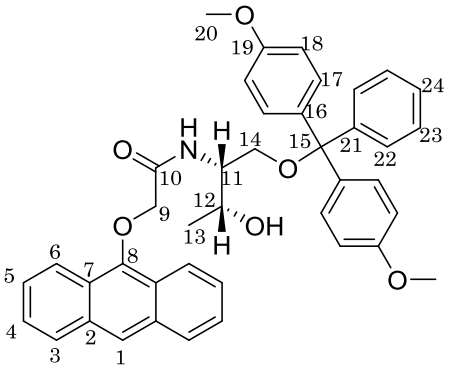


^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.29 (s, 1H, H_1), 8.22 – 8.17 (m, 2H, H_6), 8.05 – 7.98 (m, 2H, H_3), 8.02 (s, 1H, N*H*), 7.49 (m, 4H, H_5 and H_4), 4.75 (s, 2H, H_9), 4.42

– 4.33 (m, 1H, H_{12}), 4.21 – 3.90 (m, 3H, H_{11} and H_{14}), 1.42 (d, J = 6.4 Hz, 3H, H_{13}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 169.5 (C_{10}), 162.7 (C_8), 132.3 (C_2), 128.7 (C_3) 125.8 (C_5), 125.7 (C_4), 124.2 (C_7), 123.4 (C_1), 121.4 (C_6), 73.6 (C_9), 69.0 (C_{12}), 65.2 (C_{14}), 54.8 (C_{11}), 20.9 (C_{13}); m/z (ES+) calculated for $C_{20}H_{21}NO_4Na$ 362, found 362. Yield: 28%, R_f = 0.57 in CH_2Cl_2 with 10% MeOH.

8.2.2.4. n=1L DMT Protected

2-(Anthracen-9-yloxy)-N-((2R,3R)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl)acetamide

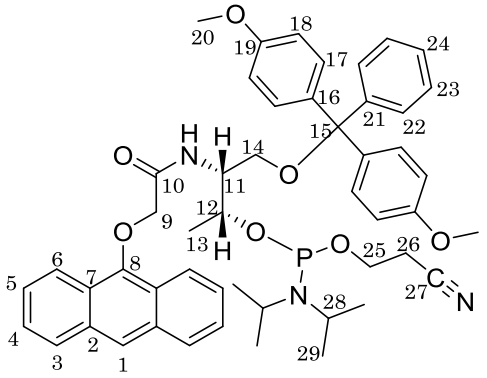


^{1}H NMR (300 MHz, Acetonitrile-*d*$_3$) δ 8.30 (s, 1H, H_1), 8.23 – 8.10 (m, 2H, H_6), 7.99 (d, J = 8.5 Hz, 2H, H_3), 7.79 (d, J = 5.8 Hz, 1H, N*H*), 7.46 – 7.41 (m, 4H, H_4 and H_{22}), 7.36 – 7.28 (m, 2H, H_5), 7.25 – 7.12 (m, 6H, H_{17} and H_{23}), 7.09 – 6.96 (m, 1H, H_{24}), 6.83 – 6.64 (m, 4H, H_{18}), 4.61 (s, 2H, H_9), 4.14 – 3.96 (m, 2H, H_{11} and H_{12}), 3.68 (s, 6H, H_{20}), 3.33 – 3.04 (m, 2H, H_{14}), 1.12 (d, J = 5.0 Hz, 3H, H_{13}); ^{13}C NMR (100 MHz, Acetonitrile-*d*$_3$) δ 158.5 (C_{10}), 145.2 (C_{19}), 139.9 (C_8), 136.1 (C_{21}), 129.7 (C_{16}), 128.5 (C_2), 128.0 (C_{17}), 127.8 (C_3), 127.6 (C_{22}), 126.8 (C_{23}), 126.7 (C_{24}), 126.0

(C_5), 125.8 (C_4), 123.2 (C_7) 121.6 (C_1), 117.5 (C_6), 112.8 (C_{18}), 96.3 (C_{15}), 74.7 (C_9), 66.2 (C_{12}), 63.4 (C_{14}), 54.9 (C_{20}), 51.1 (C_{11}), 19.8 (C_{13}); m/z (ES+) calculated for $C_{41}H_{39}NO_6Na$ 664, found 664. Yield: 50%, R_f = 0.39 in CH_2Cl_2 with 5% MeOH.

8.2.2.5. n=1L Phosphoramidite

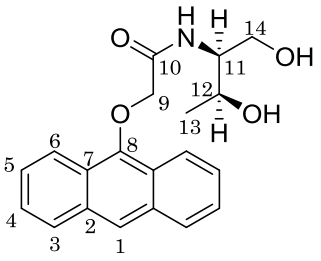
(2R,3R)-3-(2-(Anthracen-9-yloxy)acetamido)-4-(bis (4-methoxyphenyl) (phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.31 (s, 1H H_1), 8.22 – 8.18 (m, 2H, H_6), 8.00 (d, J = 8.5 Hz, 2H, H_3), 7.80 (d, J = 5.8 Hz, 1H, N*H*), 7.48 – 7.38 (m, 4H, H_4 and H_{22}), 7.38 – 7.28 (m, 2H, H_5), 7.24 – 7.19 (m, 6H, H_{17} and H_{23}), 7.09 – 7.01 (m, 1H, H_{24}), 6.78 (d, J= 2.2 Hz, 4H, H_{18}), 4.59 – 4.67 (m, 2H, H_{11} and H_{12}), 4.07 (s, 2H, H_9), 3.68 (s, 6H, H_{20}), 3.67 – 3.10 (m, 6H, H_{14}, H_{25} and H_{28}), 2.88 – 2.34 (m, 2H, H_{26}), 1.28 – 0.90 (m, 12H, H_{29}), 0.71 (d, J = 6.8 Hz, 3H, H_{13}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 148.4, 146.0; m/z (ES+) calculated for $C_{50}H_{56}N_3O_7PNa$ 864, found 864. Yield: 65%, R_f = 0.74 in 50% hexane, 50% ethyl acetate.

8.2.2.6. n=1D Diol

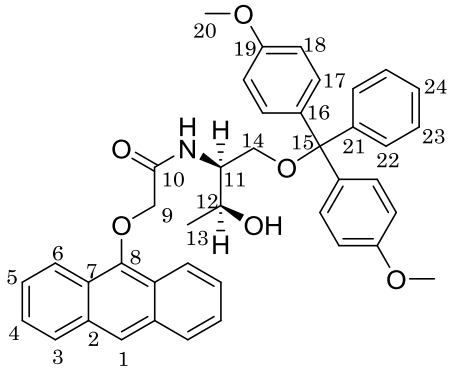
2-(Anthracen-9-yloxy)-N-((2S,3S)-1,3-dihydroxybutan-2-yl)acetamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.28 (s, 1H, H_1), 8.21 – 8.06 (m, 2H, H_6), 8.05 – 7.96 (m, 2H, H_3), 7.75 (d, *J* = 3.9 Hz, 1H, N*H*), 7.62 – 7.42 (m, 4H, H_5 and H_4), 4.73 (s, 2H, H_9), 4.84 – 4.55 (m, 1H, H_{12}), 4.37 (dd, *J* = 6.4, 2.4 Hz, 1H, H_{11}), 4.05 (d, *J* = 3.5 Hz, 2H, H_{14}), 1.40 (d, *J* = 6.4 Hz, 3H, H_{13}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 169.3 (C_{10}), 162.7 (C_8), 132.4 (C_2), 128.7 (C_3) 125.8 (C_5), 125.7 (C_4), 124.2 (C_7), 123.4 (C_1), 121.4 (C_6), 73.6 (C_9), 68.8 (C_{12}), 65.2 (C_{14}), 54.8 (C_{11}), 21.1 (C_{13}); m/z (ES+) calculated for $C_{20}H_{21}NO_4Na$ 362, found 362. Yield: 40%, R_f = 0.59 in CH_2Cl_2 with 10% MeOH.

8.2.2.7. n=1D DMT Protected

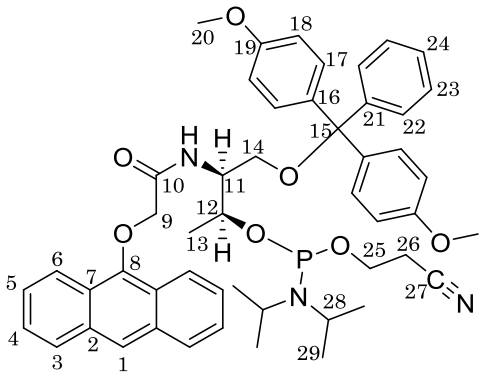
2-(Anthracen-9-yloxy)-N-((2S,3S)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl)acetamide



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.37 (s, 1H, H_1), 8.21 – 8.27 (m, 2H, H_6), 8.06 (d J = 8.4 Hz, 2H, H_3), 7.50 (d, J = 6.9 Hz, 1H, N*H*), 7.48 – 7.40 (m, 2H, H_4), 7.37 (d, J = 8.9 Hz, 2H, H_{22}), 7.32 – 7.25 (m, 2H, H_5), 7.25 – 7.15 (m, 6H, H_{17} and H_{23}), 7.13 – 7.08 (m, 1H, H_{24}), 6.80 – 6.87 (m, 4H, H_{18}), 4.68 (s, 2H, H_9), 4.36 – 3.91 (m, 2H, H_{11} and H_{12}), 3.71 (s, 6H, H_{20}), 3.25 – 2.71 (m, 2H, $H_{14)}$, 1.19 (d, J = 6.2 Hz, 3H, H_{13}); ^{13}C NMR (100 MHz, Acetonitrile-d_3) δ 158.6 (C_{10}), 145.2 (C_{19}), 139.9 (C_8), 132.1 (C_{21}), 130.0 (C_{16}), 128.5 (C_2), 128.0 (C_{17}), 127.8 (C_3 andC_{22}), 126.9 (C_{23} and C_{24}), 126.0 (C_5), 125.8 (C_4), 123.2 (C_7) 121.7 (C_1), 117.5 (C_6), 113.1 (C_{18}), 96.3 (C_{15}), 74.7 (C_9), 66.2 (C_{12}), 63.4 (C_{14}), 54.9 (C_{20}), 51.1 (C_{11}), 19.8 (C_{13}); m/z (ES+) calculated for $C_{41}H_{39}NO_6Na$ 664, found 664. Yield: 48%, R_f = 0.39 in CH_2Cl_2 with 5% MeOH.

8.2.2.8. n=1D Phosphoramidite

(2S,3S)-3-(2-(Anthracen-9-yloxy)acetamido)-4-(bis(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite

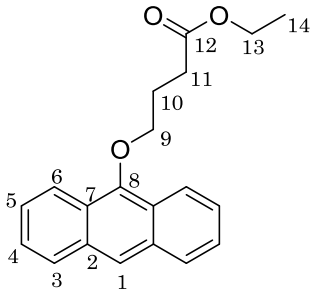


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.42 (s, 1H, H_1), 8.24 (d, J = 8.7 Hz, 2H, H_6), 8.11 (d, J = 8.4 Hz, 2H, H_3), 7.59 – 7.47 (m, 5H, N*H*, H_4 and H_{22}), 7.47 – 7.38 (m, 6H, H_5 and H_{17}), 7.37 – 7.29 (m, 2H, H_{23}), 7.28 – 7.21 (m, 1H, H_{24}), 6.93 – 6.85 (m, 4H, H_{18}), 4.84 – 4.64 (m, 2H, H_9), 4.45 – 4.25 (m, 2H, H_{11} and H_{12}), 3.79 – 3.71 (m, 6H, H_{20}), 3.70 – 3.20 (m, 6H, H_{14}, H_{25} and H_{28}), 2.89 – 2.62 (m, 2H, H_{26}), 1.45 – 0.95 (m, 12H, H_{29}), 0.81 (d, J = 6.8 Hz, 3H, H_{13}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.5, 144.5; m/z (ES+) calculated for $C_{50}H_{56}N_3O_7PNa$ 864, found 864. Yield: 30%, R_f = 0.61 in 50% hexane, 50% ethyl acetate.

8.2.3. Characterisation of n=3L and 3D

8.2.3.1. n=3 Ester

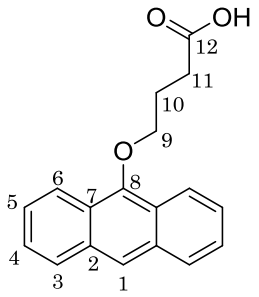
Ethyl 4-(anthracen-9-yloxy)butanoate



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.31 – 8.14 (m, 3H, H_1 and H_6), 7.99 (d, *J* = 9.7 Hz, 2H, H_3), 7.56 – 7.40 (m, 4H, H_4 and H_5), 4.26 – 4.16 (m, 4H, H_{13} and H_9), 2.79 (t, *J* = 7.4 Hz, 2H, H_{11}), 2.48 – 2.33 (m, 2H, H_{10}), 1.31 (t, *J* = 7.1 Hz, 3H, H_{14}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.3 (C_{12}), 151.0 (C_8), 132.4 (C_2), 128.5 (C_3), 125.5 (C_5), 125.2 (C_4), 124.7 (C_7), 122.2 (C_1), 122.2 (C_6), 74.6 (C_9), 60.6 (C_{13}), 31.2 (C_{11}), 26.1 (C_{10}), 14.3 (C_{14}); m/z (ES+) calculated for $C_{20}H_{20}O_3Na$ 331, found 332. Yield: 53%, R_f = 0.48 in hexane with 10% ethyl acetate.

8.2.3.2. n=3 Carboxylic Acid

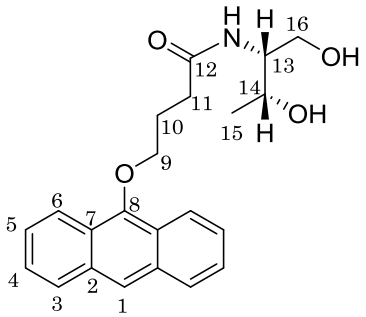
4-(Anthracen-9-yloxy)butanoic acid



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.36 – 8.13 (m, 3H, H_1 and H_6), 8.03 – 7.96 (m, 2H, H_3), 7.63 – 7.37 (m, 4H, H_4 and H_5), 4.25 (t, *J* = 6.2 Hz, 2H, H_9), 2.88 (t, *J* = 7.4 Hz, 2H, H_{11}), 2.52 – 2.28 (m, 2H, H_{10});); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 178.8 (C_{12}), 150.9 (C_8), 132.4 (C_2), 128.5 (C_3), 125.5 (C_5), 125.3 (C_4), 124.6 (C_7), 122.3 (C_1), 122.2 (C_6), 74.3 (C_9), 30.8 (C_{11}), 25.7 (C_{10}); m/z (ES+) calculated for $C_{18}H_{16}O_3Na$ 303, found 303. Yield: 93%.

8.2.3.3. n=3L Diol

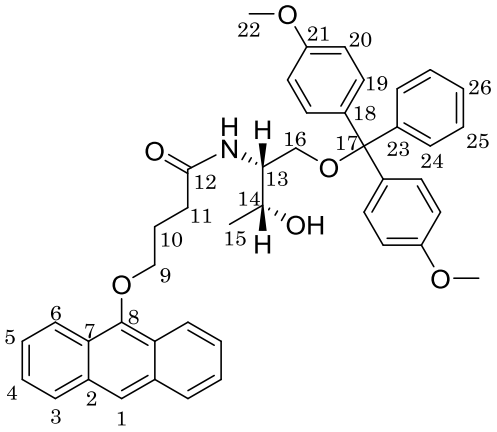
4-(Anthracen-9-yloxy)-N-((2R,3R)-1,3-dihydroxybutan-2-yl)butanamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.38 – 8.11 (m, 3H, H_1 and H_6), 8.05 – 7.88 (m, 2H, H_3), 7.57 – 7.34 (m, 4H, H_4 and H_5), 6.67 (d, *J* = 8.4 Hz, 1H, N*H*), 4.37 – 4.04 (m, 3H, H_{14} and H_9), 3.94 – 3.86 (m, 1H, H_{13}), 3.83 (d, *J* = 3.6 Hz, 2H, H_{16}), 2.73 (t, *J* = 8.4 Hz, 2H, H_{11}), 2.46 – 2.22 (m, 2H, H_{10}), 1.20 (d, *J* = 6.3 Hz, 3H, H_{15}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.6 (C_{12}), 150.9 (C_8), 132.4 (C_2), 128.5 (C_3) 125.9 (C_5), 125.5 (C_4), 125.3 (C_7), 122.2 (C_1), 122.2 (C_6), 74.9 (C_9), 68.8 (C_{14}), 65.0 (C_{16}), 54.8 (C_{13}), 31.5 (C_{11}), 26.7 (C_{10}), 20.5(C_{15}); m/z (ES+) calculated for $C_{22}H_{25}NO_4Na$ 390, found 390. Yield: 73%, R_f = 0.25 in CH_2Cl_2 with 5% MeOH.

8.2.3.4. n=3L DMT Protected

4-(Anthracen-9-yloxy)-N-((2R,3R)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) butanamide

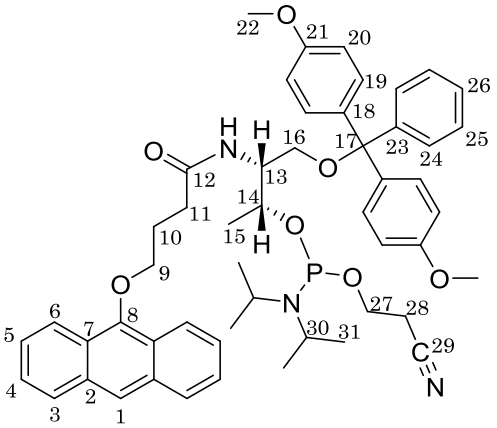


^{1}H NMR (300 MHz, Acetonitrile-*d*$_3$) δ 8.36 – 8.21 (m, 3H, H_1 and H_6), 8.14 – 7.91 (m, 2H, H_3), 7.44 – 7.34 (m, 4H, H_4 and H_{24}), 7.24 – 7.13 (m, 9H, H_5, H_{19}, H_{25} and H_{26}), 6.93 – 6.68 (m, 4H, H_{20}), 4.30 – 4.14 (m, 2H, H_9), 4.10 – 3.97 (m, 2H, H_{14} and

H_{13}), 3.70 (s, 6H, H_{22}), 3.14 (d, J = 9.1 Hz, 2H, H_{16}), 2.61 (t, J = 7.3 Hz, 2H, H_{11}), 2.31 (m, 2H, H_{10}), 1.05 (d, J = 6.3 Hz, 3H, H_{15}); ^{13}C NMR (100 MHz, Acetonitrile-d_3) δ 173.5 (C_{12}), 159.4 (C_{21}) 149.9 (C_8), 146.3 (C_{23}), 133.0 (C_{18}), 132.4 (C_2), 130.8 (C_{19}), 128.7 (C_3), 128.9 (C_{24}), 128.7 (C_{25}), 127.6 (C_{26}), 125.8 (C_5), 125.5 (C_4), 125.3 (C_7), 122.3 (C_1), 122.2 (C_6), 113.0 (C_{20}), 87.1 (C_{17}) 75.2 (C_9), 66.2 (C_{14}), 63.5 (C_{16}), 55.9 (C_{22}), 54.8 (C_{13}), 31.4 (C_{11}), 26.4 (C_{10}), 19.6 (C_{15}); m/z (ES+) calculated for $C_{43}H_{43}NO_6Na$ 693, found 693. Yield: 35%, R_f = 0.48 in CH_2Cl_2 with 5% MeOH.

8.2.3.5. n=3L Phosphoramidite

(2R,3R)-3-(4-(Anthracen-9-yloxy)butanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite

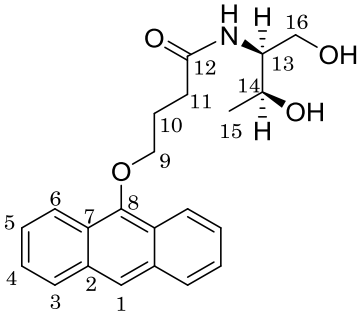


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.34 (s, 1H, H_1), 8.30 (d, J = 8.4 Hz, 2H, H_6), 8.16 – 8.00 (m, 2H, H_3), 7.57 – 7.38 (m, 4H, H_4 and H_{24}), 7.37 – 7.24 (m, 6H, H_5 and H_{19}), 6.91 – 6.75 (m, 3H, H_{25} and H_{26}), 6.39 (d, J = 9.1 Hz, 4H, H_{20}), 6.32 (d, J = 9.3

Hz, 1H, N*H*), 4.30 – 4.15 (m, 2H, H_9), 3.73 (s, 6H, H_{22}), 3.79 – 3.70 (m, 2H, H_{13} and H_{14}), 3.59 – 3.43 (m, 4H, H_{27} and H_{30}), 3.20 – 3.05 (m, 2H, H_{16}), 2.70 – 2.55 (m, 4H, H_{11} and H_{28}), 2.47 (p, J = 5.9 Hz, 2H, H_{10}), 1.35 – 1.12 (m, 12H, H_{31}), 0.99 (d, J = 6.8 Hz, 3H, H_{15}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.6, 147.0; m/z (ES+) calculated for $C_{52}H_{60}N_3O_7PNa$ 893, found 893. Yield: 38%, R_f = 0.66 in 50% hexane, 50% ethyl acetate.

8.2.3.6. n=3D Diol

4-(Anthracen-9-yloxy)-N-((2S,3S)-1,3-dihydroxybutan-2-yl)butanamide

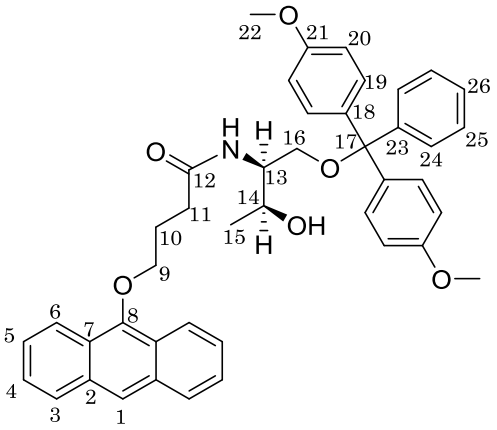


^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.28 – 8.20 (m, 2H, H_6), 8.22 (s, 1H, H_1), 8.03 – 7.95 (m, 2H, H_3), 7.53 – 7.4. (m, 4H, H_4 and H_5), 6.45 (d, J = 10.5 Hz, 1H, N*H*), 4.28 – 4.15 (m, 3H, H_9 and H_{14}), 3.95 – 3.84 (m, 1H, H_{13}), 3.86 (d, J = 4.5 Hz, 2H, H_{16}), 2.80 – 2.72 (m, 2H, H_{11}), 2.48 – 2.36 (m, 2H, H_{10}), 1.22 (d, J = 6.4 Hz, 3H, H_{15}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.4 (C_{12}), 150.9 (C_8), 132.4 (C_2), 128.5 (C_3)

125.5 (C_5), 125.3 (C_4), 124.6 (C_7), 122.3 (C_1), 122.1 (C_6), 74.9 (C_9), 69.3 (C_{14}), 65.4 (C_{16}), 54.6 (C_{13}), 33.6 (C_{11}), 26.7 (C_{10}), 20.7 (C_{15}); m/z (ES+) calculated for $C_{22}H_{25}NO_4Na$ 390, found 390. Yield: 78%, R_f = 0.25 in CH_2Cl_2 with 5% MeOH.

8.2.3.7. n=3D DMT Protected

4-(Anthracen-9-yloxy)-N-((2S,3S)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) butanamide

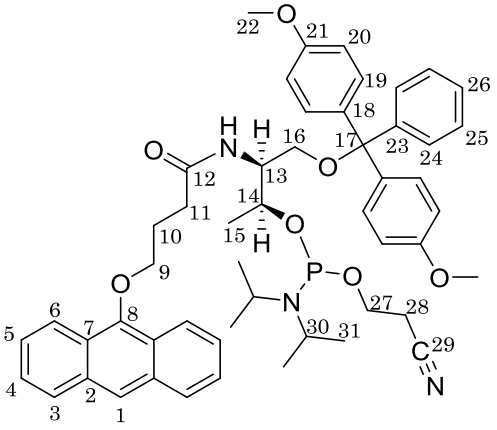


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.34 (s, 1H, H_1), 8.34 – 8.23 (m, 2H, H_6), 8.18 – 8.02 (m, 2H, H_3), 7.47 – 7.37 (m, 4H, H_4 and H_{24}), 7.31 – 7.26 (m, 6H, H_5 and H_{19}), 7.25 – 7.09 (m, 3H, H_{25} and H_{26}), 6.85 – 6.76 (m, 4H, H_{20}), 6.54 (d, J = 8.9 Hz, 1H, N*H*), 4.21 (d, J = 6.4 Hz, 2H, H_9), 4.06 – 3.89 (m, 1H, H_{14}), 3.83 – 3.75 (m, 1H, H_{13}), 3.73 (s, 6H, H_{22}), 3.17 (d, J = 9.1, 2H, H_{16}), 2.64 (t, J = 7.3 Hz, 2H, H_{11}), 2.39 – 2.28 (m, 2H, H_{10}), 1.07 (d, J = 6.3 Hz, 3H, H_{15}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.5 (C_{12}), 159.4 (C_{21}) 149.9 (C_8), 146.3 (C_{23}), 133.0 (C_{18}), 132.4 (C_2), 129.7 (C_{19}), 128.7 (C_3), 128.9 (C_{24}), 128.7 (C_{25}), 127.6 (C_{26}), 125.8 (C_5), 125.5 (C_4), 125.3 (C_7),

122.3 (C_1), 122.2 (C_6), 112.7 (C_{20}), 87.1 (C_{17}) 75.0 (C_9), 66.3 (C_{14}), 63.2 (C_{16}), 55.9 (C_{22}), 54.3 (C_{13}), 31.4 (C_{11}), 26.4 (C_{10}), 19.6 (C_{15}); m/z (ES+) calculated for $C_{43}H_{43}NO_6Na$ 693, found 693. Yield: 52%, R_f = 0.53 in CH_2Cl_2 with 5% MeOH.

8.2.3.8. n=3D Phosphoramidite

(2S,3S)-3-(4-(Anthracen-9-yloxy)butanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite



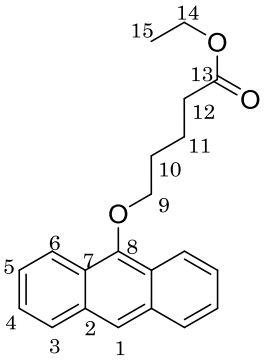
^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.42 (s, 1H, H_1), 8.24 (d, J = 8.7 Hz, 2H, H_6), 8.11 (d, J = 8.4 Hz, 2H, H_3), 7.59 – 7.38 (m, 4H, H_4 and H_{24}), 7.38 – 7.29 (m, 6H, H_5 and H_{19}), 7.28 – 7.23 (m, 3H, H_{25} and H_{26}), 6.99 – 6.81 (m, 5H, H_{20} and N*H*), 4.84 – 4.21 (m, 2H, H_9), 3.80 – 3.70 (m, 2H, H_{13} and H_{14}), 3.74 (s, 6H, H_{22}), 3.57 – 3.29 (m, 4H, H_{27} and H_{30}), 3.28 – 3.19 (m, 2H, H_{16}), 2.89 – 2.58 (m, 4H, H_{11} and H_{28}), 2.47 – 2.40 (m, 2H, H_{10}), 1.45 – 0.95 (m, 12H, H_{31}), 0.81 (d, J = 6.8 Hz, 3H, H_{15}); ^{31}P NMR

(121 MHz, Acetonitrile-d_3) δ 147.6, 147.0; m/z (ES+) calculated for $C_{52}H_{60}N_3O_7PNa$ 893, found 893. Yield: 52%, R_f = 0.67 in 50% hexane, 50% ethyl acetate.

8.2.4. Characterisation of n=4L and 4D

8.2.4.1. n=4 Ester

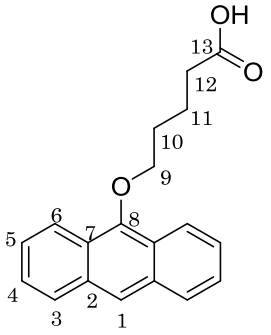
Ethyl 5-(anthracen-9-yloxy)pentanoate



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.45 – 8.20 (m, 3H, H_1 and H_6), 8.02 – 7.95 (m, 2H, H_3), 7.53 – 7.41 (m, 4H, H_4 and H_5), 4.50 – 4.00 (m, 4H, H_9 and H_{14}), 2.50 (t, *J* = 7.0 Hz, 2H, H_{12}), 2.18 – 1.99 (m, 4H, H_{10} and H_{11}), 1.29 (t, *J* = 7.2 Hz, 3H, H_{15}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.5 (C_{13}), 151.3 (C_8), 132.8 (C_2), 128.5 (C_3), 125.5 (C_5), 125.1 (C_4), 124.7 (C_7), 122.4 (C_1), 122.1 (C_6), 75.5 (C_9), 60.4 (C_{14}), 34.2 (C_{12}), 30.1 (C_{10}), 21.9 (C_{11}), 14.3 (C_{15}); m/z (ES+) calculated for $C_{21}H_{22}O_3$ 322, found 322. Yield: 42%, R_f = 0.56 in CH_2Cl_2.

8.2.4.2. n=4 Carboxylic Acid

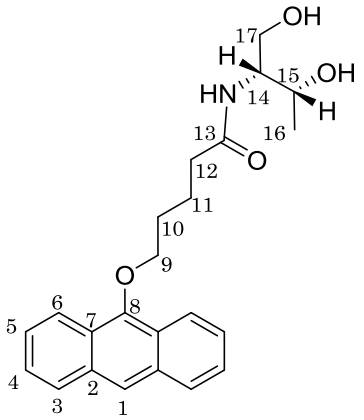
5-(Anthracen-9-yloxy)pentanoic acid



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.40 – 8.11 (m, 3H, H_1 and H_6), 8.06 – 7.93 (m, 2H, H_3), 7.55 – 7.33 (m, 4H, H_4 and H_5), 4.22 (t, J = 6.0 Hz, 2H, H_9), 2.59 (t, J= 6.0 Hz, 2H, H_{12}), 2.18 – 1.99 (m, 4H, H_{10} and H_{11}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 179.1 (C_{13}), 151.2 (C_8), 132.4 (C_2), 128.5 (C_3), 125.5 (C_5), 125.2 (C_4), 124.7 (C_7), 122.3 (C_1), 122.2 (C_6), 75.4 (C_9), 33.8 (C_{12}), 30.1 (C_{10}), 21.6 (C_{11}); m/z (ES+) calculated for $C_{19}H_{18}O_3Na$ 317, found 317. Yield: 85%.

8.2.4.3. n=4L Diol

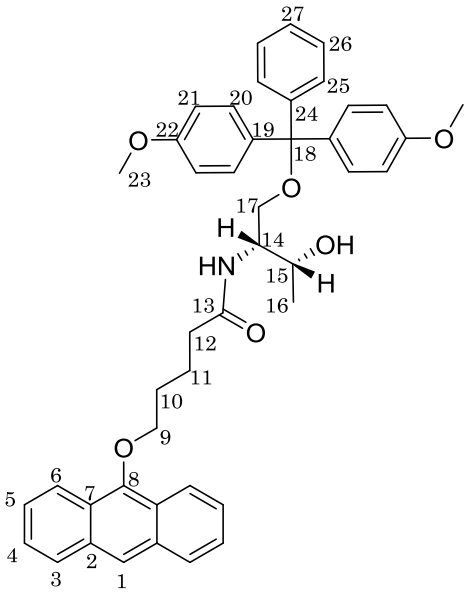
5-(Anthracen-9-yloxy)-N-((2R,3R)-1,3-dihydroxybutan-2-yl)pentanamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.39 – 8.11 (m, 3H, H_1 and H_6), 8.07 – 7.83 (m, 2H, H_3), 7.70 – 7.35 (m, 4H, H_4 and H_5), 6.41 (d, *J* = 8.3 Hz, 1H, N*H*), 4.23 – 4.13 (m, 3H, H_{15} and H_9), 3.93 – 3.66 (m, 3H, H_{14} and H_{17}), 2.73 (t, *J* = 8.4 Hz, 2H, H_{12}), 2.48 – 2.40 (m, 2H, H_{11}), 2.12 – 2.03 (m, 2H, H_{10}), 1.20 (d, *J* = 6.3 Hz, 3H, H_{16}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.7 (C_{13}), 151.2 (C_8), 132.4 (C_2), 128.5 (C_3) 125.5 (C_5), 125.2 (C_4), 124.7 (C_7), 122.3 (C_1), 122.1 (C_6), 75.6 (C_9), 68.9 (C_{15}), 65.2 (C_{17}), 54.6 (C_{14}), 36.5 (C_{12}), 31.5 (C_{10}), 22.7 (C_{11}), 20.6(C_{16}); m/z (ES+) calculated for $C_{23}H_{27}NO_4Na$ 404, found 404. Yield: 49%, R_f = 0.47 in CH_2Cl_2 with 5% MeOH.

8.2.4.4. n=4L DMT Protected

5-(Anthracen-9-yloxy)-N-((2R,3R)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) pentanamide

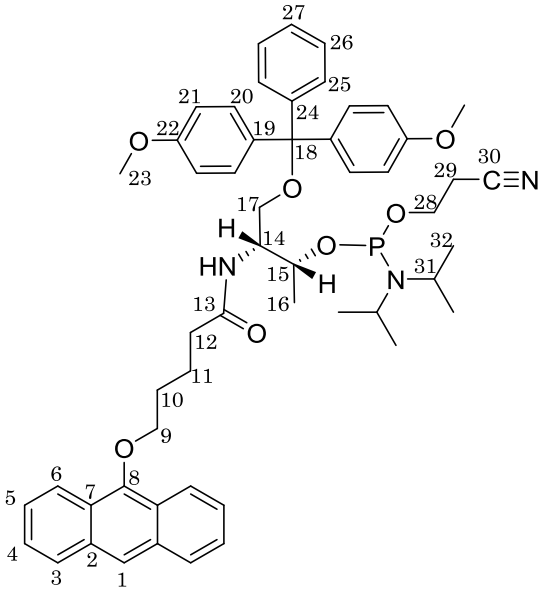


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.36 – 8.20 (m, 3H, H_1 and H_6), 8.02 – 7.85 (m, 2H, H_3), 7.48 – 7.35 (m, 4H, H_4 and H_{25}), 7.34 – 7.15 (m, 9H, H_5, H_{20}, H_{26} and H_{27}), 6.84 – 6.78 (m, 4H, H_{21}), 6.16 (d, J = 9.0 Hz, 1H, N*H*), 4.30 – 4.10 (m, 2H, H_9), 4.03 – 3.94 (m, 2H, H_{15} and H_{14}), 3.70 (s, 6H, H_{23}), 3.14 (d, J = 9.1 Hz, 2H, H_{17}), 2.43 (t, J = 6.0 Hz, 2H, H_{12}), 2.17 – 2.01 (m, 4H, H_{10} and H_{11}), 1.15 (d, J = 6.3 Hz, 3H, H_{16}); ^{13}C NMR (100 MHz, Acetonitrile-d_3) δ 173.0 (C_{13}), 158.7 (C_{22}) 144.3 (C_8), 144.2 (C_{24}), 135.5 (C_{19}), 135.3 (C_2), 132.2 (C_{20}), 129.9 (C_3), 128.4 (C_{25}), 128.1 (C_{26}), 127.9 (C_{27}), 125.5 (C_5), 125.2 (C_4), 124.7 (C_7), 122.3 (C_1), 122.1 (C_6), 113.0 (C_{21}), 87.3 (C_{18}) 75.6 (C_9), 68.9 (C_{15}), 65.5 (C_{17}), 55.2 (C_{23}), 53.3 (C_{14}), 36.7 (C_{12}), 31.4 (C_{10}), 26.4 (C_{11}),

20.0 (C_{16}); m/z (ES+) calculated for $C_{44}H_{44}NO_6Na$ 706, found 706. Yield: 38%, R_f = 0.50 in CH_2Cl_2 with 5% MeOH.

8.2.4.5. n=4L Phosphoramidite

(2R,3R)-3-(5-(Anthracen-9-yloxy)pentanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite

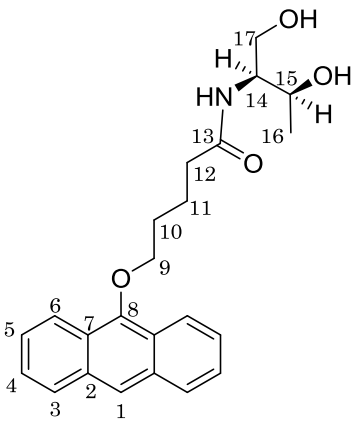


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.17 – 8.11 (m, 3H, 8.14, H_1 and H_6), 7.95 – 7.84 (m, 2H, H_3), 7.40 – 7.30 (m, 10H, H_4, H_5, H_{20}, and H_{25}), 7.19 (m, 3H, H_{26} and H_{27}), 6.78 – 6.62 (m, 4H, H_{21}), 6.20 (d, J = 8.9 Hz, 1H, N*H*), 4.05 (t, J = 5.9 Hz, 2H, H_9), 3.60 – 3.55 (m, 2H, H_{14} and H_{15}), 3.58 (s, 6H, H_{23}), 3.48 – 3.19 (m, 4H, H_{28} and H_{31}), 3.11 – 2.97 (m, 2H, H_{17}), 2.60 – 2.53 (m, 2H, H_{29}), 2.47 (t, J = 5.9 Hz, 2H, H_{12}),

2.32 – 2.24 (m, 4H, H_{10} and H_{11}), 1.12 – 0.94 (m, 12H, H_{32}), 0.86 (d, *J* = 6.8 Hz, 3H, H_{16}). ^{31}P NMR (121 MHz, Acetonitrile-*d*$_3$) δ 147.5, 147.0; m/z (ES+) calculated for $C_{53}H_{62}N_3O_7PNa$ 906, found 906. Yield: 61%, R_f = 0.70 in 50% hexane, 50% ethyl acetate.

8.2.4.6. n=4D Diol

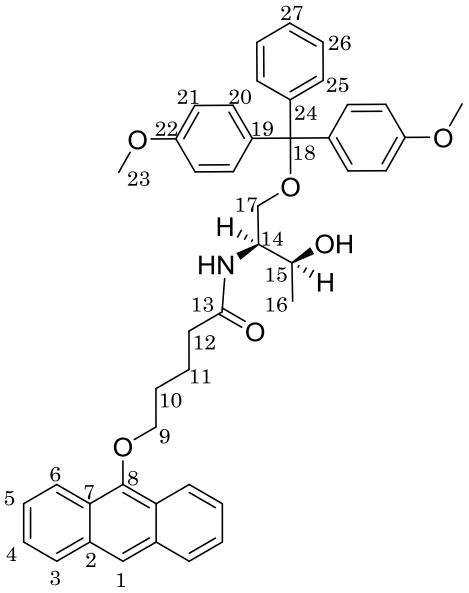
5-(Anthracen-9-yloxy)-N-((2S,3S)-1,3-dihydroxybutan-2-yl)pentanamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.35 – 8.13 (m, 3H, H_1 and H_6), 8.07 – 7.85 (m, 2H, H_3), 7.54 – 7.33 (m, 4H, H_4 and H_5), 6.37 (d, *J* = 8.3 Hz, 1H, N*H*), 4.37 – 4.06 (m, 3H, H_{15} and H_9), 3.95 – 3.69 (m, 3H, H_{14} and H_{17}), 2.87 (t, *J* = 8.4 Hz, 2H, H_{12}), 2.53 – 2.32 (m, 2H, H_{11}), 2.26 – 1.96 (m, 2H, H_{10}), 1.21 (d, *J* = 6.4 Hz, 3H, H_{16}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.6 (C_{13}), 151.2 (C_8), 132.4 (C_2), 128.5 (C_3) 125.5 (C_5), 125.2 (C_4), 124.7 (C_7), 122.3 (C_1), 122.1 (C_6), 75.6 (C_9), 69.1 (C_{15}), 65.4 (C_{17}), 54.5 (C_{14}), 36.6 (C_{12}) 30.1 (C_{10}), 22.7 (C_{11}), 20.6 (C_{16}); m/z (ES+) calculated for $C_{23}H_{27}NO_4Na$ 404, found 404. Yield: 78%, R_f = 0.46 in CH_2Cl_2 with 5% MeOH.

8.2.4.7. n=4D DMT Protected

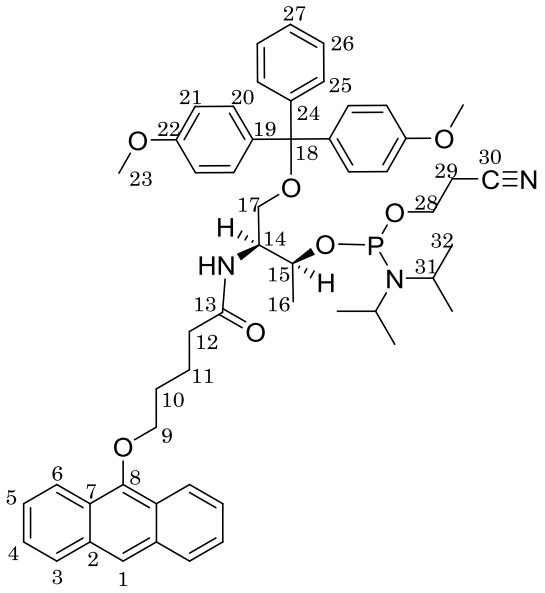
5-(Anthracen-9-yloxy)-N-((2S,3S)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) pentanamide



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.47 – 8.23 (m, 3H, H_1 and H_6), 8.12 – 8.03 (m, 2H, H_3), 7.57 – 7.41 (m, 4H, H_4 and H_{25}), 7.37 – 7.17 (m, 9H, H_5, H_{20}, H_{26} and H_{27}), 6.94 – 6.81 (m, 4H, H_{21}), 6.46 (d, J = 9.0 Hz, 1H, N*H*), 4.19 – 4.08 (m, 2H, H_9), 4.02 – 3.93 (m, 2H, H_{15} and H_{14}), 3.72 (s, 6H, H_{23}), 3.17 (d, J = 9.1 Hz, 2H, H_{17}), 2.40 (t, J = 6.9 Hz, 2H, H_{12}), 2.18 – 2.00 (m, 4H, H_{10} and H_{11}), 1.07 (d, J = 6.1 Hz, 3H, H_{16}); ^{13}C NMR (100 MHz, Acetonitrile-d_3) δ 173.0 (C_{13}), 158.3 (C_{22}) 145.3 (C_8), 145.2 (C_{24}), 136.0 (C_{19}), 135.9 (C_2), 132.2 (C_{20}), 129.6 (C_3), 128.0 (C_{25}), 127.7 (C_{26}), 127.5 (C_{27}), 126.4 (C_5), 125.3 (C_4), 125.0 (C_7), 121.9 (C_1), 117.0 (C_6), 112.7 (C_{21}), 75.4 (C_{18}) 65.5 (C_9), 63.3 (C_{15}), 60.0 (C_{17}), 54.8(C_{23}), 54.5 (C_{14}), 35.2 (C_{12}), 31.4 (C_{10}), 25.5 (C_{11}), 20.0 (C_{16}); m/z (ES+) calculated for $C_{44}H_{45}NO_6Na$ 706, found 706. Yield: 32%, R_f = 0.50 in CH_2Cl_2 with 5% MeOH.

8.2.4.8. n=4D Phosphoramidite

(2S,3S)-3-(5-(Anthracen-9-yloxy)pentanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite

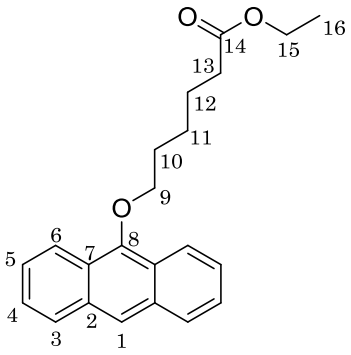


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.41 – 8.20 (m, 3H, 8.14, H_1 and H_6), 8.09 – 7.95 (m, 2H, H_3), 7.53 – 7.41 (m, 10H, H_4, H_5, H_{20}, and H_{25}), 7.21 (m, 3H, H_{26} and H_{27}), 6.86 – 6.78 (m, 4H, H_{21}), 6.31 (d, J = 8.9 Hz, 1H, N*H*), 4.08 (m, J = 7.1 Hz, 2H, H_9), 3.72 – 3.64 (m, 2H, H_{14} and H_{15}), 3.70 (s, 6H, H_{23}), 3.60 – 3.42 (m, 4H, H_{28} and H_{31}), 3.23 – 3.05 (m, 2H, H_{17}), 2.73 – 2.40 (m, 2H, H_{29}), 2.48 (t, J = 6.0 Hz, 2H, H_{12}), 2.43 – 2.34 (m, 4H, H_{10} and H_{11}), 1.27 – 1.08 (m, 12H, H_{32}), 0.98 (d, J = 6.1 Hz, 3H, H_{16}). ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.5, 147.0; m/z (ES+) calculated for $C_{53}H_{62}N_3O_7PNa$ 906, found 906. Yield: 43%, R_f = 0.63 in 50% hexane, 50% ethyl acetate.

8.2.5. Characterisation of n=5L and 5D

8.2.5.1. n=5 Ester

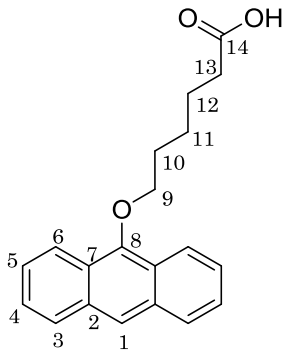
Ethyl 6-(anthracen-9-yloxy)hexanoate



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.39 – 8.13 (m, 3H, H_1 and H_6), 8.06 – 7.88 (m, 2H, H_3), 7.53 – 7.35 (m, 4H, H_4 and H_5), 4.33 – 4.03 (m, 4H, H_9 and H_{15}), 2.42 (t, *J* = 7.0 Hz, 2H, H_{13}), 2.18 – 1.94 (m, 4H, H_{10} and H_{11}), 1.87 – 1.77 (m, 2H, H_{12}), 1.28 (t, *J* = 7.2 Hz, 3H, H_{16}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.7 (C_{14}), 151.4 (C_8), 134.1 (C_2), 128.4 (C_3), 125.5 (C_5), 125.1 (C_4), 124.7 (C_7), 122.4 (C_1), 122.0 (C_6), 75.8 (C_9), 60.3 (C_{15}), 34.3 (C_{13}), 30.4 (C_{10}), 25.9 (C_{12}), 25.0 (C_{11}), 14.3 (C_{16}); m/z (ES+) calculated for $C_{22}H_{42}O_3Na$ 359, found 359. Yield: 28%, R_f = 0.22 in CH_2Cl_2 with 5% MeOH.

8.2.5.2. n=5 Carboxylic Acid

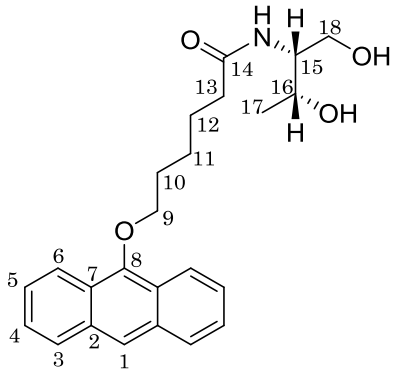
6-(Anthracen-9-yloxy)hexanoic acid



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.42 – 8.10 (m, 3H, H_1 and H_6), 8.06 – 7.89 (m, 2H, H_3), 7.52 – 7.34 (m, 4H, H_4 and H_5), 4.24 – 4.08 (m, 2H, H_9), 2.48 (t, *J*= 7.2 Hz, 2H, H_{13}), 2.16 – 2.04 (m, 4H, H_{10} and H_{11}), 1.89 – 1.70 (m, 2H, H_{12}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 179.5 (C_{14}), 151.3 (C_8), 132.5 (C_2), 128.4 (C_3), 125.5 (C_5), 125.1 (C_4), 124.7 (C_7), 122.4 (C_1), 122.1 (C_6), 75.7 (C_9), 33.9 (C_{13}) 30.4 (C_{10}), 25.8 (C_{12}), 24.7 (C_{11}); m/z (ES+) calculated for $C_{20}H_{20}O_3Na$ 331, found 331. Yield: 75%.

8.2.5.3. n=5L Diol

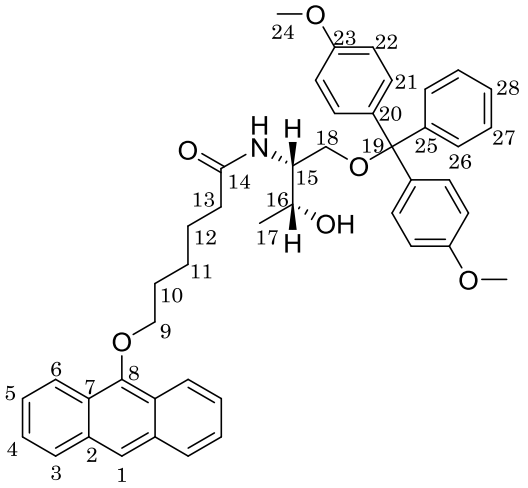
6-(Anthracen-9-yloxy)-N-((2R,3R)-1,3-dihydroxybutan-2-yl)hexanamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.36 – 8.12 (m, 3H, H_1 and H_6), 8.13 – 7.90 (m, 2H, H_3), 7.60 – 7.27 (m, 4H, H_4 and H_5), 6.38 (d, *J* = 7.5 Hz, 1H, N*H*), 4.23 – 4.14 (m, 2H, H_9), 3.90 – 3.73 (m, 2H, H_{15} and H_{16}), 2.96 – 2.78 (m, 2H, H_{18}), 2.41 – 2.34 (m, 2H, H_{13}), 2.10 – 1.94 (m, 2H, H_{10}), 1.95 – 1.58 (m, 4H, H_{11} and H_{12}), 1.20 (d, *J* = 6.4 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 174.3 (C_{14}), 151.2 (C_8), 132.4 (C_2), 128.4 (C_3), 125.5 (C_5), 125.3 (C_4), 125.2 (C_7), 122.3 (C_1), 122.1 (C_6), 75.8 (C_9), 69.0 (C_{16}), 65.3 (C_{17}), 54.6 (C_{15}), 36.6 (C_{13}), 30.4 (C_{10}), 25.9 (C_{11}), 25.7 (C_{12}), 20.5 (C_{18}); m/z (ES+) calculated for $C_{24}H_{29}NO_4Na$ 418, found 418. Yield: 79%, R_f = 0.15 in CH_2Cl_2 with 5% MeOH.

8.2.5.4. n=5L DMT Protected

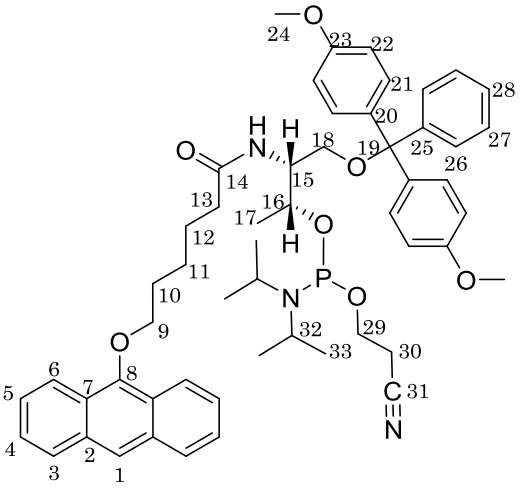
6-(Anthracen-9-yloxy)-N-((2R,3R)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) hexanamide



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.36 – 8.20 (m, 3H, H_1 and H_6), 8.02 – 7.95 (m, 2H, H_3), 7.48 – 7.40 (m, 6H, H_4, H_5 and H_{26}), 7.34 – 7.11 (m, 7H, H_{21}, H_{27} and H_{28}), 6.84 – 6.78 (m, 4H, H_{22}), 6.62 (d, J = 9.0 Hz, 1H, N*H*), 4.15 – 3.95 (m, 4H, H_9, H_{15} and H_{16}), 3.68 (s, 6H, H_{24}), 3.28 – 3.08 (d, J = 9.1 Hz, 2H, H_{18}), 2.30 (t, J = 6.9 Hz, 2H, H_{13}), 2.07 – 1.93 (m, 4H, H_{10} and H_{11}), 1.81 – 1.64 (m, 2H, H_{12}), 1.09 (d, J = 6.3 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Chloroform-d) δ 174.0 (C_{14}), 159.3 (C_{22}) 152.3 (C_8), 146.2 (C_{24}), 137.0 (C_{19}), 133.3 (C_2), 130.7 (C_{20}), 129.7 (C_3), 128.9 (C_{25}), 128.7 (C_{26}), 128.0 (C_{27}), 125.5 (C_5), 125.0 (C_4), 124.7 (C_7), 122.4 (C_6), 122.1 (C_1), 113.0 (C_{21}), 87.3 (C_{18}) 75.6 (C_9), 68.9 (C_{16}), 65.5 (C_{28}), 55.5 (C_{23}), 55.3 (C_{15}), 36.7 (C_{13}), 31.2 (C_{10}), 26.4 (C_{11}), 26.3 (C_{12}), 20.5 (C_{17}); m/z (ES+) calculated for $C_{45}H_{47}NO_6Na$ 721, found 721. Yield: 37%, R_f = 0.53 in CH_2Cl_2 with 5% MeOH.

8.2.5.5. n=5L Phosphoramidite

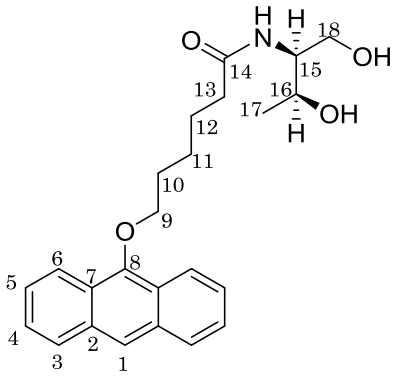
(2R,3R)-3-(6-(Anthracen-9-yloxy)hexanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.29 – 8.21 (m, 3H, 8.14, H_1 and H_6), 8.08 – 8.00 (m, 2H, H_3), 7.49 – 7.41 (m, 10H, H_4, H_5, H_{21}, and H_{26}), 7.27 (m, 3H, H_{27} and H_{28}), 6.90 – 6.73 (m, 4H, H_{22}), 6.17 (d, J = 9.1 Hz, 1H, N*H*), 4.05 (m, 2H, H_9), 3.73 – 3.67 (m, 2H, H_{15} and H_{16}), 3.69 (s, 6H, H_{24}), 3.59 – 3.40 (m, 4H, H_{29} and H_{32}), 3.13 – 3.03 (m, 2H, H_{18}), 2.60 – 2.53 (m, 2H, H_{30}), 2.47 (t, J = 6.0 Hz, 2H, H_{13}), 2.34 – 2.24 (m, 6H, H_{10}, H_{11} and H_{12}), 1.22 – 1.04 (m, 12H, H_{33}), 0.96 (d, J = 6.8 Hz, 3H, H_{17}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.6, 146.9; m/z (ES+) calculated for $C_{54}H_{64}N_3O_7PNa$ 920, found 920. Yield: 43%, R_f = 0.70 in 50% hexane, 50% ethyl acetate.

8.2.5.6. n=5D Diol

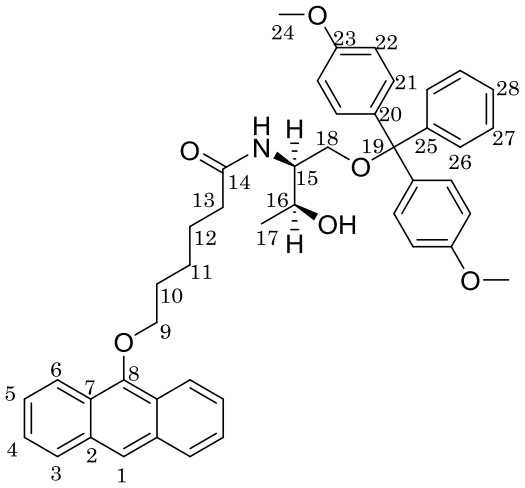
6-(Anthracen-9-yloxy)-N-((2S,3S)-1,3-dihydroxybutan-2-yl)hexanamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.33 – 8.09 (m, 3H, H_1 and H_6), 8.09 – 7.89 (m, 2H, H_3), 7.82 – 7.37 (m, 4H, H_4 and H_5), 6.24 (m, 1H, N*H*), 4.45 – 4.05 (m, 2H, H_9), 3.88 – 3.49 (m, 2H, H_{15} and H_{16}), 2.96 – 2.78 (m, 2H, H_{18}), 2.41 – 2.34 (m, 2H, H_{13}), 2.10 – 1.94 (m, 2H, H_{10}), 1.90 – 1.70 (m, 4H, H_{11} and H_{12}), 1.21 (d, *J* = 6.4 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.3 (C_{14}), 150.2 (C_8), 134.4 (C_2), 128.3 (C_3) 125.4 (C_5), 125.1 (C_4), 125.0 (C_7), 122.3 (C_1), 122.2 (C_6), 76.0 (C_9), 69.2 (C_{16}), 65.5 (C_{17}), 54.4 (C_{15}), 36.4 (C_{13}), 30.6 (C_{10}), 25.8 (C_{11}), 25.7 (C_{12}), 20.5 (C_{18}); m/z (ES+) calculated for $C_{24}H_{29}NO_4Na$ 418, found 418. Yield: 76%, R_f = 0.18 in CH_2Cl_2 with 5% MeOH.

8.2.5.7. n=5D DMT Protected

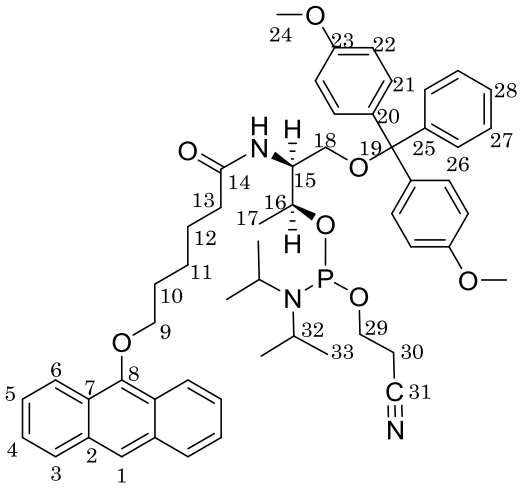
6-(Anthracen-9-yloxy)-N-((2S,3S)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) hexanamide



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.35 – 8.20 (m, 3H, H_1 and H_6), 8.13 – 7.98 (m, 2H, H_3), 7.56 – 7.40 (m, 6H, H_4, H_5 and H_{26}), 7.34 – 7.14 (m, 7H, H_{21}, H_{27} and H_{28}), 6.88 – 6.78 (m, 4H, H_{22}), 6.43 (d, J = 8.9 Hz, 1H, N*H*), 4.15 – 3.88 (m, 4H, H_9, H_{15} and H_{16}), 3.72 (s, 6H, H_{24}), 3.16 (d, J = 9.1 Hz, 2H, H_{18}), 2.32 (t, J = 6.9 Hz, 2H, H_{13}), 2.20 – 1.96 (m, 6H, H_{10}, H_{11} and H_{12}), 1.07 (d, J = 6.2 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Acetonitrile-d_3 δ 174.0 (C_{14}), 158.3 (C_{22}) 152.3 (C_8), 146.2 (C_{24}), 137.0 (C_{19}), 133.3 (C_2), 130.7 (C_{20}), 129.6 (C_3), 128.0 (C_{25}), 127.7 (C_{26}), 127.4 (C_{27}), 126.4 (C_5), 125.3 (C_4), 125.0 (C_7), 121.9 (C_6), 121.1 (C_1), 112.6 (C_{21}), 87.3 (C_{18}) 75.5 (C_9), 66.0 (C_{16}), 63.3 (C_{28}), 54.5 (C_{23}), 53.9 (C_{15}), 35.6 (C_{13}), 29.8 (C_{10}), 25.3 (C_{11}), 25.3 (C_{12}), 19.3 (C_{17}); m/z (ES+) calculated for $C_{45}H_{47}NO_6Na$ 721, found 721. Yield: 38%, R_f = 0.51 in CH_2Cl_2 with 5% MeOH.

8.2.5.8. n=5D Phosphoramidite

(2S,3S)-3-(6-(Anthracen-9-yloxy)hexanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite

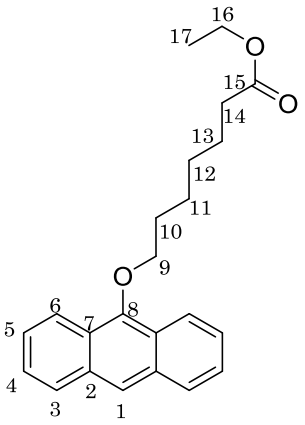


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.24 – 8.08 (m, 3H, 8.14, H_1 and H_6), 8.02 – 7.90 (m, 2H, H_3), 7.79 – 7.34 (m, 10H, H_4, H_5, H_{21}, and H_{26}), 7.19 (m, 3H, H_{27} and H_{28}), 6.77 – 6.68 (m, 4H, H_{22}), 6.12 (d, J = 9.1 Hz, 1H, N*H*), 4.20 – 3.91 (m, 2H, H_9), 3.66 – 3.61 (m, 2H, H_{15} and H_{16}), 3.62 (s, 6H, H_{24}), 3.50 – 3.28 (m, 4H, H_{29} and H_{32}), 3.10 – 3.00 (m, 2H, H_{18}), 2.62 – 2.54 (m, 2H, H_{30}), 2.54 (t, J = 5.9 Hz, 2H, H_{13}), 2.43 – 2.33 (m, 6H, H_{10}, H_{11} and H_{12}), 1.15 – 0.95 (m, 12H, H_{33}), 0.98 (d, J = 6.8 Hz, 3H, H_{17}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.6, 146.9; m/z (ES+) calculated for $C_{54}H_{64}N_3O_7PNa$ 920, found 920. Yield: 74%, R_f = 0.65 in 50% hexane, 50% ethyl acetate.

8.2.6. Characterisation of n=6L and 6D

8.2.6.1. n=6 Ester

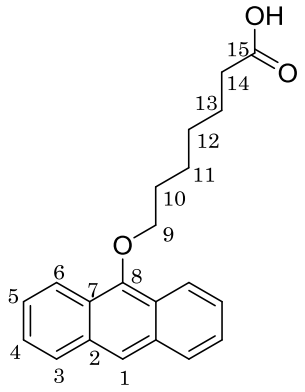
Ethyl 7-(anthracen-9-yloxy)heptanoate



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.53 – 8.15 (m, 3H, H_1 and H_6), 8.07 – 7.86 (m, 6H, H_3, H_4 and H_5), 4.24 – 4.10 (m, 4H, H_9 and H_{16}), 2.37 (t, *J* = 7.5 Hz, 2H, H_{14}), 2.17 – 1.92 (m, 4H, H_{10} and H_{11}), 1.83 – 1.43 (m, 4H, H_{13} and H_{12}), 1.28 (t, *J* = 7.1 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.8 (C_{15}), 151.5 (C_8), 134.1 (C_2), 128.5 (C_3), 125.5 (C_5), 125.1 (C_4), 124.7 (C_7), 122.4 (C_1), 122.0 (C_6), 76.0 (C_9), 60.3 (C_{16}), 34.4 (C_{14}), 30.5 (C_{10}), 29.2 (C_{13}), 26.0 (C_{12}), 25.0 (C_{11}), 14.3 (C_{17}); m/z (ES+) calculated for $C_{23}H_{26}O_3Na$ 373, found 373. Yield: 48%, R_f = 0.17 in CH_2Cl_2 with 5% MeOH.

8.2.6.2. n=6 Carboxylic Acid

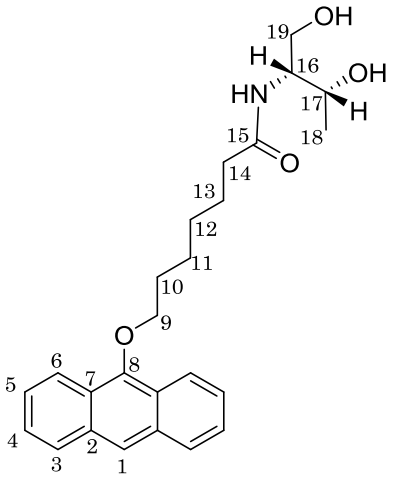
7-(Anthracen-9-yloxy)heptanoic acid



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.37 – 8.14 (m, 3H, H_1 and H_6), 8.06 – 7.38 (m, 6H, H_3, H_4 and H_5), 4.19 (t, *J* = 6.6 Hz, 2H, H_9), 2.43 (t, *J* = 7.4 Hz, 2H, H_{14}), 2.15 – 1.92 (m, 4H, H_{10} and H_{11}), 1.83 – 1.47 (m, 4H, H_{13} and H_{12}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 173.6 (C_{15}), 151.5 (C_8), 134.0 (C_2), 128.5 (C_3), 125.3 (C_5), 125.1 (C_4), 124.7 (C_7), 122.4 (C_1), 122.1 (C_6), 76.0 (C_9), 34.4 (C_{14}), 30.4 (C_{10}), 29.2 (C_{13}), 26.0 (C_{12}), 25.0 (C_{11}); m/z (ES+) calculated for $C_{21}H_{22}O_3Na$ 345, found 345. Yield: 82%.

8.2.6.3. n=6L Diol

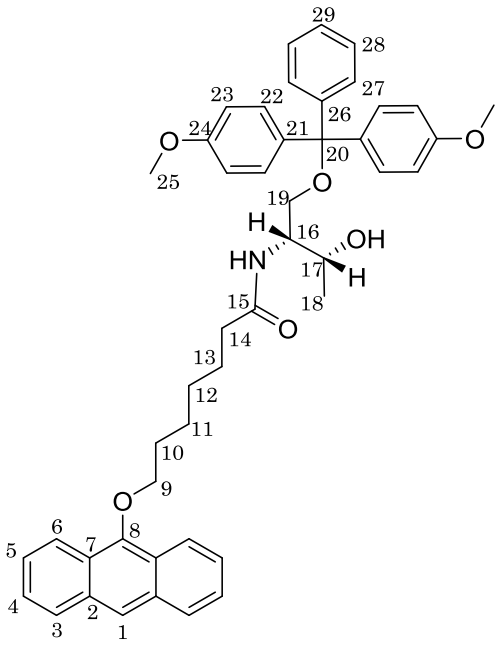
7-(Anthracen-9-yloxy)-N-((2R,3R)-1,3-dihydroxybutan-2-yl)heptanamide



^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.38 – 8.08 (m, 3H, H_1 and H_6), 8.07 – 7.84 (m, 2H, H_3), 7.58 – 7.34 (m, 4H, H_4 and H_5), 6.25 (d, *J* = 7.7 Hz, 1H, N*H*), 4.18 (t, *J* = 7.5 Hz, 2H, H_9), 3.89 – 3.78 (m, 2H, H_{16} and H_{17}), 2.92 – 2.80 (m, 2H, H_{19}), 2.31 (t, *J* = 7.5 Hz, 2H, H_{14}), 1.89 – 1.59 (m, 4H, H_{10} and H_{13}), 1.59 – 1.37 (m, 4H, H_{11} and H_{12}), 1.19 (d, *J* = 6.4 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Chloroform-*d*) δ 174.3 (C_{15}), 151.4 (C_8), 132.4 (C_2), 128.5 (C_3), 125.5 (C_5), 125.1 (C_4), 124.7 (C_7), 122.4 (C_1), 122.0 (C_6), 76.0 (C_9), 69.0 (C_{17}), 65.3 (C_{19}), 54.5 (C_{16}), 36.8 (C_{14}), 30.6 (C_{10}), 29.3 (C_{13}), 26.1 (C_{11}), 25.8 (C_{12}), 20.6 (C_{18}); m/z (ES+) calculated for $C_{25}H_{31}NO_4Na$ 432, found 432. Yield: 75%, R_f = 0.28 in CH_2Cl_2 with 5% MeOH.

8.2.6.4. n=6L DMT Protected

7-(Anthracen-9-yloxy)-N-((2R,3R)-1-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-hydroxybutan-2-yl) heptanamide

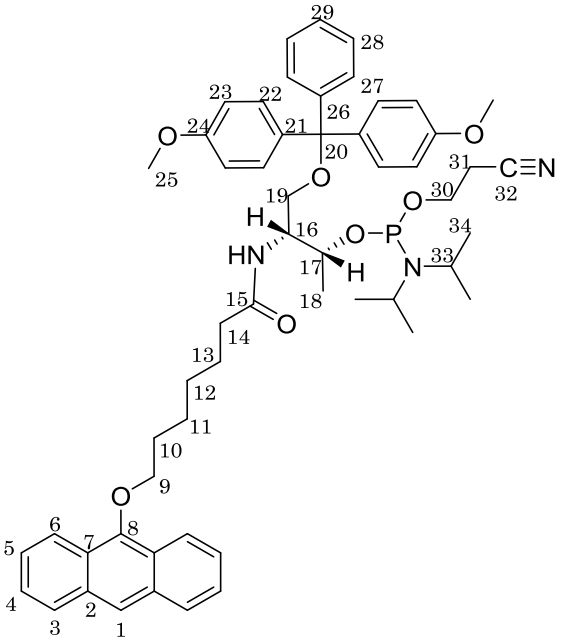


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.35 – 8.16 (m, 3H, H_1 and H_6), 8.15 – 7.88 (m, 2H, H_3), 7.52 – 7.42 (m, 6H, H_4, H_5 and H_{27}), 7.34 – 7.19 (m, 7H, H_{22}, H_{28} and H_{29}), 6.92 – 6.69 (m, 4H, H_{23}), 6.40 (d, J = 9.0 Hz, 1H, N*H*), 4.15 (t, J = 6.5 Hz, 2H, H_9), 4.05 – 3.90 (m, 2H, H_{16} and H_{17}), 3.74 (s, 6H, H_{25}), 3.31 – 2.90 (m, 2H, H_{19}), 2.24 (m, 4H, H_{12} and H_{14}), 2.08 – 1.84 (m, 4H, H_{10} and H_{11}), 1.82 – 1.57 (m, 2H, H_{13}), 1.05 (d, J = 6.2 Hz, 3H, H_{18}); ^{13}C NMR (100 MHz, Acetonitrile-d_3) δ 172.8 (C_{15}), 158.3 (C_{23}), 145.3 (C_8), 144.2 (C_{25}), 135.7 (C_{20}), 132.1 (C_2), 129.6 (C_{21}), 128.0 (C_3), 127.7 (C_{26}), 127.5 (C_{27}), 126.4 (C_{28}), 125.3 (C_5), 125.0 (C_4), 124.2 (C_7), 121.9 (C_6), 121.5 (C_1), 112.7 (C_{22}), 87.3 (C_{19}), 75.6 (C_9), 66.0 (C_{17}), 63.3 (C_{29}), 54.5 (C_{24}), 53.9 (C_{16}), 35.6 (C_{14}), 29.8 (C_{10}), 28.6 (C_{13}) 25.4 (C_{11}), 25.3 (C_{12}), 19.3 (C_{18}); m/z (ES+)

calculated for $C_{46}H_{49}NO_6Na$ 735, found 735. Yield: 35%, R_f = 0.62 in CH_2Cl_2 with 5% MeOH.

8.2.6.5. n=6L Phosphoramidite

(2R,3R)-3-(7-(Anthracen-9-yloxy)heptanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite

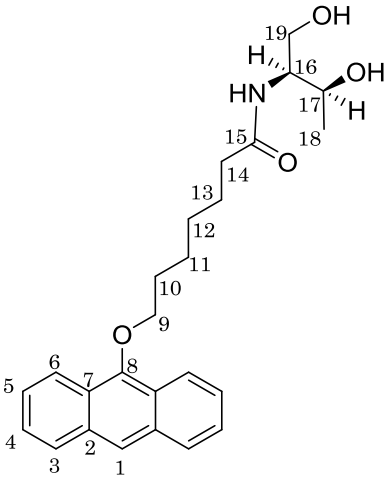


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.23 – 8.12 (m, 3H, 8.14, H_1 and H_6), 7.98 – 7.89 (m, 2H, H_3), 7.77 – 7.30 (m, 10H, H_4, H_5, H_{22}, and H_{27}), 7.25 – 7.06 (m, 3H, H_{28} and H_{29}), 6.79 – 6.71 (m, 4H, H_{23}), 6.10 (d, *J* = 9.1 Hz, 1H, N*H*), 4.07 – 3.91 (m, 2H, H_9), 3.68 – 3.60 (m, 2H, H_{16} and H_{17}), 3.63 (s, 6H, H_{25}), 3.53 – 3.26 (m, 4H, H_{30} and H_{33}), 3.15 – 2.92 (m, 2H, H_{19}), 2.62 – 2.52 (m, 2H, H_{31}), 2.49 (t, *J* = 5.9 Hz, 2H, H_{14}),

2.21 – 2.10 (m, 6H, H_{10}, H_{11} and H_{12}), 1.45 – 1.35 (m, 2H, H_{13}), 1.20 – 0.98 (m, 12H, H_{34}), 0.87 (d, J = 6.7 Hz, 3H, H_{18}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.6, 146.9; m/z (ES+) calculated for $C_{55}H_{66}N_3O_7PNa$ 935, found 935. Yield: 63%, R_f = 0.68 in 50% hexane, 50% ethyl acetate.

8.2.6.6. n=6D Diol

7-(Anthracen-9-yloxy)-N-((2S,3S)-1,3-dihydroxybutan-2-yl)heptanamide

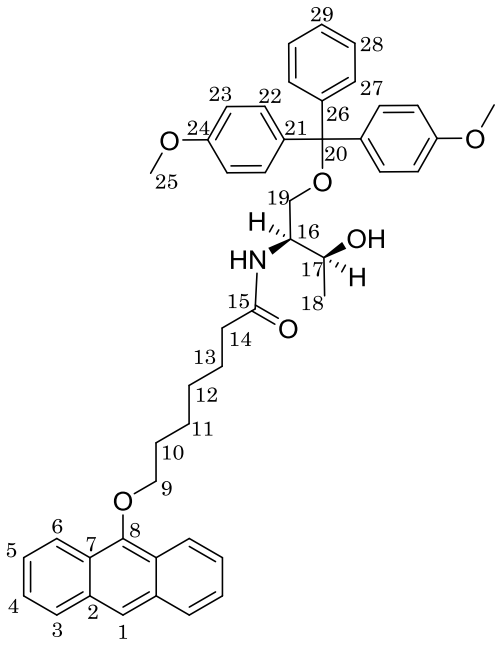


^{1}H NMR (300 MHz, Chloroform-d) δ 8.39 – 8.14 (m, 3H, H_1 and H_6), 8.07 – 7.80 (m, 2H, H_3), 7.57 – 7.33 (m, 4H, H_4 and H_5), 6.34 (d, J = 8.2 Hz, 1H, NH), 4.16 (t, J = 6.5 Hz, 2H, H_9), 3.93 – 3.64 (m, 2H, H_{16} and H_{17}), 3.39 – 3.31 (m, 2H, H_{19}), 2.29 (t, J = 7.5 Hz, 2H, H_{14}), 1.82 – 1.57 (m, 4H, H_{10} and H_{13}), 1.54 – 1.37 (m, 4H, H_{11} and H_{12}), 1.18 (d, J = 6.3 Hz, 3H, H_{17}); ^{13}C NMR (100 MHz, Chloroform-d) δ 174.0 (C_{15}), 151.2 (C_8), 132.4 (C_2), 128.5 (C_3) 125.5 (C_5), 125.2 (C_4), 124.7 (C_7), 122.3 (C_1), 122.1 (C_6), 75.8 (C_9), 68.9 (C_{17}), 65.2 (C_{19}), 54.5 (C_{16}), 36.7 (C_{14}), 30.4 (C_{10}), 29.3 (C_{13}), 25.9

(C_{11}), 25.7 (C_{12}), 20.6 (C_{18}); m/z (ES+) calculated for $C_{25}H_{31}NO_4Na$ 432, found 432. Yield: 50%, R_f = 0.38 in CH_2Cl_2 with 5% MeOH.

8.2.6.7. n=6D DMT Protected

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

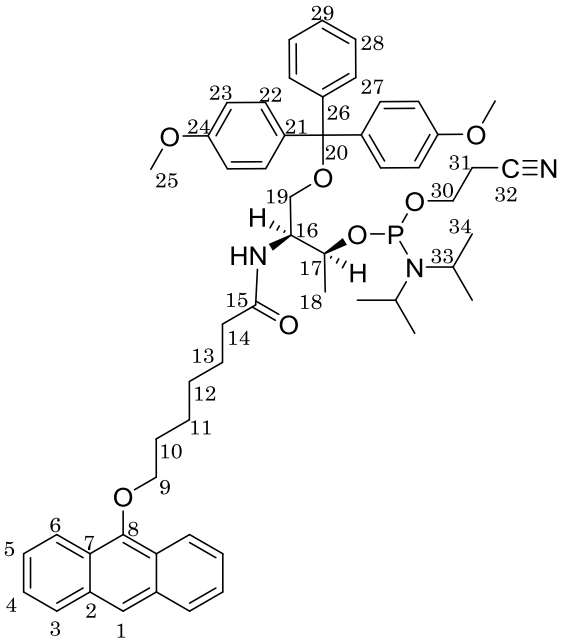


^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.35 – 8.24 (m, 3H, H_1 and H_6), 8.10 – 7.99 (m, 2H, H_3), 7.48 – 7.41 (m, 6H, H_4, H_5 and H_{27}), 7.36 – 7.17 (m, 7H, H_{22}, H_{28} and H_{29}), 6.91 – 6.79 (m, 4H, H_{23}), 6.39 (d, J=8.8 Hz, 1H, N*H*), 4.16 (t, J= 6.5 Hz, 2H, H_9), 4.10 – 3.94 (m, 2H, H_{16} and H_{17}), 3.75 (s, 6H, H_{25}), 3.15 – 2.99 (m, 2H, H_{19}), 2.27 (m, 4H, H_{12} and H_{14}), 2.19 – 1.96 (m, 4H, H_{10} and H_{11}), 1.69 – 1.42 (m, 2H, H_{13}), 1.05 (d, J= 6.2 Hz, 3H, H_{18}); ^{13}C NMR (100 MHz, Acetonitrile-d_3) δ 172.8 (C_{15}),

158.3 (C_{23}) 145.3 (C_8), 144.2 (C_{25}), 135.7 (C_{20}), 132.1 (C_2), 129.7 (C_{21}), 128.0 (C_3), 127.7 (C_{26}), 127.5 (C_{27}), 126.4 (C_{28}), 125.3 (C_5), 125.0 (C_4), 124.2 (C_7), 121.9 (C_6), 121.5 (C_1), 112.7 (C_{22}), 87.3 (C_{19}) 75.6 (C_9), 66.0 (C_{17}), 63.3 (C_{29}), 54.5 (C_{24}), 53.9 (C_{16}), 35.6 (C_{14}), 29.9 (C_{10}), 28.6 (C_{13}) 25.4 (C_{11}), 25.3 (C_{12}), 19.3 (C_{18}); m/z (ES+) calculated for $C_{46}H_{49}NO_6Na$ 735, found 735. Yield: 43%, R_f = 0.48 in CH_2Cl_2 with 5% MeOH.

8.2.6.8. n=6D Phosphoramidite

(2S,3S)-3-(7-(Anthracen-9-yloxy)heptanamido)-4-(bis-(4-methoxyphenyl)(phenyl)methoxy)butan-2-yl 2-cyanoethyldiisopropylphosphoramidite



^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 8.35 – 8.22 (m, 3H, 8.14, H_1 and H_6), 8.11 – 8.02 (m, 2H, H_3), 7.57 – 7.48 (m, 10H, H_4, H_5, H_{22}, and H_{27}), 7.38 – 7.25 (m, 3H, H_{28} and H_{29}), 6.92 – 6.79 (m, 4H, H_{23}), 6.20 (d, *J* = 9.1 Hz, 1H, N*H*), 4.27 – 3.99 (m, 2H, H_9), 3.74 (s, 6H, H_{25}), 3.64 – 3.38 (m, 6H, H_{16}, H_{17}, H_{30} and H_{33}), 3.19 – 3.01 (m, 2H, H_{19}), 2.78 – 2.65 (m, 2H, H_{31}), 2.49 (t, *J* = 5.9 Hz, 2H, H_{14}), 2.28 – 2.00 (m, 6H, H_{10}, H_{11} and H_{12}), 1.58 – 1.24 (m, 2H, H_{13}), 1.20 – 1.06 (m, 12H, H_{34}), 0.99 (d, *J* = 6.8 Hz, 3H, H_{18}); ^{31}P NMR (121 MHz, Acetonitrile-d_3) δ 147.6, 146.9; m/z (ES+) calculated for $C_{55}H_{66}N_3O_7PNa$ 935, found 935. Yield: 55%, R_f = 0.62 in 50% hexane, 50% ethyl acetate.

8.3 Oligonucleotide Synthesis

8.3.1 Synthesis

All oligonucleotides were synthesised on an Applied Bioscience 394 DNA/RNA synthesiser on a 1 μmol scale. Controlled-pore glass (CPG) resin with the first base attached was used. The standard phosphoramidites used were Ac-dC-CE, iBu-dG-CE, Bz-dA-CE and dT-CE. The resin was washed with acetonitrile prior to the first reaction. The trityl group was removed using 3% trichloroacetic acid in dichloromethane for 40 seconds. The next phosphoramidite in the sequence was then coupled using 0.5 M Ethylthiotetrazole (ETT) as the activator and appropriate phosphoramidite (0.1 M) concentration for 30 seconds for standard bases and 10 minutes for anthracene phosphoramidites. Unreacted sequences were then capped using tetrahydrofuran(THF)/pyridine/acetic anhydride (70:15:15) and 10% methyl

imidazole in THF for 15 seconds. The phosphite group was then oxidised to a phosphate using 0.02 M iodine in THF/pyridine/water (70:15:15) for 15 seconds. The cycle was repeated until the desired sequence was completed. The resin was then treated with 30% ammonia hydroxide (four 15 minute cycles) to cleave the oligonucleotide from the resin. The solution was heated to 55°C for 6 hours to remove the nucleobase protecting groups and ammonia hydroxide was removed *in vacuo* to yield a clear to yellow oil. Oligonucleotides were redissolved in water prior to purification.

For rhodamine labelling used in Chapter 6, TAMRA-dT-CE phosphoramidite was used with a 6 minute coupling time. Due to the milder cleavage conditions needed for this modification, ultra-mild nucleotide phosphoramidites, Ac-dC-CE and dmf-dG-CE were used. Cleavage and deprotection was performed using 0.05 M potassium carbonate in anhydrous methanol for 4 hours. The sample was then passed through a NAP-10 column to remove the potassium carbonate before the solvent was removed.

8.3.2 Purification

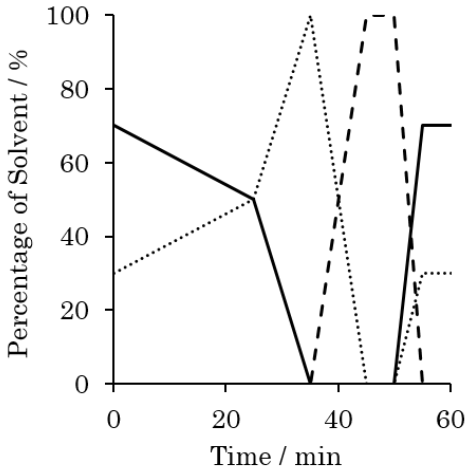
All oligonucleotides were purified using RP-HPLC on a Phenomenex Clarity 5 μm Oligo RP 150 x 10 mm C18 column. After purification, solvent removed *in vacuo,* and the sample was desalted on a NAP-10 G-25 sephadex column. Analytical RP-

HPLC was then used to check purity using a Phenomenex Clarity 5 µm Oligo RP LC column 150 x 4.6 mm.

Unmodified and TAMRA labelled oligonucleotides along with those containing an anthracene tag of carbon linker length 1, were purified using the "Oligo SP 60 min" method. Oligonucleotides containing anthracene tags with carbon linker lengths of 3-7 were purified with the "DMT-On" method, Figure 7.1.

A)

Time (min)	% Solvent (%)		
	A	B	C
0	70	30	0
25	50	50	0
35	0	100	0
45	0	0	100
50	0	0	100
55	70	30	0
60	70	30	0



B)

Time (min)	% Solvent (%)		
	A	B	C
0	0	100	0
10	0	100	0
20	0	70	30
25	0	0	100
35	0	0	100
36	0	100	0
41	0	100	0

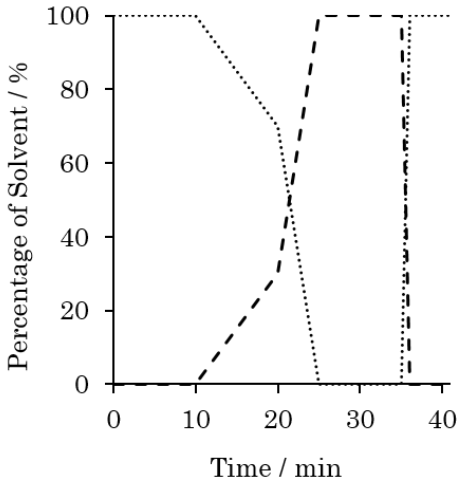


Figure 7.1 HPLC methods for oligonucleotide purification. Oligo SP 60 min (A) is used for unmodified and TAMRA labelled sequences. DMT-On (B) is used for anthracene containing sequences. The graphs (right) show the percentage of each solvent as a function of time where solvent A is 0.1 M TEAA in 5 % acetonitrile, 95% water (solid line), solvent B is 0.1 M TEAA in 5 % acetonitrile, 95% water (dotted line) and solvent C is 100% acetonitrile (dashed line).

8.3.3 Characterisation and Concentration Determination

Oligonucleotides were characterised by negative mode electrospray mass spectrometry (ESI-MS) and their concentration determined using UV-Vis spectroscopy by measuring the absorbance at 260 nm.

Molar extinction coefficients were calculated using an online oligonucleotide property calculator which utilised the nearest neighbour method.[3,4] The anthracene modifications are known to absorb at 260 nm and contributes a molar extinction value of 51 444 $M^{-1}\cdot cm^{-1}$.[1] The TAMRA modification was used as per the manufacturer's instructions with $\varepsilon_{260\ nm}$ = 32 300 $M^{-1}\cdot cm^{-1}$.

Oligonucleotides were stored at -20°C in MilliQ water. Before use they were defrosted, brought to room temperature and thoroughly mixed.

8.4 Peptide Synthesis

8.4.1 Synthesis

Ac-<u>ALKRARNTEAARRSR</u>ARKLQRMKQ<u>**KG**</u>-NH_2

The peptide was synthesized using standard solid-phase peptide synthesis (SPPS) methods[5] on rink amide MBHA resin (0.1 mM scale, 0.65 mmol/g loading) using a CEM Liberty 1 automated peptide synthesizer. Initial couplings (bold and

underlined) were performed manually to allow for the retrieval of Fmoc-Lys(Mtt)-OH. The 4-methyltrityl (Mtt) protecting group was removed with 3% TFA in DCM for 10 minutes with 3 repeats. The resin was then rinsed 5 times with 10 mL DCM and 5 times with 10 mL DMF. A ten-fold excess of 9-anthracenecarboxylic acid was then coupled using SPPS synthesis methods, with a 15 hour coupling time. The resin was then loaded into the CEM Liberty 1 automated peptide synthesizer and the remaining amino acids coupled. A standard coupling programme (25 W, 75 °C, 300 s) was used for all amino acids except for the coupling of Fmoc-Arg(Pbf)-OH and Fmoc-Ser(tBu)-OH which were double coupled along with residues towards the N-terminus (underlined). Deprotections were performed in 20% piperidine in DMF (35 W, 75 °C, 180 s). The N-terminus was manually acetylated (20% acetic anhydride, 20% DIPEA in DMF) before cleavage from the resin and deprotection of side chains (95% trifluoroacetic acid, 5% thioanisole, 3% 1, 2-ethanedithiol and 2% anisole for 2 hours).

8.4.2 Purification

The modified peptide was purified by preparative RP-HPLC (Phenomenex Jupiter 10 μm Proteo 90 Å Axia packed column, C12, 250 by 21.2 mm) using a flow rate of 10 ml/min with a gradient of 0-100% solvent B (acetonitrile with 0.05 % TFA) over 40 mins, where solvent A was 0.05% TFA in deionised water followed holding at 100 % solvent B for 10 minutes before being retuned to 100 % solvent A (Rt = 13.73 min). This product was further purified using a semi-preparative column

(Phenomenex Luna 10 μm 100 Å, C18, 250 by 4.6 mm) using a ramp of 0-20% solvent B over 15 minutes followed holding at 20% solvent B for 35 minutes (Rt = 27.80 min).

8.4.3 Characterisation and Concentration Determination

The modified peptide was characterised using positive ESI-MS. Peptide concentrations were calculated based on the anthracene absorbance at 375 nm, $\varepsilon_{375\ nm} = 6300\ M^{-1}\ cm^{-1}$ in water.[6]

8.5 UV-Vis Spectroscopy Studies

Standard measurements (including titrations and concentration determinations) were performed on a Shimadzu UV-1800 UV-Vis spectrophotometer using either a 3 mL, 1 cm pathlength or a 1 mL, 1 cm pathlength quartz cuvette. Variable temperature studies were performed on a Cary 5000 UV-Vis-NIR spectrophotometer fitted with a Peltier temperature controller. These studies were performed in 500 μL, 1 cm pathlength, low head space quartz cuvettes. The temperature was increased 0.5°C/min and held for 3 minutes at the extremes of temperature. Each experiment contained 3 dissociations and 2 annealing ramps.

8.6 Fluorescence Spectroscopy Studies

All measurements taken on Shimadzu RF-5301 PC Spectrofluorophotometer using a 1 mL, 1 cm quartz cuvette apart from those identified in Chapter 3.3.2 which were performed on an Edinburgh FLSP920 fluorescence spectrometer using a 50 μL, 1 cm quartz cuvette. All concentrations, buffer conditions and fluorescence parameters are detailed within the main body of this thesis.

8.7 Irradiation studies

Samples were irradiated with a water cooled 125 W mercury arc lamp, purchased from Photochemical Reactors Ltd, Reading, UK, placed in a housing with a small outlet fitted with a 365 nm filter purchased from Edmunds Optics Ltd, York, UK.

8.8 Circular Dichroism Spectroscopy Studies

Circular dichroism (CD) spectroscopy studies were performed on a Jasco J-810 spectropolarimeter using a 1 cm pathlength quartz cuvette. A data pitch of 0.2 nm was used at a scan rate of 200 nm/min with a response time of 1 second, with the signal averaged over 10 accumulations per sample.

Peptide molar ellipticity values were calculated using Equation 1;

$$\Theta_m = \frac{\Theta}{10 \times l \times c \times N_a} \qquad (1)$$

where Θ_m is the molar ellipticity, Θ is the ellipticity, l is the pathlength in cm, c is the concentration of anthracene tagged peptide in mol dm^{-3} and N_a is the number of amino acid residues in the peptide. Percentage folded values were calculated based on the molar ellipticity at 222 nm using Equation 2;[7]

$$\% \ folding = \left(\frac{\Theta_m - 640}{\left(-42500 \times \left(1 - \left(\frac{3}{N_a} \right) \right) \right) - 640} \right) \times 100 \qquad (2)$$

All experiments were performed in 100 mM NaCl, 10 mM sodium phosphate buffer pH7. DNA, peptide and any addition buffer or salt concentrations are detailed in the body of the thesis. For studies on the folding of peptide in the presence of DNA, the signals from the DNA were subtracted to give a difference spectrum for the peptide.

8.9 Peptide Denaturing Gel Electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) studies were performed with a 20% polyacrylamide separating gel with 0.5 M tris HCl buffer pH 8.8, 0.5% SDS and a 4% polyacrylamide stacking gel with 0.1 M tris HCl pH 6.8, 2% SDS. Samples were prepared in 250 mM Tris HCl pH 6.8, 10% glycerol and 2% SDS and equilibrated for 1 hour. The gel was run with 1x tris tricine running buffer for 15 minutes at 100V and increased to 200V for the remainder of

the gel running. The gels were stained with coomassie brilliant blue R-250 protein stain.

8.10 Gel Electrophoretic Mobility Shift Assays

Gel EMSA studies were performed using a 10% polyacrylamide gel with 0.5x TBE buffer. 0.5x TBE buffer was also used as a running buffer. Peptide and DNA were prepared in 20 mM Tris HCl pH 7.5, 2 mM $MgCl_2$, 2 mM EDTA, 90 mM KCl, 10% glycerol, 2.25% Igpeal CA630 and 0.1 mg/mL BSA. Samples were then irradiated with 365 nm light for 10 minutes where stated, and equilibrated for 1 hour. The gels were run at 100V for 1 hour. Gels were visualised with SYBR Gold stain and imaged using an excitation wavelength of 480 nm and a camera fitted with an ethidium bromide filter.

8.11 Native DNA Gel Electrophoresis

Native DNA gels were performed using a 20% polyacrylamide gel with 25 mM sodium chloride and 1 x TBE and a 25 mM sodium chloride and 1 x TBE running buffer. Samples were also prepared in 25 mM sodium chloride and 1 x TBE. The gels were run at 100V for 1 hour. Gels were visualised by anthracene and TAMRA fluorescence as well as using SYBR gold nucleic acid stain.

8.12 Denaturing DNA Gel Electrophoresis

Denaturing DNA gels were performed using 20% polyacrylamide gel with 1 x TBE and a 1 x TBE running buffer. Samples were prepared in 50% formamide and heated to 90 °C for 5 minutes before being quickly cooled on ice prior to loading the gel. Gels were pre-run at 15W for 10 minutes prior to loading and then at 15W for 30 minutes with samples. Gels were visualised by anthracene and TAMRA fluorescence as well as using SYBR gold nucleic acid stain.

8.13 Clauss Assay

The Clauss assay was performed by monitoring the transmittance of the sample at 450 nm on a Cary 5000 UV-Vis-NIR spectrophotometer fitted with a Peltier temperature controller. 3 μM fibrinogen was prepared in 50 mM tris HCl pH 7, 100 mM KCl and equilibrated at 25 °C for 10 minutes prior to the experiment. 0.1 NIH units of thrombin were incubated at 25 °C for 10 minutes with 37 nM aptamer in 50 mM tris HCl pH 7, 100 mM KCl. The two solutions were mixed and the measurement started with the temperature maintained at 25 °C.

8.14 References

1 J.-L. H. A. Duprey, PhD Thesis, University of Birmingham, 2010.
2 J. Manchester, PhD Thesis, University of Birmingham, 2013.
3 C. R. Cantor, M. M. Warshaw and H. Shapiro, *Biopolymers*, 1970, **9**, 1059–1077.
4 G. D. Fasman, *Handbook of Biochemistry and Molecular Biology, Volume 1: Nucleic Acids*, CRC Press, 1975.
5 W. C. White and P. D. Chan, *Fmoc solid phase peptide synthesis*, Oxford University Press: New York, New York, 2000.
6 J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer US, 3rd edn., 2006.
7 J. K. Myers, C. N. Pace and J. M. Scholtz, *Proc. Natl. Acad. Sci.*, 1997, **94**, 2833–2837.

Appendices

9.1 SNP Detection Using an Anthracene Probe

Table 9.1 MS characterisation and analytical HPLC retention times of oligonucleotides used for SNP and base modification studies. "Oligo Gen 60" refers to the Oligo General 60 minute method described in the experimental section along with the DMT On method. Those marked with "*" were performed by Dr. Zheng-Yun Zhao.

Sequence	Predicted Mass (m/z)	Observed Mass (m/z)	HPLC Method	Retention Time (Rt)
Alz-Pr(1L)	4641	4642	Oligo Gen 60	38.992
Alz-Pr(1D)	4641	4641	Oligo Gen 60	37.003
Alz-Pr(3L)*	4669	4669	DMT On	15.001
Alz-Pr(3D)	4669	4669	DMT On	15.097
Alz-Pr(4L)	4683	4683	DMT On	15.888
Alz-Pr(4D)	4683	4683	DMT On	16.080
Alz-Pr(5L)	4697	4697	DMT On	16.336
Alz-Pr(5D)	4697	4697	DMT On	17.025
Alz-Pr(6L)	4711	4710	DMT On	17.670
Alz-Pr(6D)*	4711	4711	DMT On	17.441
Alz-Pr(7L)*	4725	4725	DMT On	19.052
Alz-Pr(7D)*	4725	4725	DMT On	19.130
Alz-Pr(UM-T)*	4543	4543	Oligo Gen 60	30.092
Alz-Pr(UM-G)*	4568	4568	Oligo Gen 60	26.721
Alz-TarA	4578	4577	Oligo Gen 60	25.852
Alz-TarC	4602	4601	Oligo Gen 60	24.685
BM-Pr(1L)*	4720	4720	Oligo Gen 60	39.852
BM-Pr(1D)	4720	4720	DMT On	37.259
BM-Pr(3L)*	4748	4748	DMT On	15.250
BM-Pr(3D)	4748	4748	DMT On	15.415
BM-Pr(4L)	4762	4762	DMT On	16.215
BM-Pr(4D)	4762	4762	DMT On	16.240
BM-Pr(5L)	4776	4776	DMT On	16.768
BM-Pr(5D)	4776	4776	DMT On	17.050
BM-Pr(6L)*	4790	4790	DMT On	16.872
BM-Pr(6D)*	4790	4790	DMT On	17.863
BM-Pr(7L)*	4804	4805	DMT On	18.458
BM-Pr(7D)*	4804	4804	DMT On	18.497
BM-TarG*	4536	4536	Oligo Gen 60	33.392
BM-TaroxoG*	4552	4552	Oligo Gen 60	30.051

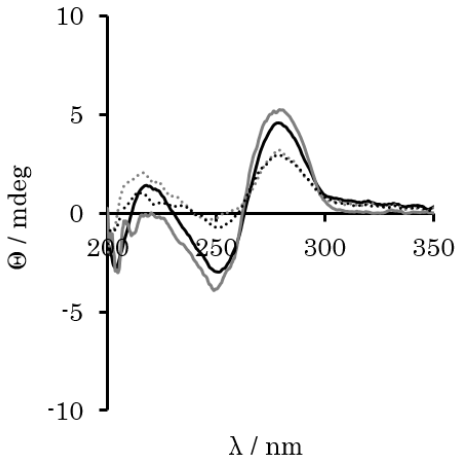


Figure 9.1 Circular dichroism spectra of **Alz-Pr(UM-G)** (grey) and **Alz-Pr(UM-T)** (black) as ssDNA (dotted) and duplexed with target **Alz-TarC** and **Alz-TarC** respectively (solid). All spectra shown are recorded with 1 µM probe and target, where applicable, in 10 mM sodium phosphate buffer pH 7 and 100 mM NaCl at 293 K.

9.2 Controlling DNA Binding using an Anthracene Tagged Peptide

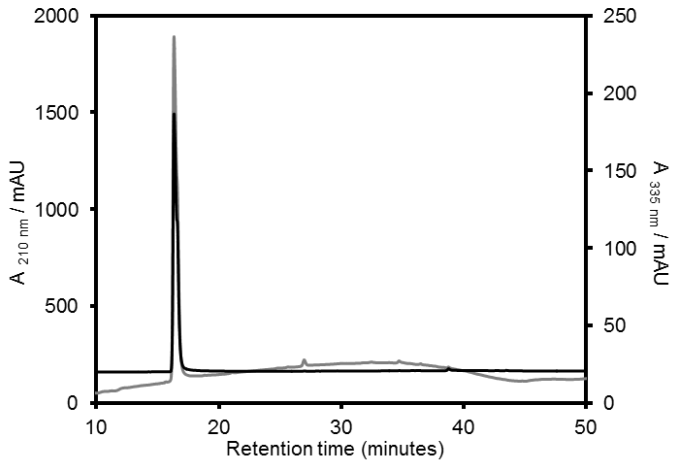


Figure 9.2 Analytical HPLC of anthracene tagged peptide. Run using gradient of 0-100% water-MeCN with 0.1% TFA over 40 minutes, followed by a 10 minute hold at 100% MeCN with 0.1% TFA. Monitored at 210 nm (black) and 335 nm (grey) for peptide and anthracene detection respectively.

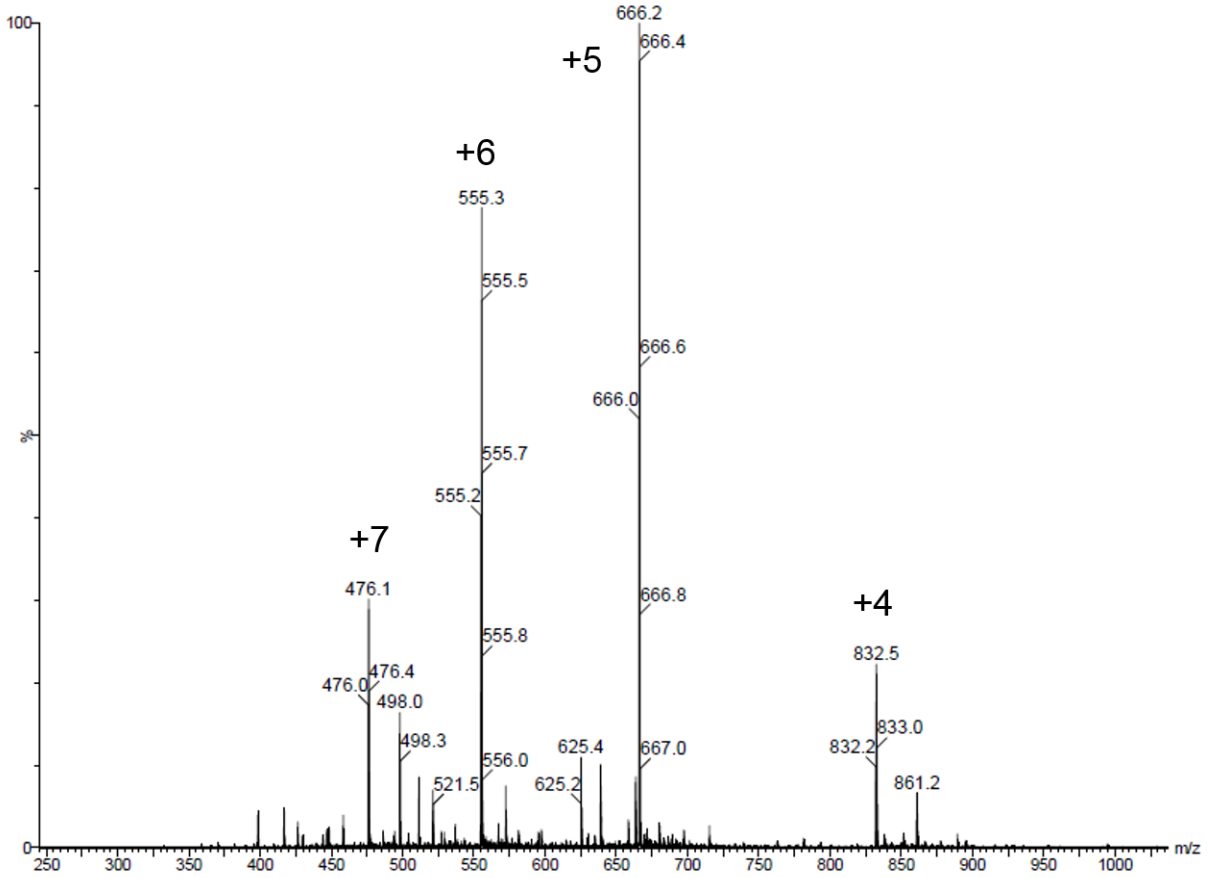


Figure 9.3 ESI-MS spectrum of purified anthracene tagged peptide, molecular weight 3326.

Table 9.2 MS characterisation and analytical HPLC retention times of oligonucleotides used for studies in anthracene tagged peptide binding work. "Oligo Gen 60" refers to the Oligo General 60 minute method described in the experimental section.

Sequence	Predicted Mass (m/z)	Observed Mass (m/z)	HPLC Method	Retention Time (Rt)
CRE 1	6086	6087	Oligo Gen 60	32.256
CRE 2	6148	6148	Oligo Gen 60	32.779
Half CRE 1	6179	6179	Oligo Gen 60	37.327
Half CRE 2	6055	6055	Oligo Gen 60	36.148
NS 1	6179	6179	Oligo Gen 60	31.775
NS 2	6055	6057	Oligo Gen 60	32.491

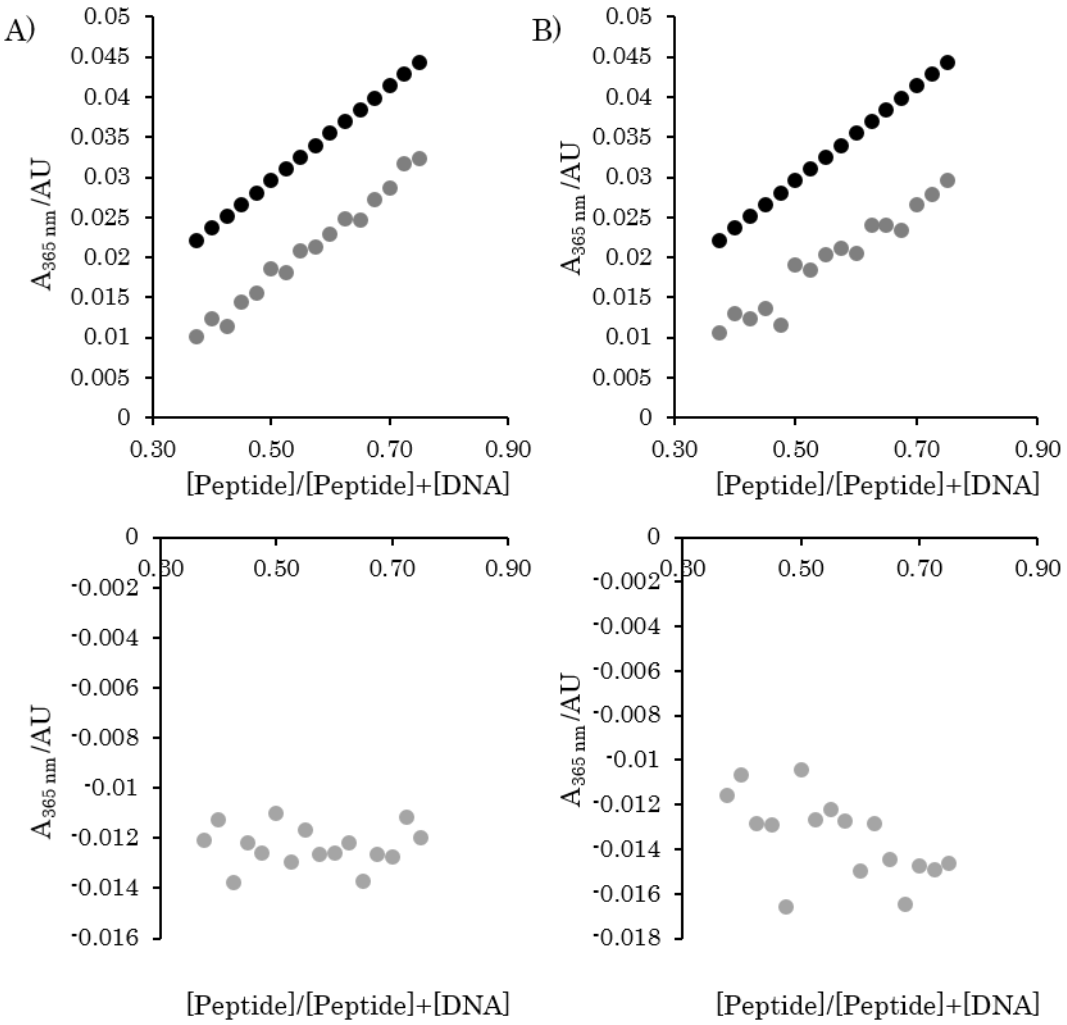


Figure 9.4 Job plot to deduce the stoichiometry of anthracene tagged peptide binding to CRE and NS DNA. A) CRE DNA, B) NS DNA. The top graphs show the theoretical absorbance at 365 nm due to the change in anthracene concentration (black) and the raw data from the Job Plot (grey). The bottom graphs show the raw data with the theoretical absorbance subtracted. All performed in 100 mM NaCl and 10 mM sodium phosphate buffer at pH7.

9.3 Photo-Triggered Release of DNA

Table 9.3 MS characterisation and analytical HPLC retention times of oligonucleotides used for photo-triggered release studies. "Oligo Gen 60" refers to the Oligo General 60 minute method described in the experimental section along with the DMT On method. Those marked with "*" were performed by Jack Manchester.

Sequence	Predicted Mass (m/z)	Observed Mass (m/z)	HPLC Method	Retention Time (Rt)
SL*	8685	8688	DMT On	21.201
OH1	5836	5837	Oligo Gen 60	32.332
OH2	5234	5235	Oligo Gen 60	25.872
OH3	5203	5204	Oligo Gen 60	29.084
T1	8361	8363	DMT On	16.021
T2	8361	8363	DMT On	15.938
PT1	6130	6129	DMT On	16.654
PT2	6114	6114	DMT On	17.399
PT3	6130	6129	DMT On	17.233
PT4	6114	6114	DMT On	16.352
C1-S	5966	5965	Oligo Gen 60	16.200
C1-M	7739	7738	Oligo Gen 60	39.419
C1-L	8028	8026	Oligo Gen 60	39.388
C2	5971	5970	Oligo Gen 60	21.721
TCon	8195	8195	Oligo Gen 60	14.972

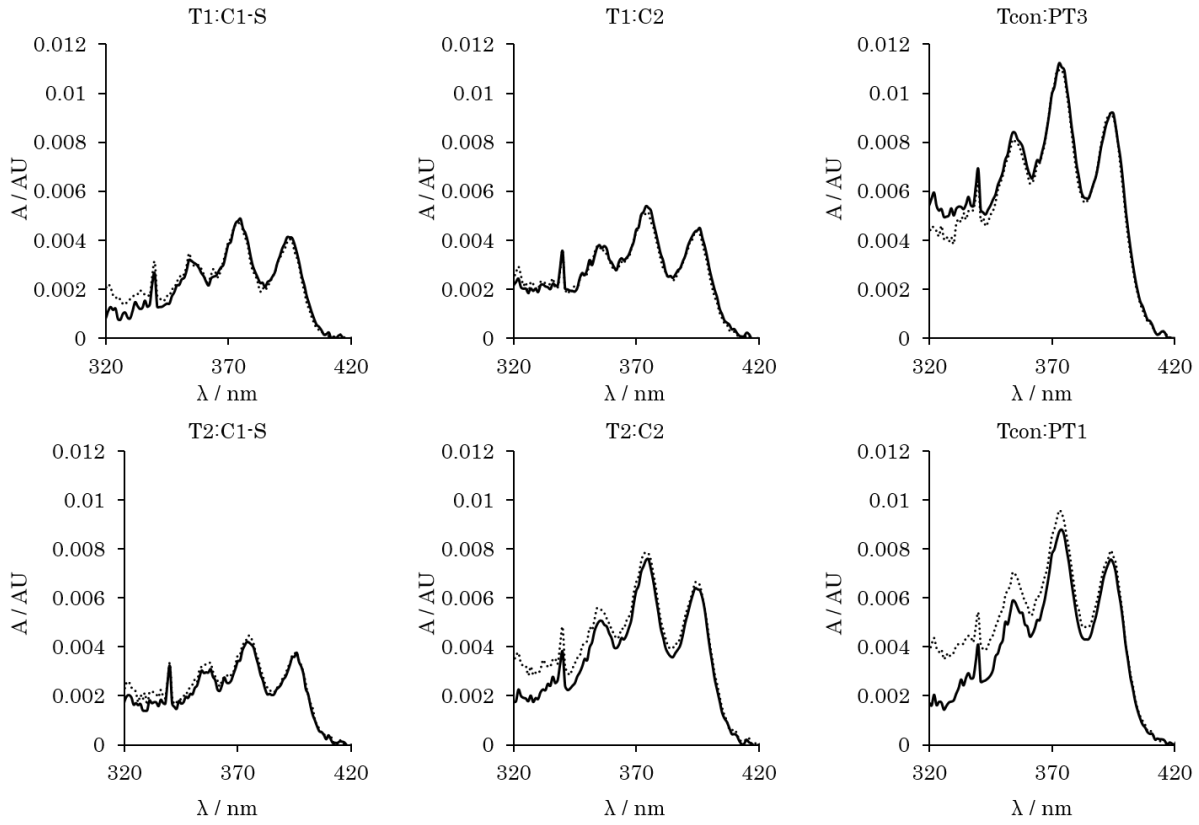


Figure 9.5 UV-Vis spectra of anthracene containing duplexes. All performed at 1 µM duplex in 100 mM NaCl and 10 mM sodium phosphate buffer pH 7 at 293 K. Samples were recorded at 0 minutes (solid) following degassing with argon for 10 minutes they were then irradiated with 365 nm light for 30 (dotted).

9.4 Controlling DNA Folding using Anthracene Photodimerisation

Table 9.4 MS characterisation and analytical HPLC retention times of oligonucleotides used for thrombin binding aptamer work. "Oligo Gen 60" refers to the Oligo General 60 minute method described in the experimental section along with the DMT On method. All were performed by Jack Manchester.

Sequence	Predicted Mass (m/z)	Observed Mass (m/z)	HPLC Method	Retention Time (Rt)
TBA	4726	4726	Oligo Gen 60	26.901
M1-TBA	6277	6277	DMT On	22.823
M2-TBA	5060	5060	DMT On	20.912
M3-TBA	5060	5060	DMT On	20.513