UNIVERSITY OF Southampton

Faculty of Natural and

Environmental Sciences

School of Chemistry

Functionalised DNA - Introducing & Applying a Versatile Porphyrin Molecular Ruler

Jonathan R. Burns

Thesis for degree of doctor of philosophy

March 2012

Abstract

Porphyrin moieties were rigidly attached to DNA to generate an accurate molecular ruler. Molecular ruler analysis was conducted using steady-state fluorescence, circular dichroism and small angle X-ray scattering spectroscopic techniques, in an attempt to analyse the FRET, exciton coupling and scattering intensity between different porphyrin-porphyrin labelled DNA combinations. A 21-mer test sequence was labelled with a porphyrin in one position on one strand, and seven different positions on seven complementary strands, to overall give seven porphyrin-porphyrin inter-strand combinations. Steady-state fluorescence and circular dichroism spectroscopic analysis of the Soret band revealed individual Watson-Crick bases pair molecular ruler sensitivity. Small angle X-ray scattering attempts between metallated-porphyrin entities did not reveal sufficient scattering at low concentrations, in contrast, an iodinated analogue of the porphyrin system did displayed scattering correlating to different iodine-iodine distances.

After calibration of the porphyrin system, the moieties were applied to study protein-DNA interactions between Tus, a 36 KDa DNA binding protein, and *Ter*, a specific 21-mer DNA sequence. Molecular ruler analysis of the complex required an extended version of the *Ter* DNA sequence to which modifications were attached. Established FRET pairs FAM and TAMRA were applied to investigate protein-DNA complexation. Native PAGE analysis revealed Tus binds to the extended DNA *via* a sliding mechanism. Fluorescence analysis of the established FRET pairs identified changes in fluorescence not correlating to changes in FRET, and instead was attributed to emission quenching upon protein binding. Applying the zinc and free base porphyrin version displayed subtle changes in the Soret band circular dichrosim upon complexation, indicating small DNA helical change upon complexation.

A 45-mer DNA sequence was designed to form multiple hairpin-duplex conformations with the addition of an appropriate complementary strand. Attaching FRET pairs to the extremes of the DNA sequence enabled multiple DNA conformations, and hence FRET distances to be obtained from one doubly modified DNA sequence. The combinations were characterised by UV-Vis, fluorescence and circular dichroism spectroscopy.

Finally, terpyridine labelled DNA sequences selectively formed DNA nanotubes through orthogonal hydrogen bonding and metal complexation interactions. Short DNA strands were designed to self-assemble into long duplexes through a sticky-end approach. Addition of weakly binding metals such as zinc induced the formation of tubular arrays consisting of DNA bundles 50-200 nm wide and 2-50 nm high. TEM displayed additional long distance ordering of the terpyridine-DNA complexes into fibers.

1 Introduction

2 Zinc and 2H porphyrin energy transfer

Author's Declaration

I, Jonathan R. Burns declare that the thesis entitled:

Functionalised DNA - Introducing & Applying a Versatile Porphyrin Molecular Ruler

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

1) Bandy, T. J.; Brewer, A.; Burns, J. R.; Marth, G.; Nguyen, T.; Stulz, E., DNA as supramolecular scaffold for functional molecules: progress in DNA nanotechnology. *Chemical Society Reviews* **2011,** *40* (1), 138-148.

2) Burns, J. R.; Zekonyte, J.; Siligardi, G.; Hussain, R.; Stulz, E.; Direction Formation of DNA Nanoarrays through Orthogonal Self-Assembly. *Molecules* **2011**, 16, 4912-4922.

3) Singleton, D. G.; Burns, J. R.; Fox, K. R.; Siligardi, G.; Hussain, R.; Varnai, P.; Stulz, E.; Modulation of G-quadruplex structure and stability through selective covalent modification with porphyrins. *Manuscript in preparation*.

4) Burns, J. R.; Rindermann, J.; Preus, S.; Singleton, D. G.; Hussain, R.; Siligardi, G.; Lagoudakis, P.; Stulz, E.: DNA-scaffold modulated energy transfer in mixed-metal porphyrin arrays. *Manuscript in preparation.*

5) Burns, J. R.; Singleton, D. G.; Preus, S.; Siligardi, G.; Hussain, R.; Neylon, C.; Stulz, E.: DNA based adjustable strap analysed by FRET and CD. *Manuscript in preparation.*

Signed: ……..

Date:…….

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Eugen Stulz for his excellent supervision and guidance. A special thanks to Dr. Cameron Neylon for his supervision in the expression and analysis of Tus. Thanks to Dr. Giuliano Siligardi and Dr. Armin Hoell for CD and SAXS beamline support respectively. Special thanks to Dr. Dorota Bartczak for support throughout and TEM assistance. Thanks to Dr. Luke Clifton for help with SAXS analysis and protein purification, Dr. Thomas Rühl for purifying the terpyridine-DNA strands, Junis Rindermann for modelling the porphyrin FRET prerequisite values, Dr. Søren Preus for modelling the porphyrin FRET distances and Dr. Jurgita Zekonyte for AFM analysis. I thank Dr. Warren Duffy for proof reading this thesis. Sincere thanks to the EPSRC and ATDBio Ltd for financial support.

A big thank you to past and present Stulz group members for making a great lab atmosphere, in particular Dr. Ashley Brewer, Thomas J. Bandy, Dr. ThoaNguygen Nguygen and Daniel Singleton.

Finally, I thank my all my friends and family, in particular my parents for their continued support throughout.

Abbreviations

2H	Free base
A	Adenine
Abs	Absorption
aq.	Aqueous
AIBN	Azo-*bis*-isobutyronitrile
AS	Adjustable Strap
br. s	Broad singlet
C	Cytosine
CD	Circular dichroism
Concn.	Concentrated
CEP	β-Cyanoethyl phoshoramidites
d	Douplet
DCM	Dichloromethane
DDQ	2,3-Dicayno-5,6-dichloro-1,4-benzoquinone
DIPEA	*N*-Diisopropanolethylamine
DNA	Deoxyribonucleic acid
DMF	*N*,*N*-Dimethylformaldehyde
DMT	4-4'-Dimethoxytrityl
EA	Ethyl acetate
ε	Extinction coefficient
EDTA	Ethylene diamine tetraacetic acid
E_{FRET}	FRET efficiency
E.coli	Escherichia coli
FAM	Fluorescein
FRET	Förster resonance energy transfer
F_D	Corrected fluorescence constant
G	Guanine
HFIP	Hexafluoroisopropanol

His^6	Histidine tag
HPLC	High pressure liquid chromatography
HMTA	Hexamethylenetetramine
hr	Hour
IPTG	Isopropyl-thiogalactopyranoside
κ	dipole-dipole orientation variable
λ	Wavelength
M	Concentration
m	Multiplet
mdeg	Millidegrees
m / z	Mass / charge
MALDI	Matrix assisted laser desorption / ionisation
mRNA	Messenger ribonucleic acid
NBS	N-Bromosuccinimide
n	Refractive index of the medium
NMR	Nuclear magnetic resonance
NoE	Nuclear Overhauser effect
Q	Scattering intensity
Q_D	Donor quantum yield
PPh_3	Triphenylphoshine
PCR	Polymerase chain reaction
r	distance
RNA	Ribonucleic acid
R_f	Retention factor
R_0	Forster distance
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
s	singlet
S_0	Ground state
S_1	First excited state
SSF	Steady state fluorescence

T	Thymidine
t	Triplet
TAMRA	Tetramethylrhodamine
TCAA	Trichloroacetic acid
Ter	Termination DNA sequence
TEA	Triethyl amine
TEAA	Triethylammonium acetate
TFA	Trifluoroacetic acid
TEM	Transmission electron miscropy
TLC	Thin layer chromatography
ToF	Time of flight
τ_D	Lifetime of the donor excited state
UV-Vis	Ultra violet-visible
U	Uridine

Introduction

1.1 Primary research objective

The primary aim of the research was to generate an accurate DNA molecular ruler using porphyrin moieties. High precision would be achieved by rigidly attaching porphyrins to DNA, as opposed to conventional long and flexible linkers which can result in large errors and poorly defined molecular distances. Steady-state fluorescence, circular dichroism and small angle X-ray scattering spectroscopic techniques were employed to identify the porphyrin molecular ruler distances. After calibration of the molecular ruler, the ruler was applied to investigate a protein-DNA complex to identify information about the binding mechanism and protein induced conformational change. The objectives of the research were categorised into five areas:

- Design accurate molecular ruler probes using porphyrin moieties.
- Porphyrin monomer synthesis and incorporation into DNA.
- Attempt to obtain versatile base-pair molecular ruler resolution using three spectroscopy techniques.
- Apply the porphyrin DNA molecular ruler to monitor complex formation and identify protein induced conformational change.

1.2 Introduction overview

Firstly, the definition and concept of a molecular ruler is explained (section 1.3). The ruler was applied to investigate DNA and any structural changes induced onto DNA, therefore, the opening section is followed by a detailed introduction into the components of DNA, and the structure it can form in solution (section 1.4). After calibration of the molecular ruler, the ruler was further tested and applied to study protein-DNA interactions. Therefore, the DNA section is followed by an introduction to proteins and their elementary building blocks, the amino acids (section 1.5).

The porphyrin DNA molecular ruler was analysed using multiple light sensitive techniques (section 1.6), including: ultraviolet-visible (UV-Vis), steady-state fluorescence (STF), circular dichroism (CD) and small angle X-ray scattering spectroscopy (SAXS). After an introduction to absorption and emission phenomena, a detailed section on the primary form of molecular ruler analysis is discussed - Förster resonance energy transfer (FRET). The key concept of FRET is discussed and how it is applied to cutting edge research through the use of single molecule analysis. The effect of the linker between the light sensitive moiety and the scaffold on the FRET efficiency is explained through the dipole-dipole orientation factor, the κ^2 term. Finally, the four methods used to extract the FRET term to obtain a molecular ruler distance is discussed. The FRET section is proceeded by introductions to the two other forms of molecular ruler analysis applied within; CD and SAXS spectroscopy techniques. CD molecular ruler analysis was conducted by analysing the strong exciton coupling effect between porphyrin moieties. SAXS molecular ruler analysis was applied to the porphyrin DNA constructs in attempts to identify the direct metal-to-metal scattering between two zinc porphyrin labels.

After the different forms of molecular ruler analysis section, an introduction to the probes positioning and method of attachment to DNA is discussed (1.7). This section is followed by an introduction to the phosphoramidite chemistry method of making and purifying synthetic DNA (section 1.8). The molecular ruler probes used within are introduced (section 1.9). Firstly, the porphyrin molecules are introduced, including their properties, use in nature and use by scientists, followed by how the porphyrins can be applied for use as accurate molecular ruler probes. The well-

established FAM and TAMRA FRET pairs introduction is followed, including spectral details and typical Förster distances. FAM and TAMRA probes were used as FRET pair analogous to the porphyrin system. These FRET pairs enabled comparison of the rigid and flexible linker effect on the molecular ruler accuracy. An introduction to the Tus-*Ter* complex is followed (section 1.10). The Tus-*Ter* complex is the protein-DNA complex to which the molecular ruler was applied to study. The X-ray crystal structure of the complex is shown, along with relevant distances highlighted, therefore giving direct comparison between crystals distances and those obtained in solution using FRET.

1.3 The molecular ruler

A molecular ruler is a device that allows distances to be calculated between two coupling moieties at the molecular scale[1, 2, 3, 4]. These two positions were traditionally labelled with fluorophores which interact with each other *via* the Förster resonance energy transfer (FRET) mechanism, as pioneered by Stryer *et al.* in 1967[1]. Figure 1 displays a schematic representation of two coupling entities separated by a fixed distance (r). The two coloured probes are attached to a backbone scaffold using linkers. If the linkers and scaffold are rigid, and FRET molecular ruler experimentation is performed, FRET analysis should yield a fixed distance. This information can then be used to deduce the structure of the scaffold. If the scaffold is for example a DNA duplex or a protein, then distance information about the scaffold can be obtained, which can be useful to identify binding mechanisms or key conformational changes[5].

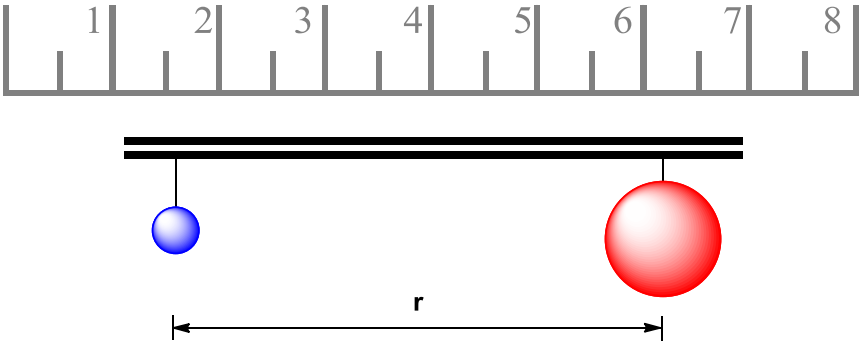


Figure 1 Example of a molecular ruler, two coupling moieties (blue and red spheres), attached to a scaffold (scaffold = thick black lines, linkers = thin black lines). Coupling moieties are separated by a fixed distance (r), with ruler analogy (grey).

The distance information between atoms and molecules can be identified using other techniques, for instance, within a crystal lattice using X-ray crystallography[6]. However, for X-ray crystallography to be applied, crystals are required. Due to the large number of rotational bonds present in biological systems, for example: proteins, DNA and membranes, crystallisation attempts are challenging. Also, the crystal phase does not allow for the study of dynamic interactions, such as the identification of host-

guest binding information. Whereas, a molecular ruler based on two coupling entities attached to a host using FRET enables the probes to be monitored over time, for example, as a function of guest added. Any changes imposed on the system upon binding could potentially enable binding information to be identified. The probes could be monitored as a function of temperature, to potentially reveal thermodynamic information, as the constructs melt and anneal[7]. NMR can be used to identify the solution phase structure of atoms in proteins and DNA, for example Liao and co-workers analysed a protein-DNA complex and were able to determine the complexes structure[8]. Other techniques where any readout changes depending on the distance and / or angle separation can also be regarded as a molecular ruler, for example, exciton coupling between two spectroscopically similar chromophores, as analysed by circular dichroism (CD), or small angle X-ray scattering (SAXS) involving electronic scattering between two electron rich entities, amongst many other techniques including: NMR (NoE), Dexter (electron transfer) etc., all have distance and angle dependencies.

Deoxyribonucleic acid (DNA) is ubiquitous in nature, and is the source of storing genetic information. This has resulted in extensive research into the structure of DNA[10]. The results of which are now being exploited by scientists from many fields, including: supramolecular chemists, biochemists and biophysicists[11]. DNA has a biologically stable and predictable structure; the duplex of DNA can be considered a rigid cylinder up to 100 nm. This has lead to DNA's use as a scaffold link between two coupling entities for use a molecular ruler[2, 3]. The smallest possible increments using probes attached to DNA is the individual base-pairs themselves, sensitivity of which was applied in this research project.

Attention was paid to the modifications attachment to DNA, the linker. The link between the coupling moiety and scaffold plays an important part for maximum molecular ruler accuracy[12]. If the linker is long and flexible, resolution is reduced since the modification occupies a large area, and therefore distances obtained are approximated over a greater area. Small and rigid linkers are therefore highly desirable, for example, an acetylene linker compared to a hexyl linker[13]. Figure 2 (α) shows this effect; the hexyl linker is relatively long and flexible due to the high number of relatively unhindered rotational CH_2 sp^3 hybridised bonds. The hexyl linker of the deoxyuridine based monomer shown on the left has an approximate cone shaped area

where the modification occupies. The population of occupation can be calculated to better define the linkers location, as shown by Brown *et al.*[12]. Applying the short acetylene linker enables cylindrical rotation around this triple sp hybridised bond, and so the modification has a smaller area of occupation (Figure 2 (β)), ideal for maximum molecular ruler resolution[7].

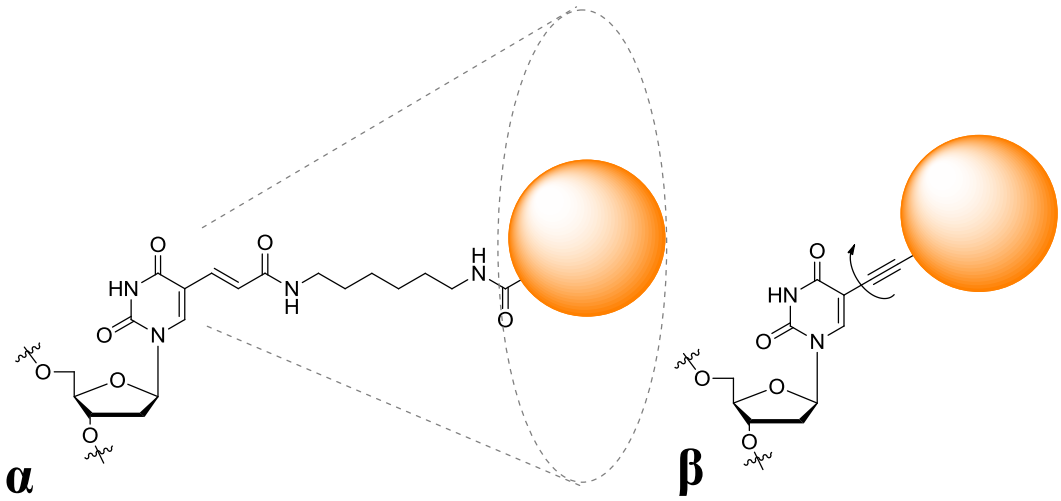


Figure 2 The 'flexible' hexyl and 'rigid' acetylene linkers of deoxyuridine (α and β respectively), approximate modification area of occupation represented by dashed lines for the hexyl linker. Modifications represented as orange spheres.

1.4 Deoxyribonucleic acid

The molecular ruler was applied to study conformational changes induced on deoxyribonucleic acid (DNA). DNA is found in nature and is used to store genetic information; DNA is either replicated or transcribed into messenger ribonucleic acid (mRNA) which is in turn translated into peptides and proteins[10, 14]. The generic structure of DNA was discovered in 1953 by Francis Crick, James Watson in 1953[15]. DNA is constructed from four heterocyclic (aromatic) bases, two of the bases are purines, and the other two are pyrimidines (see 1.4.1). The purines can form Watson-Crick base pairs to the pyrimidine via hydrogen bonds interactions (see 1.4.2). Each hydrophobic base can interact with another base above and below the bases π system via an attractive π-π stacking interaction (see 1.4.3). The bases are attached to a furanose ring which contains 3' and 5' oxygen atoms, known as deoxynucleosides. These oxygens can link together via a phosphodiester bond to other deoxynucleosides, which are known as deoxynucleotides. Multiple deoxynucleotides can be joined together to form a DNA single strand (see 1.4.4). If strands with complementary bases come into close proximity, then the two strands can combine via Watson-Crick base pairs and π stacking interactions to generate the double helix structure (1.4.4).

1.4.1 DNA bases

DNA is comprised of four heterocyclic bases: Adenosine (A), Guanine (G), Cytosine (C) and Thymidine (T). A and G are purines whilst C and T are pyrimidines. The bases are comprised of carbon, oxygen and nitrogen atoms, which give rise to different functional groups and hence different properties. Each of the four bases are attached to a furanose ring (monosaccharide), the bases are attached to the 1' position of the sugar, also known as the glycosidic bond[10]. At the 5' and 3' positions are two hydroxyl groups as shown in Figure 3.

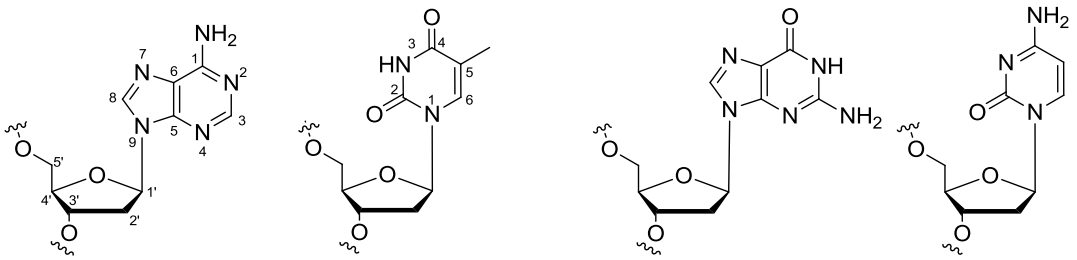


Figure 3 DNA nucleosides: Adenosine, Thymidine, Guanine and Cytidine, with labelled positions for purines, pyrimidines and furanose ring (left to right).

1.4.2 Watson-Crick base pairs

The four bases contain amine and carbonyl functional groups. These functional groups on the bases can interact with each other non-covalently *via* hydrogen bonding and electrostatic interactions[10]. A hydrogen bond is the interaction of a hydrogen atom with an electronegative atom such as: oxygen or nitrogen; maximum hydrogen bonding occurs when the interaction angle between the two entities is 180°. The amine functional group is a good hydrogen bond donor, whilst the carbonyl group is a good hydrogen bond acceptor. The two can combine to form a hydrogen bond, as shown in Figure 4. The average hydrogen bond strength correlates to approximately 6-10 kJ mol^{-1}. The amines and carbonyls from the bases can form hydrogen bonds; adenosine forms

Watson-Crick base-pairs with thymidine to give two hydrogen bonds, whilst guanine binds to cytosine to give three hydrogen bonds, hence G-C base pairs are stronger than A-T base pairs; this effect explains why A-T rich regions melt before G-C rich regions[10].

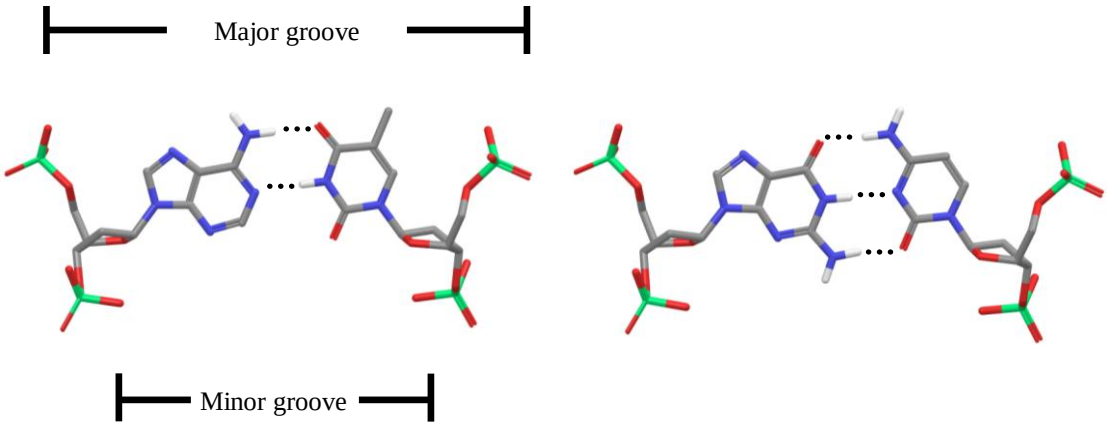


Figure 4 Watson-Crick base pairs, A-T (left) and G-C (right), hydrogen bonds (black dashed line), carbon (grey), nitrogen (blue). Oxygen (red), phosphorus (green) and polar hydrogen (white). Major and minor grooves positions annotated.

1.4.3 Base pair stacking

The stacking interaction between Watson-Crick base pairs is related to each individual bases π system. The bases are aromatic, and are comprised of double and single bonds. Double bonds have electric charge situated in a π orbital which is located above and below the σ bond [16]. The π bond in general is able to delocalise around an aromatic molecule, giving rise to a resonance form. If the molecule is homoatomic, such as benzene, then the electronic orbitals are symmetrical around the molecule, as shown in the left-hand side of Figure 5. The net effect of this electron density induces polar regions on the molecule. More electron density is located above and below the plane of the molecule than at the sides. This arrangement induces a quadrapole moment into the molecule, as shown in the right-hand side of Figure 5[10, 17]. Induced negative

charge is located in the π system above and below the ring, leaving electron deficient regions perpendicular. For the Watson-Crick bases, the actual electron density will be different to that shown in Figure 5, the nitrogen and oxygen atoms perturb the even distribution, but the induced quadrupole moment remains.

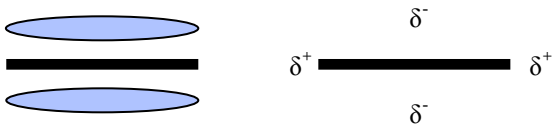


Figure 5 Electron density of the π orbital (blue ovals) and σ orbital looking through the plane of an homoatomic molecule (thick black lines, left), and subsequent quadrupole moment (right).

Each Watson-Crick base can interact with another base via a through space π-π stacking interaction, as shown in Figure 6[10]. The bases are staggered and do not lie directly above, this leads to a favourable electrostatic interaction between multipole moments. Each δ^-- δ^+ interaction is weak, however this interaction is additive, and a large stabilisation energy is produced if multiple bases stack.

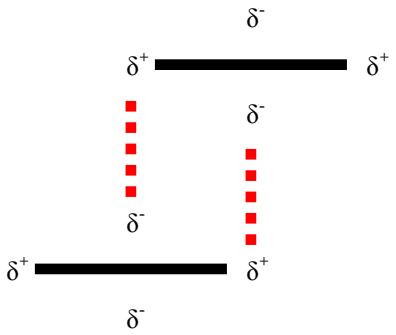


Figure 6 Higher ordered structural arrangement of bases (thick black lines), favourable electrostatic interactions shown in red.

1.4.4 Phosphodiester backbone

In DNA, the oxygen atoms on the sugars at the 5' and 3' positions are involved a larger DNA structural alignment property; the phosphodiester backbone, as shown in blue in Figure 7. Each phosphate group contains negatively charged oxygen atoms which can form ionic bonds with positively charged counterions such as sodium, potassium or magnesium[10]. Magnesium is a divalent cation and can bind to two oxygen anions to stabilise the DNA's duplex (see below) to a greater extent as opposed to potassium, which in monovalent, and only binds to one oxygen. The phosphodiester backbone and DNA bases are arranged in a 5' → 3' configuration along the phosphodiester bond to form a DNA single-strand.

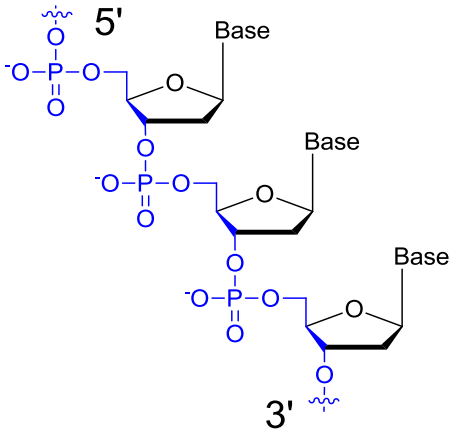


Figure 7 DNA phosphodiester directionality shown in blue.

1.4.5 DNA double helix

The Watson-Crick base-pairing and stacking interactions, and the phosphodiester backbone of DNA gives rise to a higher ordered structural arrangement; the DNA double helix, Figure 8. If two DNA single-strands are combined consisting of the correct Watson-Crick base paring, then the complementary strands can combine to form the double helix structure. Each strand is arranged in an anti-parallel 5' → 3' direction. The phosphodiester backbone forms the external hydrophilic structure (phosphorus atoms shown in purple), which are linked to the sugars *via* the 5' and 3' oxygen positions, approximately perpendicular to the phosphodiester backbone. The internal hydrophobic bases are attached to the sugars at the 1' position. The hydrophobic centre is stabilised by Watson-Crick base pairing, π-π stacking, and dipole-dipole interactions from the bases situated above and below[10]. Each base-pair is rotated between the base-pairs above and below by approximately 36° to maximise the π-π stacking interaction[10]. The distance between each stack is approximately 3.6 Å (the actual distance depends on the base-pairs and solvent environments in question), this correlates to about 10.5 base-pairs per full helical turn (~ 38 Å)[10]. The double helix contains major and minor grooves, as shown in Figure 8, which are derived from the alignment of the bases as shown in Figure 4.

The sugars linking the bases and phosphate groups can adopt multiple conformations, with the extremes being the C2' endo conformation and the C3' endo conformation[10, 12]. In nature, double stranded DNA typically adopts a C2' endo-conformation termed the B form of DNA. However, in high salt concentrations, DNA can adopt a C3' endo conformation, which is described as the A form of DNA. B-DNA is wider and denser than A-DNA. DNA can therefore have the same sequence but adopt different structural conformations depending on the solvent conditions employed[10, 12].

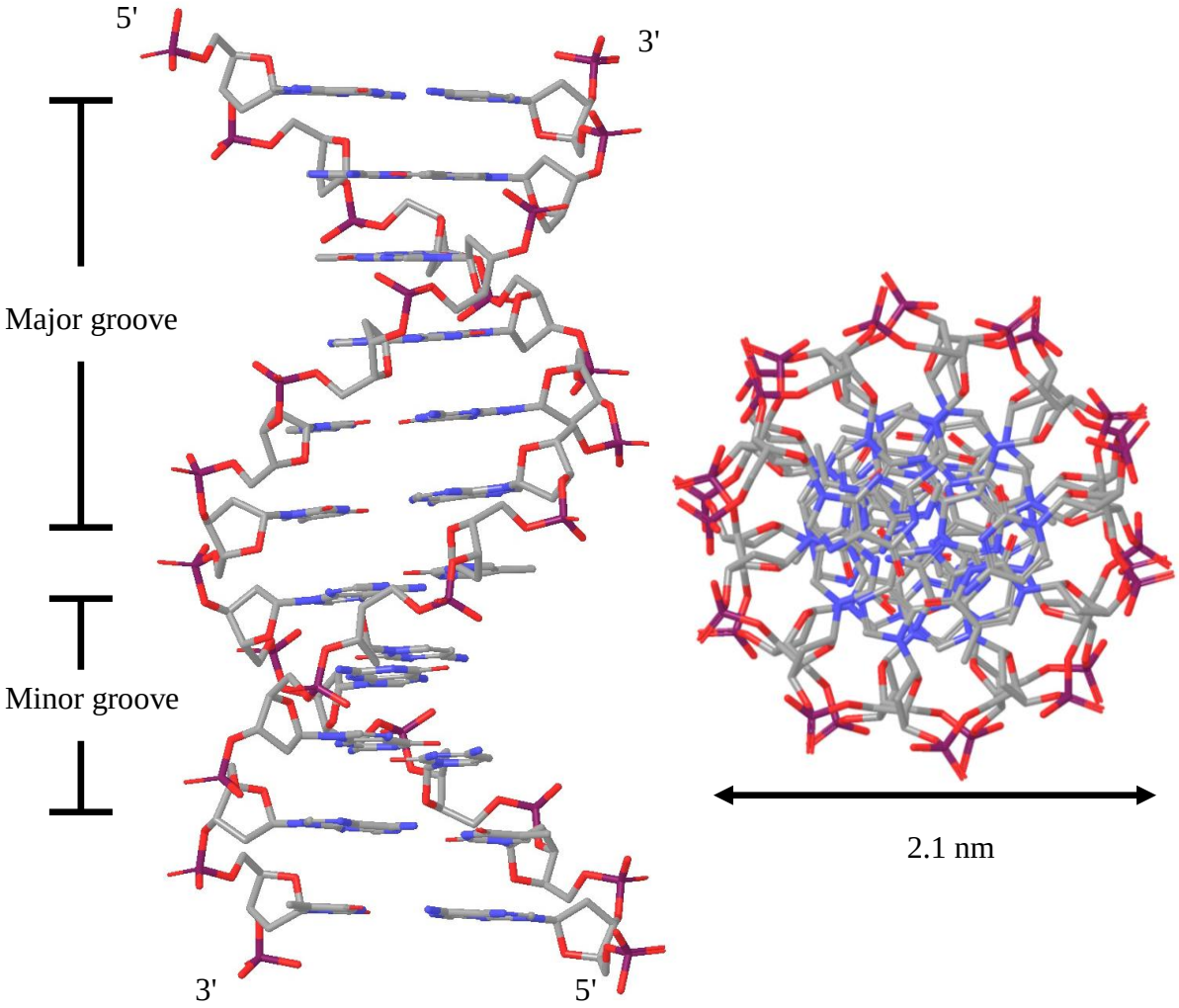


Figure 8 DNA double helix side and top views, phosphodiester backbone shown as tubes, bases shown as wires, major and minor grooves are annotated, hydrogen atoms are omitted for clarity, images generated using Maestro 9.2 modelling software from Schrödinger.

1.5 Proteins

Proteins are biology's molecular machines[14]. Proteins are comprised of a linear chain of amino acids; there are twenty different natural amino acids which are linked in a protein via a peptide bond. Proteins are created by enzymes, after multiple amino acid addition, the peptide chain generated begins to obtain a secondary structure, turning the peptide chain into a functioning protein. The twenty naturally occurring amino acids have the same generic structure and differ only by an intermediate R group (with the exception of proline). Each amino acid contains a carboxylic acid terminus and an amine terminus, as shown in Figure 9. The R groups have different functionality, some are hydrophobic such as alanine and tryptophan, others are hydrophilic, such as histidine and lysine, and some have specialised functionality, like proline, which has a secondary amine and can be used for conformational rigidity in proteins[9].

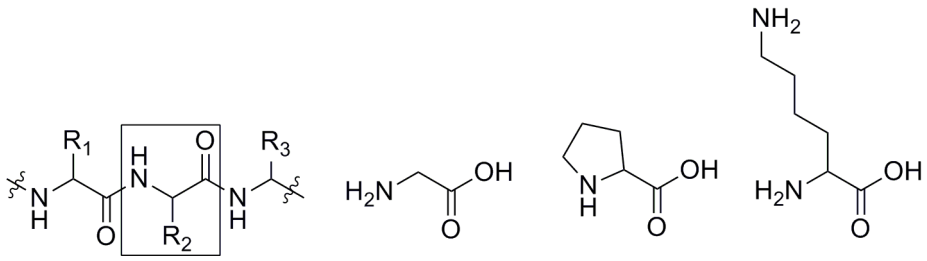


Figure 9 Generic amino acid structure (boxed region), glycine, proline and lysine amino acids (left-to-right) as examples. Amino acid amine terminus (NH_2, left of each structure as drawn) and carboxy terminus (COOH, right of each structure as drawn).

Sequential amino acids can give rise to succinct folding conformations *via* electrostatic interactions through the use of hydrogen bonds, hydrophobic and Van der Waal interactions, or with the use of covalent bonds *via* disulfide bridges. A carbonyl group from the amide bond can interact with an amine at every i + 4 amino acids to form a hydrogen bond (where i is the amino acid of interest). If multiple interactions occur, then the structure can form a helical secondary structure, described as an α-helix. The α-helix is found in many proteins and generally has a right-handed helical shape. Alternatively, the amino acids can interact through hydrogen bonds in a sheet format by stacking in a parallel or anti-parallel configuration; this structure is described as a β-

sheet. The α-helix and β-sheet are essential in giving proteins their structure, whereby positioning relevant amino acids, and hence relevant functional groups in the correct position for efficient protein-DNA complexation. Even a single amino acid mis-match during expression could seriously reduce a proteins activity (or conversely increase activity)[14].

1.6 Methods of molecular ruler analysis

Steady-state fluorescence, circular dichroism and small angle X-ray scattering spectroscopy techniques were used to identify the intramolecular distances between the molecular ruler probes. These techniques are spectroscopic in nature, therefore an introduction to the principle of spectroscopy - light absorption and emission is explained (see section 1.6.1). This section is followed by an introduction to FRET (see 1.6.2), CD (see 1.6.3) and SAXS (see section 1.6.4) spectroscopy. Each of these sections define their concepts and how these techniques can be used to extract the molecular ruler distances.

1.6.1 Absorption and emission

For a molecule to absorb light, the light applied to the system must be of the same frequency at which absorption takes place, i.e. the system and applied light (blue arrow) must come into resonance, see Figure 10[17]. In doing so, the system's electrons go from a ground state to an excited state, $S_0 \rightarrow S_1$. The newly formed excited state is of higher energy, this energy can be released by the system radiatively (red arrow) to return the molecule to the ground state. This process is known as fluorescence, the light emitted is of lower energy and longer wavelength. Organic molecules which absorb and emit light are known as fluorophores (note, chromophores are the same as fluorophore except they do not necessarily have to emit light, energy is lost *via* other relaxation pathway mechanisms).

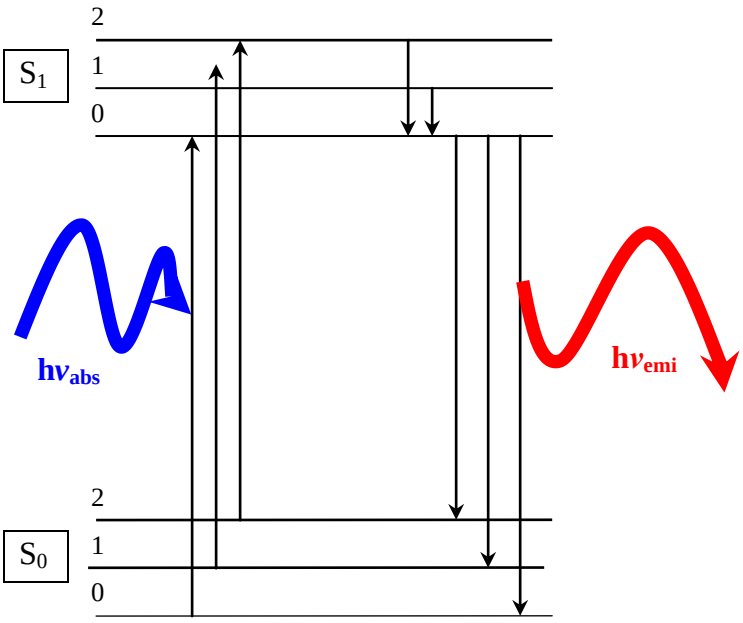


Figure 10 Absorption and emission diagram, ground state excitation of the molecule from S_0 → S_1 upon absorption of photonic energy of high energy / short wavelength (blue), vibrational sub-energy levels 0, 1 and 2 also shown. Radiative emission with emission of low energy / longer wavelength light S_1 → S_0 (red).

The frequency at which absorption and emission takes place can be changed for a molecule, for example, free base (2H) porphyrin has a main absorption around 424 nm, and main emission band at 656 nm. These spectroscopic properties are the result of 18 π electrons. Merely changing the solvent can drastically affect the absorption and emission intensity, or introducing electron withdrawing / electron donating groups to the aromatic system will change these spectroscopic values[16]. Absorption and emission properties can be monitored with UV-Vis and fluorometer spectroscopic instruments respectively. Both techniques are quantitative and can be used to identify the concentration of a sample[5, 10]. UV-Vis spectroscopy is used ubiquitously to identify the concentrations of DNA and proteins at 260 and 280 nm respectively (see section 1.8).

1.6.2 Förster resonance energy transfer

Förster resonance energy transfer (FRET) enables the distance between two coupling fluorophores to be identified. FRET occurs between 10 and 100 Å. The coupling efficiency has a strong distance dependency, which can be monitored by using fluorescence spectroscopy. The concept of FRET was first proposed by Förster in 1947[5, 18, 19]. He proposed that if two chromophores; a donor and acceptor chromophore; if the donor's emission overlaps with the acceptor's absorption, then energy can be transferred from the donor domain to the acceptor domain via a non-radiative energy transfer process. For energy to be transferred the two domains must come into resonance with each other. FRET occurs typically in the 10^{-12} s^{-1} timescale which is orders of magnitude slower than absorption and emission timescales (at approximately 10^{-9} S^{-1})[5]. Förster proposed that this energy transfer process has a r^{-6} distance dependency; when the distance separation is small, energy transfer in large, and therefore, acceptor emission is high. When the separation of the two chromophores is large, energy transfer is low, and therefore acceptor emission is reduced. This FRET distance range (10 to 100 Å) is suitable for analysis molecular biology interactions. This has lead to extensive research in developing FRET as an analytical tool to identifying biological distances and interactions, for example, its ubiquitous use in imaging, FRET in broad sense, i.e. FRET or no FRET (on or off), such as the work conducted by Brown *et. al* for FRET use a molecular beacon[20, 21], or to identify FRET distances in greater detail, for example, the work done by Clegg *et al.*, to identify the structural organisation of the Holliday junction, whose results gave way to precise conformational determination[22, 23, 24]. Since then multiple groups have successfully shown the molecular ruler aspects of FRET[3, 13, 25].

Figure 11 displays the typical transition mechanism pathways during the FRET process; firstly, the donor molecule absorbs photonic energy (blue arrow), the donor state goes from the ground state to the first excited state ($S_0 \rightarrow S_1$) by the absorption of high energy and short wavelength light. Provided the FRET prerequisites are satisfied (see below), energy is transferred without the emission of photons (non-radiatively, orange arrow) to the acceptor's singlet state (S_1). This requires the donor's emission orbitals and acceptor's absorption orbitals to resonate at the same frequency[5]. After

energy transfer, the donor returns to the ground state and the acceptor is promoted to the first excited state. The acceptor then relaxes to its ground state with the emission (red arrow) of this energy radiatively of lower energy and longer wavelength light.

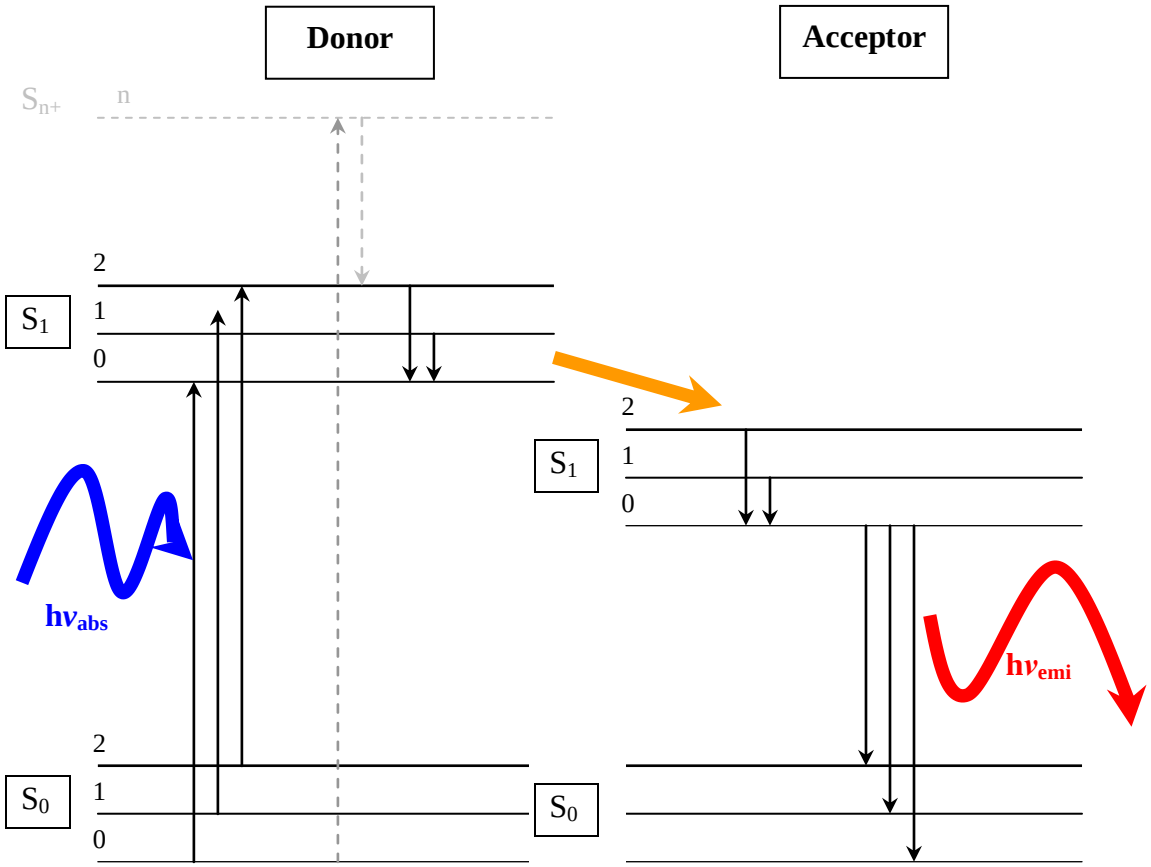


Figure 11 FRET energy profile (Jablonski diagram), ground state excitation of the donor from S_0 → S_1 upon absorption of photonic energy of high energy / short wavelength (blue), vibrational sub-energy levels 0, 1 and 2 also shown. Non-radiative energy transfer step between donor excited state and acceptor excited state S_1 → S_1 (orange). Radiative emission of the acceptor excited state with emission of low energy / longer wavelength light S_1 → S_0 (red).

1.6.2.1 FRET equation

The FRET efficiency is calculated using equation 1[5]. As predicted by Förster, the non-radiative energy transfer process has an r^{-6} distance dependency, and therefore FRET is effective over the 10 to 100 Å distance range. The exact FRET range depends on the spectral overlap value which unique for each FRET pair (see section 1.6.2.1.4). The energy transfer process is affected by other variables including: the relative dipole-dipole orientation factor of donor and acceptor (κ^2 - see section 1.6.2.1.1), the quantum yield of the donor (see section 1.6.2.1.2), the extinction coefficient of the acceptor (see section 1.6.2.1.3) and the refractive index of the medium, as shown in equation 1. With these values determined the Forster distance can be established (see section 1.6.2.2).

$$k_T(r) = \frac{Q_D \kappa^2}{\tau_D r^6}\left(\frac{9000(\ln 10)}{128\pi^5 N_A n^4}\right) \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \qquad \textbf{equation 1}$$

Q_D = Quantum yield of the donor in the absence of the acceptor

κ = Dipole-dipole orientation factor

τ_D = Lifetime of the donors excited state

r = Distance

N_A = Avogadro's number

n = Refractive index of the medium

$F_D(\lambda)$ = Corrected fluorescence instrument constant

ε_A = Extinction coefficient of the acceptor

κ_T = Transfer rate

1.6.2.1.1 κ^2 variable

Fluorophores have absorption and emission orbitals which are directly related to the dipole moment of the chemical structure[5]. These dipole moments of donor and acceptor vary the FRET efficiency through the dipole-dipole relative orientation

variable (κ^2) term as stated above. If the emission dipole of the donor and the absorption dipole of the acceptor are aligned in a head-to-tail fashion in space, then κ^2 equals 4. If the dipoles are aligned parallel, then κ^2 is 1 and if they are perpendicular then κ^2 is 0, as shown in Figure 12.

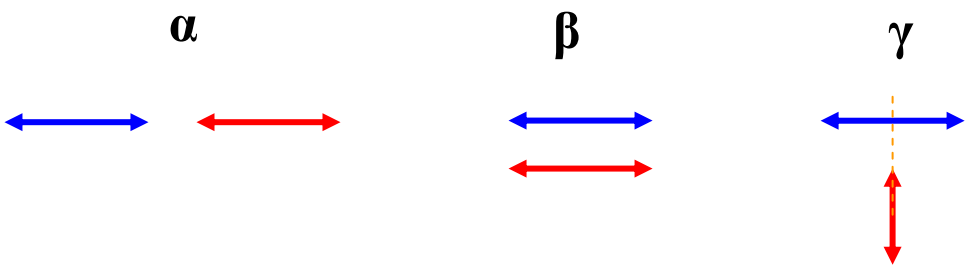


Figure 12 Dipole-dipole alignment between two FRET coupling entities, head-to-tail $\kappa^2 = 4$ (α), parallel $\kappa^2 = 1$ (β) and perpendicular $\kappa^2 = 0$ (γ).

If the dipoles and hence chromophores are attached to DNA *via* a long and flexible linker, then an approximated κ^2 value can be applied[3, 5, 26], in particular when applying FAM and TAMRA FRET pairs attached to DNA *via* a hexyl linker[3]. However, if the chromophores are attached *via* rigidified linkers, then the κ^2 value deviates from the approximate behaviour and so must be applied for. Figure 13 displays the approximated and not approximated κ^2 values by Lilley and co-workers for E_{FRET} versus distance[3]. The peaks and troughs can be clearly seen as the distance and angle between the donor and acceptor is varied around the helical double stranded DNA. This result also means more than one E_{FRET} values of the same intensity is obtainable.

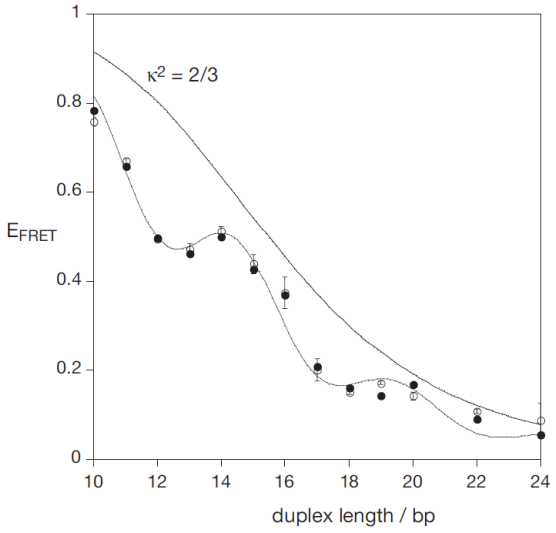


Figure 13 Comparing the relative dipole-dipole orientation values for defined (dots) and approximated to 2 / 3 for Cy3 and Cy5 terminally labelled FRET pairs, from Lilley *et al.*[3].

1.6.2.1.2 Quantum yield

The FRET transfer efficiency is dependent on the quantum yield of the donor. The quantum yield is the ratio of light absorbed to light emitted[5]. In order to measure the quantum yield accurately and prevent inner filter effects distorting the result, the absorption in the region of excitation wavelength must be kept below 0.05, after which fluorescence analysis can be applied. The optical density (absorption intensity) and fluorescence integrals were applied to equation 2, from which the quantum yields were established with respect to known standards (for example quinine sulphate or fluorescein).

$$Q = Q_R \frac{I}{I_R} \frac{OD_R}{OD} \frac{n^2}{n_R^2} \qquad \textbf{equation 2}$$

Q = Quantum yield of samples

Q_R = Quantum yield of standard

I = Integrated fluorescence emission of sample

I_R = Integrated fluorescence emission of standard

OD = Optical density of samples

OD_R = Optical density of standard

n = Refractive index of sample solvent

n_R = Refractive index of standard solvent

1.6.2.1.3 Molar extinction coefficient

The molar extinction coefficient (ε) of the acceptor affects the FRET transfer efficiency. The Beer-Lambert law enables identification of the extinction coefficient of the acceptor chromophore in question (equation 3)[5]. The extinction coefficient is calculated using predetermined acceptor concentrations (C) and known cell path lengths (l), to which UV-Vis absorption spectroscopy is applied. Absorptions (Abs) must be kept below 1 to prevent non-linear Beer-Lambert behaviour of absorption with concentration.

DNA concentrations can be identified using the Beer-Lambert law. However, extinction coefficients have to be determined using extinction coefficient software, which utilise the bases nearest-neighbour approach, by calculating the absorption at 260 nm.

$$\varepsilon = \frac{Abs}{Cl} \qquad \textbf{equation 3}$$

ε = Extinction coefficient (M^{-1} cm^{-1})

Abs = Absorption (dimensionless)

C = Concentration (M)

l = Path length of cell (cm)

1.6.2.1.4 Spectral overlap

The spectral overlap integral (J) affects the FRET transfer efficiency between donor and acceptor; J is the region of donor emission and acceptor absorption overlap[5]. Using equation 4, J is calculated by the sum of the integral of the normalised fluorescence intensity, extinction coefficient of the acceptor and the wavelength to the 4^{th} power.

$$J = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \qquad \textbf{equation 4}$$

F_D = Corrected fluorescence intensity at λ (dimensionless)

ε_A = Extinction coefficient of the acceptor at λ (M^{-1} cm^{-1})

λ = Wavelength (nm)

1.6.2.2 Förster distance

The Förster Distance (R_0) is the distance between donor and acceptor when the FRET efficiency (E_{FRET}) is 50 %[5]. The transfer efficiency is related to R_0 through equation 5, which shows the r^{-6} distance dependency. A plot of r / R_0 versus E_{FRET} is shown in equation 5 for an R_0 value of 50 Å. For this R_0 value, the actual distance (r) is equal to R_0, which equals 1 and hence an E_{FRET} value of 0.5. Either side of this distance the E_{FRET} term varies linearly until either saturation (approaching 1 on the E_{FRET} term) or no E_{FRET} (approaching 0 on the E_{FRET} term) occurs, i.e. small and large donor-acceptor distances respectively, see Figure 14. This equation only holds true for randomly orientated chromophores, where κ^2 is 2 / 3. For rigidly held chromophores on DNA, the relative orientation must be calculated. Sören Preus *et al.* have developed software which calculates the finite positions of the dipoles, the results are then applied to the E_{FRET} equation from which non-approximated κ^2 values can be determined[12].

$$E_{FRET} = \frac{1}{1+\left(r/R_0\right)^6}$$ **equation 5**

E_{FRET} = FRET efficiency

r = Donor-acceptor distance

R_0 = Förster distance

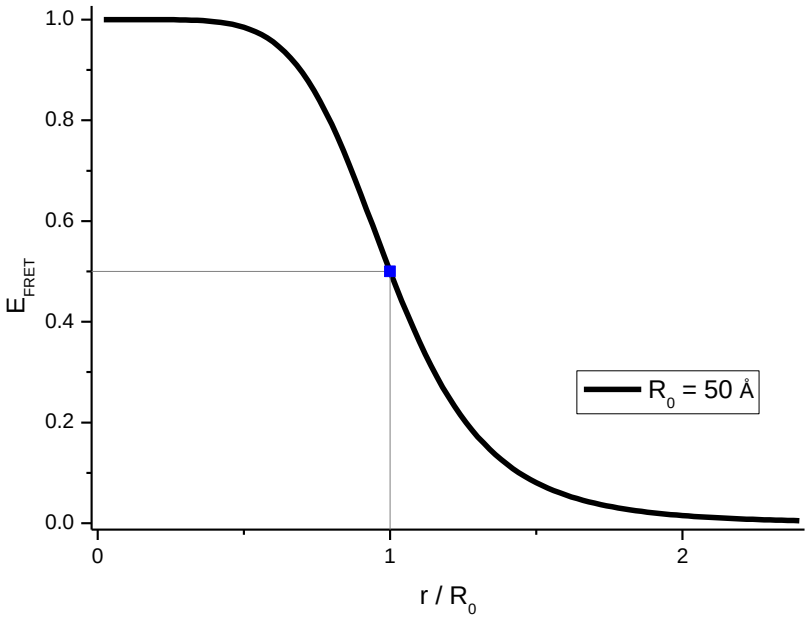


Figure 14 r / R_0 versus E_{FRET} for a R_0 value of 50 Å (black line), the 50 % energy transfer is shown (grey line).

1.6.2.2.1 Calculating the E_{FRET} term

The donor emission (F_D) is dependent on the concentration ([S]), the extinction coefficient (ε) and the quantum yield (Q_D) of the donor, and also the degree of FRET occurring from donor to acceptor, as shown in equation 6[5, 7, 27]:

$$F_D = [S]\varepsilon_A Q_D(1 - E_{FRET})$$ **equation 6**

The acceptor emission is dependent on the concentration, the extinction coefficient and the quantum yield of the acceptor, and the degree of FRET occurring from donor to acceptor, as shown in equation 7[5, 7, 27]:

$$F_A = [S]\varepsilon_A Q_D E_{FRET} \qquad \textbf{equation 7}$$

Taking the donor and acceptor components (F_D and F_A), the E_{FRET} values can be identified from three different techniques:

1. Steady-state fluorescence spectroscopy (see section 1.6.2.2.2).
2. Time resolved fluorescence spectroscopy (see section 1.6.2.2.3).
3. Single molecule fluorescence spectroscopy (see section 1.6.2.2.3).

1.6.2.2.2 Steady-state FRET

The simplest steady-state measurement to identify the extent of FRET, and hence donor-acceptor distance, is to analyse the change in donor emission in the absence and presence of the acceptor, equation 8. This equation alleviates complications associated with other relaxation mechanistic pathways associated to the acceptor, for example dark states. However, it should be noted that the acceptor chromophore does not necessarily have to emit light for FRET to be applied, for example non-emissive FRET can be conducted to a quencher chromophore[5].

$$E_{FRET} = 1 - \frac{F_{DA}}{F_D} \qquad \textbf{equation 8}$$

E_{FRET} = FRET efficiency

F_{DA} = Fluorescence of donor in the presence of the acceptor

F_D = Fluorescence of donor

If the donor and acceptor emission overlap, then peak deconvolution is required to identify the donor only value, and hence E_{FRET} term. The Clegg *et al.* method[7, 27] for peak deconvolution is shown in equation 9. The fluorescence emission for FAM at 495 nm (F_{495}) is comprised of donor and acceptor parts (AF_{D495} and BF_{A560} respectively). A and B are the weighting factors of F_{D495} and F_{A560} respectively, B is defined using equation 10, which takes into account the relative donor and acceptor extinction coefficients at 495 and 560 nm. Applying these equations the donor component can be identified, after which accurate E_{FRET} values can be determined.

$$F_{495} = AF_{D_{495}} + BF_{A_{560}} \qquad \textbf{equation 9}$$

$$B = \left[E_{FRET} \left(\frac{\varepsilon_{D_{495}}}{\varepsilon_{A_{560}}} \right) + \frac{\varepsilon_{A_{495}}}{\varepsilon_{A_{560}}} \right] \qquad \textbf{equation 10}$$

1.6.2.2.3 Time resolved and single molecule FRET

The FRET efficiency can be measured by analysing the decay emission ratios of the donor in the presence and absence of the acceptor, as shown in equation 11[5, 28]. Experimentation is conducted using a pulsed light source, after which the emission of donor and acceptor is monitored over time. Time-resolved analysis is more informative than steady-state fluorescence experimentation, time resolved spectroscopy contains information on the decay rates deriving from the donor and acceptor emission components, from which the transfer rate can be identified. Ansari and co-workers have shown time resolved fluorescence analysis of DNA wrapping around a protein[29]. Other techniques employed previously did not reveal kinetic information to a high resolution; stopped-flow instruments enabled millisecond analysis, whereas time-resolved fluorescence spectroscopy enabled nanosecond resolution on the binding event. This analysis enabled them to identify further information about the binding mechanism.

$$E_{FRET} = 1 - \frac{\tau_{DA}}{\tau_D}$$ **equation 11**

E_{FRET} = FRET efficiency

τ_{DA} = Fluorescence lifetime of donor in the presence of the accept

τ_D = Fluorescence lifetime of donor

Single molecule FRET applies the same theory as time-resolved FRET, however, instead of observing an ensemble of molecules, and therefore an average of FRET intensities, individual molecules are monitored[30, 31]. Single molecule FRET enables direct observation of dynamic interactions between donor and acceptor moieties, for example, a range of FRET distances from one molecule can be identified and monitored, of which can then be translated into a range of conformational structures from that one molecule. Single molecule FRET alleviates the problem of particle-particle interactions associated at higher concentrations, as required for time resolved and steady-state FRET. Single molecule analysis is usually monitored over seconds to minutes, therefore dye stability is of paramount importance. Time sensitive detectors are used to identify the relative intensities of donor and acceptor components, from which the inter-dye distance information can be extracted. Single molecule FRET analysis is a ratiometric process in which the donor and acceptor intensities are monitored simultaneously, this process reduces the effects of background noise levels since only valid FRET interactions will correspond to additive changes in both donor and acceptor components. Oesterhelt and co-workers have recently applied time resolved and single molecule fluorescence spectroscopy to analyses bulges and kinks in DNA[32]. Their FRET pairs were attached to the DNA using flexible linkers, and produce a model to accurately analyse the experimental values (greater molecular ruler resolution can still be achieved if the FRET pairs are attached rigidly, see chapter 2).

1.6.3 Circular dichroism

A porphyrin DNA circular dichroism (CD) molecular ruler was proposed by analysis the exciton coupling extent of the porphyrins Soret band. CD is a ground state absorption technique analogous to UV-Vis absorption spectroscopy[33, 34]. A CD signal ($\Delta\varepsilon$) occurs when there are differences in the extent of circularly polarised absorption, where Δ_L and Δ_R are the extent of left and right circular polarised absorption respectively (see equation 12). For this to occur, the molecule absorbing light must be chiral. A chiral molecule is a molecule which is non-superimposable on its mirror image, i.e. a molecule which does not have a plane of symmetry. The molecule in question however does not have to be inherently chiral, induced chirality also gives rise to a change in CD signal. CD is widely used for molecular biology analysis to the study of proteins and nucleic acids, both of which are inherently chiral[33].

$$\Delta\varepsilon = \Delta_L - \Delta_R \qquad \textbf{equation 12}$$

The CD study of proteins enables the approximate conformational arrangement to be identified, for example the extent of α-helices and β-sheets present. The analysis is conducted in an empirical analysis by comparison to known standards. Any changes upon, for example, the addition of a substrate, may lead to a change in conformational rigidity of the protein's backbone, which may give rise to a change in CD signal to potentially reveal a binding transition. The substrate itself does not have to be chiral, but once added to a chiral environment, the moiety has induced chirality. If this substrate is a porphyrin for example, then upon binding non-covalently, the induced porphyrin CD signal is typically isolated from the protein signal (> 350 nm compared to < 300 nm respectively), and so any change in this region arises from the porphyrin only, making binding information easier to obtain and analyse.

Proteins have small tryptophan, tyrosine and phenylalanine aromatic absorptions at around 280 nm[10, 33]. The range 260-180 nm is the secondary structure absorption region, arising from the peptidic backbone CD absorption. Below 180 nm the CD signal is normally saturated by the solvent (even with the use of very small path length

cells), especially if the solvent contains chlorides or fluorides. Potentially proteins can be taken into different salt environments to reduce the background absorption in this region. However, the non-native behaviour perturbed on system upon changing the solvent must be analysed, requiring the use of extensive controls to maintain biological and conformational activity.

The CD study of nucleic acids enables the identification of different nucleic acid conformations, for example single, double or triple strand, A-, B- or Z-DNA etc.[33, 34]. The CD spectra is generated by the absorption of the aromatic Watson-Crick bases; G, C, A and T. The extent of CD absorption depends on the base's environment, for example, double stranded or single strand DNA has a strong effect, as do solvent conditions. Thermodynamic parameters for DNA can be obtained, CD can be monitored as a function of temperature. Temperature studies of DNA can give detailed information about these melting / annealing, whereas UV-Vis temperature studies may not necessarily identify transitions. Adding internal or external modifications to DNA may disrupt the native behaviour of the system; CD can be used to identify the modifications perturbation extent, by comparing the native and modified systems directly.

1.6.3.1 Exciton coupling

Covalently attaching an aromatic molecule, for example, a porphyrin or stilbene to DNA, results in the aromatic molecule becoming intrinsically chiral (deriving from the chiral DNA)[35, 36, 37]. If two porphyrins are attached to DNA and are within a relatively short distance, then the two aromatic residues may interact *via* exciton coupling, which can split the CD spectra leading to positive or negative Cotton effects (+ / - or - / + respectively), see Figure 15 (β), and lead to UV-Vis peak broadening[32]. Exciton coupling has an r^{-2} distance dependency, and like FRET depends on the relative orientation of the coupling entities. Exciton coupling has a sin(2θ) dipole-dipole angle dependency, maximum coupling occurs when the angles between the dipoles are 45 and 135° and minimum when the dipoles are 0 and 90°, see Figure 15 (α). The Davydov

splitting is the difference between the two oppositely signed CD maxima, which is related to the vectors joining the two chromophore dipoles, Figure 15 (β)[2, 33, 37].

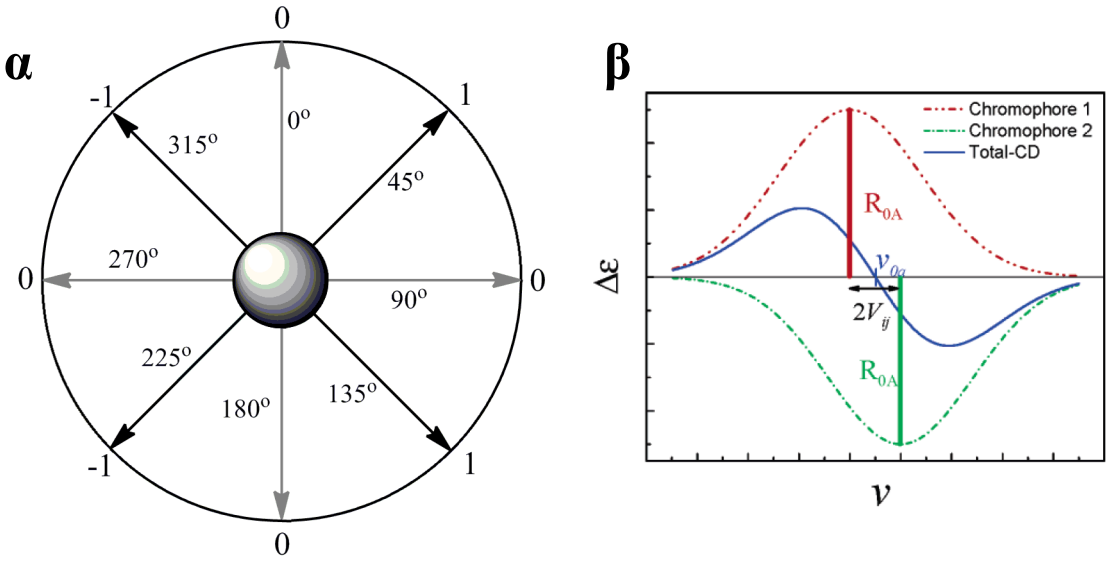


Figure 15 Exciton coupling intensity versus the angle between coupling dipoles, maximum exciton coupling occurs in between parallel and orthogonal alignments (α) and wavenumber (ν) versus molar ellipticity (Δε) for two exciton coupling entities (β): net positive Cotton CD signal observed (blue) with negative and positive components (green and red respectively), separated by Davydov splitting (ν_{ij}), image from Lewis *et al.*[3].

Exciton coupling is inherently sensitive to small changes in distance or angle (or both) between two coupling aromatic molecules, therefore the technique is ideal for use as a molecular ruler (see chapter 4). Berova and co-workers have shown porphyrin moieties to undergo strong exciton coupling interactions[35]. Simulated models were fitted to experimental CD curves, after which distance and angle information was extracted. Therefore, a porphyrin based CD molecular ruler was proposed analogues to the FRET version. Complications arise when generating models to compare with experimental results, which require high levels of technical expertise.

1.6.4 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) can be used to calculate exact experimental distances by detecting the electron-rich metal-to-metal centre scattering between heavy metals[4, 9, 38, 39]. SAXS is an elastic scattering technique which allows structural information in the 1-25 nm distance range to be obtained. SAXS is an ideal technique for studying the shape and size of small to large biological molecules. Unlike X-ray crystallography, crystalline samples are not required, experimentation can be conducted in the solution phase. X-rays are applied to the sample in question, which interact with the molecule's electrons, producing elastic scattering at small angles (between 0.1 and 10°). This scattering reveals structural information about the shape and size of the sample. The scattering obtained is inversely proportional to the structural distance of the scattered light. A scattering plot is obtained when plotting the intensity (Q) versus inverse distance. After computation, the data is converted into a structure. However, most small angle scattering (SAS) data does not yield well defined structures of high resolution, and so shapes are fitted to the experimental data. If two gold nanoparticles are held at a fixed distance apart within the SAXS spectral window, Gang *et al.* has shown that the scattering information, and hence distance separation between two nanoparticles can be extracted[9]. If this methodology can be applied to porphyrin DNA metalled with heavy metals or iodinated DNA analogues, then potentially a SAXS DNA molecular ruler can be generated, whereby the SAXS metal-to-metal centre scattering can be extracted.

SAXS has experimental difficulties; SAXS requires high sample concentrations, which must be free of aggregation. Aggregates scatter electrons greater than smaller mono-dispersed entities, to mask the scattering and perturb the analysis. Relatively high sample concentrations are required for sufficient scattering, however, apart from aggregation, non-native particle-particle interactions can also occur. Precautions must be applied for when conducting experimentation, including extensive purification using size exclusion chromatography[38, 39].

1.7 Synthesising the DNA scaffold

The DNA scaffold was synthesised using solid support chemistry on a DNA synthesiser. The key steps of solid phase DNA synthesis is explained, including mechanistic details and deprotection steps (see 1.7.1). The methods commonly used to purify synthetic DNA is followed (see 1.7.2).

1.7.1 Solid phase DNA synthesis

Porphyrin labelled and native DNA sequences were required to test the molecular ruler. Sequence specific DNA can be synthesised using automated solid support chemistry[10]. The DNA can be synthesised from nmol to mmol scales, with 99 % coupling efficiency per base added. Therefore DNA strands up to and over 100 bases in length can give rise to practical yields. Modified nucleobases can be incorporated into the DNA strands, and be placed at selected locations ideal for DNA molecular ruler synthesis and application. Solid phase DNA synthesis works step-wise, base by base through a cyclic synthetic protocol. Each cycle contains a deprotection, activation coupling, capping and oxidation step. The bases are coupled to a starting base pre-attached to resin. The resin is then contained within a plastic column, the column allows solvents and reagents in the solution phase to pass though, whilst keeping the resin *in situ*. The column is then placed within a DNA synthesiser, to which DNA synthesis can commence. The synthesis utilises phosphoramidite chemistry through the use of the β-cyanoethyl group. Each of the bases is coupled in turn, in the 3' → 5' alcohol direction along the phosphodiester backbone (converse to nature's biosynthesis: 5' → 3').

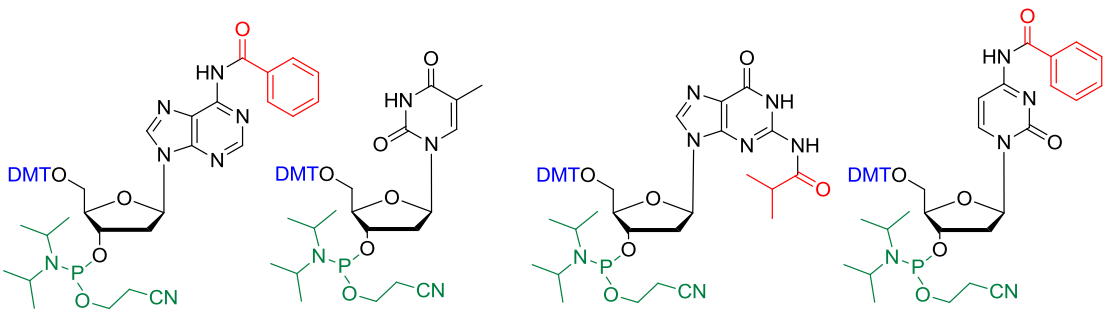


Figure 16 DNA phosphoramidite monomers with amine (red), DMT (blue) protecting groups and β-cyanoethyl phosphoramidite groups (green) for A, T, G and C (left to right).

Figure 16 shows the four bases with the relevant protecting groups required for solid phase DNA synthesis. The β-cyanoethyl group contains the reactive phosphorus (III) group required for DNA elongation. Phosphorus (III) reacts with water or oxygen exothermically, to give phosphorus (V), therefore a dry and oxygen free environment is required during coupling conditions (the diisopropylamine group is added to reduce the phosphorus reactivity). The dimethoxy trityl (DMT) group is a hydroxyl protecting group required for the protection of the 5' alcohol of DNA monomers. This alcohol is protected to prevent unwanted side reactions, such as 5' OH nucleophilic attack on the base added. Nucleophilic amines on the bases too are protected for the same reason, A and C amines are protected with the benzoyl group (at the N^6 and N^4 positions respectively), whilst G's amine is protected with isobutyryl (at the N^2 position).

Solid phase DNA synthesis is initiated by deprotection of the DMT group of the base tethered to the resin, this step is known as the deblocking step (Figure 17). The DMT group is cleaved with the addition of 2 % trichloroacetic acid (TCAA), to yield an orange DMT cation. The DMT cation is used to approximately identify the coupling yield of the DNA base by UV-Vis absorption spectroscopy. The column, and hence resin is then washed in anhydrous acetonitrile to prevent subsequent acid deprotection of the base to be coupled. Further intermittent steps are also washed with anhydrous acetonitrile (note: ethyl acetate can be used as a substitute). The following base to be coupled is activated using tetrazole. Tetrazole is a weak organic acid (pka approximately 4.6, acidity analogous to a carboxylic acid), which is required to protonate the diisopropylamine group, and hence convert it into a good leaving group,

but not strong enough to deprotonate the DMT group[10]. The reaction undergoes an S_N2 reaction between the 5' alcohol on the resin, and the phosphors from the β-cyanoethyl group to be coupled at the 3' position. Coupling occurs with the elimination of the diisopropylamine group. Coupling takes up to 20 seconds on the µmol scale, with coupling efficiency approaching 99 % per base. Any 5' hydroxyl's which do not couple are then capped to prevent further reaction with the use of capping reagents, such as acetic anhydride and n-methylimidazole. Phosphorus (III) is oxidised to phosphorus (V) with the use of pyridine and iodine, after which the DNA can be extended or deprotected, cleaved and purified (as shown in Figure 17).

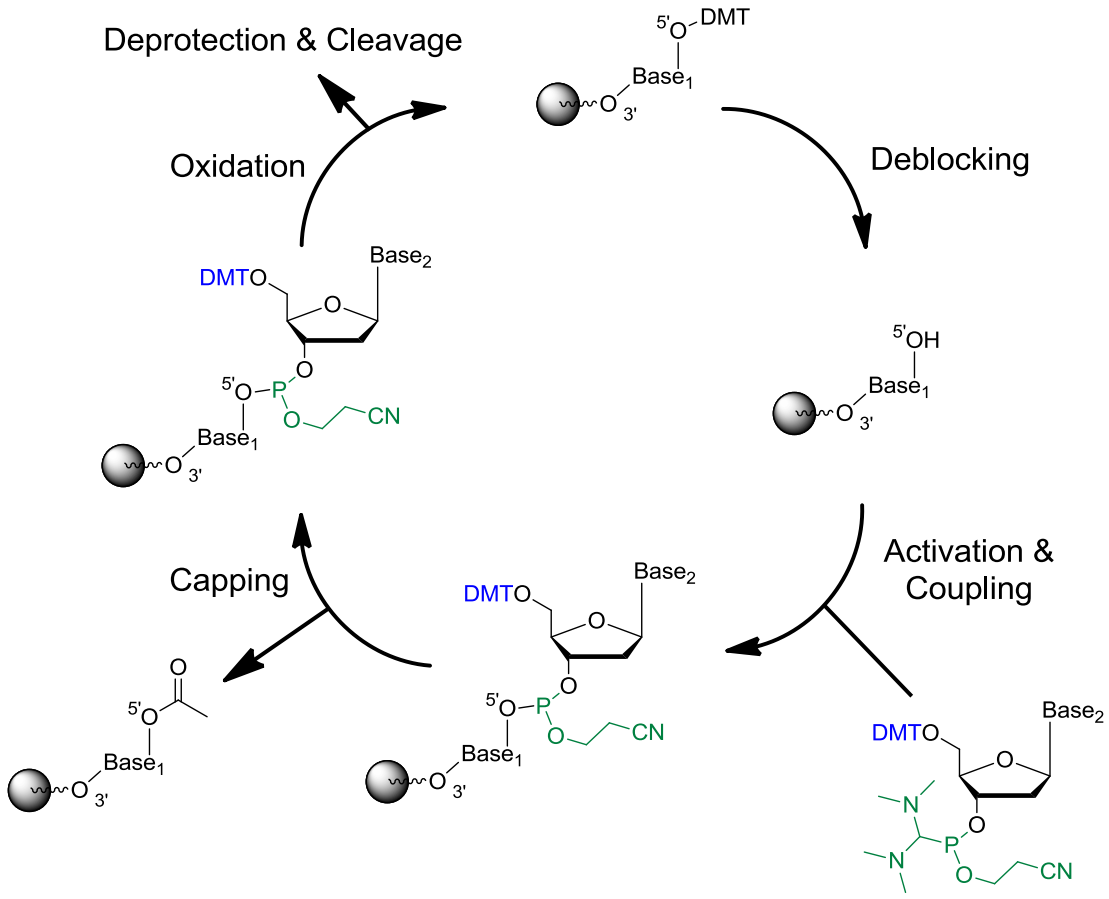


Figure 17 Solid phase DNA synthesis cycle, initiated by deprotecting the DMT group (TCAA) of base attached to resin, after which activation (tetrazole) and coupling with the subsequent base in the 3' → 5' direction is conducted. Capping (acetic anhydride) is applied to prevent side products reacting, followed by oxidation (pyridine and iodine) and strand deprotection and cleavage (ammonia) or subsequent DNA elongation.

Once the desired DNA has been synthesised, the amine protecting groups, β-cyanoethyl groups and resin must be removed. All of these groups are removed with the addition of a weak base; ammonia is typically applied at approximately 50 °C for 4 hours to remove these groups. Figure 18 shows the resin-DNA cleavage *via* an ester hydrolysis step (α). Ammonia attacks the carbonyl group of the ester to undergo an addition-elimination reaction, yielding the new amide bond and the eliminated 3' alcohol DNA strand (note the R group on the DNA eliminated, the identity of this group is important for DNA purification). The benzoyl protecting groups of the amine's of A and C (β and δ) are deprotected when ammonia initiates an addition-elimination reaction at the N^6 and N^4 positions of A and C respectively. T has no amine group and therefore no protection and deprotection is required (γ). The cyanoethyl groups on the phosphates are removed by ammonia deprotonation (ζ) at the acidic proton position adjacent to the cyanide bond. Deprotection occurs *via* a conjugate base elimination (E_{1CB}) mechanism.

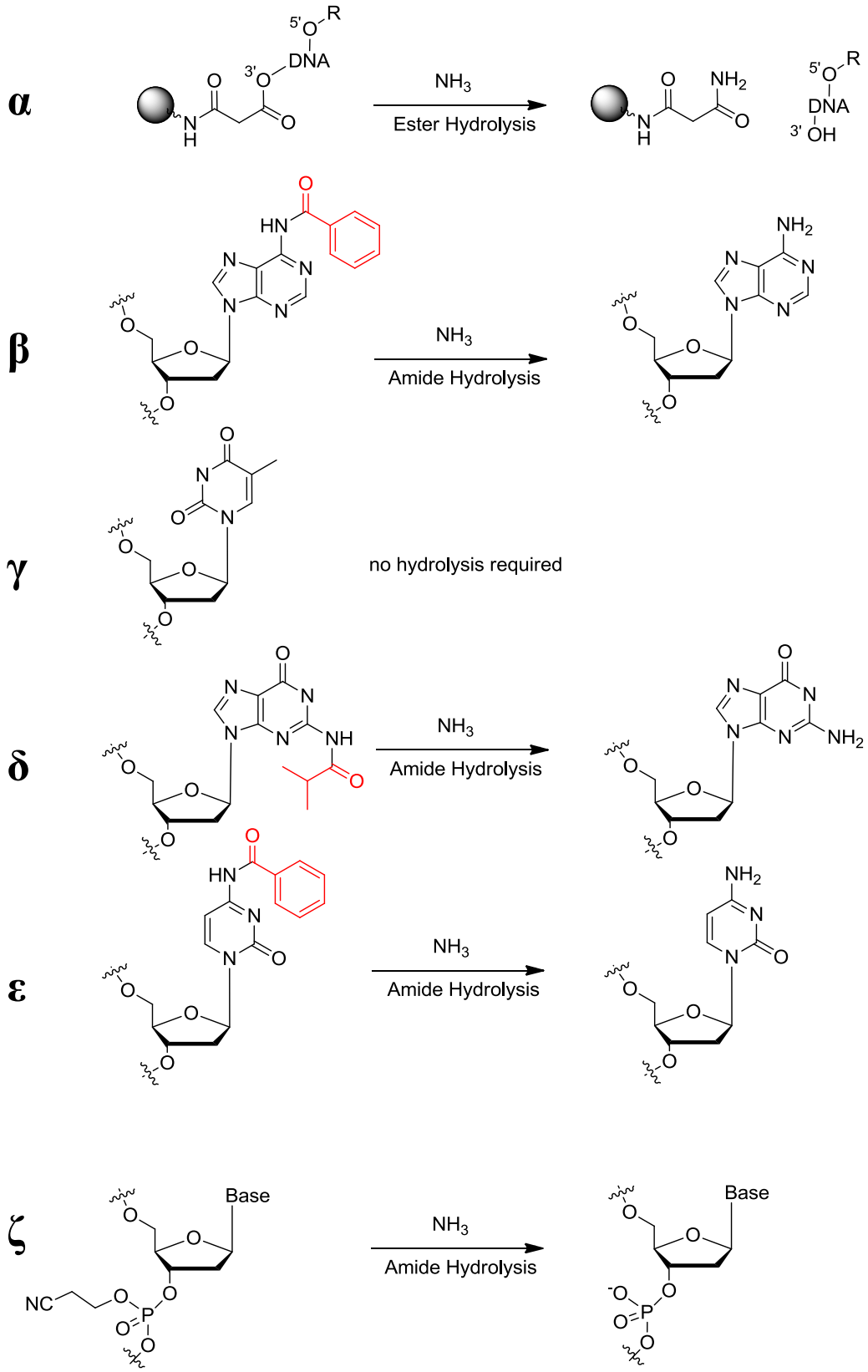


Figure 18 Ester hydrolysed DNA strand cleavage from resin (α), benzoyl group hydrolysis of A and C (β and ε), isobutyryl amide hydrolysis (δ) and β-cyanoethyl elimination (ζ).

1.7.2 DNA purification

After DNA synthesis, cleavage and deprotection, the DNA must be purified from any impurities present, including: failure sequences, capped DNA strands and eliminated protecting groups[10]. There are various methods available to purify DNA, each of which exploits different properties of the DNA for purification, for example: size, polarity or affinity[40]. Size exclusion chromatography (SEC) utilises the relative hydrodynamic radius of components. Purification between small and large molecules is easily conducted with high success rates. For example, gel filtration chromatography allows the efficient separation of DNA > 10-mers, from single-nucleotides. Apolar pores inside the silica resin allow small molecules to enter, whereas larger molecular cannot. Overall the small molecules travel a larger column distance and experience a longer time in the pores compared to the larger molecules, enabling elution resolution between reagents. SEC does not efficiently separate molecules of similar size, and so other techniques are employed in conjunction with SEC.

High performance liquid chromatography (HPLC) is a technique which enables a high degree of DNA purification, up to single nucleotide resolution is possible. Different columns can be incorporated into the instrument for appropriate purification, for example a reverse-phase column. This type of column is used extensively for the purification of DNA[40]. Reverse-phase columns typically contain non-polar alky chains which enable polar components to elute first, followed by the reduced polar components. Introducing a solvent concentration gradient between non-polar and polar solvents can help to further increase resolution, for example, see section 2.7. Steep gradients can reduce resolution, therefore a trial and error approach is required to identify optimum conditions.

Affinity chromatography utilises a molecules high affinity for a substrate, for example DMT or fluorous tag affinity. When conducting DNA synthesis, if the final DMT group is left on (final deblocking step is omitted), then only DNA sequences with the full DNA sequence contain a DMT group (due to previous deblocking and capping steps). Glen pac purification utilises DMT affinity, the DMT group bound to DNA in aqueous solvents is sufficiently hydrophobic to enable intermolecular DMT interaction, i.e. π-π stacking and aggregation. Failure DNA sequences do not contain this group,

therefore does not have DMT π-π stacking and aggregation, thus enabling resolution and hence DNA purification. Fluorous affinity chromatography also adopts a similar method[41]. Fluorine's are hydrophobic and lipophobic, and therefore have a high tendency to aggregate with other fluorines in water. In fluorous-DMT DNA synthesis the last base coupled in solid phase DNA synthesis is a fluorous-DMT base analogue. The final deblocking step is missed to enable subsequent fluorous affinity purification. After purification the fluorous-DMT group is cleaved with the addition of acid, yielding the purified DNA product.

1.8 DNA modification positioning†

Deoxyribonucleic acid (DNA) is a versatile scaffold to attach functional molecules. Functionalised moieties including porphyrins[42], pyrenes[43, 44], metallated bipyridines and terpyridines[45, 46], amongst various other alkyl and aryl groups[47] have all been incorporated into DNA. The moieties are attached to the exterior of the DNA to maintain the duplexes helical character. The modifications can be positioned in the major and minor grooves; the major groove is larger and can accommodate bulkier substituents more appropriately than the minor groove, therefore major groove is the site more commonly chosen for modification. Synthetic reasons play an important factor in deciding where to introduce modifications. All nucleobases are commercially available as iodinated analogues, thereby enabling easy Sonogashira cross-coupling or Stille cross-coupling induced functionalisation[48, 49]. The most commonly used starting nucleobase is 5-iodo-deoxyuridine. Thymidine and deoxyuridine does not require protection of the nucleobase, unlike the other bases which contain relatively reactive amines, and therefore must be protected to prevent unwanted side reactions. Using these iodinated bases, cross-coupling is facile and allows for a huge verity different substituents to be attached to DNA. Modifications are usually coupled to monomers before solid phase DNA synthesis, utilising phosphoramidite chemistry. However, Richert *et al.* demonstrated that Sonogashira coupling to an iodinated DNA single-strand attached to a solid support is possible for a variety of substituents[47].

If the site of functionalisation on the nucleobase is chosen appropriately, the modification will protrude from the double helix into the major groove, whilst maintaining their Watson–Crick base pair interaction. Therefore to maintain duplex character, pyrimidines should be modified at the 5-position, and purines should be replaced with 7-deazapurines with modifications made at the 7-position. Famulok *et al.* showed that all nucleotides in the DNA strand can be functionalised with groups that are derived from amino-acid side chains[50]. These modified DNA strands are suitable for polymerases, and can therefore be amplified *via* PCR, as shown by Seela[51].

† This section was published in *Chemical Society Reviews*, by the author and co-workers[52]

1.9 Molecular ruler probes

Porphyrins and FAM and TAMRA probes were applied for molecular ruler analysis throughout. Firstly, the porphyrins are introduced, their spectroscopic details explained, and method of monomer synthesis discussed (see section 1.9.1). Following the porphyrins section, FAM and TAMRA FRET pairs are introduced, their spectroscopic details are defined and their appropriateness as FRET pairs explained (see section 1.9.2).

1.9.1 Porphyrins

Porphyrin moieties were attached to DNA for use as novel molecular ruler probes. From haem in nature, to man-made supramolecular arrays, the porphyrins have been used extensively to transfer energy and electrons from one domain to another[42, 53, 54, 55, 56, 57, 58, 59, 60]. Porphyrins are ideally suited to energy transfer; the porphyrin moiety is relatively flat, electron rich, and aromatic. These properties give rise to sharp and intense absorption and emission bands, essential for efficient energy transfer[36, 53, 55]. Porphyrins typically have an intense Soret band (extinction coefficient > 200000) and high quantum yield, with multiple modification sites available (central, β-pyrrole and

methine positions), overall they are an excellent choice for the functionalisation of DNA and application as a molecular ruler.

Tetra phenyl porphyrin consists of four pyrrole's, each linked *via* 4 methine carbons. Each methine is covalently linked to four phenyl groups. Porphyrins are aromatic molecules due, they have eighteen delocalised π electrons, which satisfies Hückel's rule for aromaticity (where n = 4 for 4n + 2π electrons). The absorption and emission bands of the porphyrins are readily customised and tweaked, for instance adding metals to the central porphyrin cavity, or electron withdrawing or donating groups to the phenyl or β-pyrrole positions, the porphyrins electronic properties can be change, as shown in Figure 19 (α). Tetra phenyl porphyrin has a main Soret band absorption peak at around 424 nm (blue line), as shown in Figure 19 (β). This main absorption peak is flanked by four smaller peaks called the Q bands. Excitation at any of these peaks gives rise to the three emission peaks shown around 655 nm (red line).

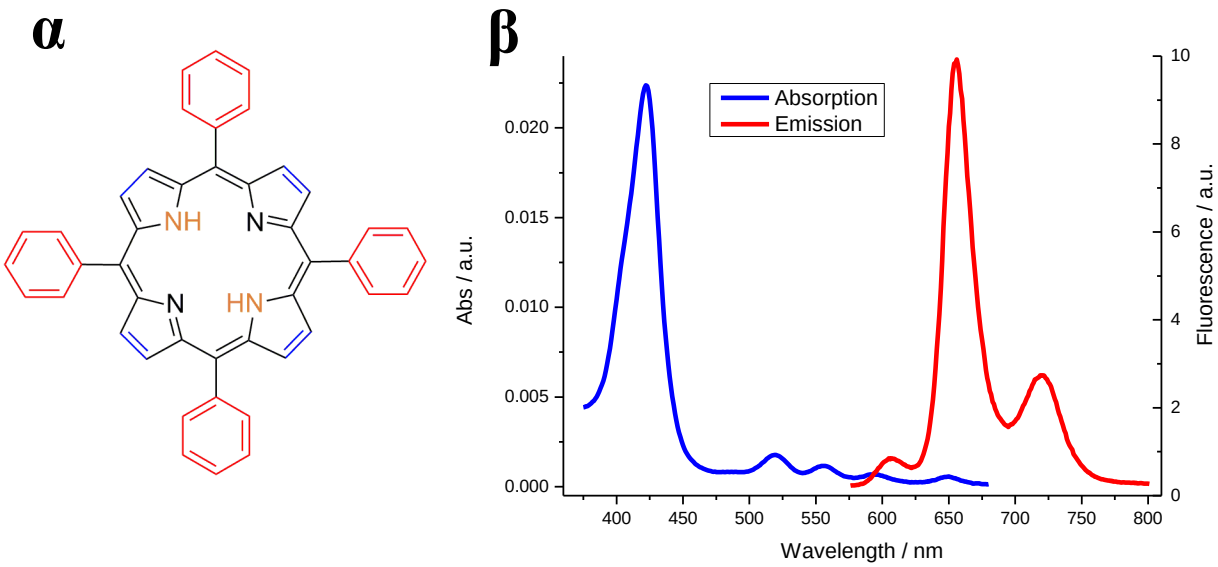


Figure 19 Tetraphenyl porphyrin chemical structure (α), functionalisation positions coloured: β-pyrroles (blue), phenyls (red) and central porphyrin cavity (orange). Tetraphenyl porphyrin normalised UV-Vis absorption and emission spectra (β).

In order to generate a accurate molecular ruler, the porphyrin must be rigidly attached to DNA. Stulz *et al.* attached porphyrins to DNA *via* the 5 position of deoxyuridine (see Figure 3). The porphyrin was tethered to the DNA base using a rigid

acetylene bond[42, 60]. The triple bond is sp hybridised and only rotates symmetrically, and therefore is ideal to extract the maximum molecular ruler resolution.

Porphyrins are ideally suited for molecular ruler probes; they are versatile probes; they can be used for FRET pairs due to their intense and sharp absorption and emission bands. Energy transfer has been shown previously to go from zinc to 2H porphyrin moieties in supramolecular arrays in organic solvents[61, 62, 63]. Exciton coupling between porphyrins has been conducted by Berova and co-workers[35]. They have shown this coupling to be distance and angle dependant. SAXS using zinc metallated porphyrins, since electron rich metals are required for intense SAXS (see section 2.1). Very few molecular ruler probes have this much versatility, for instance FAM and TAMRA are used for FRET analysis since they do not exhibit exciton coupling and cannot accommodate a transition metal.

1.9.2 FAM and TAMRA

Fluorescein (FAM) and tetramethylrhodamine (TAMRA) fluorescent dyes (Figure 20) have been used extensively in FRET experimentation for solving biological problems[3, 22]. FAM and TAMRA labelled DNA sequences were used as established FRET pair comparisons for porphyrin molecular ruler analysis. The two fluorophores are ideally suited as FRET pairs due to the high quantum yield of the donor (FAM), high extinction coefficient of the acceptor (TAMRA) and excellent spectral overlap of donor emission and acceptor excitation.

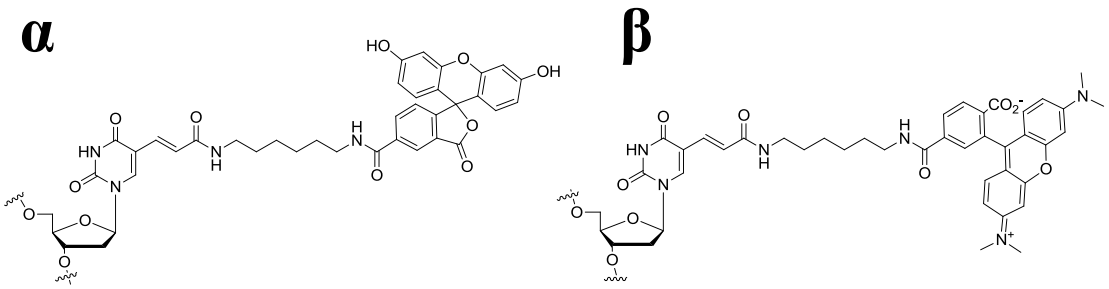


Figure 20 FAM deoxyuridine modification (α) and TAMRA deoxyuridine modification (β).

The spectral overlap of FAM and TAMRA absorption and emission spectra is shown in Figure 21. The donor's emission covers 62 % of the acceptor absorption upon excitation of the donor wavelength maximum at 495 nm. This spectral overlap corresponds to an approximate R_0 value of 54 Å. The r / R_0 values between 0.5 and 1.5 theoretically enables reliable distances to obtained, whereas values above and below 0.5 and 1.5 respectively results in non-linear E_{FRET} versus distance relationship (using the κ^2 2 / 3 value). The ratios cover approximately 30-80 Å, which covers distances ideal for extended Tus-DNA study by having FRET pairs positioned either side of the binding domain (see chapter 5).

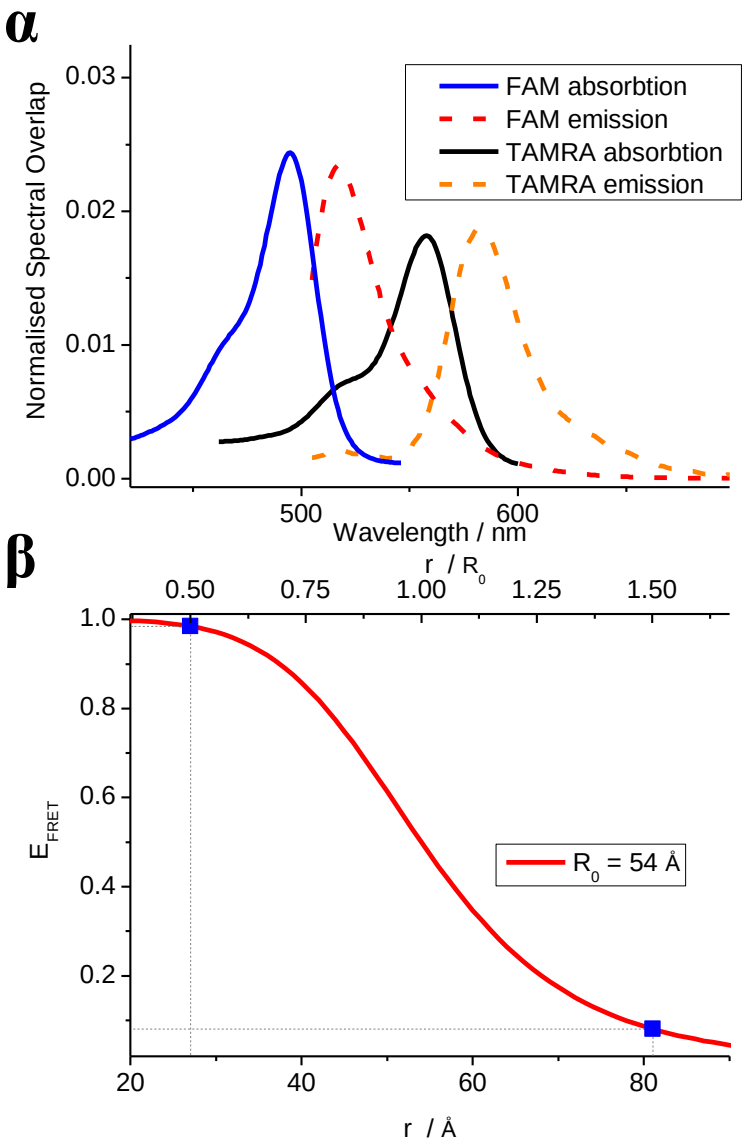


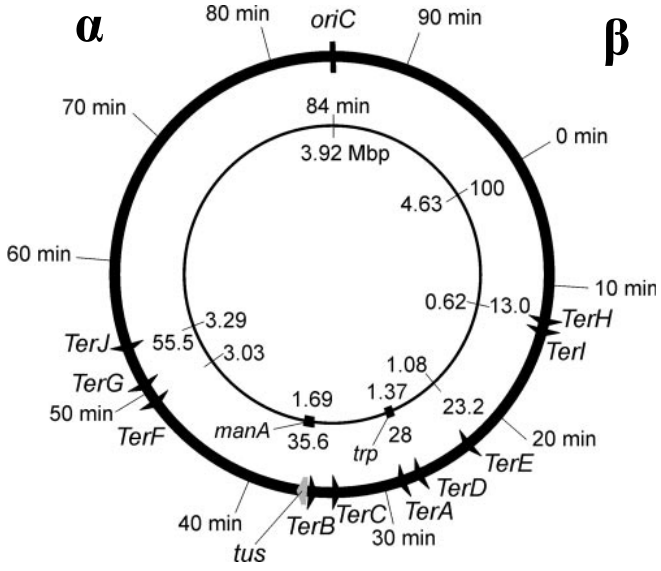
Figure 21 Normalised spectral overlap between FAM (donor) emission and TAMRA absorption (acceptor, α). Region between blue dots, FRET regions from which reliable distances can be obtained: E_{FRET} versus distance r and r/R_0 for FAM and TAMRA R_0 values of 54 Å (β).

1.10 Applying the molecular ruler

The Tus-*Ter* complex is a protein-DNA complex to which the porphyrin DNA molecular ruler was applied to study, to ultimately identify protein induced conformational changes. The complex was selected since a crystal structure is available, enabling comparison between the crystal and solution phase structure. The complex is introduced and defined, along with X-ray crystal structure analysis.

1.10.1 Tus-*Ter* complex

Tus is a 36 kDa protein which plays an important role in the replication of chromosomes in *Escherichia coli* (*E.coli*)[64, 65, 66]. Tus binds to specific DNA sites and enables replication forks to enter from one direction only. Replication is initiated at an origin site (*oriC*) of the chromosome to which replication precedes bidirectionally by replication forks (see Figure 22). Opposite the origin site are ten termination sites (*Ter*) where the Tus protein binds to sequence specific DNA, to form a strong 1 : 1 complex. Five of the binding sites are located on the leading strand, with the other five being located on the lagging strand. The complex allows the replication forks to proceed from one direction only; a *permissive* side, but not the reverse; the *non-permissive* side. Replication in this termination region is strictly mono-directional for each strand, generating what has been described as 'replication fork trap by polar arrest'[65].



Ter site	Sequence
A	AATTAGTATGTTGTAACTAAAGT
B	AATAA-T----------AGT
C	ATATA-G----------TAT
D	CATTA-T----------ATG
E	TTAAA-T----------GCA
F	CCTTC-T---------G-CAGT
G	GTCAA-G----------CCG
H	CGATC-T----------TCTC
I	AACAT-G-A----------AGT
J	ACGCA-T-A----------AGT

Figure 22 *E. coli* chromosome, replication proceeds from an origin site (*oriC*) to proceed bidirectionally using replication forks. Ten termination sites (*TerA-F*) allow replication forks to enter from one side (permissive) but not the reverse (non-permissive) as shown by the anti-arrow head direction (α), image from Neylon *et al.* [63]. The ten termination sequences, direct binding region shown in yellow region (β).

Tus binding affinity is very strong, the dissociation constant for the Tus-*TerB* complex is 3.4×10^{-13} M[64]. The direct protein-DNA binding comprises of a 16 base-pair unit, some of which differ subtly inside this region, up to 3 bases for *TerF* and *TerI*, as shown in Figure 22 (see appendix section 8.2 for DNA and Protein sequences). Some of the *Ter* sites have the same DNA sequence in this direct binding region, but have different bases outside this 16 base region, for examples, *TerA*, *TerB*, *TerD* and *TerE*. However, the binding constants for these complexes vary, therefore the region outside the binding site must play an important role for Tus binding.

The Tus-*Ter* complex has been studied by X-ray crystallography[65, 66, 67]. The crystal structure analysis shows the protein wrapping around the DNA to within 2.7 Å resolution, see Figure 23. The crystal structure contains the 16 base-pair binding region, termed the Kamada sequence. The complex was selected for study and application of the porphyrin molecular ruler since the crystal structure is known, therefore giving direct comparison between the crystal phase and the solution phase, i.e. distances from

the crystal structure and those obtained from FRET / CD / SAXS can be compared and analysed.

The region outside the binding site plays an important role, the binding constants vary even though the intermediate sequence changes subtly, the effect of which could potentially be identified with an extended DNA molecular ruler containing the binding region within. Tus*'s* Van der Waal radius is electropositive, this helps the protein attract the negatively charged backbone of duplex DNA[65, 66]. After initial protein-DNA non-specific binding, the protein is expected to scan across the DNA until the protein finds a *Ter* binding sequence, generating the stable 1 : 1 complex crystallised. Potentially, the DNA regions outside the binding sequence interacts attractively to the electropositive shell of the protein, further stabilising the complex[68, 69, 70]. If this is the case, then molecular ruler analysis using an extended DNA sequence with probes attached either side of the binding sequence will identify DNA wrapping around Tus, reducing the probes distances, for FRET analysis this will lead to increased acceptor emission.

Figure 23 Tus-Kamada x-ray crystal structure, side and top-down images, DNA (blue), α-helices (pink) and β-sheets (yellow), taken from protein data bank[66].

The Tus protein contains 309 amino residues (see appendix, section 8.2 for sequence). Tus can be expressed from the phcs151 plasmid using the *E.coli* BL21(DE3)

strain (see chapter 2). The plasmid contains the *bla*, Tus, histidine tag and sortase-recognition coding genes[65, 66]. *Bla* codes for the ampicillin digestion enzyme β-lactamase (antibiotic). β-Lactamase hydrolyses the conformationally strained four membered β-lactam ring of ampicillin, which ensures the successful expression of the Tus gene. The histidine tag is used to purify the protein using metal affinity chromatography. The tag consists of 6 histidines located on the amine terminus of the protein, each of which has a high affinity for nickel and cobalt. Any impurities should not contain the histidine tag, since impurities should not contain either the *bla* or the Histidine tag coding genes and can be separated from the tagged protein due to the lower metal affinity of the impurities. The sortase recognition sequence consists of seven amino residues. The sortase enzyme binds to this sequence, and enables functionalisation of the protein by nucleophilic displacement (see section 1.23).

Zinc and 2H porphyrin molecular ruler

2.1 Introduction

Two porphyrin moieties were rigidly attached to DNA in an attempt to obtain a highly accurate and versatile molecular ruler. The novel ruler was analysed using steady-state fluorescence (STF), circular dichroism (CD) and small angle X-ray scattering (SAXS) spectroscopy techniques. To test the accuracy of the ruler, the porphyrins were applied to study DNA first. The probes were applied to study duplex DNA since the structure is well defined, which forms predictable and reliable structures (see section 1.2). The porphyrins were incorporated into two 21-mer complementary single-strands. One porphyrin was labelled on one strand (zinc), and another porphyrin modification (2H porphyrin) was incorporated into seven different positions on seven complementary strands, in total giving rise to seven porphyrin-porphyrin duplex DNA combinations. Each of the combinations had varying base-pair separations between porphyrins, which were designed to give a full range of intramolecular porphyrin-porphyrin distances and angles. The different distances and angle were expected to result in varying fluorescence resonance energy transfer, exciton coupling, elastic scattering intensity, for FRET, CD and SAXS molecular ruler analysis respectively.

Porphyrins moieties are highly suitable molecular ruler probes, as discussed in section 1.9.1. Previous groups have shown efficient energy transfer from zinc- to 2H-porphyrins based on supramolecular arrays in organic solvents[61, 62, 63]. Zinc- and 2H-porphyrins are well suited as FRET pairs due to the high quantum yield of the zinc-porphyrin (donor), high extinction coefficient of the 2H-porphyrin (acceptor), and good spectral overlap of donor emission and acceptor excitation[53, 55]. Attaching porphyrins to DNA *via* rigid acetylene linkers theoretically enables exploitation of the FRET relative dipole-dipole alignment property between donor and acceptor porphyrin moieties - the κ^2 variable. This alignment property potentially enables greater accuracy and reliability in FRET intensities and hence probes distances obtained, as opposed to the conventional approach of using an averaged 2 / 3 κ^2 variable, which leads to large errors in distances calculated and therefore poor structural determination[3, 5, 26].

Firstly, the porphyrin DNA synthesis is explained in detail, including the key steps and mechanisms, such as, porphyrin DNA phosphoramidite synthesis and post porphyrin DNA synthesis purification methods (section 2.1). The porphyrin DNA

sequences selected for molecular ruler analysis are introduced, which includes modification positioning and strand design (section 2.2). The duplex was modelled using two established models from which experimental FRET distances were identified. Following the modelling section, the spectral properties of the zinc and 2H porphyrin DNA is introduced (section 2.3). Any potential anomalous behaviour and destabilising effects of the porphyrin moiety on the DNA duplex was identified by native PAGE, UV-Vis melting, CD melting and fluorescence melting analysis (section 2.4). After which the spectroscopic prerequisite values for FRET were identified: quantum yield of the donor, extinction coefficient of the acceptor and spectral overlap between donor emission and acceptor absorption (section 2.5). Following porphyrin DNA stability analysis, FRET experimentation was conducted and analysed to yield steady-state FRET intensities from which zinc- and 2H-porphyrin distances and hence the DNA conformation was identified (section 2.6).

CD spectroscopic analysis of the Soret band region for each porphyrin DNA combination was performed and analysed to monitor the extent of exciton coupling between intramolecular porphyrins (section 2.7). The experiment was conducted to identify the comparison between exciton coupling and the distance and angle separation between each exciton coupling entity. Exciton coupling, like FRET is strongly distance and angle dependant between the coupling moieties; from previous literature experimentation, fluctuation in the CD intensity and sign should correspond to different porphyrin DNA combinations, hence its use as a molecular ruler[2, 34, 36, 37].

Small-angle X-ray scattering (SAXS) studies of doubly zinc metallated porphyrin DNA was performed and analysed in an attempt to identify the zinc-to-zinc intra-strand elastic scattering (section 2.8). This scattering intensity can potentially be translated into experimental distances, and therefore used to identify the DNA's structural conformation. However, upon experimental analysis, sufficient resolution was not attainable at the concentration ranges applied. These results lead to SAXS experimentation using iodinated DNA analogous of the porphyrin system. Iodine scatters light to a greater extent than zinc due to the higher number of electrons present. The monomer and DNA synthesis of which is discussed, and experimental analysis presented. Finally, a conclusion section is followed (section 2.9); the evolution of the ruler is discussed and potential future experiments and applications.

2.2 Porphyrin DNA synthesis

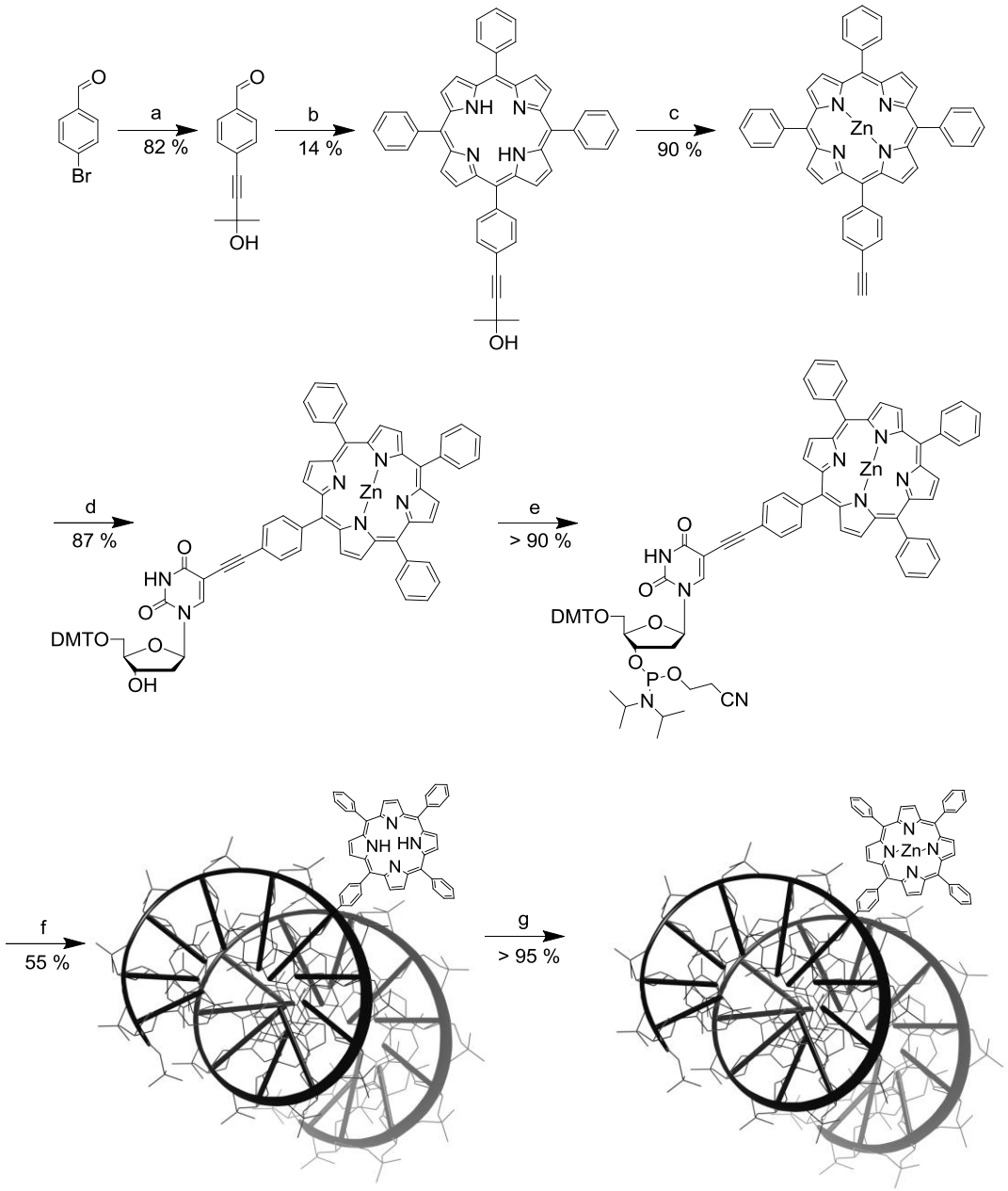


Figure 24 Porphyrin DNA synthesis using modified protocols from Stulz *et al.*[42, 60].

a i) Pd-C, PPh_3, K_2CO_3, DME, CuI ii) 2-methylbut-3-yn-2-ol

b i) Pyrrole, Benzaldehyde, $BF_3.O(CH_2CH_3)_2$ ii) DDQ

c i) NaOMe, Toluene ii) Zinc acetate dihydrate $_{(aq)}$, DCM, Methanol

d DMT-iodo-deoxyuridine, CuI, TEA, DMF, $Pd(PPh_3)_4$

e DCM, DIPEA, CEP-Cl

f Solid Phase DNA synthesis

g i) Zinc acetate dihydrate $_{(aq)}$, Δ, ii) EDTA $_{(aq)}$

The porphyrin moiety is ideally suited for light studies in general, including: FRET and light scattering studies (see chapter 4). Porphyrins have a wide variety of positions available for functionalisation. The absorption and emission properties of the porphyrin can be customised either by inserting metals into the central porphyrin cavity, or by attaching electron withdrawing or electron donating substituents to the surrounding phenyl groups or β-pyrrole positions (see section 1.21). Attaching the porphyrin to DNA *via* a rigid bond should theoretically enable more accurate translation of the porphyrin position, and hence more precise identification of the DNA conformation, for example, through the investigation the of relative dipole-dipole variable of donor and acceptor (κ^2) in FRET (see section 1.6), or for SAXS studies (metal-to-metal scattering between to two metallated porphyrins).

2.2.1 Porphyrin DNA synthesis - Step a

The rigid acetylene porphyrin modification was synthesised using modified protocols established from Stulz and co-workers (see appendix for synthetic protocol, section 8.18-8.22)[42, 60]. Tetraphenyl porphyrin was attached to DMT-iodo-deoxyuridine *via* an acetylene bond to the 5' position of uridine, as shown in Figure 24 (reaction d). The unsymmetrical acetylene containing porphyrin monomer required the synthesis of the protected acetylene first: hydroxy-dimethyl acetylene benzaldehyde was synthesised using the well-established Sonogashira palladium catalysed cross-coupling reaction (step a, see below for reaction mechanism). 4-Bromobenzaldehyde was cross-coupled to hydroxy-dimethyl acetylene at 90 °C, under a N_2 atmosphere, in a water / ethylene glycol solvent. The reaction was conducted on a relatively large scale due to the following low-yielding porphyrin cyclisation step (step b).

2.2.2 Porphyrin DNA synthesis - Step b

The unsymmetrical hydroxy-dimethyl acetylene tetraphenyl porphyrin was synthesised using 2-methylbut-3-yn-2-ol benzaldehyde, benzaldehyde, pyrrole and 0.9 equivalents of boron trifluoride diethyletherate in chlorofrom, under a N_2 atmosphere in the dark. After formation of the porphyrinogenic molecule, in equilibrium, as indicated by an intense blue emission under UV light by TLC after 45 minutes, the dehydrogenation and oxidising agent dicyano-dichloro-benzoquinone (DDQ) was added to aromatise the porphyrinogen to the thermodynamically stable porphyrin molecule. Porphyrins with different substituents were obtained due to the two benzaldehyde reagents used; multiple porphyrin side products were observed: tetraphenyl porphyrin, mono-acetylene and two different di-acetylene tetraphenyl porphyrins, as shown in Figure 25. The thermodynamically most stable porphyrin is tetraphenyl porphyrins, for instance, if DDQ is added after periods longer than one hour, increasing quantities of tetraphenyl porphyrin are obtained. However, at least one hour is required to form sufficient quantities of the unsymmetrical porphyrinogenic species[71, 72, 73, 74]. The desired product can be favoured statistically by using appropriate quantities of the starting reagents; pyrrole : benzaldehyde : acetylene benzaldehyde in 6 : 6 : 1. However, even with favourable kinetics, a yield of only 14 % was obtained, which for unsymmetrical porphyrin synthesis is acceptable. Purification of the desired porphyrin was challenging; the product had to be purified from different porphyrin products, poly-pyrroles and DDQ side-products which combine to form a black think tar-like substance. Filtration over celite, alumina and silica chromatography was required multiple times to remove these impurities.

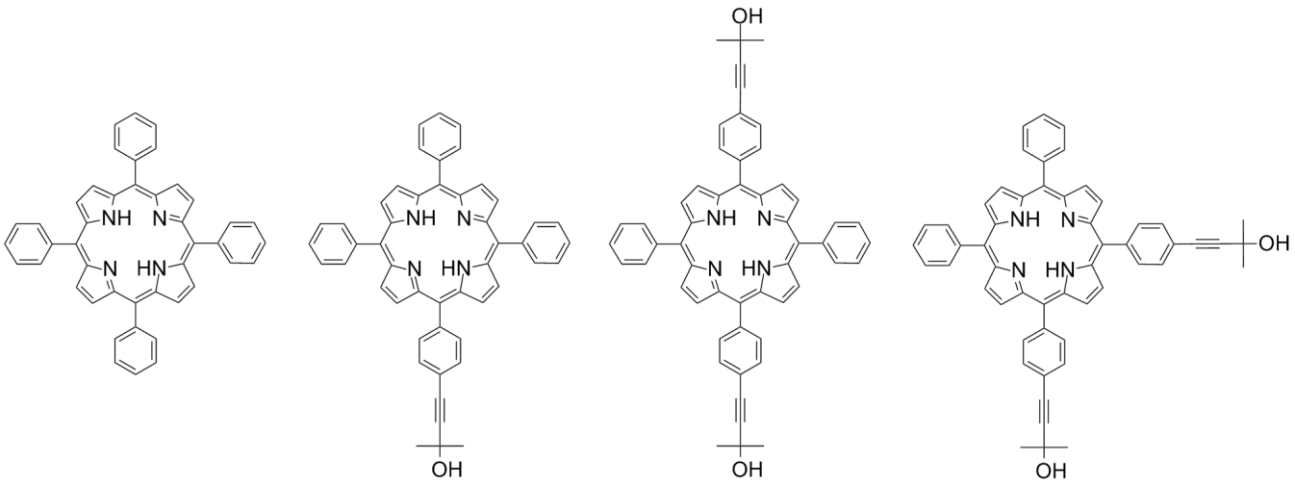


Figure 25 The four major porphyrin products obtained from porphyrin synthesis in step b, tetraphenyl porphyrin, hydroxy-acetylene tetraphenyl porphyrins, *para*-dihydro-dimethyl-acetylene tetraphenyl porphyrin and *ortho*-dihydro-dimethyl-aceytlene tetraphenyl porphyrin (left to right).

Porphyrin synthesis is initiated by a series of boron trifluoride induced hydrolysis steps under equilibrium, as shown in Figure 26. A lone pair of electrons from the amine of the pyrrole molecule undergoes an internal resonance displacement to attack the carbonyl group from a benzaldehyde molecule. Attack is mediated at the 5' position of pyrrole, forming a new pyrrole-benzaldehyde bond, generating an iminium ion and oxonium anion. The proton at the 5' position of the pyrrole unit is eliminated, the electrons shift back into the pyrrole ring to remove the positive charge and regenerate the amine. The oxonium anion intermediate formed is then protonated twice to form a hydronium cation. The hydronium cation is eliminated *via* the pyrrole's amine lone pair of electrons. The second iminium ion formed is then reduced by a second pyrrole molecule which undergoes an aromatic rearrangement *via* a resonance shift by attacking the newly formed methine position. Again, the proton at the 5' position of the second pyrrole unit is then eliminated, the electrons shift back into the pyrrole ring to remove the positive charge and regenerate the amine, and therefore phenyl-dipyrrolemethane (these polypyrrolemethane intermediates are intensely coloured, initially they appear dark red on TLC plates, after an hour however, the red turns to a brown colour through oxidation. Under short wavelength UV light they appear blue). This process is then repeated until ring cyclisation occurs to generate the porphyrinogen species formed. The porphyrinogen emits a green colour.

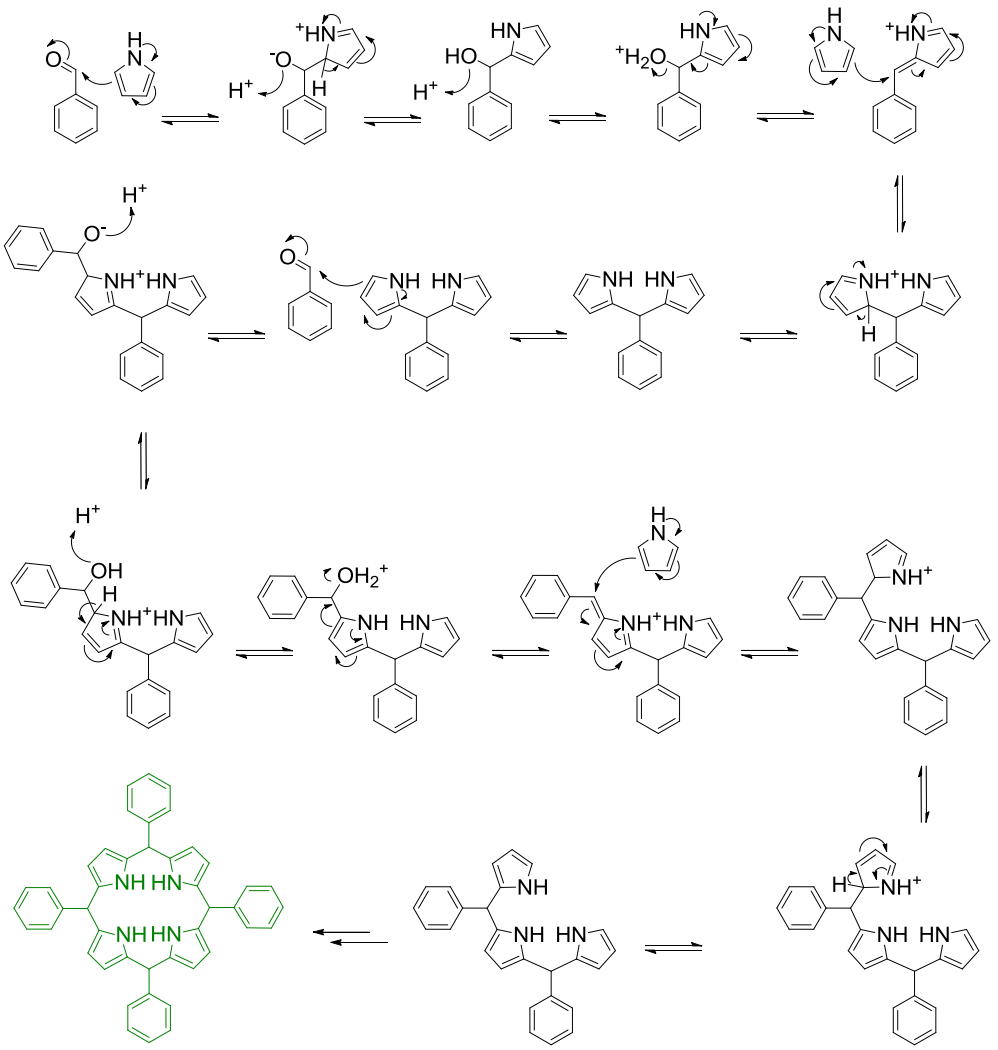


Figure 26 Porphyrinogen mechanism, porphyrinogen shown in green[71, 72, 73, 74].

To oxidise the porphyrinogen and convert it to the thermodynamically more stable aromatic porphyrin structure, the oxidising agent DDQ is added *in situ*, as shown in Figure 27. DDQ also converts the polypyrrolemethanes intermediates to polypyrrolemethenes (the intense red colour is converted to a black tar-like substance). Hydrogen atom abstraction occurs at the methine position by radicalisation with DDQ present. The radical resonates around to the amine, to eliminate a second hydrogen radical,. This process is repeated at the other three methine positions and *para* amine position to yield the final porphyrin product. The porphyrin has 18 delocalised π electrons (shown in red), each of the pyrrole sub-units have different resonance forms.

53

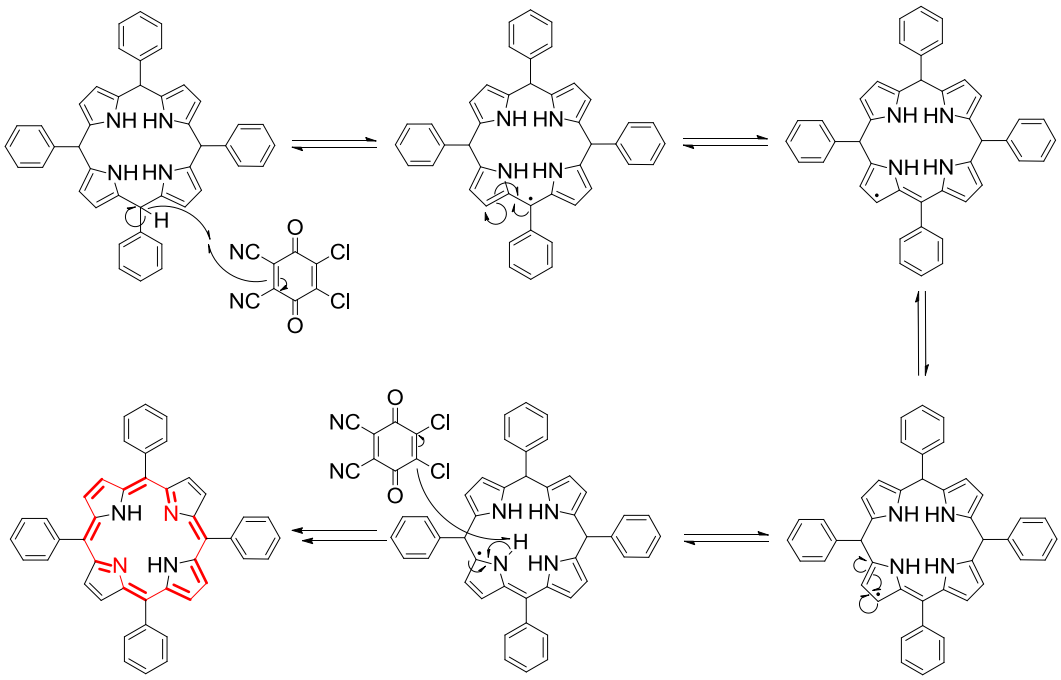


Figure 27 Porphyrin oxidation mechanism, porphyrin aromatic π system shown in red[71, 72, 73, 74].

2.2.3 Porphyrin DNA synthesis - Step c

Deprotection of the acetone protecting group was achieved in near-quantitative yield using excess NaOH in dry toluene, under N_2 and heating to reflux (step c). The reaction went to completion in 30 minutes, as compared to the several hours conditions required by Stulz and co-workers using the less basic NaOMe. The acetylene 2H-tetraphenyl porphyrin was zinc metallated to prevent complications with the Sonogashira coupling reaction in the next step, i.e. to prevent Cu-porphyrin complexation[75]. Excess zinc acetate in DCM with MeOH was required to dissolve zinc acetate and insert the metal into the porphyrin cavity. After washing, filtration yielded an intense light-purple product (reaction C).

2.2.4 Porphyrin DNA synthesis – Step d

The zinc metallated acetylene tetraphenyl porphyrin was coupled to DMT-iodo-deoxyuridine using the Sonogashira coupling conditions described by Stulz and co-workers[42, 60]. For high yielding coupling conditions (> 80 %), it is crucial the reaction environment is free of water and oxygen, requiring the use of oven dried glassware, molecular sieves and N_2 purged solvents. Separation of the monomer from the starting DMT-iodo-deoxyuridine reagent using silica column chromatography required cautious solvent polarity increases, due to the similar retention values of reagent and product.

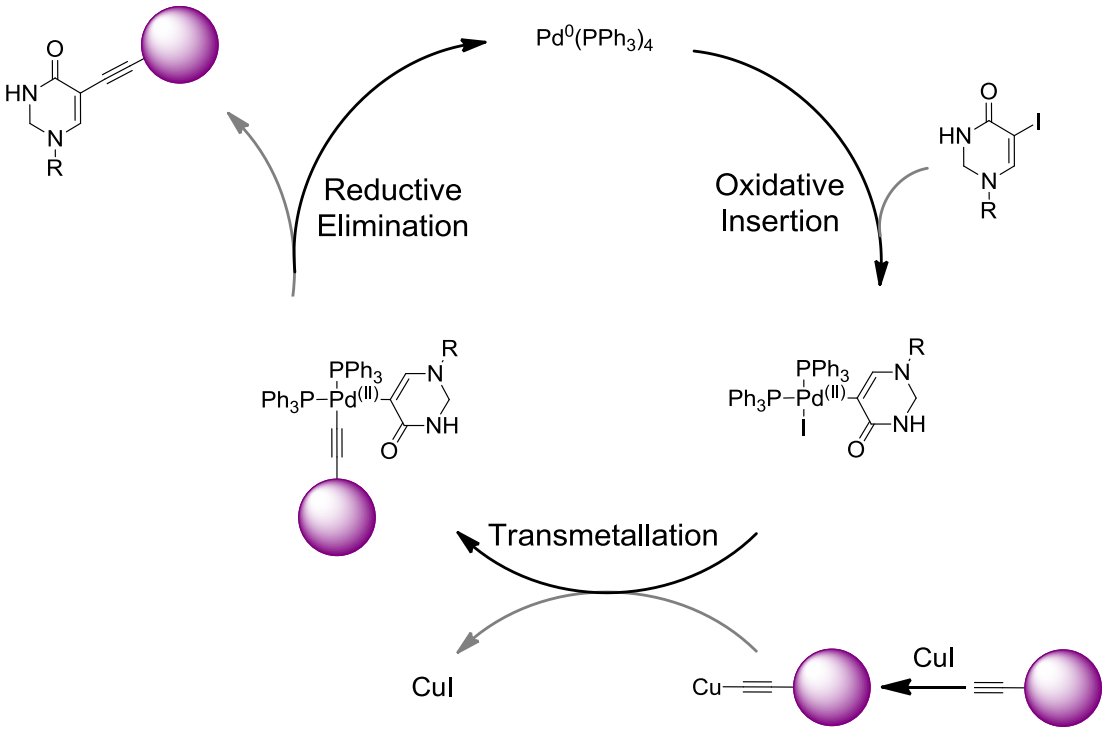


Figure 28 Porphyrin-DMT-deoxyuridine Sonogashira coupling mechanism (R = DMT deoxyUridine and purple sphere = Zn-tetra phenyl porphyrin)[16].

The Sonogashira reaction couples halogen-alkenes to alkynes, as shown in Figure 28. The reaction utilises cross-coupling chemistry, in the form of palladium. $Pd^{(0)}$ is the active oxidation state required for oxidative insertion between an aryl-halide bond. $Pd^{(0)}$ is converted into $Pd^{(II)}$ upon insertion, with a loss of two triphenyl phosphine

groups. Copper iodide associates with the alkyne, to displace the terminal proton. The copper-alkyene intermediate formed then undergoes a transmetallation at the palladium centre, with the loss of iodine. Finally the aryl and alkyne bond is formed by a reductive elimination step, which also regenerates the $Pd^{(0)}$ catalyst.

2.2.5 Porphyrin DNA synthesis - Step e

The porphyrin monomer was incorporated into DNA *via* the 3' phosphoramidite modification (step e). The active phosphorus (III) oxidation state was maintained using water, light and an oxygen-free environment, again with the use of N_2, molecular sieves and oven dried glassware. After the reaction went to completion, as monitored by TLC, the product was purified by precipitation in degassed hexane, filtered and then washed with degassed hexane multiple times. DNA synthesis was conducted on a 1 μmol scale using solid phase supports (step f) using established DNA synthesis protocols (see DNA synthesis below, section 2.7)[42, 60]. The relative porphyrin phosphoramidite coupling times were increased to allow for the reduced coupling efficiencies of modified nucleobases, from 2 to 6 minutes.

2.2.6 Porphyrin DNA synthesis - step f

The porphyrin monomer was inserted into DNA using a solid phase synthesiser (see 1.7). The reaction was monitored by observing the trityl deprotection absorption (see 2.2.6.1). Once solid phase synthesis was complete, the porphyrin DNA was cleaved from the supporting resin and the protecting groups removed (see 2.2.6.2). Finally, the porphyrin DNA was purified using fluorous affinity chromatography[41], purity was confirmed with HPLC analysis (2.2.6.3), after which the labelled DNA was analysed spectroscopically (2.2.6.4).

2.2.6.1 Solid phase synthesis

The porphyrin monomer (step f) was incorporated into the DNA sequence along with unmodified DNA bases in the 3' → 5' direction. Trichloroacetic acid (TCAA) is added to deprotect the trityl leaving group. The quantity of trityl obtained is proportional to quantity of coupled DNA, and hence can give a good indication of coupling efficiency. However, these values obtained should not be taken as absolute values for yield determination, but used instead as an approximate guide. Figure 29 displays a typical dimethoxytrityl cation (trityl) reading for the synthesis of porphyrin modified DNA. The porphyrin monomer does not affect coupling efficiencies after the modified base is incorporated, indicating efficient modified monomer coupling[42, 60].

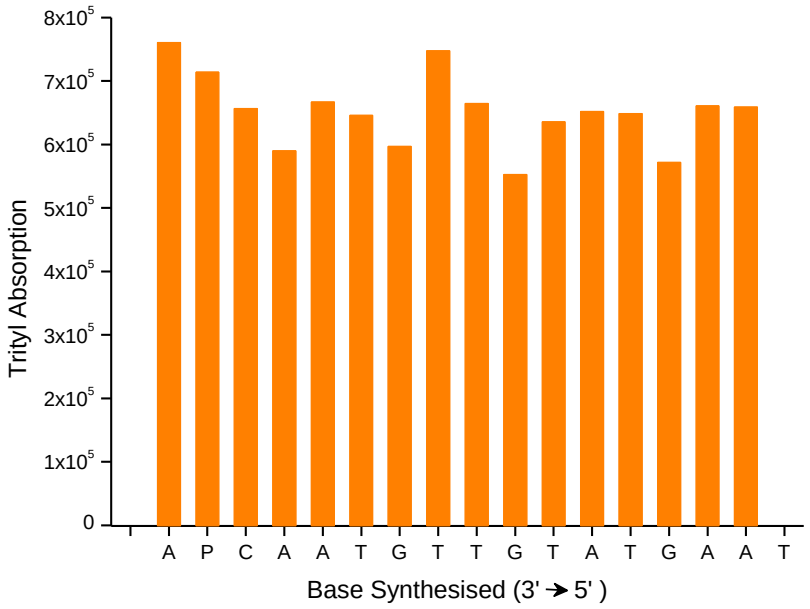


Figure 29 Trityl UV reading for DNA synthesis of porphyrin modified DNA *TerB* sequence, 3' → 5'. P = 2H porphyrin modification, synthesis conducted on a 1 µmol scale.

2.2.6.2 Porphyrin DNA deprotection

After porphyrin modification coupling, the remaining bases were coupled, after which the DNA single-strands were deprotected and cleaved from the resin using ammonia at 50 °C for 6 hours. The ammonia was removed under vacuum overnight, the resultant solid was dissolved in loading buffer for subsequent purification.

2.2.6.3 Porphyrin DNA purification

Porphyrin modified DNA was purified using fluorous affinity chromatography, whereby the last base synthesised in the 3' → 5' direction contained a C_8F_{17}-DMT protecting group (see appendix section 8.1 for protocol)[41]. Fluorine affinity was sufficiently strong to overcome any π-π stacking interactions of failure sequences containing porphyrins. Porphyrin modified DNA purity was confirmed by High Pressure Liquid Chromatography (HPLC) by monitoring the UV-Vis absorption at 260 and 424 nm, the DNA and porphyrin absorption maximums respectively. The solvent conditions for porphyrin modified DNA HPLC analysis have been described by Berova *et al.*[76]. A hexafluro-isopropanol (HFIP) and MeOH gradient was required to gain resolution and elute the modified DNA. The chromatogram in Figure 30 shows typical purity of porphyrin modified DNA for the porphyrin *TerB* DNA sequence. One should note the solvatochrosim effects of the relative DNA and porphyrin ratio. The approximate 1 : 2 ratio arises from the high temperatures and high concentration of HFIP involved during HPLC purification, as apposed to the approximate 1 : 1.6 ratio observed in salt buffer at room temperature (taking extinction coefficients of 20400 and 120000 M^{-1} cm^{-1} at 260 and 424 nm respectively).

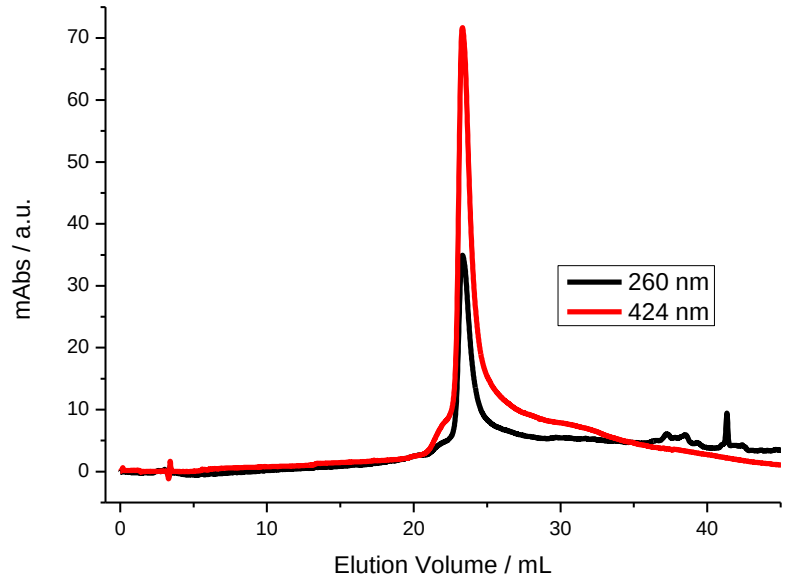


Figure 30 HPLC chromatogram of 2H porphyrin DNA (*TerB* sequence, see appendix section 8.1), 0.2 μmol total DNA injection, monitoring at DNA and 2H porphyrin absorption peak maximums at λ = 260 (black) and 424 nm (red).

Multiple attempts to obtain mass spectrometric analysis of the porphyrin modified strands were unsuccessful. Analysis was attempted at the University of Southampton Mass Spectrometry department and at the EPSRC National Mass Spectrometry Centre at the University of Swansea. Results did not yield the desired m / z calculated, even though trityl and HPLC analysis indicated the correct strands had been synthesised and purified. Porphyrin DNA Porphyrin modified DNA containing over approximately 10 bases consistently failed to be detected on MALDI-ToF and ESI instruments. These effects have been experienced by various other groups working with porphyrin modified DNA. For example, Brown *et al.* synthesised a control porphyrin poly-thymidine strand containing 12 bases[77]. A successful MALDI-ToF confirming the structure was obtained, however for the zinc porphyrin 14-mer sequence applied to conduct FRET studies, a MALDI-ToF was not obtainable. The results of Brown *et al.* indicate the failure of a porphyrin modified DNA 14-mer to ionise in mass spectrometry instruments may not only be size dependant but also sequence dependant. These abnormalities could be potentially related to the hydrophobic nature of the porphyrin itself, deriving from its large and flat aromatic structure, the result of which may lead to strong π-π stacking interactions forming intermolecular porphyrin DNA aggregates.

These π-π interactions may dominate at high concentrations / dry conditions as required for ESI and MALDI-ToF analysis respectively. Potentially, ionisation of the sample could lead these complex aggregations passing through the detector, not seen in the spectral window of the instrument due to significantly greater molecular mass, or this is the result of porphyrin DNA failing to ionise. Fragmentation of DNA can also lead to ambiguous mass spectrometry results, whereby fragments of the desired product are observed rather than the desired monomeric species. The glycosidic bond is susceptible to fragmentation upon protonation[78, 79]. Fragmentation can result in cleavage of the phosphodiester bond at the 3' O-C. Terminal bases too are susceptible to fragmentation upon laser exposure.

2.2.6.4 Porphyrin DNA analysis

Zinc and free-base (2H) porphyrins attached to DNA were required for FRET experimentation. Demetallation of the zinc porphyrin to 2H porphyrin was achieved directly from solid phase DNA synthesis due to the relatively strong acidic conditions required for DMT deprotection (see section 1.18)[10], which consequently resulted in zinc demetallation of the starting material. Figure 31 shows the UV-Vis and fluorescence spectra's of 2H porphyrin DNA in duplex form (single-strand form can result in quenching of fluorophores absorption and emission spectra). The intense Soret band and the four Q bands of the 2H porphyrin were observed at 424 nm and 505-655 nm respectively (see introduction for details). Excitation at 424 nm yields the fluorescence emission spectra typical for 2H porphyrins. The emission comprises of three bands at 606, 656 and 717 nm. The absorption and emission bands are related to the orbitals and the symmetry of the porphyrin, as described by Gouterman *et al.*[53, 54, 55].

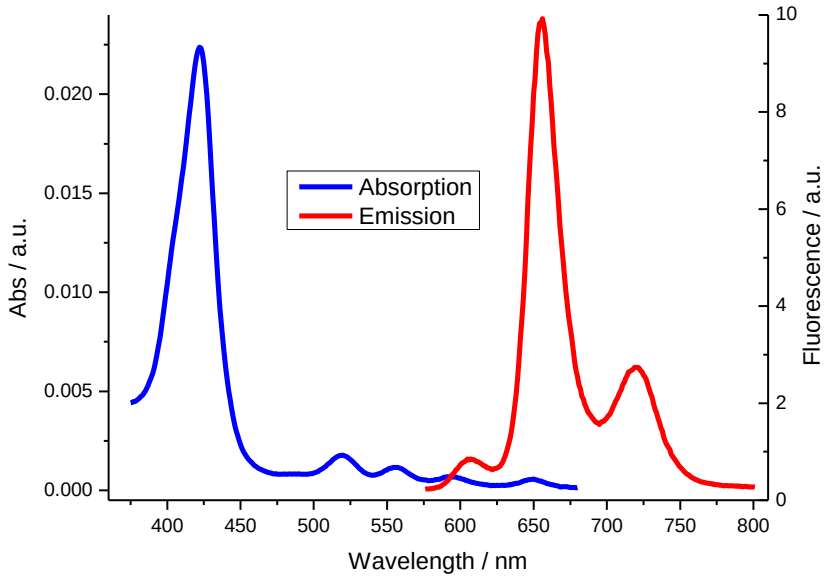


Figure 31 UV-Vis and fluorescence spectra of duplex 2H porphyrin Reduced DNA, Soret band absorption at 424 nm and 4 Q bands from 505-655 nm (blue line). Excitation at 424 nm yields the fluorescence emission (red line).

The purified 2H porphyrin DNA was re-metallated using excess zinc acetate. A solution of 2H porphyrin DNA was heated to 80 °C in water, the reaction was monitored by UV-Vis and fluorescence spectroscopy to observe metallation progression (step G). Excess zinc was required for efficient metallation, which consequentially resulted in DNA precipitation, possibly due to zinc-phosphate coordination. The excess zinc was removed using the metal chelator ligand EDTA in excess, followed by NAP size exclusion chromatography. The DNA is orders of magnitude larger than zinc, EDTA and zinc-EDTA complex, and so was purified effectively. Figure 32 displays the UV-Vis absorption and fluorescence (excitation at 426 nm) spectra of zinc porphyrin DNA in duplex form. Zinc metallation of the 2H porphyrin resulted in a Soret band red-shift from 424 nm to 428 nm; the number of Q bands reduce from four to two due to the change in orbitals and hence symmetry. The overall intensity of absorption of the zinc porphyrin DNA decreased with respect to 2H porphyrin DNA by 2.7 times, from 150000 to 57000 M^{-1} cm^{-1}. The spectra displays typical emission for zinc metallated porphyrins, comprising of two peaks at 606 and 656 nm[53, 55, 56].

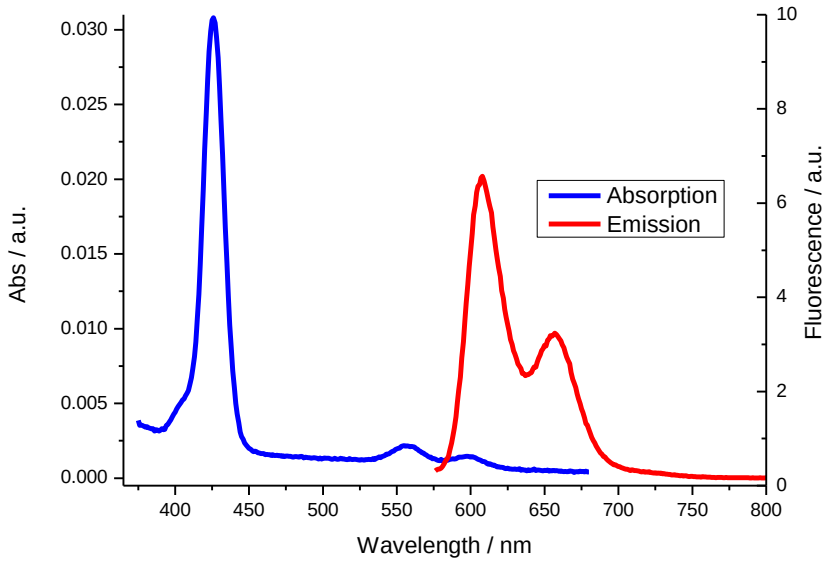


Figure 32 UV-Vis and fluorescence spectra of duplex zinc acetylene porphyrin Reduced DNA, Soret band absorption at 426 nm and 2 Q bands from 555-605 nm (blue line). Excitation at 426 nm yields the fluorescence emission with major peaks at 605 and 656 nm (red line).

2.3 Zinc and 2H porphrin *TerB* DNA sequences

The DNA sequence selected for porphyrin molecular ruler application is a 21-mer from the Tus-*TerB* DNA recognition sequence (see Figure 33)[65, 67]. The sequence was selected to test and calibrate the porphyrin DNA molecular ruler. The calibration was conducted in duplex form only, in the absence of Tus (to study Tus binding interactions, an elongated DNA version would be required since the vast majority of the *TerB* DNA sequence is responsible for strong-binding interactions, any porphyrin labelled within will have a direct effect on binding, and therefore perturb the system).

For FRET analysis, simplification is crucial to interpret results. Steady-sate FRET efficiencies are normally calculated with respect to the donor emission, therefore it is desirable to keep the donor position constant throughout the assay, and hence move the position of the acceptor. This enables simpler and more reliable FRET and tr-FRET distances by preventing complications surrounding the analysis of the E_{FRET} term as discussed previously (see section 1.6.2). It is also desirable to have the donor not positioned at the extremes of the DNA sequence to prevent the possibility of fraying

and intercalation anomalies. However, it is also advantageous to not have the donor at the centre of the DNA, potentially giving rise to approximately symmetrical results, since a broad range of distances and angles is required to characterise the FRET pairs. Concluding these remarks, the strand side Z^1, and the 8^{th} base was selected for donor incorporation (5' → 3') as shown in Figure 33. The H^1 strand gives rise to a range of acceptor modification sites without resorting to changing the native sequence, and therefore is ideally suited to expose the full extent of the porphyrin molecular ruler. The iamge also shows Macromodel images of the seven zinc and 2H porphyrin DNA combinations selected. Helix side and top views are shown, the zinc porphyrin moiety position is fixed, whilst the 2H porphyrin position is varied for the seven combinations. The side view shows the distance variation between intramolecular porphyrins, whilst the top view shows the angular variation.

H$^{\#}$ = 2H-poprhyrin

Z$^{\#}$ = Zn-porphyrin

5′ A **7**AA G**6**A **5**G**4** **3**G**2** AAC **1**AA AG 3′

3′ T ATT CAT ACA ACA **1**TG ATT TC 5′

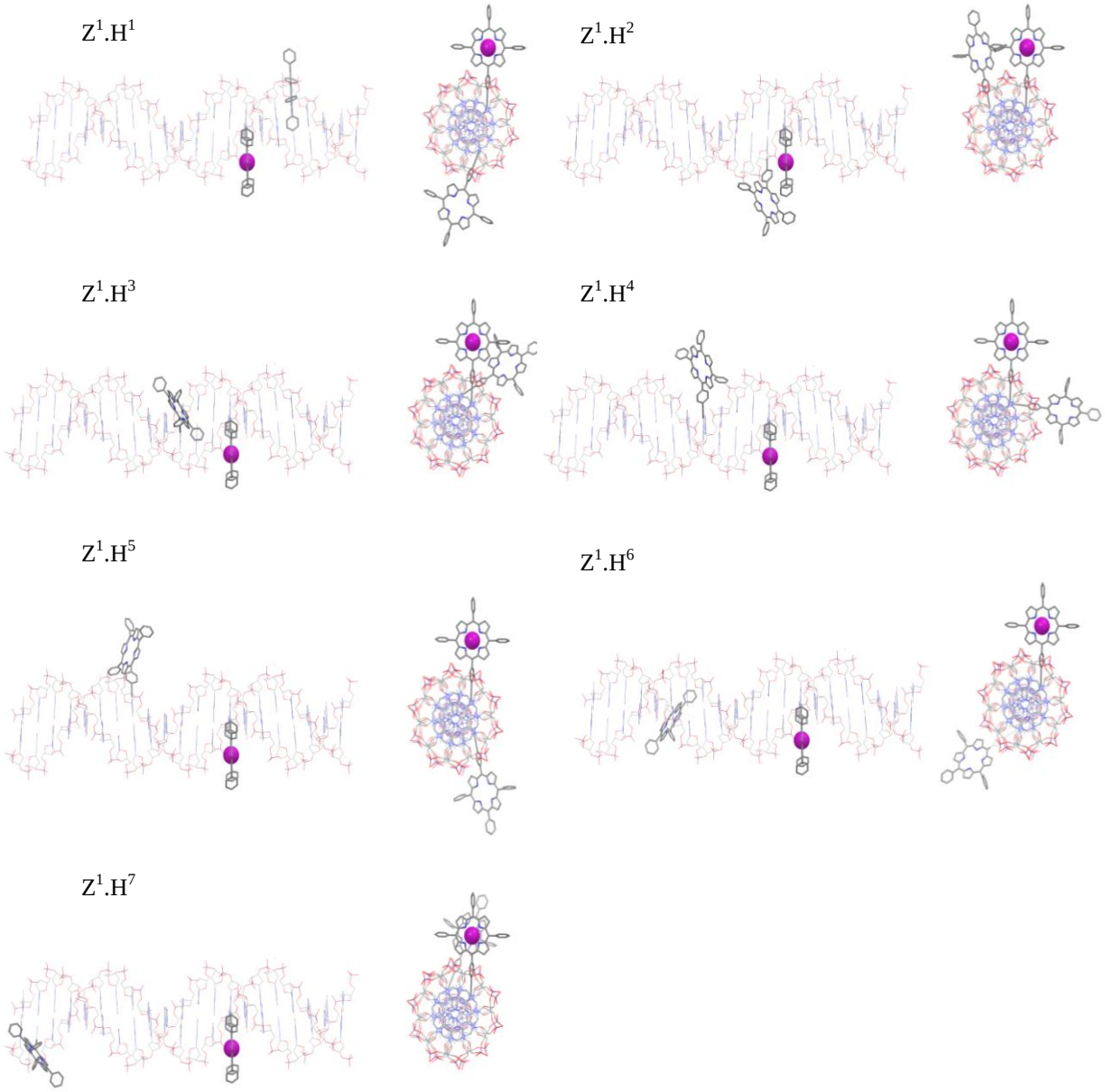


Figure 33 Zinc and 2H porphyrin modified DNA *TerB* sequences in letter format (top), where red numbers show 2H-porphyrin modification and blue number shows Zn-porphyrin modification, and the seven Macromodel Zn and 2H porphyrin DNA combinations side and down the duplex views (bottom), zinc is shown as a purple sphere.

2.3.1 Modelled distances and angles expected

To test the validity of the molecular ruler experimental distances, in particular for FRET analysis - which requires distances to be calculated from a reliable model, two appropriate DNA models were applied (section 1.6.2). Firstly, an approximate 3.6 Å distance separation between modification positions was applied, as used by Liley *et al.*, this model assumes a constant distance separation between modifications, and does not account for the helical rotation of the duplex (see Table 1)[3]. This model was therefore expected to be less accurate over shorter distances. Macromodel was used to generate the other model. Macromodel is ideally suited to obtain a satisfactory model without requiring extensive preparation and processing time. Macromodel has been used by previously in similar areas, and have produced complementary results between computational models and experimental data[80, 81]. The main limitation for modelling using Macromodel is its solvent shell approximation. Macromodel does not approximate for counter-ions along the DNA phosphate back-bone, one can only apply water or chloroform as a solvent shell. This can potentially have implications for a true representation of biological systems; therefore caution must be applied to the models and distances and angles obtained.

Table 1

FRET Pair	Base-pair separation	3.6 Å approximation	Macromodel template distance (Å)
Z^1. H^1	2	10.8	30.2
Z^1. H^2	0	3.6	10.9
Z^1. H^3	2	10.8	13.7
Z^1. H^4	3	14.4	21.9
Z^1. H^5	5	21.6	34.9
Z^1. H^6	7	28.8	39.8
Z^1. H^7	11	43.2	41.1

The Macromodel template model and the 3.6 Å model of the *TerB* DNA vary considerably at shorter distances, table 4. The 3.6 Å model does not take into account the helical curvature as the base-pair separation between the porphyrins change. When applying models to the rigidly attached modifications for zinc and 2H porphyrin analysis, the modifications have well defined relative dipole-dipole orientations, accurate modelling was required to interpret the experimental FRET distances, and so both models were compared, see section 2.6.3.

2.4 Spectroscopic analysis of donor-acceptor duplex

For subsequent FRET analysis, UV-Vis (see 2.4.1) and fluorescence (see 2.4.2) spectroscopy was performed on the zinc and 2H porphyrin DNA (in duplex format) to identify the peaks responsible from the donor and acceptor entities.

2.4.1 UV-Vis analysis of donor-acceptor duplex

UV-Vis absorption of the donor only, acceptor only and donor with acceptor in duplex format is shown in Figure 34, with all combinations at the same concentration to give relative results. Analysis was required for FRET molecular ruler analysis of the porphyrin region in order to determine the FRET prerequisites values (see introduction 1.6.2). The UV region in Figure 34 (α), displays the DNA absorption region, which has small variations in peak height due to background porphyrin absorption which varies for zinc and 2H porphyrins containing entities. This assignment is in agreement with the order of the peak heights for the main Soret band absorption in Figure 34 (β): donor < acceptor < donor and acceptor. The donor only peak displays significant quenching compared to the acceptor only peak; a direct result of the addition of zinc to the central porphyrin cavity. The combination of donor and acceptor in duplex format gives rise to the combined spectrum (black line), which displays the expected additive behaviour, increases to the intensity of the Soret band and to two of the Q bands.

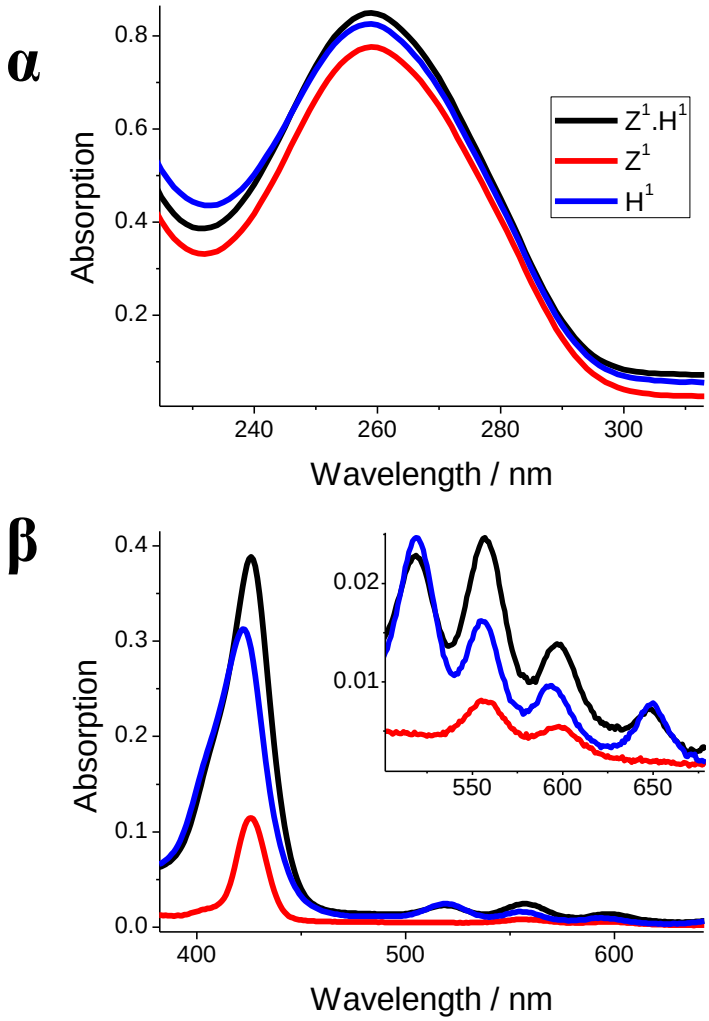


Figure 34 UV-Vis absorption of zinc porphyrin duplex DNA (Z^1 red line), 2H porphyrin duplex DNA (H^1 blue line) and zinc and 2H porphyrin duplex DNA ($Z^1.H^1$ black line), DNA region (α), porphyrin region (β) and Q band region (inset). Donor only (red line), acceptor only (blue line) and donor and acceptor combined (black line) at 2 μM in Tus buffer, 1 cm quartz cell, scanning at 600 nm per minute.

2.4.2 Fluorescence analysis of donor-acceptor duplex

To the same samples analysed by UV-vis, fluorescence spectroscopy was performed to identify the emission peaks deriving from the donor and acceptor entities in preparation for steady-state FRET experimentation. Fluorescence excitation at 426 nm of the donor only, acceptor only and donor with acceptor gives rise to the fluorescence spectra observed in Figure 35, with all combinations at 2 μM for relative

comparison. Comparing the donor only and the acceptor only (red and blue lines respectively); the donor only exhibits a higher quantum yield; the zinc porphyrin absorption at 426 nm is approximately 30 % the intensity of the 2H porphyrin, yet the total fluorescence emission of the donor is 65 % of that compared to the acceptor. The fluorescence emission of the donor only plus acceptor only is greater than the two combined experimentally by approximately 23 %. The reduction in fluorescence upon annealing can arise from various excited state reaction pathways, for example, non-emissive FRET, induced exciton coupling and thermal mechanistic pathways. All factors must be considered for accurate FRET and CD analysis.

Initial FRET analysis of the $Z^1.H^1$ combination shows reduced emission of the predominately donor only peak at 605 nm, indicating the expected FRET from zinc to 2H porphyrin. However, analysis of the acceptor only peak at 717 nm does not give the expected increase in acceptor emission corresponding to FRET, but instead a reduction in fluorescence emission, this process is further investigated below (see 4.11 steady-state FRET analysis).

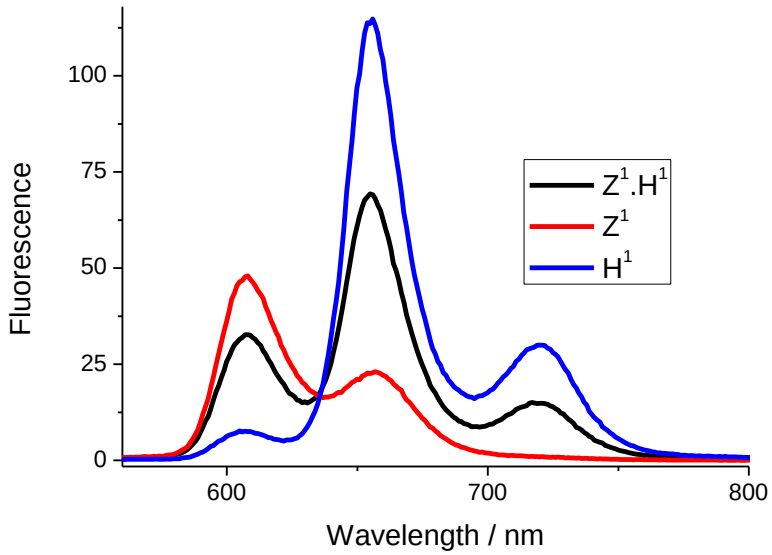


Figure 35 Fluorescence emission of zinc porphyrin duplex DNA (Z^1 red line), 2H porphyrin duplex DNA (H^1 blue line) and zinc and 2H porphyrin duplex DNA ($Z^1.H^1$ black line), excitation at 424 nm, at 2 μM in Tus buffer, 1 cm quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

2.5 Duplex formation and stability

The extent of the modifications effect on the DNA duplex had to be identified to ensure an accurate representation of duplex DNA. Owczarzy and co-workers reported that long linkers have a reduced destabilising effect on duplex stability compared to shorter linkers[82]. Since the porphyrin modifications are attached to the DNA on short rigid acetylene linkers, the destabilising effect of the moiety was investigated thoroughly. Native poly-acrylamide gel electrophoresis (PAGE) and thermal melting studies are simple methods to help identify DNA duplex analysis (see appendix section 2.27 for porphyrin modified native PAGE). UV melting studies at 260 nm enabled direct comparison of entropic and enthalpic values of the DNA between unmodified and porphyrin modified DNA, i.e. the direct modification effect. CD studies enabled the helical nature of the DNA to be identified as a function of temperature; the annealed and denatured forms were compared directly[80, 81]. With UV-Vis, CD and fluorescence studies, the changes associated with the modification and hence the modifications environment can be monitored to identify any structural changes from the native structure. Native DNA and modified DNA was annealed using protocols defined in the appendix (see appendix section 8.1 for DNA hybridisation).

2.5.1 UV-Vis melt at 260 and 424 nm

Monitoring the absorption change at 260 nm, the UV melt of the unmodified version of the *TerB* sequence gives rise to a melting temperature of approximately 50 °C (see Figure 36 and table 2). With single and double porphyrin modified DNA the melting temperature decreases by approximately 3 °C, indicating the porphyrin moieties have a small destabilisation effects on the overall duplex stability in comparison to the native version. The porphyrin moiety was likely to have a destabilising affect on the uridine base it is directly attached to, resulting in a reduction of base-pairing strength and hence base-pairing recognition. This perturbed local interaction will have subtle repercussions either side of the modification site, reducing the base-paring strength for a few of the surrounding base-pairs. However, the UV melting study indicates the effect

is not significant, a 3 °C reduction in thermal stability. Incorporation of the porphyrin modifications does not lead to a significant change in the melting temperature when comparing unmodified, singly and doubly porphyrin modified DNA.

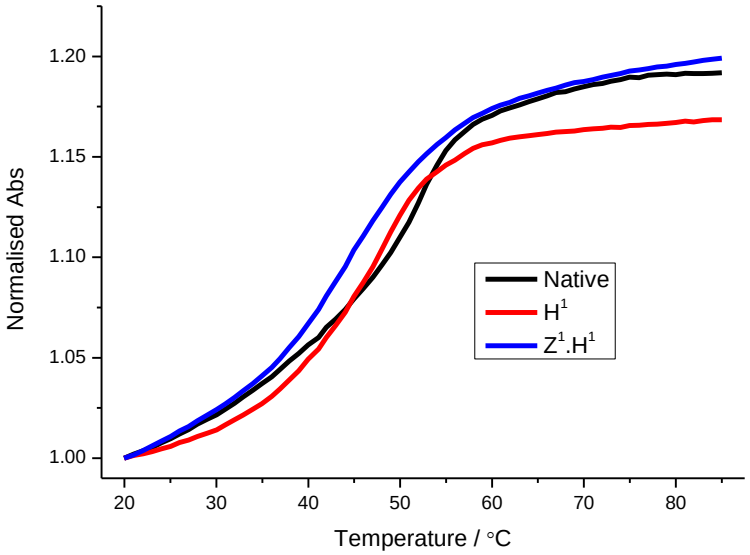


Figure 36 Normalised UV melting spectrum of unmodified comps (black line), 2H modified DNA (red line) and zinc and 2H modified DNA at 2 μM monitoring at 260 nm in Tus buffer, melting at 1 °C / min, heating from 20 to 85 °C, normalised with respect to the absorption intensity at 20 °C.

Table 2

DNA *TerB* sequence	**UV_{260} T_m (°C)**
TerB - unmodified	50
H^1 - 1x porphyrin modified	47
$Z^1.H^1$ - 2x porphyrin modified	47

Monitoring the Soret band region absorption at approximately 426 nm as a function of temperature enabled the direct porphyrin environment to be monitored in

duplex and single-stranded form above and below the melting temperature. The spectra in Figure 37 (α) shows the melting and annealing curves for the doubly porphyrin modified reduced DNA at 426 nm (collected simultaneously to the DNA region melt from above). Differentiating the visible melting curve yields a T_m 6 °C above the DNA region T_m. The change in T_m from DNA melting to porphyrin melting is most probably the result of the weaker A-T base-pairs melting. A-T Watson-Crick base-pairs have 2 hydrogen bonds compared to the 3 from G-C base-pairs. The reduced DNA sequence contains both types of base-pairs; when heating is applied the A-T regions melt first, followed by the G-C regions, generally differentiation of the melting curve cannot distinguish the individual melting regions since they are not generally well resolved. The T_m is the midpoint value between duplex DNA and single-strand DNA. At the DNA region T_m one expects 50 % duplex DNA and 50 % single-strand DNA, hence there is still strong association of the duplex, which may give rise to the increased porphyrin region T_m.

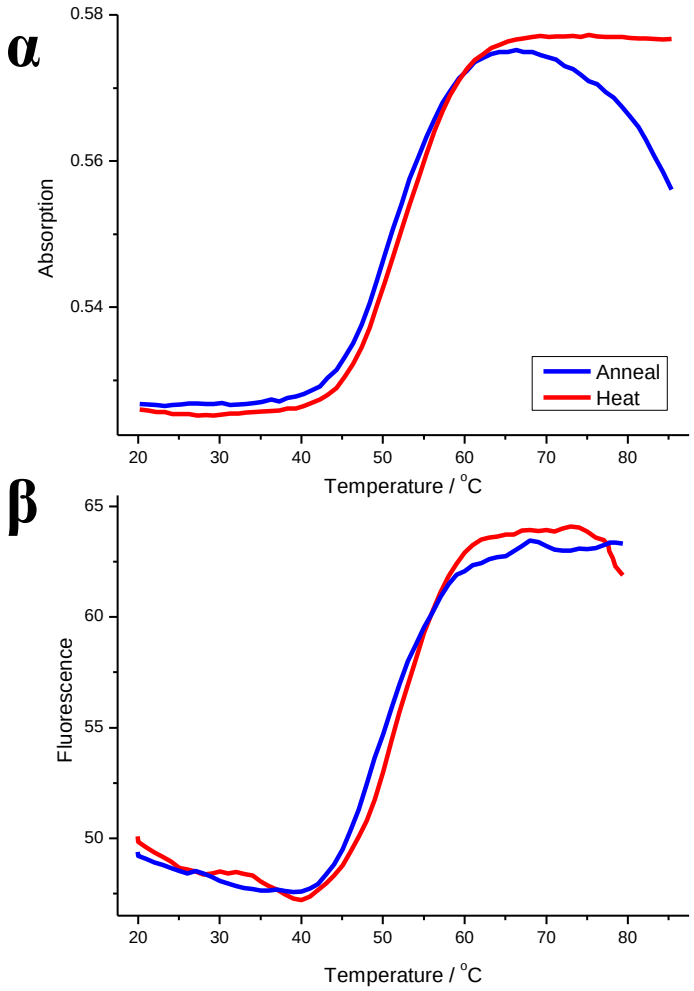


Figure 37 Zinc and 2H porphyrin modified DNA ($Z^1.H^1$) visible temperature melting analysis, monitoring at 426 nm (α) and fluorescence temperature analysis excitation at 426 nm monitoring at 656 nm (β), at 2 μM in Tus buffer, heating and annealing at 1 °C per minute, 1 cm quartz cell, 500 PMT voltage.

Fluorescence melting analysis of the porphyrin modified DNA enabled the fluorescence of the porphyrin moieties to monitored upon denaturing and annealing. Fluorescence is typically more sensitive to the surrounding environmental change compared to UV-Vis spectroscopy. Monitoring fluorescence whilst heating enables the fluorophores environment to be identified, effects changing fluorescence include; double or single-stranded surrounding base environment, reduced or increased FRET as a result of changes to donor and acceptor fluorophore distance separation (DNA melting or annealing), and exciton coupling (also arising from coupling fluorophores which

change depending on the distance separation of coupling moieties). The fluorescence spectra in Figure 37 (β) shows zinc and 2H porphyrin DNA heating and cooling whilst monitoring the predominantly zinc region emission at 605 nm. An approximate sigmoidal curve gives rise to a melting temperature of 50 °C in excellent agreement with the UV-Vis melting from above. The increase in fluorescence can arise from reduced quenching; from double-strand to single-strand melting, whilst also increases the distance separation of the two porphyrin entities, therefore a reduction in FRET and exciton coupling is also expected, both effects can give rise to the increased fluorescence observed at 605 nm above the melting temperature.

2.5.2 CD temperature analysis

CD melting studies enabled the helical structure of the DNA to be monitored as a function of temperature. Analysis of the native and singularly porphyrin modified DNA also enabled the identification of the porphyrin moieties effect on the surrounding duplex DNA. At temperatures lower than the melting temperature (T_m), the DNA CD curvature was expected to be constant; changes to the helical structure were expected as the temperature approaches the T_m (47 °C for H^1 as determined from UV melting above). Figure 38 shows CD set temperature analysis of the unmodified version of the *TerB* sequence at temperatures below and above the T_m. This CD curvature is typical for B-type duplex DNA at 5 and 20 °C; the + / - / + CD curvature and their relative ratios indicate the B form of DNA[80, 81]. Heating above the T_m the intensities and their relative ratios decrease; the 220 nm positive peak disappears, indicating dominating single strand character. The CD of the single porphyrin modified version also displays similar CD curvature below and above the T_m. These results indicate the porphyrin moiety has negligible effect on duplex character.

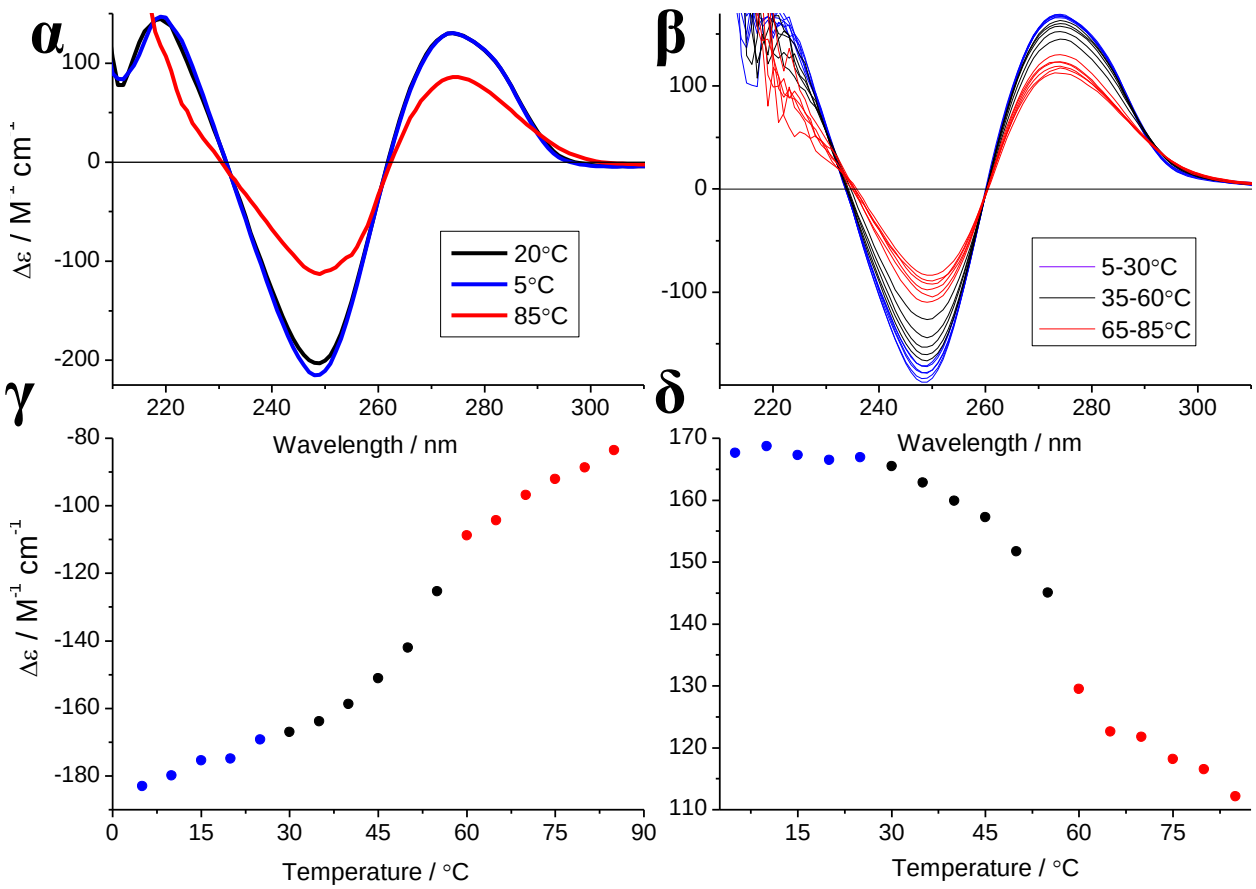


Figure 38 CD melting temperature analysis of unmodified *TerB* DNA (α) and 2H porphyrin modified duplex DNA (H^1) melting spectrum (β), 250 nm cross-section of melt (γ) and 275 nm cross-section of melt (δ), analysis at 4 μM in Tus buffer, heating at 1 °C per minute, using a 1 cm quartz cell, scanning at 2 seconds per nm.

2.6 Determining the prerequisite FRET values

To identify the E_{FRET} term from which experimental distances can be identified, the Förster distance must be determined first (see section 1.6.2). The Förster distance is dependent on the quantum yield of the donor, the extinction coefficient of the acceptor and the spectral overlap of donor emission and acceptor absorption. These values are determined below, after which the Förster distance for the forward and reverse (zinc-to-2H porphyrin, and the 2H-to-zinc porphyrin) values are identified. Applying the forward Förster distance value, the calculated E_{FRET} term can be identified. From which a plot of E_{FRET} versus donor-acceptor distances is plotted. Two distance models are

plotted which take into account rigid and flexible dipoles, Sören Preus model and the averaged κ^2 term respectively (see section 2.6.3).

2.6.1 Quantum yield of donor

The quantum yield of the donor must be identified to calculate the distance between donor and acceptor when defining the Förster distance (see introduction, 1.6.2). The quantum yield of zinc porphyrin tethered to DNA in aqueous solvent was calculated with respect to the quantum yield of a known standard; quinine sulphate, provided the absorption intensity at the desired excitation wavelength is below 0.1 whereby preventing any inner filter problems. The quantum yield was calculated using equation 2. A value of 0.12 was obtained for the acetylene zinc porphyrin quantum yield with respect to quinine sulphate. The value is low compared to other established fluorophores such as FAM (0.95) and TAMRA 101 (1.00). The low quantum yield reduces the FRET efficiency, and shift the Förster distance to smaller distances (r / R_o).

2.6.2 Spectral overlap and Förster distance

To identify the spectral overlap between donor emission and acceptor absorption, and hence determine the Förster distance of zinc and 2H porphyrin DNA, Junis Rindermann applied the spectral overlap equation (see section 1.6.2) to the data obtained in Figures 34 and 35. The calculated value obtained for spectral overlap from donor to acceptor was 3.5×10^{14} M^{-1} cm^{-1} nm^4, which, as expected is greater than the reverse energy transfer process, where energy is transferred from acceptor to donor (3.5×10^{13} M^{-1} cm^{-1} nm^4). These results are in agreement with the previous literature results for energy transfer analysis between zinc and 2H porphyrin supramolecular arrays in organic solvents. Applying the Förster distance equation to the quantum yield of the donor, extinction coefficient of acceptor, a κ^2 value of 3 / 2, the refractive index of the medium and spectral overlap, the R_0 value for donor to acceptor was identified. As expected the donor to acceptor value is greater than the acceptor to donor value, at 30.3

Å and 20.6 Å respectively. This result correspond well to the design of the seven 21-mer porphyrin labelled DNA combinations, therefore, energy transfer is expected for the majority of the zinc and 2H porphyrin combinations, from which distances and angles can be experimentally identified, see FRET results section 2.7.

2.6.3 Calculated E_{FRET} versus distance

The E_{FRET} efficiencies for the seven zinc and 2H-porphyrin FRET pair combinations were calculated to compare experimental and theoretical results. Two models were applied to calculate the E_{FRET} terms; a model which accounts for rigid dipoles through the κ^2 term, and an averaged dipole approximation model which does not. The rigid model theoretical results were calculated by Sören Preus using his customised software[12]. This model is more relevant to the rigid porphyrin molecular ruler since the porphyrins are attached to the DNA rigidly, therefore non-averaged type behaviour of the FRET probes was expected. Figure 39 shows the two calculated E_{FRET} models versus distance separation for the seven zinc and 2H porphyrin DNA combinations. When calculating the rigid model, the relative dipole moments for each donor and acceptor pair was aligned to the effective dipole moment through the acetylene bond, as described in the literature by Berova et al.[85]. For small donor-acceptor distance separations: $Z^1.H^1$, $Z^1.H^2$ and $Z^1.H^3$, base-pair separations of 2, 0 and 2 respectively, the fluorophores should conduct FRET approaching 1. As the distance separation increases, the FRET efficiency decreases, as which point the κ^2 effect becomes more pronounced; $Z^1.H^4$ to $Z^1.H^5$ decreases sharply due to dipoles being aligned perpendicular, whilst $Z^1.H^6$ to $Z^1.H^7$ increases as the dipoles are aligned parallel. The calculated results show deviation from the 2 / 3 dipole-dipole approximation model. After conducting the steady-state fluorescence experiment, deconvolution of the donor emission peaks was expected to yield FRET efficiencies in agreement with the rigid model

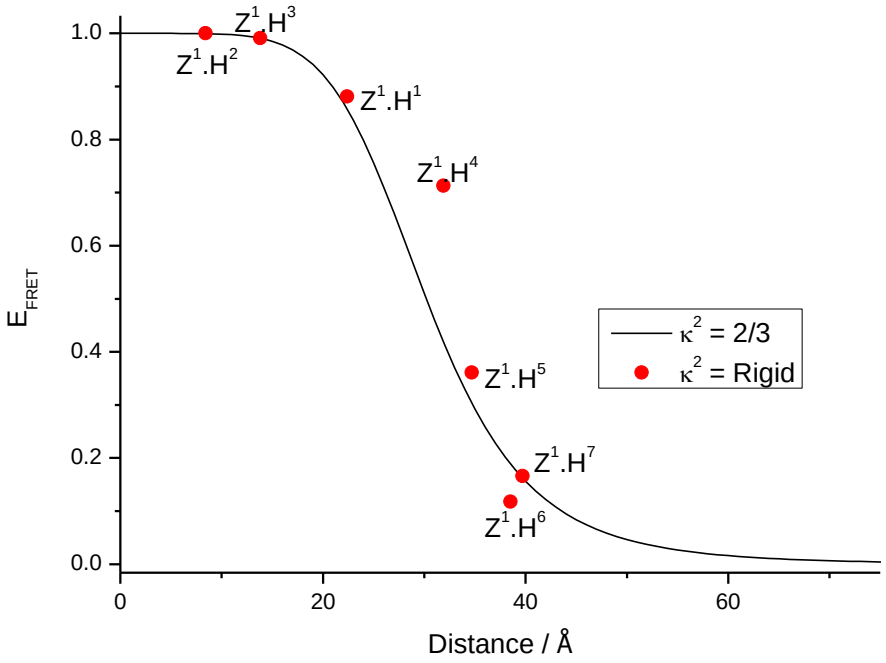


Figure 39 Calculated E_{FRET} versus Distance for zinc and 2H porphyrin combinations, κ^2 rigid (red dots) and averaged (grey line), rigid values calculated by Sören Preus, averaged model distances calculated using the 3.6 Å model.

2.7 Steady-state fluorescence

Steady-state fluorescence experimentation was applied to the seven zinc and 2H porphyrin combinations to analyse the FRET molecular ruler properties. Two normalisation methods were employed to identify the E_{FRET} values from the donor emission, normalised absorption and normalised concentration.

2.7.1 Matched absorption steady-state fluorescence analysis

After annealing the donor-acceptor combinations in a 1 : 1 molar ratio, plus donor only and acceptor only duplexes (the latter two were required for FRET deconvolution, see FRET introduction section 1.6.2), the absorption at 426 nm was matched to the same intensity. To give sufficiently intense results, the absorption of the

Soret band was matched to 0.58. Normalising the absorption at the excitation wavelength (426 nm) enables the same number of photons entering each system therefore giving comparative analysis of each systems fluorescence emission, and hence comparative FRET results. Figures 40 and 41 show the matched absorption spectra. The $Z^1.H^7$ combination exhibits an anomalous absorption peak; the Soret band peak width is wider than the other combinations, whilst the fluorescence emission is significantly decreased. This result for $Z^1.H^7$ indicates potential single-strand quenching, potentially arising from the H^7 porphyrin modifications terminal DNA positioning. Therefore, the $Z^1.H^7$ combination was omitted from the matched absorption steady-state fluorescence experiment.

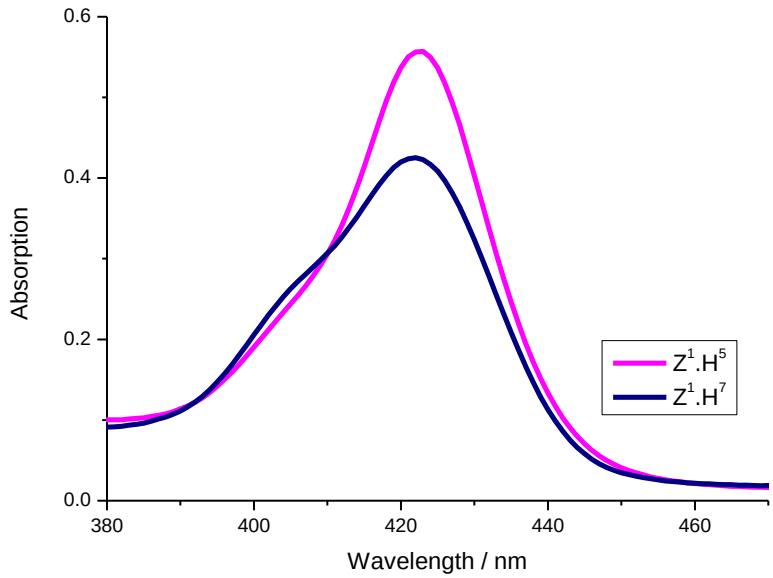


Figure 40 $Z^1.H^5$ and $Z^1.H^7$ porphyrin absorption region (pink and blue respectively), matched absorption at 426 nm, in Tus buffer, scanning at 600 nm per minute, 1 cm quartz cell.

The remaining FRET pairs absorption peaks were successfully matched to 0.58 intensity at 426 nm. Figure 41 (α) displays the matched visible absorption spectra, with Q band zoomed in regions. Interestingly the Q bands display subtle changes potentially arising from exciton coupling and / or surrounding environmental effects, which could be independent of the main Soret band absorption - the Soret belongs to the second excited state, whilst the Q bands belong to the first excited state. To the matched

absorption FRET pairs, excitation at 426 nm yielded the fluorescence emission spectrum shown in Figure 41 (β). The fluorescence spectra display large variation in intensity and peak height ratio for the remaining six zinc and 2H porphyrin DNA combinations.

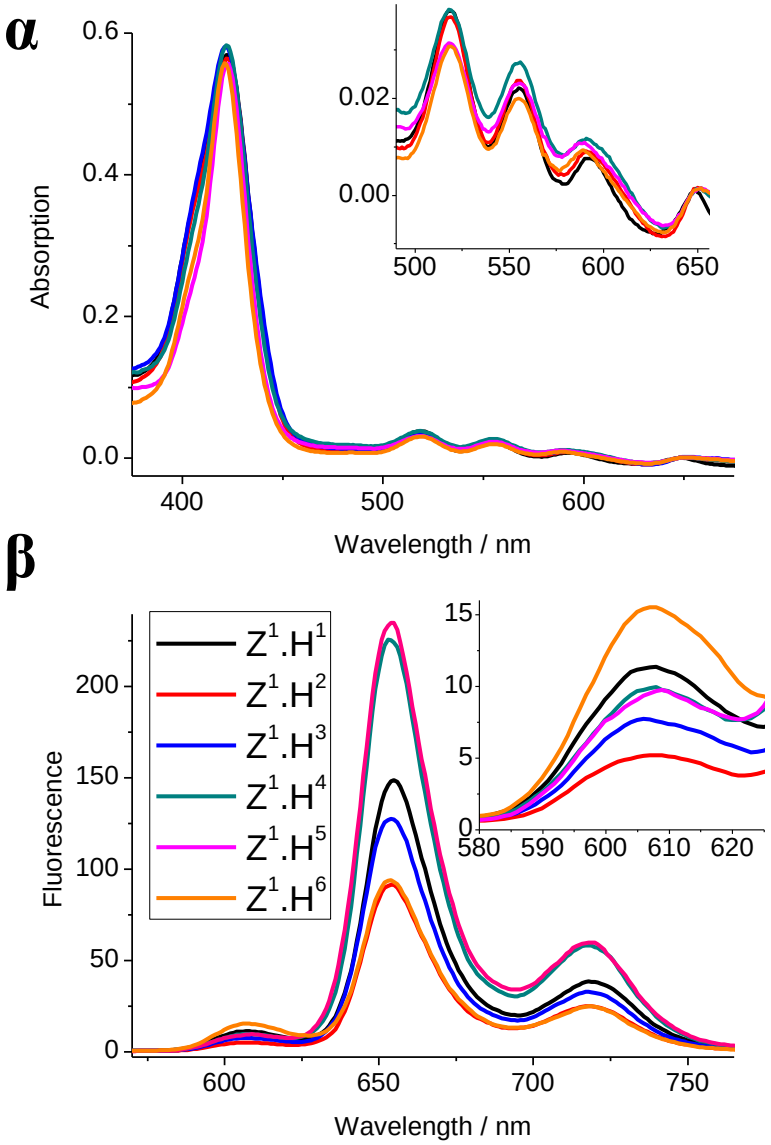


Figure 41 Z^1 with H^1 to H^6 porphyrin combinations visible absorption region (α) and g band region (inset) and fluorescence emission (β) and donor region (inset) at matched absorption at 426 nm, in Tus buffer, 1 cm quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

Observing the fluorescence emission peaks of the predominantly donor region at 607 nm for the six FRET pair combinations (see Fluorescence analysis of the donor only and acceptor only, Figure 35) and applying the FRET equation, equation 5, normalising and plotting E_{FRET} versus distance yielded Figure 42. The calculated values (shown in red from above), have an approximate agreement with the experimental data points (blue); there is a general decrease with increasing distance, however, accurate analysis of the dipole-dipole orientation is not observed. For combinations $Z^1.H^4$, $Z^1.H^5$ and $Z^1.H^6$, where the orientation factor was expected to be pronounced - E_{FRET} was expected to increase going from $Z^1.H^5$ and $Z^1.H^6$, these results are not in agreement with the calculated values. An alternative approach was to match the concentration rather than absorption, this experiment means the same number of entities are being analysed. The steady-state fluorescence experiment was repeated using matched DNA concentrations at 2 μM. Matching the concentration means the same number of porphyrin entities are measured, rather than number of photons entering the system, an approach used by other groups [10].

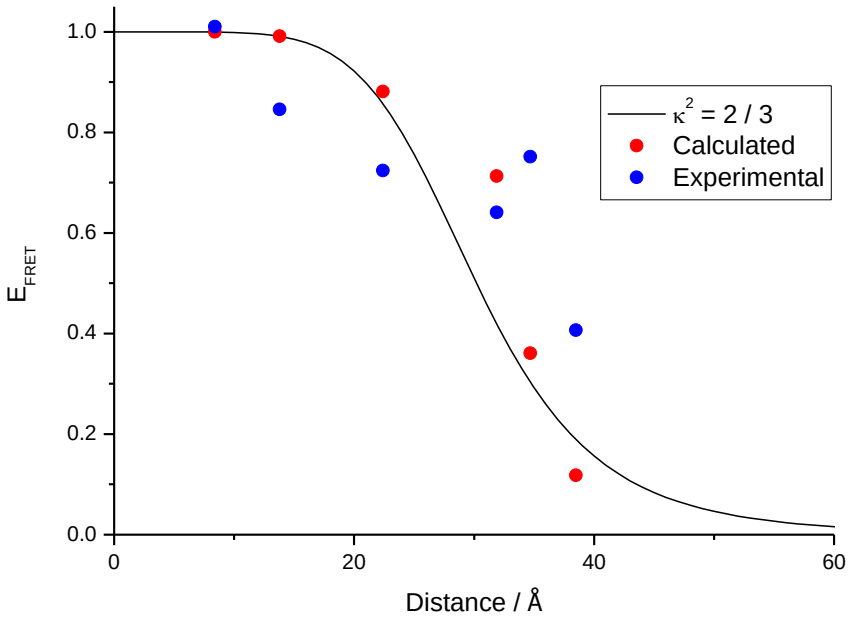


Figure 42 E_{FRET} versus Distance for Z^1 with H^1 to H^6 porphyrin combinations, matched absorption, averaged (black) and calculated (red) κ^2 dipole-dipole orientation variable, experimental (grey) and normalised experimental (blue), distances calculated using the 3.6 Å model.

2.7.2 Matched concentration steady state fluorescence analysis

Annealing the $Z^1.H^1$ to $Z^1.H^6$ porphyrin combinations at 2 μM then applying steady-state fluorescence excitation at 426 nm yields the fluorescence emission spectra as shown in Figure 43 and donor region (inset). Taking the fluorescence emission peaks of the predominantly donor region (607 nm) for the six FRET pair combinations and applying the FRET equation, normalising and plotting E_{FRET} versus distance, yields Figure 44. The results correlate well with the calculated values; the trend with distance and angle are in good agreement. Two combinations do show some deviation; $Z^1.H^2$ and $Z^1.H^6$. Potentially, this deviation could be related to other forms of energy transfer; steady-state analysis does not give information about the lifetimes of excited state. Potentially, energy is transferred to and from donor and acceptor, as shown previously in Figure 35 analysis. This effect will complicate steady-state analysis; time resolved and single molecule FRET analysis is required to clarify the zinc and 2H porphyrin energy transfer mechanism. The analysis could further verify the steady-state fluorescence analysis, and overall the use of zinc and 2H porphyrins as an accurate molecular ruler. The results show a fairly accurate FRET molecular ruler has been achieved by steady-state fluorescence analysis.

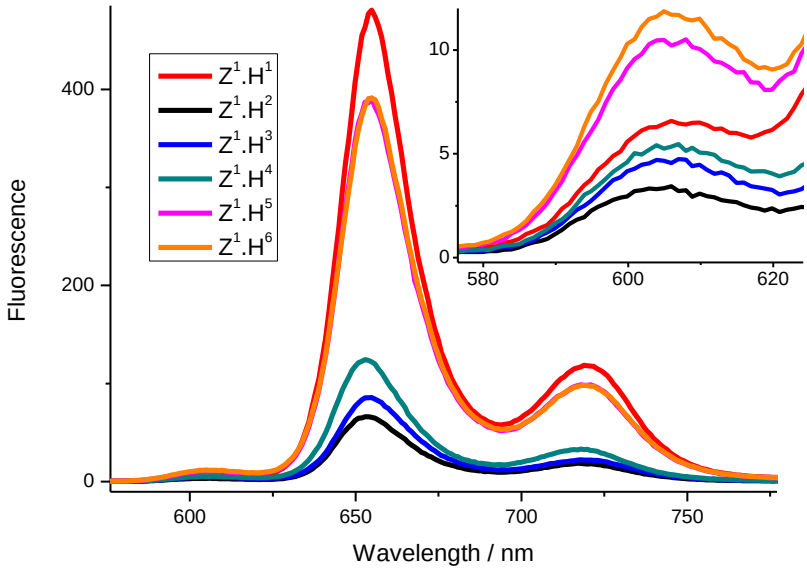


Figure 43 Z^1 with H^1 to H^7 porphyrin combinations fluorescence emission and donor emission peak region (right inset) at 2 μM, in Tus buffer, 1 cm cell, 500 PMT voltage.

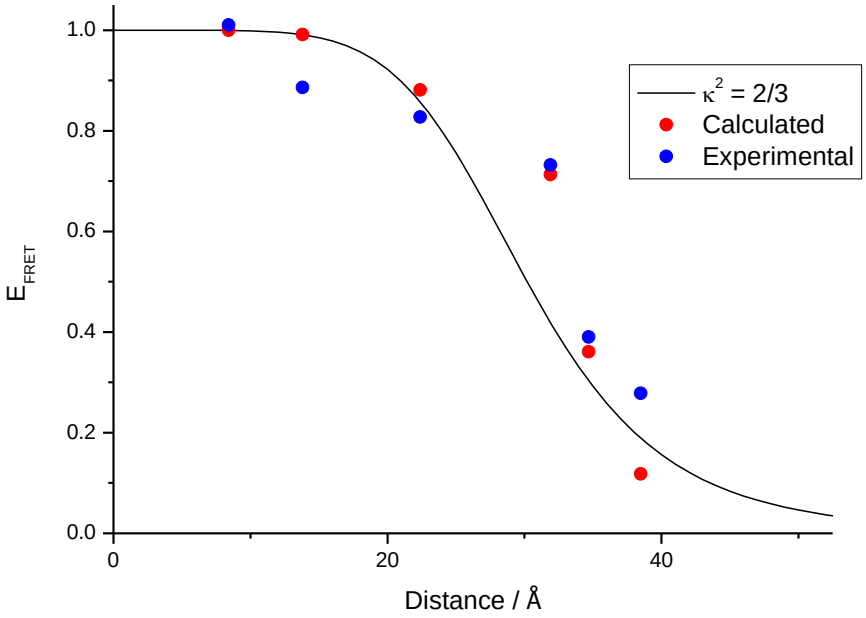


Figure 44 E_{FRET} versus Distance for Z^1 with H^1 to H^7 porphyrin combinations at 2 μM, averaged (black) and calculated (red) κ^2 dipole-dipole orientation variable, experimental (grey) and normalised experimental (blue).

Both donor and acceptor peaks absorb at the same wavelength (acceptor only emits at 717 nm) which adds to the complexity of steady-state FRET analysis. Potentially, the deconvolution from donor and acceptor emission is further complicated by exciton coupling between porphyrin moieties. CD is a ground state effect whilst fluorescence is a relaxation of the first excited state phenomena, exciton coupling can effect the absorption bands and fluorescence emission depending on the distance and angle between porphyrins; Anderson showed peak broadening of the Soret band with increasing porphyrin moieties[36]. Another factor to consider when deconvoluting the donor and acceptor peaks is the surrounding base quenching affect resulting from the relative proximity to guanine's (guanine base has the greatest quenching properties)[86]. These effects cannot be identified by steady-state fluorescence analysis, and so time-resolved fluorescence was employed to identify relaxation pathways by collaborators Junis Rinderman and Prof. Pavlos Lagoudakis.

2.8 Porphyrin DNA CD molecular ruler

The seven porphyrin DNA combinations were applied to circular dichroism (CD) spectroscopy to identify the porphyrins exciton coupling molecular ruler properties. The intensity and sign of exciton coupling signal is strongly dependant on the distance and angle between the coupling porphyrins (see chapter 1). CD analysis of the Soret band region at approximately 426 nm gave rise to the exciton coupling shown in Figure 45. Coupling porphyrins typically give rise to a bisignet coupling pattern. The Figure shows the intensity and sign of the exciton coupling between the porphyrin moieties change significantly depending on the distance and angle between the porphyrins. Combinations $Z^1.H^2$ and $Z^1.H^3$ show potential deviation from typical exciton coupling; a result of the close proximity of the porphyrins, which can perform electron transfer *via* a Dexter mechanism. The $Z^1.H^7$ CD spectrum yields a low exciton coupling spectra, potentially a result of the large distance between the two coupling moieties. However, the H^7 2H porphyrin entity from $Z^1.H^7$ undergoes an intercalation interaction rather than the desired duplex formation, as observed in UV analysis previously (see Figure 40). Therefore, this low exciton coupling could be the result of a dominating quenching interaction arising from intercalation, rather than displaying the distance cut-off of exciton coupling.

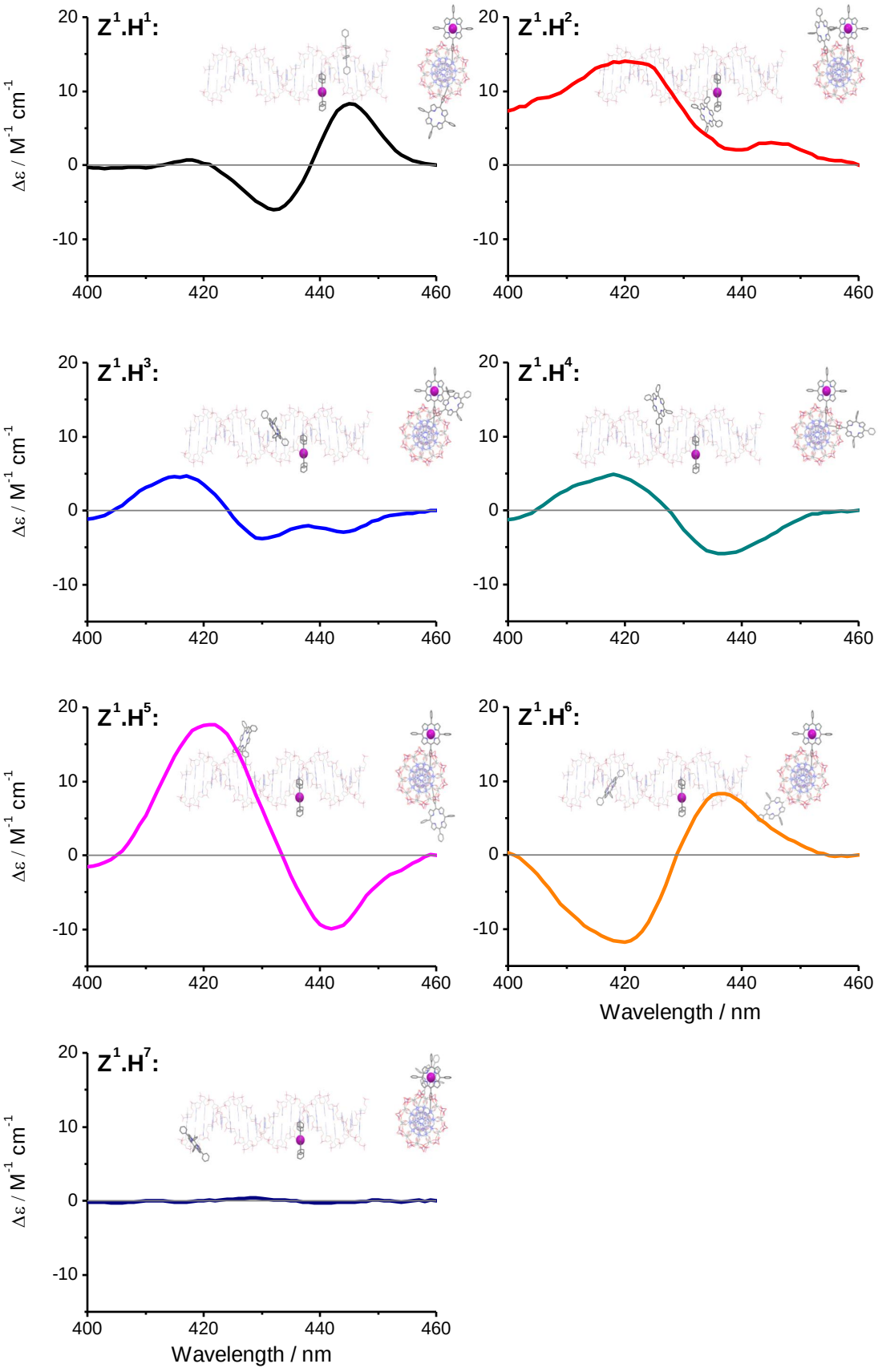


Figure 45 Zinc and 2H porphyrin Soret band region CD spectra for the seven *TerB* DNA combinations, samples at 6 μM, in Tus buffer, 1 cm quartz cell, scanning at 2 seconds per nm.

The porphyrin transition dipole moment is ascribed through the unsymmetrical linker of the porphyrin moiety[85]. Exciton coupling is distance and angle dependant; maximum exciton coupling occurs at short distances and when dipoles are in-between parallel and perpendicular alignments (see introduction, CD section 1.6.3). Combinations: $Z^1.H^4$, $Z^1.H^5$ and $Z^1.H^6$ contain a sequential one Watson-Crick base-pair change. The CD spectra display a corresponding change in sign of exciton coupling; a - / +, - / + and + / - respectively. These three combinations show the strong angle dependency of exciton coupling, the inversion of CD sign arises from the sin (2θ) dependency. This angle dependency dominates up to at least 30 Å, as shown when comparing $Z^1.H^1$ and $Z^1.H^6$ combinations; these porphyrins are separated by approximately 10.8 and 28.8 Å respectively, a 20 Å difference. Yet the angles and hence the sign of the dipoles are calculated to be quite similar; 155° and 135° (angles calculated from the template model from Macromodel). The experimental data shows the same CD sign is observed, yet the intensity is subtly reduced in the expected order; $Z^1.H^6$ is greater than $Z^1.H^1$. There is also a sizable Davydov splitting change for these two combinations; the point at which the exciton coupling crosses zero is approximately 440 nm and 430 nm for $Z^1.H^1$ and $Z^1.H^6$ respectively.

The CD spectra for the porphyrins exhibit strong excitonic coupling, which translates to a sensitive distance and angle dependency, in agreement to the results obtained by Lewis *et al.*[2]. These results show that the porphyrins have a strong potential for use as a CD molecular ruler. Even the $Z^1.H^7$ may use in future experiments, this combination exhibits no exciton coupling as a result of the H^7 porphyrins positioning. This phenomenon can be viewed as a positive effect, i.e. the porphyrin CD ruler can be used to identify porphyrin stacking interactions within DNA; it can highlight the presence of a single-stranded DNA environment via the absence of exciton-coupling. In order to identify the distance and angles between the zinc and 2H porphyrins, and thus identify the DNA conformation, a CD model is required from which the results can be compared, analogous to the method employed by Lewis *et al.*. The porphyrin CD molecular ruler was applied to study the Tus-*Ter* protein-DNA complex in chapter 4.

2.9 Iodine SAXS molecular ruler

A novel molecular ruler technique based on the metal-to-metal scattering between two zinc porphyrin moieties attached to DNA was synthesised using the strands described above (see section 2.3). However, to obtain sufficient signal-to-noise in SAXS experimentation, the DNA concentration was required to be greater than 100 μM. Therefore, a porphyrin based system would not be applicable for SAXS molecular ruler studies. In conclusion of these results, an iodinated version of the porphyrin *TerB* DNA sequence was synthesised and applied. For complete molecular ruler analysis, both the singularly iodinated and doubly iodinated *TerB* combinations was required, both in duplex form, from which the appropriate subtractions could be made to potentially identify the iodine-to-iodine scattering, and hence obtain a iodine-iodine distance.

2.9.1 Iodinated DNA synthesis

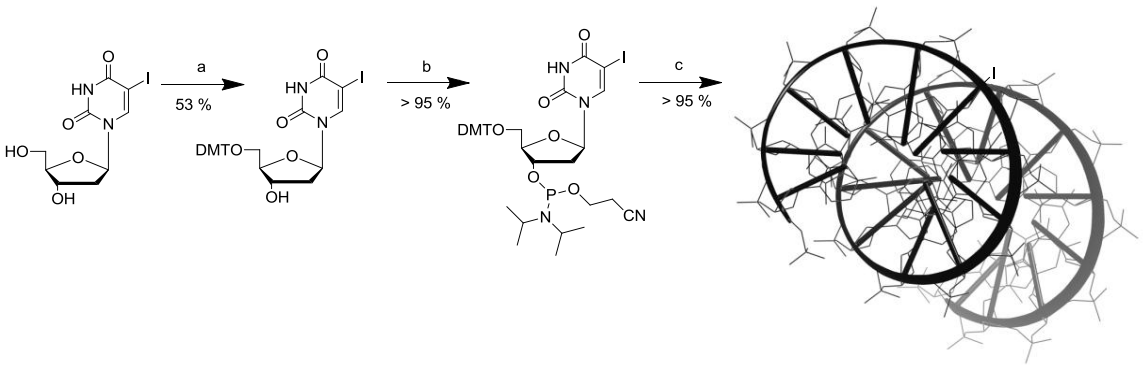


Figure 46 Iodinated DNA Synthesis[48].

a DMT-Cl, Pyridine, DMAP

b DCM, DIPEA, CEP-Cl

c Solid phase DNA synthesis

Iodinated DNA was required for SAXS studies (see introduction, chapter 1). 5-iodo-deoxyuridine was protected at the 5' position of the sugar using protocols established from Stulz *et al.*, Hobbs, and Richert *et al.*[42, 47, 48].

2.9.1.1 Iodine DNA synthesis - step a

Small portion-wise addition of tritylchloride over six hours was required for selective 5' DMT protection. Excess or extensive addition resulted in DMT addition at the 5' and 3' hydroxyl positions, the primary and secondary alcohols on the sugar respectively. M*ono*- and di-DMT monomers have very similar retention values, the more direct *mono*-DMT addition obtained in the reaction vessel, the easier the purification and the better the yield.

2.9.1.2 Iodine DNA synthesis - step b

DMT 5-iodo-deoxyuridine was incorporated into DNA *via* the 3' phosphoramidite modification (see appendix sections 8.6-8.8 for experimental protocol). After the reaction went to completion, as verified by TLC, the product was purified by precipitation and filtration, followed by washing with degassed hexane multiple times. Due to the oxygen sensitivity of the product, only the molecular weight confirmation was performed. However, activity of the product, and hence confirmation of the phosphorus (III) oxidation state was confirmed by the trityl reading from the DNA synthesiser. HPLC and mass spec analysis of the DNA product further verified purity and product relative mass. DNA synthesis was conducted on a 1 μmol scale using solid support chemistry. The relative iodine modified uridine phosphoramidite concentration and coupling times were doubled to allow for the reduced coupling efficiencies of modified nucleobases.

2.9.1.3 Iodine DNA synthesis - step c

Attention was paid to the method of synthesis of the multiple *TerB* iodinated DNA strands. The active phosphoramidite has a limited lifetime before oxidation of phosphorus III to V. To ensure efficient coupling the monomer was synthesised and consumed in one day. Since 8 iodinated positions were incorporated into 8 DNA

sequences, in total requiring a synthesis time of approximately twelve hours. The DNA combinations were synthesised up until the position of modification before the iodine phosphoramidite synthesis, whereby reducing the total iodine phosphoramidite coupling time. After iodine modification coupling, the remaining bases were coupled, to which deprotection and cleavage of the single-strands to solution was conducted in ammonia at approximately 55 °C for six hours. The ammonia was removed under vacuum overnight. Small organic molecules and small failure sequences were removed using a size exclusion column in deionised water, the first 1.5 mL fraction was collected and removed under vacuum. The samples were redissolved in HPLC loading buffer (10 mM TEAA / acetonitrile 99 : 1), to which HPLC analysis and purification was performed. The mass of the pure samples was confirmed using negative mode ESI micro-ToF analysis. Micro-ToF analysis required high temperatures to ionise the DNA efficiently from the TEAA buffer (55 °C).

To the pure iodinated DNA strands, each of the required combinations in 100 nanomole quantities were measured and annealed slowly at 1 °C per minute to room temperature as required for SAXS studies at BESSY II in Berlin. The combinations were passed through a S75 size exclusion column to purify double-stranded DNA from single-stranded DNA predominantly arising from errors associated with the extinction coefficient and measurement of the samples. However, double-stranded DNA is in an equilibrium with the single-stranded form, although the equilibrium is far over to the double-stranded side, passing the annealed DNA through a size exclusion column may distort the equilibrium and lead to peak broadening. Figure 47 shows the chromatogram for Iodine *TerB* duplex DNA eluting at approximately 32 mL. Purification was performed in deionised water to prevent problems associated with salt concentrations when concentrating the sample (extremely accurate solvent baselines are required for SAXS experimentation). Conducting the purification in deionised water reduced the melting temperature of duplex, therefore additional attention was paid to the single-stranded region (elution volumes > 40 mL). Negligible single-stranded DNA and small nucleotide fragments appear at approximately 44 and 67 mL. The pure peaks were collected and concentrated under vacuum using a freeze drier.

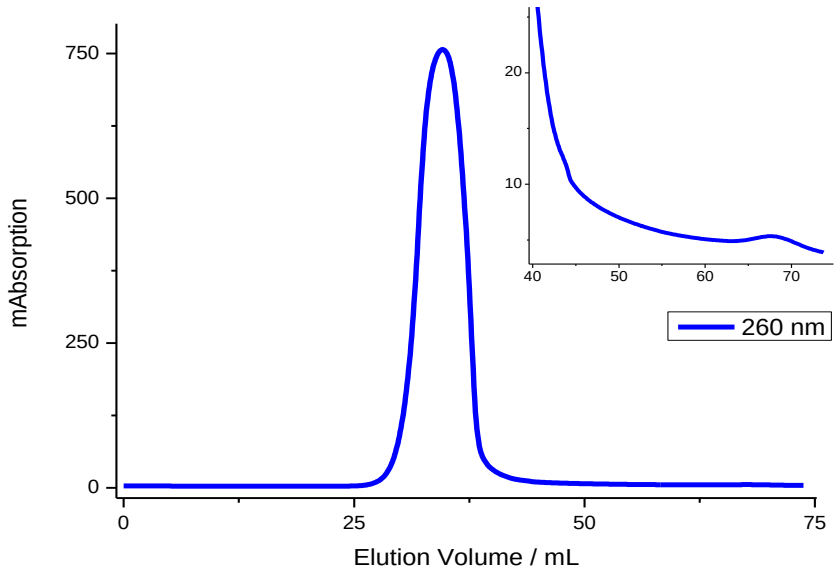


Figure 47 Size exclusion chromatogram of Iodine DNA in duplex form, monitoring at 260 nm, 100 nmol injection, flow rate 1 mL per minute, inset zoomed in single-strand and single nucleobase elution peaks.

The strands were purified by HPLC, annealed in duplex form, and then further purified by gel filtration to remove any single-strands (see 2.4). The pure samples were concentrated under vacuum, dissolved in pure water, and finally re-annealed before SAXS analysis.

2.9.2 Iodine DNA SAXS analysis

The iodine combinations: $X^1.Y^1$ and $X^1.Y^6$ (as shown in Figure 48, top) consist of two iodine-iodine inter-strand distances, a small distance separation where the iodine atoms are separated by approximately 10.8 Å (calculated using the 3.6 Å model in chapter 2), and a larger distance separation where the iodine's are separated by approximately 39.6 Å. These distances should correspond to a Q peak scattering at approximately 2 nm^{-1} and 1.5 nm^{-1} respectively, since scattering is proportional to 2 multiples of π over the actual distance[38, 39]. The experiment was conducted at BESSY II in Berlin using a SAXS beamline instrument. Firstly, the background scattering was

subtracted from the data (buffer and cell scattering), then normalised at a low Q range where the scattering is expected to be the same, i.e. where no iodine scattering was expected. The remaining scattering is plotted in Figure 48 (bottom). The results show deviancies between the small and larger iodine-iodine distances towards higher Q, correlating to different scattering distances. This change is in the expected order; $X^1.Y^1$ has a greater scattering intensity at higher Q than $X^1.Y^6$. However, the iodine-iodine scattering is not a well defined peak which clearly displays a strong correlation. To identify the metal-to-metal scattering in greater detail, a more intense beamline could be applied, with even higher sample concentration, however, additional precautions are required to ensure no DNA aggregation at this level.

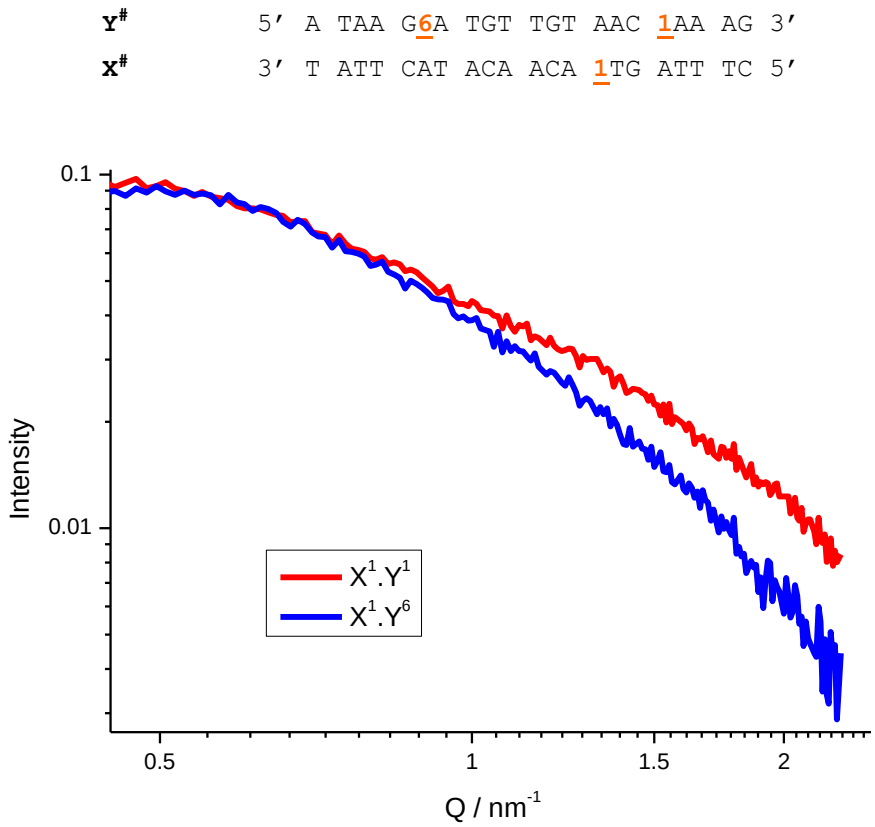


Figure 48 $X^1.Y^1$ and $X^1.Y^6$ iodinated DNA duplexes (top), iodine modifications shown as orange numbers. Iodine SAXS spectra obtained at 15000 eV, 3443 cm detector distance, 0.1 cm cell, at 400 μM (bottom).

2.10 Summary

The porphyrin DNA molecular ruler concept was introduced, along with synthetic details and key mechanisms (section 2.1 and section 2.2). The DNA sequence selected was defined and modelled. Two models were applied to identify the different porphyrin-porphyrin DNA intramolecular distances (section 2.3). UV-vis, fluorescence and CD melting studies and native PAGE analysis identified the formation of stable porphyrin DNA duplexes with respect to the native form (section 2.4 and section 2.5). These results indicated the porphyrin moiety has a negligible effect on the duplexes native character, which is important for native DNA analysis.

The spectroscopic properties of the zinc and 2H porphyrin DNA was introduced (section 2.6). Fluorescence analysis of the donor in the absence and presence of the acceptor showed the porphyrins do conduct energy transfer in aqueous conditions. However, conventional energy transfer from donor to acceptor (zinc to 2H porphyrin domains) was not observed, as shown by decreases in both the donor and acceptor emission peaks at 607 and 717 nm respectively. The result suggests a complex energy transfer mechanism, i.e. energy could be transferred interchangeably between zinc and 2H porphyrin moieties. To identify this energy transfer process in greater detail time resolved FRET analysis was performed by collaborators Junis Rinderman and Prof Pavlos Lagoudakis at the University of Southampton.

Analysis of the donor peak by steady-state fluorescence spectroscopy, conversely, did display good molecular ruler properties (section 2.7); the normalised concentration experiment gave results in good correlation to those calculated by Sören Preus. Sören's model accounts for the relative dipole-dipole orientation of the donor and acceptor, rather than using an averaged value, which is a necessity when attempting to generate a highly accurate FRET molecular ruler. To further develop the porphyrin ruler, using spectroscopically more different FRET pairs will ease FRET analysis, and also increase the Förster distance (provided there is greater spectral overlap) to give a greater spread of FRET results and therefore distances covered, for example, applying rigid FRET pairs: terpyridine (chapter 5) and 2H porphyrin; both are rigidly attached to DNA via the acetylene bond, they have excellent spectral overlap, and the quantum

yield of the terpyridine is high. Therefore, terpyridine and 2H porphyrin FRET pairs look like ideal candidates to further the highly accurate molecular ruler field.

CD analysis of the Soret band region of the zinc and 2H porphyrin DNA combinations displayed strong ground state exciton coupling interactions (section 2.8). Individual base-pair CD molecular ruler resolution was established; exciton coupling displayed sensitive distance and angle dependencies depending on the different porphyrin-porphyrin DNA alignments. Individual base-pair resolution was obtained. To further characterise and extract a distance and angle between the zinc and 2H porphyrins, an appropriate model is required from which direct comparisons could be made, analogous to the method Lewis *et al.*. With a valid model and experimental results to complement, the area of a CD molecular ruler should expand. CD has some big advantages to a FRET molecular ruler; CD typically is more sensitive to distance and angle changes, and therefore is more susceptible to small conformational changes. When studying biological systems, the DNA and protein region can also be monitored to complement results, and give further information about binding and induced conformational change.

SAXS analysis of the zinc porphyrin DNA did not reveal scattering at low concentrations (section 2.9). Applying significantly higher concentrations would of resulted in extensive porphyrin intermolecular interactions and subsequent aggregation. In conclusion, an iodinated version of the porphyrin system was synthesised and applied to the SAXS beamline at BESSY II. Two iodine-iodine distances were analysed at high concentrations, yielding scattering patterns at long and short iodine-iodine distances. The differences in scattering correlated to different iodine centre-to-centre scattering. However, to characterise the SAXS molecular ruler properties in detail, other iodine-iodine DNA combinations are required With a range of distances analysed, the data must be fitted to a model from which distance information can be extracted and a SAXS molecular ruler established.

Overall, the porphyrins have been shown to have applications as a versatile molecular ruler. Individual base-pair resolution was possible using FRET and CD techniques. Complete interpretation of the experimental results are complex, time resolved fluorescence spectroscopy should be employed to identify any complex energy transfer mechanisms, and a simulated model is required to extract the distances and

angles between the exciton coupling CD results. The results, however, are very promising; using a porphyrin labelled system, two types of molecular ruler analysis is possible; after annealing the zinc and 2H porphyrin DNA, instant steady-state fluorescence and CD analysis can be conducted without even changing the cell, from which vast information about the porphyrin environment can be extracted.

The porphyrin DNA molecular ruler was applied to investigate protein induced conformational change, within a protein-DNA (Tus-*TerB*) complex (see chapter 4). For appropriate controls, established FRET pairs FAM and TAMRA labelled DNA was applied in parallel. Firstly, an elongated *TerB* DNA version was designed and synthesised. The FAM and TAMRA control version was analysed by FRET in the absence of Tus (see chapter 3), and then in the presence of Tus (see chapter 4).

FAM and TAMRA control experiment

3.1 Introduction

To test the reliability and the linker effect of the zinc and 2H porphyrin FRET molecular ruler, FAM and TAMRA FRET pairs were tested in parallel. Fluorescein (FAM) and tetramethylrhodamine (TAMRA) fluorescent molecules have been used extensively as FRET pairs to study dynamic DNA interactions[20, 21, 18]. The fluorophores are ideally suited as FRET pairs due to the high quantum yield of the donor moiety (FAM), high extinction coefficient of the acceptor moiety (TAMRA), and excellent spectral overlap between donor emission and acceptor absorption. The FRET pairs have an approximate Förster distance (R_0) of 54 Å (see section 1.9.2). Typically the probes are tethered to DNA via relatively long and flexible hexyl linkers. The hexyl linker means the fluorophores occupy a large area compared to rigidly linked fluorophores, and therefore any FRET distances obtained will contain large errors; typically base-pair resolution is not possible. This large area of occupation means the dipoles associated to the donor and acceptor moieties no longer have defined orientations on the FRET timescale, which is why an averaged κ^2 variable can be applied. To compare the linker effect between rigid and flexible linkers, FAM and TAMRA FRET pairs were applied to first study duplex DNA, analogous to the porphyrin system. After DNA only analysis, the FRET pairs were applied to study protein-DNA binding interactions in chapter 4.

For FAM and TAMRA Tus-DNA molecular ruler analysis, an elongated *TerB* DNA duplex sequence was designed from which modifications could be attached (the *TerB* sequence was introduced in chapter 1 (section 1.10), and applied to the porphyrin molecular ruler in chapter 2). The sequence was extended to prevent perturbing the direct protein-DNA binding interaction in subsequent Tus-DNA experiments. Also the large Förster distance value of FAM and TAMRA means very few distances would be covered if the FRET pairs were applied to the *TerB* sequence only. To identify which regions to expand, the different regions responsible for direct and indirect protein-DNA binding were analysed; the flanking permissive and non-permissive regions. With an extended DNA sequence selected, the FRET pair modifications positioning were identified. The positions were chosen to obtain maximum FRET information in the absence and presence of Tus. In the duplex only format, the FRET pairs were positioned to test their helical sensitivity, to give comparison to the zinc and 2H

porphyrin versions. For the complex analysis, with Tus bound, the flanking DNA sequence can potentially wrap around the protein, further stabilising the complex, to overall increase binding affinity, therefore having FRET pairs positioned either side of the binding domain was imperative to test this hypothesis. Any ambiguity associated with the DNA sequence, especially related to the non-permissive region, which was shown to display non-characteristic B-type behaviour could potentially be identified by FRET, therefore to also test this hypothesis, FRET pairs were also positioned within the same side of the binding domain.

Preceding the design and synthesis of the FAM and TAMRA molecular ruler, steady-state fluorescence analysis was conducted to identify the FRET values and therefore DNA conformation. To analyse these FRET values the quantum yield of the donor, extinction coefficient of the acceptor, and spectral overlap, and therefore Förster distance for each FRET pair combination was identified first. Three models were applied to identify the FRET distances from the E_{FRET} values; the conventional 3.6 Å model was compared to two Macromodel models; a template version, and a conformationally search version. The 3.6 Å model was tested since this model does not account for the helical rotation of DNA, which plays an important role at smaller FRET pair distance separations (see sections 1.6.2.1 and 2.6.3). After steady-state fluorescence experimentation, the E_{FRET} results were compared to the various distance models from which conclusions were made. After duplex DNA FRET analysis, the strands were analysed in the presence of Tus; with increasing protein concentration to monitor protein binding, and to observe any changes in DNA upon binding (see section 4.7.1).

3.2 Sequence screening

To identify an appropriate FAM and TAMRA labelled DNA sequence for subsequent protein-DNA experiments, the Kamada, termination and flanking regions of the Tus-*TerB* sequence were analysed.

3.2.1 Kamada sequence

```
5' A GTA TGT TGT AAC TAA 3'
3' T CAT ACA ACA TTG ATT 5'
```

Figure 49 Kamada sequence, direct protein-DNA interactions shown in yellow.

The 16-mer Kamada DNA sequence is from the crystal structure of the Tus-DNA complex obtained from Kamada *et al.* (see introduction, section 1.10.1). The Kamada sequence contains the region responsible for strong protein-DNA interactions (yellow). These interactions have been shown to make direct or water mediated protein-phosphate and protein-base contact. 17 of these contacts are sequence specific whilst 31 others are not[65, 66, 67]. The actual DNA complexed in the crystal structure contains more base-pairs. Purely thinking about DNA length alone, this DNA will be problematic for any modification study, modifications on such a short strand of DNA will distort the duplex's native character, leading to considerable negative effects on duplex stability (calculated T_m approximately 45 °C from Mfold[87]). Deviancies from the native form could potentially have large implications for subsequent protein binding study by disrupting these protein-DNA contacts. This 16-mer region is a necessity for any Tus-DNA studies and so was implemented in the final sequence, but the final sequence must be elongated.

3.2.2 *TerB* sequence

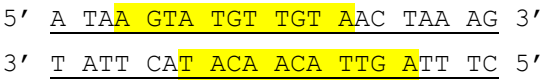


Figure 50 Termination B DNA sequence (*TerB*), direct protein-DNA interactions shown in yellow.

The 21-mer termination (*TerB*) sequence contains the region responsible for direct protein-DNA interactions (yellow). The Ter sequence contains the essential recognition Kamada sequence from above. It encompasses a 1.5 helical turn and is approximately 72 Å in length (calculated using the 3.6 Å model). The native *TerB* sequence exhibits a T_m of 58.2 °C in Tus buffer as identifed by UV melting studies (see appendix section 8.24). The *Ter* sequence contains a good number of thymidines which are exchangeable for FRET pair modification sites (see introduction, section 1.17). The sequence contains 38 % thymidine bases, with 7 thymidine's on one strand and 9 on the other strand. However, these modification sites encroach upon the binding region, since the majority are located within the strong protein-DNA binding region. Another factor to consider is that the overall distance of the DNA is not ideal for extensive FRET pair characterisation for FRET pairs containing a Förster distance greater than 30 Å. The maximum distance separation available for modifications on either strand is around 65 Å (since terminal labelling will lead to intercalation and quenching of fluorescence), and so applying FAM and TAMRA's with an approximate R_o of 54 Å will not test their full Förster distance range. Also he DNA sequence is probably not long enough to test the hypothesis of extended DNA wrapping around the protein, which can further stabilise the complex and increase binding affinity. Combining these factors, the *TerB* sequence is not ideal for FAM and TAMRA FRET experimentation.

3.2.3 *TerB* and flanking sequence

```
5′ GCC AGC TCC GAA TAA GTA TGT TGT AAC TAA AGT GGA TCA ATT CAT AA 3′
3′ CGG TCG AGG CTT ATT CAT ACA ACA TTG ATT TCA CCT AGT TAA GTA TT 5′
```

Figure 51 Termination and flanking region sequence *T&F*, direct protein-DNA interactions shown in yellow.

The 47-mer termination and flanking (*T&F*) sequence contains the *TerB* sequence with flanking regions termed the non-permissive (left of the *TerB* site as shown) and permissive sides (right of the *TerB* site as shown). The DNA total length is approximately 165 Å, the *T&F* DNA forms a stable duplex, with a melting temperature of 77.4 °C in Tus buffer (see appendix section 8.24). The sequence is ideal for FRET studies between Tus-DNA, modifications can be positioned outside the direct protein-DNA interaction region (yellow), thereby not interfering with protein binding (provided the mechanism of binding is a 'clamping' mechanism and not a 'sliding' one, see Tus-DNA chapter 5). The only limitation for the sequence is the yield when synthesising DNA of this length[10]. Using a 47-mer with a fully optimised solid phase DNA synthesiser, one would expect a maximum yield of approximately 62 % (assuming a 99 % coupling efficiency per base). With modifications present, the yield could decrease by approximately 20 % per modification coupling, as a result of steric hindrance of the modification coupling to the protruding base, and the effect of this modification on subsequent bases yet to be coupled (see section 1.18)[88]. In order to maximise the yield whilst still maintaining sufficient modification sites outside the binding region, the flanking regions were reduced. To decide the region to be edited, both flanking regions were analysed independently.

3.2.4 *TerB* and permissive sequence

```
5' T GGA TCA ATT CAT AA 3'
3' A CCT AGT TAA GTA TT 5'
```

Figure 52 Permissive flanking sequence (*Per*).

The permissive flanking region (from the right side of the *T&F* sequence above), the 15-mer DNA, has been shown to allow replication forks to pass from either direction, hence the term permissive region (see section 1.16). The *Per* sequence has a T_m of 51.5 °C and has typical UV melting characteristics of short DNA (see appendix section 8.24).

3.2.5 *TerB* and non-permissive sequence

```
α-Non  5' GCC AGC TCC GA 3'
β-Non  3' CGG TCG AGG CT 5'
```

Figure 53 Non-permissive flanking region (*Non*).

The non-permissive flanking region (*Non*) is an 11-mer DNA sequence from the left side of the T&F sequence. This region however, has been shown to be involved in the replication mechanism of *E.Coli*, it is from this region replication forks pass from one direction but not the reverse (see introduction 1.16)[65, 66]. The synthesis and purification of this 11-mer was problematic, extremely low yields for the β-non strand ranging from 2-5 % were obtained repeatedly. These results suggest uncharacteristic behaviour of the strand itself. HPLC analysis of the pure strand indicated a shoulder peak of the DNA region at 275 nm atypical for single-strand DNA. CD melting analysis (see appendix section 8.24) of the β-non strand yielded information of the strand in solution. At low temperatures an additional negative peak at approximately 255 nm and a positive peak shift to the 280 nm peak to 285 nm suggested possible G-quadruplex character[89].

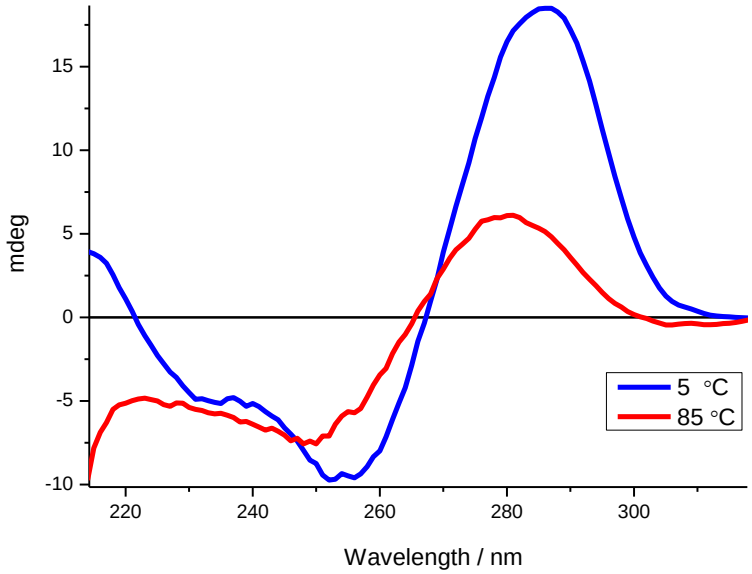


Figure 54 CD temperature spectrum of non-permissive flanking region, at 15 μM in Tus buffer, 1 cm quartz cell, scanning a 2 seconds per nm.

The non strand contains 5 guanines (G) which are the most likely source of this anomalous behaviour, guanines have the potential to form G-quadruplex structures as shown in Figure 54. In the G-quadruplex the guanines interact *via* Hoogsteen hydrogen bonds[10]. The G-quadruplex can form from intermolecular or intramolecular strands. Figure 55 shows a G-quadruplex deriving from two strands or from four single-strands. The intramolecular structure can form many different G-quadruplex structures, for example parallel or anti-parallel conformations. The non sequence contains 5 G's, four of which are in pairs. Potentially the strand in solution is forming an intermolecular or intramolecular G-quadruplex, or potentially the strand could adopt a single-stack version of the intramolecular G-quadruplex structure, however this may require extensive distortion in the intermediate DNA sequence, or a combination of the above[97-98].

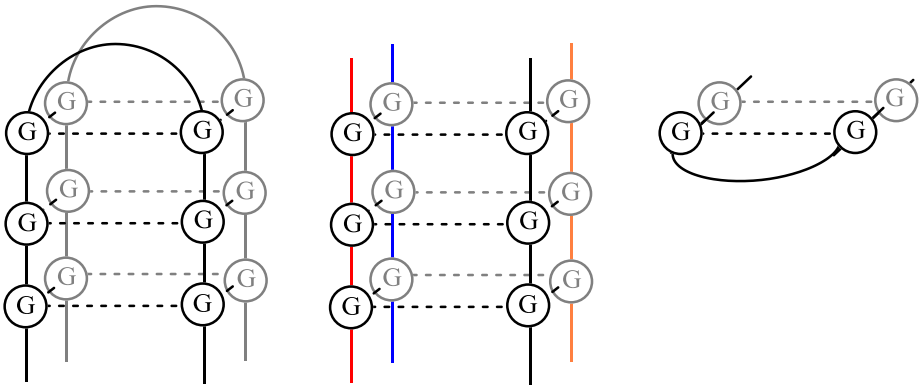


Figure 55 Examples of intermolecular G-quadruplex from two or four G rich strands, intermolecular and intermolecular strands respectively, and the hypothesised G-quadruplex single-stack structure based on the non sequence (left to right). Linear chains represented by solid lines, Hoogsteen bonds shown as dashed lines.

A mutated version replacing two of the G's in the β-non strand, bases 2 and 8 (3' → 5'), for a T and a C respectively, verified the G's dependency of the β-non strand (see appendix section 8.24). However, whether this G-quadruplex type structure is intermolecular or intramolecular has not been determined. These results could indicate sequence effects of the natural DNA sequence, potentially the region has a pre-formed alignment atypical for normal DNA, and therefore plays a role in DNA-Tus-fork recognition. This non-permissive region was therefore included within the FRET system for analysis. Additional attention was paid to non region, and to any potential conformational effects resulting when applying the FRET pairs.

3.2.6 Extended sequence

```
5′ GCT GCC AGC TCC GAA TAA GTA TGT TGT AAC TAA AGT GGA 3′
3′ CGA CGG TCG AGG CTT ATT CAT ACA ACA TTG ATT TCA CCT 5′
```

Figure 56 The Extended DNA sequence, direct protein-DNA contact region shown in yellow.

The 39-mer strands contain the ter sequence with flanking regions either side of the binding region, with the full non-permissive region and reduced permissive region.

The DNA total length is approximately 137 Å. The strands form a stable duplex with a T_m of 75.3 °C in Tus buffer (see appendix section 8.24). This sequence should be ideal for FRET calibration of FAM and TAMRA FRET pairs in the absence of the protein, and then in the presence of the protein. Modifications can be positioned outside the direct protein-DNA interaction region (yellow), and therefore not interfering with protein binding (provided the mechanism of binding is a 'clamping' mechanism and not a 'sliding' one, see below). Following these conclusions, the exact modification sites and hence FRET distances expected were identified.

3.3 FAM and TAMRA modification positioning

The Extended DNA sequence below is adapted from the Tus-*Ter* recognition sequence. The 39-mer DNA contains the direct protein-DNA interactions region (yellow), with flanking regions either side, termed the permissive and non-permissive sides (left → right as shown in Figure 56). This sequence was selected to study a range of FRET distances using the FRET pairs FAM and TAMRA. Modifications using the FAM and TAMRA deoxyuridine monomers described previously were selected to replace some of the native thymidine (T) positions. Any modifications incorporated into the direct protein-DNA region may have implications for protein binding and so was omitted from modification. Hence, only T positions in the permissive and non-permissive regions were modified. Concluding these remarks, having one donor position on one strand to simplify the FRET analysis (see section 1.6.2), and varying the acceptor position, would leave seven donor-acceptor combinations outside the binding region. To obtain greater FRET resolution, and hence more FRET combinations, if the donor and acceptor were modified in multiple donor and acceptor positions on the same sequence but on different strands, i.e. only having one modification per strand, the result would enable multiple combinations of donor-acceptor pairs and thus give rise to multiple FRET distances. In total seven pairs were selected, three from the donor and four from the acceptor, giving rise to twelve FRET pair combinations. The twelve combinations cover distances from approximately 3.6 to 155 Å, with FRET pairs on the same side and either side of the direct binding region, which could potentially give

information on the binding mechanism, for instance how Tus approaches and binds to the binding sequence. But also information on the non-permissive flanking region, and whether this region plays a role in protein binding, be it a pre-formed atypical duplex structure as suggested in CD studies (see Figure 54).

Multiple donor sites on various strands however can complicate FRET analysis, first deconvolution of the donor and acceptor peak is required, followed by normalisation of the donor intensity for each of the same donor combinations, i.e. F^1 with T^1, T^2, T^3 or T^4, and then comparison with other donor combinations (F^2 and F^3 with T^1, T^2, T^3 and T^4). The reason for normalisation before comparison is the result of the multiple donor positions each having different local environments, which derive from the different surrounding neighbouring bases, and in some cases the terminal proximity of the fluorophore. These environmental discrepancies have an effect on the donor's absorption and emission orbitals, and hence FRET efficiencies and therefore must be applied for through appropriate normalisation.

Non-Permissive **Permissive**

T# 5′ GC1 GCC AGC 2CC GAA 3AA GTA TGT TGT AAC TAA AG4 GGA 3′

F# 3′ CG1 CGG 2CG AGG CTT ATT CAT ACA ACA TTG ATT 3CA CCT 5′

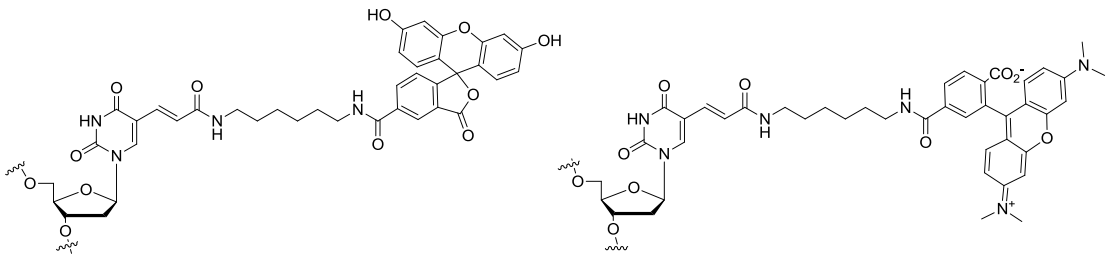


Figure 57 FAM and TAMRA Extended modified DNA sequence (top), binding region highlighted in yellow. FAM deoxyuridine modification (bottom left) and TAMRA deoxyuridine modification (bottom right).

The donor positions, as shown in Figure 57 (highlighted in green) were selected for FAM modification sites, note that F^1 is actually a T-T mis-match with its complementary strand. The reason for the mis-match base-pair was due to the interest in this region, potentially involving the non-permissive DNA end upon protein binding (see section 1.10.1). This perturbation however should not affect recognition greatly since this region is not responsible for direct protein-DNA binding, but any destabilising effects on the duplex must be analysed. One also has to note F^1 and T^1 close proximity to the DNA terminus (DNA tends to fray towards the extremes of the duplex, these bases are known to have significantly single-strand type behaviour) [10] .

When conducting studies on modified Tus-DNA using the sequences above, the effect of the modification upon protein binding has to be attributed, depending on the mechanism of protein binding to DNA, either a direct 'clamping' mechanism or a 'sliding' mechanism to the recognition sequence, or a combination of both, any sliding mechanism will be affected by the presence and position of modifications. This could potentially complicate FRET results with Tus for 1 : 1 complex studies, however these results may also prove fruitful and yield information of the mechanism of protein binding. Native PAGE can be applied to screen for modification effect on binding.

3.4 DNA modelled distances

To calculate the FRET efficiencies and verify the FRET distances obtained from experimental results, a reliable model was required from which the experimental distances can be compared and validated. The approach by previous groups is to use a 3.6 Å approximation for each base-pair separation of the FRET pairs. This model shall be adopted and compared to computer models obtained from Macromodel. Macromodel enabled template DNA and conformationally searched template DNA models to be analysed and compared. The template DNA from Macromodel is shown in Figure 58, the DNA structure contains the double helix structure, from which the modification distances can be measured directly. The distances are calculated using the points at which the modifications are directly attached to their uridine bases. With the modifications present, the linker length (the distance between the fluorophore and the

DNA) can be applied to the measured distance, in order to establish the area of modification occupation, and hence the errors associated with linker flexibility. For FAM, the linker distance is 16 Å (from uridine base modification point to the central fluorophore position), and for TAMRA the linker distance is 14 Å. In reality the modifications should rotate approximately perpendicular to the DNA, and the errors associated with the linker should be minimal, unless for example, the modification has a high preference for intercalation. Any deviancies from typical Förster behaviour should become visible when comparing all the FRET combinations. The energy minimised version of the Extended DNA (Figure 58), has structural deviancies in comparison to the template version. The structure is slightly bowed in the centre, the effects of which will generally reducing the longer FRET pair distances measured, and the extremes of the duplex are slightly frayed.

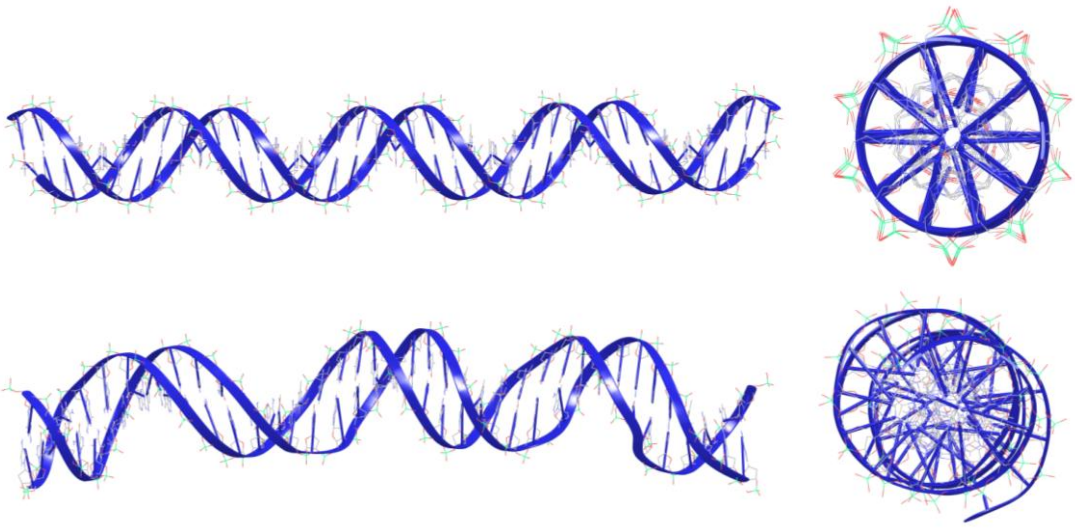


Figure 58 Side and top views of template (top row) and energy minimised (bottom row) Extended DNA sequence, phosphate backbone highlighted in blue ribbon, energy minimised model calculated using the Amber* conformational search function in Macromodel software from Schrödinger.

Table 3

FRET Pair	3.6 Å Approximation	Macromodel template distance (Å)	Conformational search distance (Å)
$F^1.T^1$	0	10	8
$F^1.T^2$	25	23.5	24
$F^1.T^3$	47	43	45
$F^1.T^4$	119	111	112
$F^2.T^1$	14.5	14.5	12
$F^2.T^2$	11	14	13
$F^2.T^3$	32.5	31	32
$F^2.T^4$	104.5	97	97
$F^3.T^1$	119	105.5	100
$F^3.T^2$	86.5	82	82
$F^3.T^3$	65	62	58
$F^3.T^4$	7	12.5	11.5

The three models applied; the 3.6 Å, Macromodel template and Macromodel energy minimised are not in complete agreement with each other, see table 3. The 3.6 Å model does not take into account the helical rotation of DNA when changing the distance separation of the moieties, this effect is more dramatic at smaller distance separations. The Macromodel models are slightly different as a result of the bow in the energy minimised model as compared to the template version. However, overall the models are largely in agreement and therefore should give rise to reliable models from which experimental values are compared. Figure 130 (see appendix section 8.25) shows the 3.6 Å model E_{FRET} values expected for the FRET combinations. The expected E_{FRET} values cover the full distance range for the FAM and TAMRA combinations. The combinations cover small distances, where FRET pairs are the same side of the binding region (both the permissive and non-permissive ends) and large

distances which bridge the binding region, i.e. FRET pairs either side of the binding region.

3.5 Spectroscopic analysis of FAM and TAMRA DNA

To apply FAM and TAMRA as FRET pairs to the Tus-*Ter* system, spectroscopic analysis of the donor and acceptor independently was required to be analysed, i.e. donor only and acceptor only in duplex format. UV-Vis and fluorescence spectroscopy enabled identification of the FRET prerequisites, quantum yield of the donor, extinction coefficient of the acceptor and the spectral overlap of the donor emission and acceptor absorption. Taking these values, the Förster distance was obtained using the Förster equation. Once established, the FRET experiment was conducted and analysed to yield FRET pair distances and finally DNA conformation.

3.5.1 FAM DNA

The FAM only duplex DNA strands: F^1, F^2 and F^3 show different absorption and emission properties, Figure 59. The majority of the difference is not due to discrepancies resulting from extinction coefficient calculation errors (see UV-Vis graph, Figure 64), instead the discrepancy is the result of modification positioning within the DNA. Previous groups have reported significant absorption and emission effects related to neighbouring DNA groups[82, 100]. Depending on the base the fluorophore is directly attached to, and to those that are adjacent, fluorescence emission can be significantly quenched. G is the strongest quencher; quenching decreases in the order: G > A > C > T. Comparing the FAM strands, F^1 is adjacent to a G and C (3' → 5'), F^2 is adjacent to G and C, and F^3 is adjacent to T and C. Hence, one would expect F^3 to be quenched the least, with F^1 and F^2 having about the same effective quenching, however, for F^1 this is not observed. F^1 exhibits an 8 nm red-shift in the absorption spectrum of λ_{MAX}, and emits about 20 % emission of F^2 and F^3. The red-shift and quenching effect is not purely related to positioning near a G. A second point to note is the fluorophores

close proximity to the terminal position of the DNA. DNA is known to fray towards the terminal bases (a result of reduced hydrogen bond strength between extreme bases arising from the absence of a neighbouring base-pair). F^1 fluorophore is located 3 bases from the non-permissive terminus. One would expect the 3 bases to experience a degree of fraying (combined with the decreased duplex stability of the fluorescein modification itself). It is therefore plausible to expect the fluorescein moiety to experience a single-stranded environment, potentially leading to intercalation and hence quenching of the fluorophore, resulting in a red-shift in absorption and decreased quantum yield.

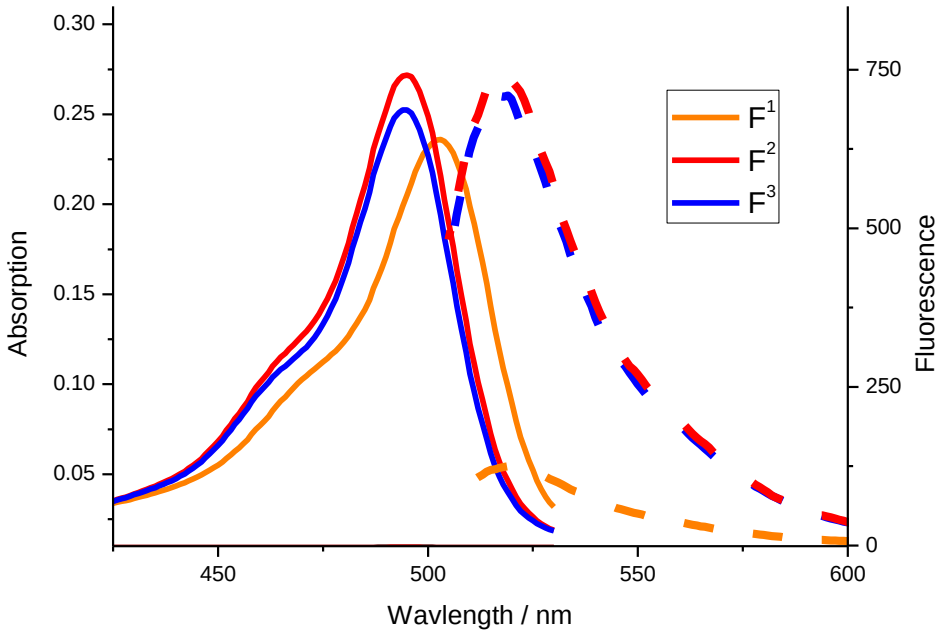


Figure 59 UV-Vis and Fluorescence spectrums of FAM modified Extended DNA duplexes at 2 μM in Tus buffer, using a 1 cm quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

Since steady-state FRET has a donor emission and acceptor absorption dependency, any significant change in either of these (resulting from the fluorophore's local environment) has to be taken into account. This complicates the FRET analysis considerably, meaning if FRET is calculated with respect to steady-state donor emission (or acceptor emission), then FRET can only be compared directly using the same donor (or acceptor) for each FRET pair. That is, valid steady-state FRET can only be analysed

comparing like-for-like, as in this case: $F^1.T^1$ with $F^1.T^2$, $F^1.T^3$ and $F^1.T^4$, and not with F^2 or F^3 combinations. F^2 and F^3 could possibly be compared, because differences are relatively small. Another factor to consider is the presence of any excitonic coupling of fluorophores, as discussed in the introduction chapter. These are potential complications for steady-state FRET, which may be overcome using time-resolved FRET, as discussed in section 1.13.

3.5.2 TAMRA DNA

UV-Vis analysis of the TAMRA modified DNA strands: T^1, T^2, T^3 and T^4, in double-stranded form display fluctuations in the intensity and wavelength of λ_{MAX} (Figure 60). Fluorescence analysis of the strands at the same concentration yields extensive fluctuations in emission intensity. Again the positioning of the modification within the DNA strand and the surrounding bases must be applied for. T^1 is adjacent to C and G (5' → 3'), T^2 is adjacent to two C's, T^3 is adjacent to two A's, and T^4 is adjacent to two G's respectively. From this basis alone T^4 was expected to be quenched the most, T^2 the least, and T^1 and T^3 to have similar quenching. Interestingly this is not what is observed, T^1 is quenched the most and T^3 the least. It is possible that for T^1, with the modification again attached 3 bases from the terminus experiencing significant fraying of the strand, and hence leading to intercalation, and significant quenching. The fluorescence results show how significant these environmental effects are on the TAMRA modification. These factors affecting the fluorophores need to be considered in detail for accurate interpretation of results as discussed above[27].

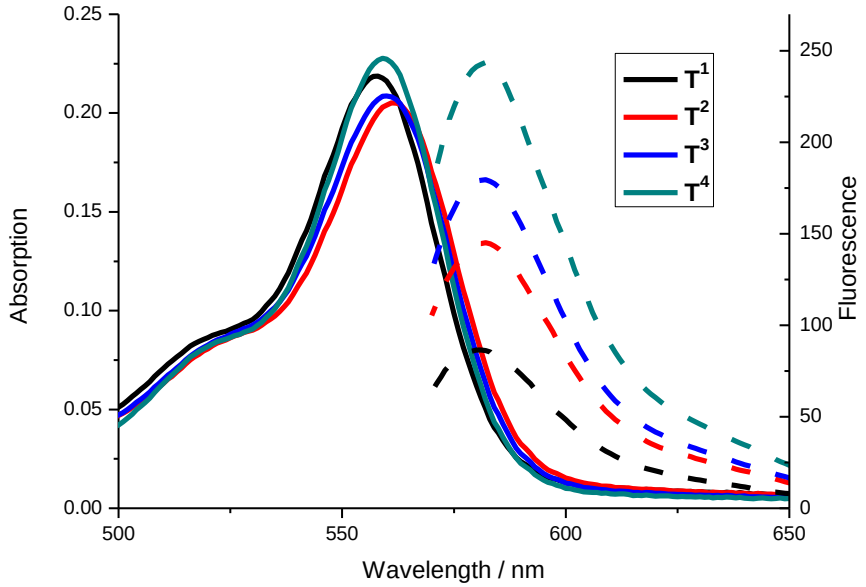


Figure 60 UV-Vis (solid lines) and fluorescence (dotted lines) spectrums of FAM modifications at 2 μM, in Tus buffer, using a 1 cm quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

3.6 FAM and TAMRA Extended DNA stability

The FAM and TAMRA modification's effect on the Extended DNA sequences were analysed to ensure DNA duplex character was maintained. Native duplex behaviour was important for subsequent protein-DNA analysis, any perturb DNA structure could disrupt native protein-DNA binding interactions. Gel electrophoresis and thermal melting studies are simple methods for DNA duplex analysis (see Tus complex analysis, for Gel Electrophoresis analysis)[90]. UV melting studies enabled direct observation of the DNA and modification absorption regions at 260 and 495 nm's respectively. The values obtained were compared to the native version of the Extended DNA sequence, which gave rise to a T_m of 75.3 °C in Tus buffer (see appendix section 8.4). The F^1 combination was analysed separately to the F^2 and F^3 combinations due to the previous F^1 combinations anomalous UV-Vis and fluorescence results.

3.6.1 F^1 Extended DNA stability

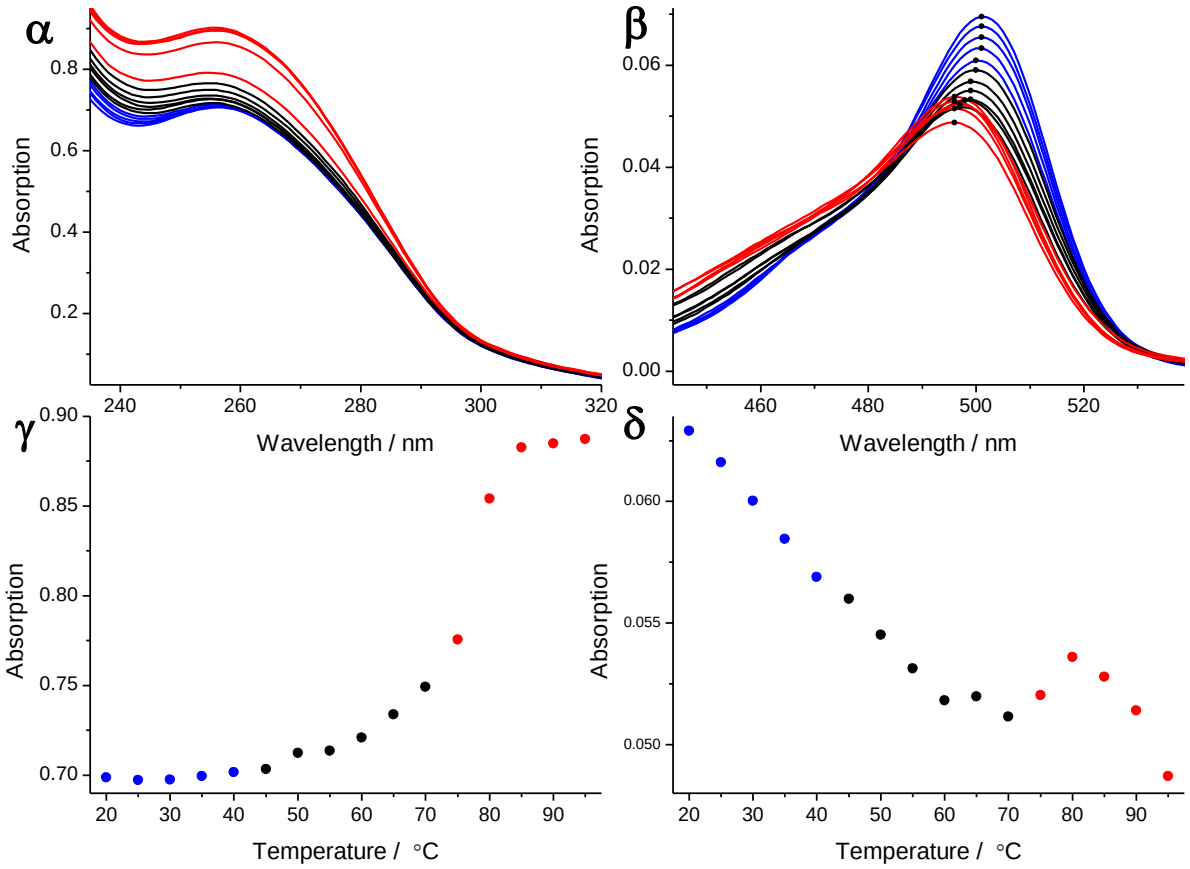


Figure 61 UV-Vis melting spectra of F^1 Extended duplex DNA; DNA absorption spectrum (α), FAM absorption spectrum (β) and melting cross-sections at 260 nm (γ) and 495 nm (δ), at 2 μM, melting at 5 °C intervals, scanning at 600 nm per minute, 1 cm quartz cell, blue lines 10-40 °C, black lines 45-70 °C and red lines 75-95 °C.

Figure 61 displays the UV-Vis melting analysis of the F^1 combination in duplex format. The DNA region at around 260 nm (α) displays an increase in absorption with increasing temperature, as expected for the denaturing of duplex DNA. Plotting the UV cross-section at 260 nm gave rise to an expected sigmoidal melting curve (β). The curve indicates the duplex is stable from 20 to around 45 °C, after which the melting transition takes place. The melting transition is sharp, which levels off after 80 °C, indication complete melting of the duplex. Differentiation of the melting curve gives rise to a T_m of 75.4 °C, indicating no destabilising effects of the modification compared

to the native form. The hyporchromicity of this UV absorption peak increases by 22 % from 20 to 95 °C. The FAM absorption region at λ 495 nm yields information on the fluorophore environment (β). Interestingly the FAM absorption decreases with increasing temperature, rather than increase in fluorescence. The decrease is possibly due to increased salt and single-strand quenching at elevated temperatures. Plotting the cross-section at 495 nm yields a melting transition from 60 to 80 °C (δ). Upon closer inspection of the FAM absorption (β), the peak undergoes a blue peak height shift of 6.5 nm from $\Delta\lambda_{MAX}$ 501 to 495.5 nm between 20-95 °C. Interestingly, UV change was observed λ 450 nm, which is consistent with the change associated when observing purely single-stranded FAM modified DNA, which has main peak absorption λ_{MAX} 450 nm (see appendix section 8.25).

3.6.2 F^2 and F^3 Extended DNA stability

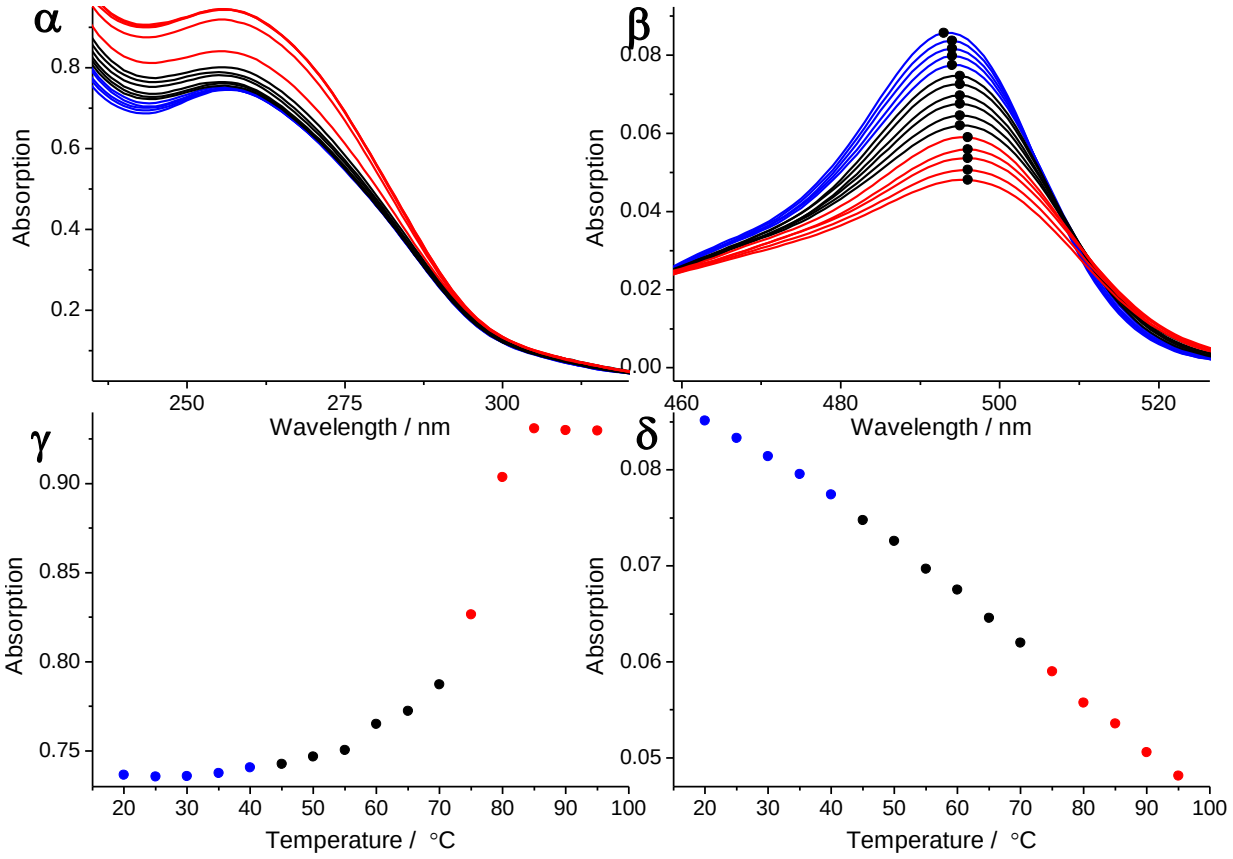


Figure 62 UV-Vis melting spectrums of F^3 Extended DNA with unmodified complement; DNA absorption (α), FAM absorption (β) and melting cross-sections at 260 nm (γ) and 495 nm (δ), at 2 μM, melting at 5 °C intervals, scanning at 600 nm per minute, 1 cm quartz cell, blue lines 10-40 °C, black lines 45-70 °C and red lines 75-95 °C.

Figure 62 displays the UV-Vis melting of the F^2 Extended duplex. The DNA region at around 260 nm (α) also shows absorption increases with increasing temperature. Plotting the cross-section at 260 nm (γ) the melting curve was obtained, differentiation yielded a T_m of 75.0 °C, again in agreement with the native version. The DNA region yields a 21.5 % change in hyporchromicity from 20 to 95 °C. The FAM absorption region (at λ 495 nm) yields information on the fluorophore environment (β), interestingly the FAM absorption decreases with temperature, possibly due to increased salt and single-strand quenching at elevated temperatures. Plotting the cross-section at 495 nm yields a subtle change in absorption above 70 °C (δ), very dissimilar to the F^1

FAM Extended combination. Upon closer inspection of the FAM absorption (β), the peak undergoes a red peak height shift of 2 nm from $\Delta\lambda_{MAX}$ 493.5 to 495.5 nm between 20-95 °C, however the same incremental change of the λ 495 nm cross-section is still observed. The F^3 Extended combination also yields the same changes as the F^2 Extended combination (see appendix section 8.25).

3.6.3 FAM Extended DNA fluorescence analysis

Fluorescence melting analysis of the FAM modified DNA combinations with their unmodified complementary strands are shown in Figure 63 (all at 2 μM, exciting at 495 nm and monitoring at 515 nm). The variation in fluorescence intensity is in agreement with the steady-state fluorescence from above, the F^1 combination is significantly lower than F^2 and F^3 FAM combinations. Interestingly F^2 and F^3 combinations decrease with increasing temperatures (in agreement with decreasing absorption in the FAM regions), both exhibit subtle changes upon DNA melting at around 75 °C. Conversely, the F^1 combination increases in fluorescence emission until the duplex has melted, after which the fluorescence decreases. These changes are consistent with the changes in the visible red and blue shifts for F^1 and F^2 or F^3 respectively as reported above in the absorption spectra.

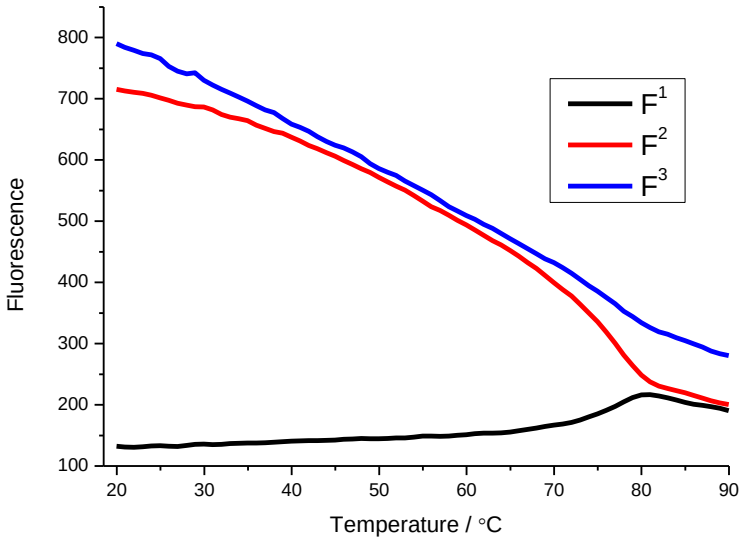


Figure 63 FAM modified DNA duplexes fluorescence melt, at 2 μM, excitation at 495 nm, emission at 515 nm in Tus buffer, 1 cm quartz cell, 500 PMT voltage.

3.7 Determining the prerequisite FRET values

To identify the E_{FRET} term from which experimental distances can be identified, the Förster distance must be determined first (see section 1.6.2). The Förster distance is dependent on the quantum yield of the donor, the extinction coefficient of the acceptor, the spectral overlap of donor emission and acceptor absorption and the relative orientation of donor and acceptor dipoles. These values are determined below, after which the Förster distance was identified and steady-state fluorescence analysis performed.

3.7.1 Quantum yield of FAM DNA

The quantum yield (the ratio of light absorbed to light emitted) of the donor (FAM) sites in duplex form was measured with respect to known standards (see section 1.6.2). The UV-Vis absorption was kept below 0.05 to prevent inner filter effects at 495 nm, to which fluorescence analysis was applied. Care was taken to ensure all

conditions were constant throughout. The absorption intensities and fluorescence integrals were applied to equation 2, from which the quantum yields were established (see appendix section 8.25), see table 4.

Table 4

Fluorophore	Quantum yield (*Q*)
Fluorescein free acid	0.94
F^1	0.11
F^2	0.71
F^3	0.77

As expected, the quantum yield of the F^1 combination is significantly less than the other two FAM modifications. As previously discussed this modifications position has a peak height absorption red-shift of 6 nm and significantly reduced fluorescence emission, which explains the reduced quantum yield. This reduction is most probably due to the terminal proximity of the donor's position, either fraying of the strands leading to single-strand quenching, multiplied by the fact that this base-pair is a mutant T-T mis-match. However, potentially the abnormal behaviour is a result of conformational deviances deriving from the non-permissive region, as discussed in section 3.1.

To identify the T-T mis-match effect, the complete matching complementary was synthesised and analysed (Figure 64). The fluorescence emission actually decreased further than compared to the mis-match version, suggesting the T-T mis-match has little effect on the conformational abnormality of the terminal FAM modification, and is rather a result of terminal DNA fraying or pre-formed conformational deviances as observed in the CD analysis (see section 3.2).

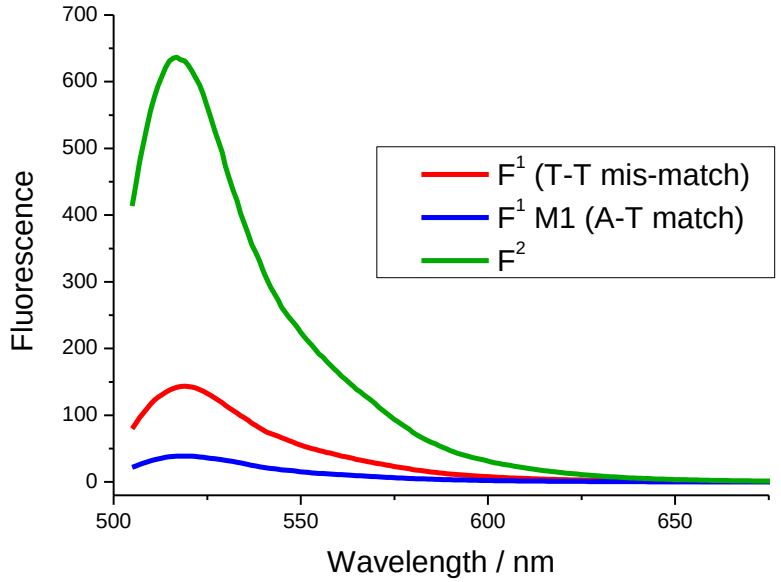


Figure 64 F^1 duplex match and mis-match fluorescence spectra, 2 μM, in Tus buffer, scanning at 600 nm per minute, 1 cm quartz cell.

3.7.2 Extinction coefficient of TAMRA DNA

The extinction coefficients for the four acceptor positions were established using the Beer-Lambert Law (see introduction 1.6.2). The values were identified using the UV-Vis data obtained in Figure 59.

3.7.3 Spectral overlap of donor and acceptor

The spectral overlap (J(λ)) of the donor emission and acceptor absorption was determined using equation 4, as shown in Table 3. All the values are in general agreement with each other since the spectral overlap does not deviate significantly, even for the four F^1 combinations. However, the R_0 values do vary significantly; for combinations with F^2 and F^3, the R_0 values are in good agreement with the literature values (R_0 = 54 Å)[4]. For F^1 combinations, the R_0 values are significantly lower due to the abnormalities deriving from the donor quantum yield quenching discussed in 3.7.1.

3.7.4 FAM and TAMRA κ^2 approximation

When applying the FRET theory, the relative dipole-dipole orientation of the donor and acceptor species must be accounted for through the κ^2 term. Since, FAM and TAMRA modifications are attached to DNA *via* relatively long and flexible hexyl linkers, they protrude up to 16 Å (calculated using Maestro software from Schrödinger), an averaged κ^2 value of 2 / 3 is representative of the random rotation of the dipoles (see section 1.6.2).

3.7.5 Förster distances for FAM and TAMRA combinations

The Förster distance (R_0) for each FRET pair combination was calculated using the Förster equation, using the values described above for donor quantum yields and acceptor extinction coefficients (Table 5). Table 3 displays the $J(\lambda)$ and R_0 values for the FAM and TAMRA combinations. One has to note the anomalous value for the F^1 combinations, which is over 6 times less than the F^2 and F^3 combinations (and literature values).

Table 5

DNA sequence	$J(\lambda)$	R_0 / Å
$F^1.T^1$	4.26426E+15	9
$F^1.T^2$	4.00349E+15	9.5
$F^1.T^3$	4.14998E+15	9
$F^1.T^4$	4.23323E+15	9
$F^2.T^1$	4.18223E+15	56.5
$F^2.T^2$	3.91716E+15	53
$F^2.T^3$	4.06377E+15	55
$F^2.T^4$	4.14183E+15	56
$F^3.T^1$	4.18728E+15	61
$F^3.T^2$	3.92434E+15	57.5
$F^3.T^3$	4.07026E+15	60
$F^3.T^4$	4.14867E+15	61

3.8 Steady-state fluorescence

Steady state fluorescence analysis was performed on the FAM and TAMRA FRET pairs to obtain FRET distances, and therefore identify the conformational structure of the DNA. The FRET values were identified using the E_{FRET} equation (see section 1.6.2). FRET efficiencies at the steady-state can be calculated from either the donor emission or acceptor emission (or respective lifetimes using time-resolved measurements). The simplest method to obtain these values is to monitor the fluorescence emission of donor. Monitoring the acceptor emission, the emission intensity can be the result of various types of quenching mechanisms such as dark states etc, the net effect of which is seen in the steady-state fluorescence[17, 27]. The acceptor emission is generally not used for establishing FRET distances in steady-state measurements, however, for time-resolved experiments the different quenching

mechanisms can be identified. Monitoring the donor emission directly alleviates these measurements dependencies.

The FRET experiment with the FAM and TAMRA combinations was conducted at the same concentration in order to obtain normalised results. That is, the DNA combinations peak heights should match in the UV-Vis spectrophotometer at 260 nm (DNA peak maximum), as should the donor and acceptor peaks, ensuring the same quantity of photons are being absorbed by the system at the excitation wavelength (in this case λ at 495 nm to get normalised results). However, as previously discussed, F^1 FAM combination has a λ_{MAX} red-shift, and markedly reduced quantum yield, thus not making these combinations comparable, and therefore was omitted from analysis.

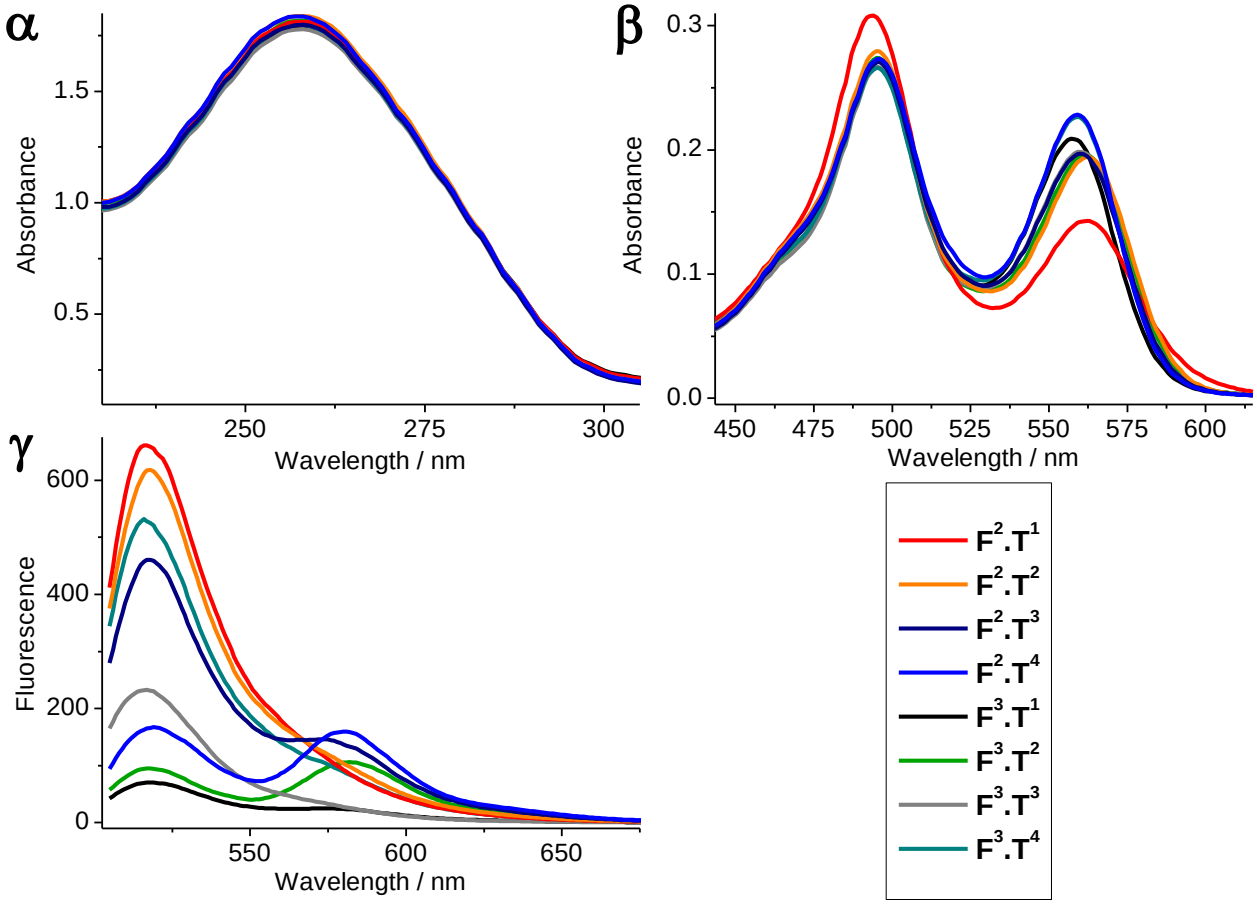


Figure 65 FAM and TAMRA combinations UV-Vis absorption spectrum, DNA region (α), FAM and TAMRA region (β) and fluorescence emission exciting at 495 nm (γ), at 2 μM in Tus buffer, 1 cm quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

Observing Figure 65 the DNA region at λ 260 nm (α), peak heights are within 5 % fluctuation error deriving from pipetting and ε calculation errors, confirming the validity of sample measurement (since the DNA is pure and all combinations have the same ε values). The F^1 combinations were omitted due to the anomalous behaviour of the FAM moiety (FAM's terminal positioning resulting in single-strand quenching and intercalation, hence not comparable to the F^2 and F^3 FAM and TAMRA FRET combinations). Viewing the donor-acceptor absorption region (400-600 nm), Figure 65 (β), fluctuation in peak heights resulting from either the environmental effects of the surrounding bases, excitonic coupling effects or direct electron transfer variations is observed. Applying excitation at the donor peak maximum at λ_{MAX} = 495 nm, the steady-state fluorescence emission spectra is obtained, Figure 65 (γ). The emission contains all the donor-acceptor combinations deactivation pathways.

3.9 Steady-state fluorescence analysis

To obtain donor-acceptor distances from the experimental fluorescence values described above, the E_{FRET} equation was employed to obtain E_{FRET} (fluorescence resonance energy transfer efficiency, see section 1.6.2) versus distance graph. To the fluorescence emission from Figure 65, due to overlapping donor and acceptor emission peaks, the Clegg method donor and acceptor peak deconvolution was applied (see section 1.6.2). After which the donor peaks were normalised with respect to the donor only value, to yield the E_{FRET} values. Each of the combinations exhibited the expected sigmoidal curvature from low to high donor-acceptor distance separations. However, the values do not tend towards 1 and 0 at low and high distance separations as expected, and instead converge towards 0.9 and 0.3 for F^2 and F^3 combinations, and 0.5 to 0 for F^1 combinations. The deviation from the expected E_{FRET} behaviour, i.e. not from 1 to 0 as expected from the calculated distances could potentially be explained by limitations in the Förster model at the steady-state by not approximating for all the different de-excitation pathway mechanisms. However, another factor to consider is that potentially the donor may only have a finite ability to conduct energy transfer to the acceptor

fluorophore. By normalising the data points, taking what is calculated from the various models to be the lowest and highest E_{FRET} points (no FRET and maximum FRET respectively), the following curve in Figure 66 was obtained. The data points are in good agreement with the Förster model, and display the expected FRET sigmoidal behaviour. These results show that FAM and TAMRA have been successfully applied to study duplex DNA; Förster's theoretical model for FRET and experimentally obtained E_{FRET} values are in good agreement. Theses results show the averaged κ^2 value is reliable and representative of experimental results - the fluorophores are attached via long and flexible linkers, they occupy large areas and therefore their dipoles are approximated. If this was not the case, for example, if the FRET pairs were held rigidly, their dipoles would have fixed positions, and then non-averaged κ^2 behaviour would dominate, which would be clearly seen in the plot of E_{FRET} versus distance.

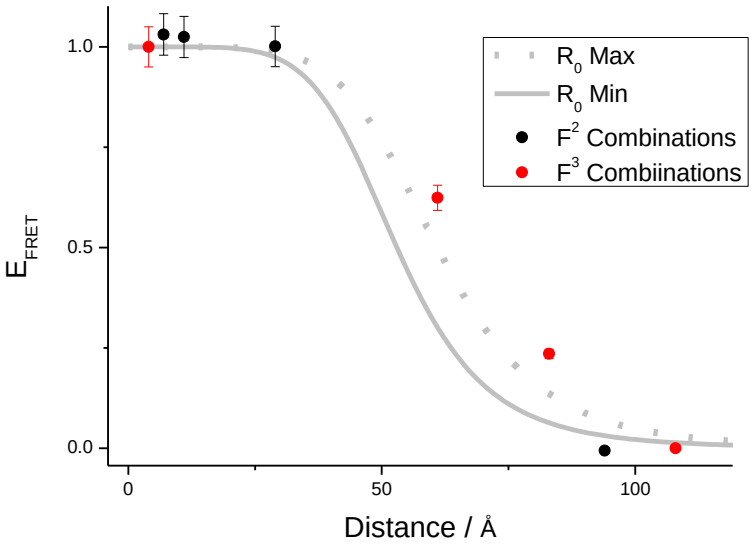


Figure 66 Normalised E_{FRET} versus Distance for the steady-state F^2 and F^3 and TAMRA combinations, using distances derived from 3.6 Å model, Förster model using 3/2 κ^2 values for 54 Å R_0 (grey line).

The distances calculated for the FRET pairs depend on the method chosen for calculating each of the combination's distance. The methods selected to calculate the distance separation of the chromophore separation include: Förster's theory using the

Förster distance (R_0) values obtained above, the 3.6 Å model and computational chemistry (Macromodel) using a DNA template version and a conformational energy minimised version (see section 3.4). The various distance methods applied are shown in Figure 67. The R_0 values are the theoretical distances calculated using the Förster equation (see section 1.6.2), the values of which are based on experimental spectra derived from the donor emission and acceptor absorption, the minimum and maximum R_0 are shown (grey solid and dashed lines respectively). If the theory is in agreement with the distances calculated using either the 3.6 Å base-pair separation approximation, Macromodel template and Macromodel Template or the Macromodel Conformational Search, then the results will be agreement. The 3.6 Å and both Macromodel models are in good agreement with each other, i.e. all values are similar. However, these models do subtly derivate from the Förster model towards the middle of the sigmoidal curve (grey lines). These values suggest a slightly larger R_0 value for each of the applied models. The conformationally searched model is in greatest agreement with the R_0 values. Following these conclusions, the control FAM and TAMRA FRET pair combinations were applied to study the Tus-DNA binding interactions (see chapter 5).

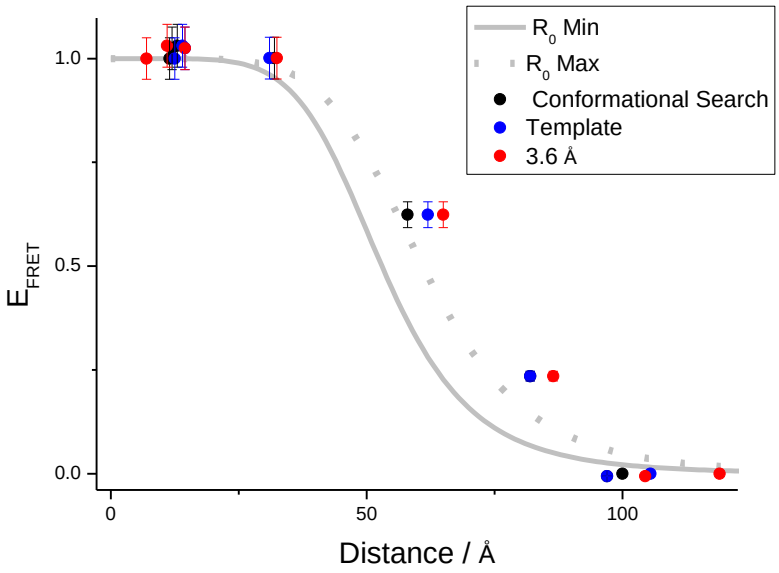


Figure 67 E_{FRET} versus Distance for the steady-state FAM and TAMRA combinations, using variable R_0 conformational search. Grey lines represent the R_0 values derived by the Forster model for each of the FAM and TAMRA combinations.

3.10 Summary

Established FRET pairs FAM and TAMRA were applied to study the Tus-*Ter* complex in parallel to the zinc and 2H porphyrin version. These FRET pairs also enabled comparison of the linker effect between rigid and flexible linkers. Firstly, the FRET pairs were tested and analysed in duplex form only to ensure reliable and predictable FRET behaviour for subsequent protein-DNA application (see chapter 4). To conduct this experiment, the DNA sequence and modification positioning had to be identified (section 3.2); the sequence had to contain the essential Tus binding region, therefore the DNA was elongated to ensure a sufficient quantity of FAM and TAMRA FRET distances was obtained. The two flanking regions are known as the permissive and non-permissive regions. During replication, the non-permissive region allows replication forks to enter from one side only, whilst the permissive region does not discriminate. Solid phase DNA synthesis of the non-permissive region repeatedly produced poor yields (2-5 %). This region also exhibited anomalous CD behaviour at around 255 nm, suggesting G-quadruplex behaviour. These results potentially indicated a pre-formed alignment property of this non-permissive region involved in protein-DNA recognition. The final strands were modified in seven positions, three from the donor and four from the acceptor, giving rise to twelve FRET pair combinations, and therefore twelve FRET distances. The modifications were selected to not interfere with protein binding, therefore the direct protein-DNA contact region was omitted. The FRET combinations selected cover the majority of the DNA sequence, and were expected to cover a full E_{FRET} range from 1 to 0. Three models were applied to calculate the final FRET distances: the 3.6 Å model, Macromodel templated DNA and Macromodel conformationally searched templated DNA (section 3.3).

Upon spectroscopic analysis of the four F^1 combinations, the absorption and fluorescence deviated substantially from the eight F^2 and F^3 combinations (section 3.4). This result was due to the terminal positioning of the FAM modification in the F^1 combinations. The terminal positioning resulted in a FAM single-strand environment due to terminal DNA fraying, and subsequent fluorophore quenching. Interestingly the FAM moiety in the F^1 duplex did not give rise to reduced thermodynamic stability in comparison to the unlabelled version. UV melting analysis of the FAM absorption

region at around 495 nm gave rise to a λ_{MAX} blue-shift upon heating, whereas the F^2 and F^3 combinations gave rise to a λ_{MAX} red-shift upon heating. The quenching of the F^1 combinations subsequently resulted in a reduced quantum yield of the donor moiety, and hence reduced spectral overlap between donor emission and acceptor absorption.

The FRET values as identified upon fluorescence analysis at 495 nm, exhibited spectral deviancies in donor emission corresponding to different donor-acceptor distances (section 3.5 to section 3.8). The different FAM combinations were not directly comparable due to the F^1 combination's quenching, a direct result of this moieties terminal positioning. The F^1 combinations yielded E_{FRET} values from 0.5 to 0, whilst F^2 and F^3 combinations vary from 0.9 to 0.3. Normalisation of the deconvoluted F^2 and the F^3 combination's fluorescence emission gave rise to the expected sigmoidal curvature for FAM and TAMRA FRET pairs. Comparison of the E_{FRET} results with the different distance models gave results largely in agreement with each other. The two Macromodel models are in agreement with the averaged κ^2 model, and therefore the κ^2 model is an accurate description of the fluorophores dipole location; FAM and TAMRA are tethered to DNA via long and flexible hexyl linkers, therefore their relevant dipole moments are randomly orientated, and hence an averaged value of κ^2 is an accurate representation of the FRET process. This result means that these FRET pairs can be used as a molecular ruler to detect large conformational changes, for example DNA wrapping around a protein. FAM and TAMRA FRET pairs were applied to study the Tus-*Ter* binding mechanism in chapter 4.

Tus-DNA molecular ruler analysis

4.1 Introduction

Zinc and 2H porphyrin, and FAM and TAMRA molecular rulers were applied to study the Tus-*Ter*, protein-DNA complex. The probes were incorporated into the DNA sequence to identify protein induced conformational change. Potentially, the flanking DNA regions of the main binding sequence wrap around the electropositive shell of the protein, to increase the complexes stability and affinity[68, 69, 70]. However, the non-permissive flanking region of the DNA sequence may adopt an unorthodox secondary structure, as indicated by results shown in Figure 54. The FAM and TAMRA FRET molecular ruler established in chapter 3, was applied to investigate potential DNA wrapping. These FRET pairs were appropriate since they are sensitive to conformational changes occurring over a large distance (R_0 = 54 Å). A zinc and 2H porphyrin CD molecular ruler was applied to study the non-permissive region of the binding domain. These porphyrin moieties are appropriate to analyse the non-permissive region since exciton coupling is strongly susceptible to small distance and angle changes (see section 1.6.3 and 2.8).

Firstly, the Tus protein is introduced and the method of Tus biosynthesis explained (section 4.2). After expression and purification, the binding activity of the protein was identified. The Tus-*TerB* binding constant is known from pervious literature results, however, how binding is effected using an elongated and labelled DNA sequence is not known. Native poly-acrylamide gel electrophoresis was employed to identify initial Tus-DNA complexation (section 4.3). The use of labelled DNA enabled the identification of the Tus binding mechanism (section 4.4); Tus binds either directly, *via* a 'clamping' mechanism, or *via* a 'sliding' mechanism. The binding constant was eventually identified by conducting a fluorescence titration between TAMRA labelled DNA and Tus (section 4.5).

With the binding information established, analytical gel filtration, small angle neutron scattering and circular dichroism techniques were employed to identify the solution phase structure of Tus and its complexes (section 4.6). These experiments were conducted to complement the subsequent molecular ruler experiments, and thus identify the flanking DNA regions. Thermal circular dichroism studies yielded information on Tus and its complex's stability, with and without DNA modifications,

and the effect of the histidine tag. Finally, molecular ruler analysis of the complex was performed (section 4.7); firstly, steady-state fluorescence analysis using FAM and TAMRA labelled DNA displayed protein induced changes; secondly, the porphyrin CD molecular ruler was applied to analyse the non-permissive region, (section 4.8), and thirdly, a new type of FRET molecular ruler is introduced and analysed; FAM and 2H porphyrin FRET between FAM labelled Tus and 2H porphyrin labelled DNA (section 4.9). The molecular ruler is not specific to DNA, and can be applied to other biomolecules too. FAM was attached to the amine terminus of Tus using a sortase catalyst[91, 92].

4.2 Tus

Tus is a 36 kDa protein which plays a crucial role in terminating replication of plasmid DNA from *Escherichia coli*. The protein forms an asymmetric protein-DNA complex to one of ten DNA sequence specific sites, known as termination (*Ter*) sites. Tus binding is very specific for each of these 16 base-pair sites. Tus allows replication forks to enter the complex region from the non-permissive direction but not leave, unlike the permissive region which does not have any hindrance on replication (see introduction, section 1.10). The Tus protein was selected since it forms an extremely strong interaction to *TerB* DNA, to which modifications can be attached and the molecular ruler experimentation employed[64, 65, 66]. The protein was expressed from the plasmid pHCS151 using the *E.coli* BL21(DE3) strain (Figure 68). The phcs151 plasmid from Dr. Cameron Neylon contains the *bla*, Tus, sortase-recognition and histidine tag coding genes[91, 92]. *Bla* codes for the ampicillin digestion enzyme β-lactamase. β-Lactamase hydrolyses the conformationally strained 4 membered β-lactam ring of ampicillin[16], which also ensures the successful expression of the Tus gene. The sortase recognition sequence consists of 7 amino acids which bind to the sortase enzyme enabling functionalisation of the carboxy terminus. The histidine tag is used to purify the protein using metal affinity chromatography. The tag consists of 6 histidines located on the amine terminus of the protein, each of which has a high affinity for nickel or cobalt. Any impurities should not contain the histidine tag, since impurities should

not contain either the *bla* or the Histidine tag coding genes and can be separated from the tagged protein due to the lower metal affinity of impurities.

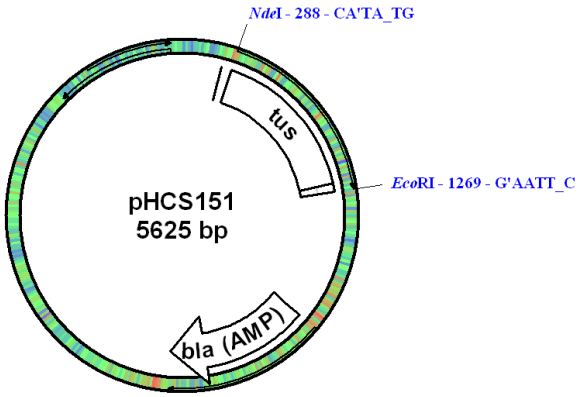


Figure 68 Schematic representation of pHCS151 Tus Plasmid.

The cell stock was transferred to a small volume of Luria broth growth medium containing ampicillin. Overnight a concentrated solution of cell stock was obtained by heating at 37 °C whilst stirring. Ampicillin was added to prevent expression of any plasmids not containing the *bla* gene, i.e. impurities. Once the visible scattering at 500 nm had reached approximately 0.4 absorption, transcription was initiated with the addition of isopropyl-thiogalactopyranoside (IPTG). After 4 hours the cell pellet was collected by centrifugation. The cell walls were broken by the lysozyme enzyme and freeze-thawing process, the preferred method to sonication, since sonication produces local temperature maximums > 1000 °C, potentially leading to DNA fragmentation. Tus is a strong DNA binding enzyme, hence sonication (and fragmentation) would lead to extensive Tus-DNA complexation. Purification of Tus was achieved using the histidine tags high affinity for nickel (II)[[14]. Proteins not containing a histidine tag will have a lower affinity for nickel and hence pass through the column to elute first. High concentrations of imidazole were passed though the column competing effectively with Tus for nickel, to elute pure Tus.

4.2.1 Tus analysis

To ensure protein purity, size exclusion chromatography was performed. Figure 69 shows Tus eluting at around 65 mL.

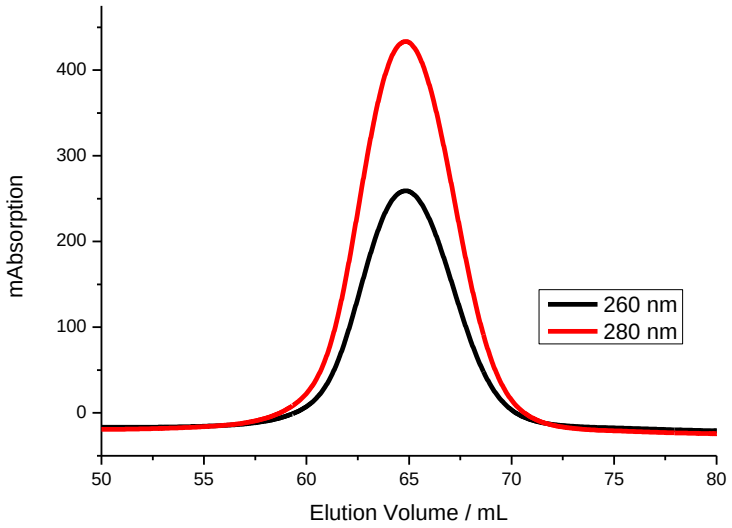


Figure 69 HPLC Tus chromatogram, 100 nmol injection, in Tus buffer, S75 size exclusion column monitoring UV absorption at 260 and 280 nm, 0.7 mL per minute flow rate.

Denaturing poly acrylamide gel electrophoresis (PAGE) was conducted on the isolated protein to ensure purity and confirm the approximate molecular mass, Figure 70. A 12 % gel with 5 % loading stack was generated[65, 66, 67]. A wide marker (6.5-200 KDa) was added to the gel for approximate molecular weight comparison.

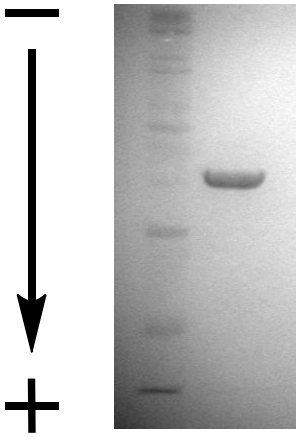


Figure 70 Denaturing PAGE of wide marker (left lane) and pure Tus (right lane), , 1 nanomole loading, stained using coomassie, see appendix section 8.1 for experimental procedure.

The molecular mass of the protein was confirmed by positive electrospray mass spectrometry. The protein was transferred into 25 mM TEAA buffer and concentrated to 30 μM before being sent for analysis. The calculated molecular mass of the Tus expressed with histidine tag was 37673 Da. The observed peak was 37542 Da, which gives a difference potentially corresponding to a methionine elimination from the amino terminus.

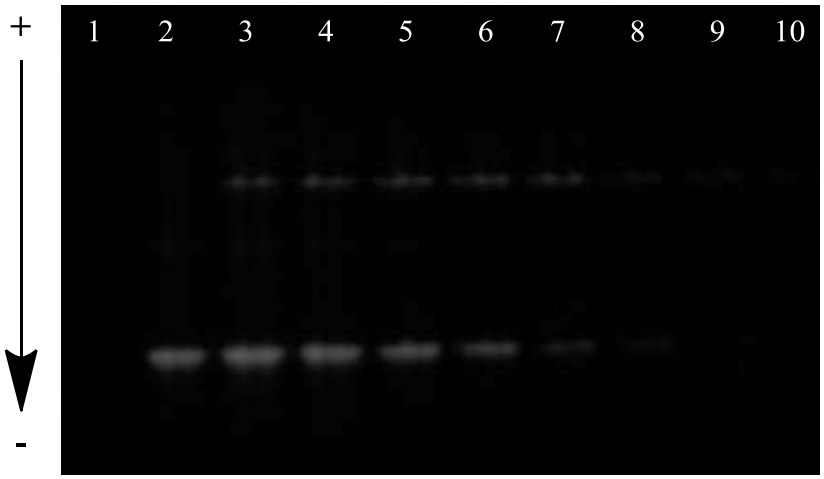
4.2.1 Tus activity – native PAGE

Native PAGE was conducted to analyse the Tus only, Extended DNA only, and the Tus-DNA complex in their folded, duplex and complexed state (native PAGE protocol is discussed in the appendix section 8.1)[93]. Tus has an electropositive Van der Waals radius (in Tus buffer,)[65, 66, 67], and is attracted to the negative electrode at the top of the gel, as shown in Figure 71. Due to the negative character of the phosphate backbone of DNA, the DNA is attracted to the positive electrode at the bottom of the gel. A combination of the two, i.e. the complex, gave rise to a mixed combination of charges, and hence is found in-between the Tus only and DNA only bands. To verify the activity of Tus, native PAGE was performed with Tus and the *Kamada* sequence

(21-mer, binding only sequence), as studied previously by Neylon *et al*[65, 66, 67]. Analysis revealed efficient complexation in agreement with previous results.

To ensure Tus-Extended DNA complexation, native PAGE analysis of Tus with varying Extended DNA concentration was conducted. Figure 71 shows an approximate 1 to 1 complexation of Tus with varying Extended DNA concentration. The ethidium bromide (EB) stain binds to DNA in the duplex form, *via* an intercalation mechanism[3]. The stain fluoresces under UV light to enable visualisation of the DNA present, bound and unbound to Tus. Tus (lane 1) does not fluoresce, and is visualised using the coomassie stain (organic blue dye which binds only to proteins). Observing the complex region (upper band), the bands remain constant for lanes 3 to 7, suggesting saturation of Tus when between 0.5 and 1 equivalents of Extended DNA. This approximate 1 to 1 complex is in agreement with the literature binding constant value of 0.5 nM[2].

Native PAGE has been used previously by other groups to yield binding constants using the fluorescence emission of ethidium bromide. However, due to the relatively harsh conditions of native PAGE (high voltage and TBE solvent conditions), caution must be applied to make absolute conclusions and comparisons for the study of protein-DNA interactions. Other techniques such as circular dichroism and fluorescence were used to identify the binding constants accurately in solution (see section 4.5 and 4.6).



Lane	1	2	3	4	5	6	7	8	9	10
Molar ratio Tus : Extended DNA	1 : 0	0 : 2	1 : 3	1 : 2	1 : 1.5	1 : 1	1 : 0.5	1 : 0.25	1 : 0.175	1 : 0.0625

Figure 71 Native PAGE analysis of Tus with varying Extended DNA concentration, ethidium bromide stain (DNA only observed). Controls: Tus only (lane 1 not observed, 0.5 nanomole loading) and DNA only (lane 2, 0.5 nanomole loading), see appendix section 8.1 for experimental protocol.

4.3 Tus binding mechanism analysis

Native PAGE analysis of fluorescein (FAM) and tetramethylrhodamine (TAMRA) modified Extended DNA was conducted to establish whether modification positioning effects binding. Tus can potentially bind to the DNA sequence *via* a direct 'clamping' or 'sliding' mechanism, or a combination of both (see introduction). FAM and TAMRA linkers protrude up to approximately 16 and 14 Å from the DNA respectively (calculated using Maestro software from Schrödinger). If DNA combinations are selected containing modifications on the non-permissive side (left of the binding sequence, as shown in Figure 72, and on the permissive (right of the binding sequence, as shown in Figure 74) and on either side of the binding region; potentially the extent of binding and the mechanism of binding can be identified. If Tus complexation is predominately from one side of the DNA region, then 'blocking' one region should significantly reduce complexation and hence reducing the middle band

fluorescence in native-PAGE experimentation. If the mechanism takes no preference to binding direction, i.e. a direct clamping mechanism, then blocking either side of the binding region should have little effect on complexation, and hence have no effect on the complex band fluorescence. This experiment assumes the linkers and fluorophores are sufficient to prevent binding from the modifications direction. This experiment assumes that the protein cannot pass from the modified direction, the linkers are long but relatively flexible.

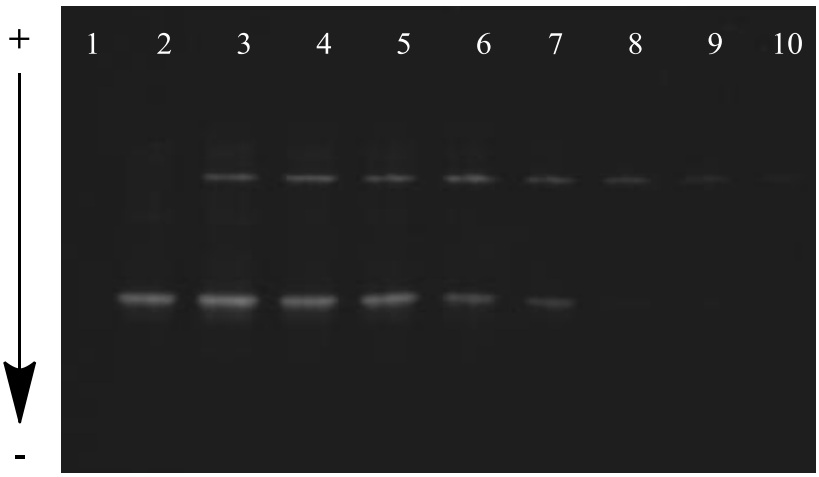
Neylon *et al.* obtained Biacore Tus binding information using *TerB* DNA tethered to a surface *via* a Biotin linker[94]. The linker was attached to the *TerB* DNA sequence at the non-permissive side, their result show tight binding is maintained whilst access to this region is blocked. Therefore, native PAGE analysis was expected to show no non-permissive region dependency.

4.3.1 Non-permissive side blocked

Figure 73 shows blocking the non-permissive region of the Extended DNA has little effect on binding, in comparison to the unmodified version above. Maximum complex formation occurs between molar ratios 1 and 0.5 (lanes 6 and 7), suggesting the non-permissive region has little effect in protein binding. Hence, either the protein binds directly to the binding region, or *via* the permissive region.

T^2 Ext 5′ GCT GCC AGC 2CC GAA TAA GTA TGT TGT AAC TAA AGT GGA 3′
F^2 Ext 3′ CGA CGG 2CG AGG CTT ATT CAT ACA ACA TTG ATT TCA CCT 5′

Figure 72 Non-permissive side blocked DNA sequence, FAM modification shown in green, TAMRA modification shown in red, direct binding region shown in yellow, Kamada DNA sequence shown underlined.



Lane	1	2	3	4	5	6	7	8	9	10
Molar ratio Tus : $F^2.T^2$	1 : 0	0 : 2	1 : 3	1 : 2	1 : 1.5	1 : 1	1 : 0.5	1 : 0.25	1 : 0.175	1 : 0.0625

Figure 73 Native PAGE analysis of Tus with varying FAM and TAMRA Extended DNA concentration (permissive region blocked), ethidium bromide stain (Only DNA observed). Controls: Tus only (lane 1 not observed, 0.5 nanomole loading) and DNA only (lane 2, 0.5 nanomole loading), see appendix section 8.1 for experimental protocol.

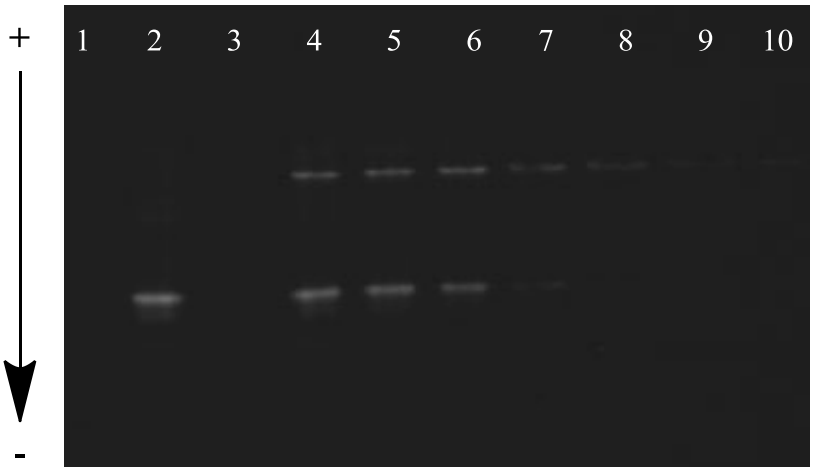
4.3.2 Permissive side blocked

Blocking the permissive region of the Extended DNA using FAM and TAMRA DNA (see Figure 75), had little effect on binding compared to the native and blocked non-permissive versions. Again, maximum complexation occurs between molar ratio 1 and 0.5 (lanes 6 and 7). To conclude, there is no substantial reduction in binding when blocking either the permissive or non-permissive regions, suggesting Tus adopts a direct 'clamping' mechanism rather than a 'sliding' one. To verify these results, solution phase titrations were performed using conditions as close to the native system as possible, i.e. matching salt, temperature, pH etc..

T^4 Ext 5′ GCT GCC AGC TCC GAA TAA GTA TGT TGT AAC TAA AG4 GGA 3′

F^3 Ext 3′ CGA CGG TCG AGG CTT ATT CAT ACA ACA TTG ATT 3CA CCT 5′

Figure 74 Permissive side blocked DNA sequence, FAM modification shown in green, TAMRA modification shown in red, direct binding region shown in yellow, Kamada DNA sequence shown underlined.



Lane	1	2	3	4	5	6	7	8	9	10
Molar ratio Tus : $F^3.T^4$	1 : 0	0 : 2	N / A	1 : 2	1 : 1.5	1 : 1	1 : 0.5	1 : 0.25	1 : 0.175	1 : 0.0625

Figure 75 Native PAGE analysis of Tus with varying FAM and TAMRA Extended DNA concentration (non-permissive region blocked), ethidium bromide stain (Only DNA observed). Controls: Tus only (lane 1 not observed, 0.5 nanomole loading) and DNA only (lane 2, 0.5 nanomole loading), see appendix section 8.1 for experimental protocol.

4.4 Fluorescence titration – Tus with TAMRA DNA

Fluorescence has been used extensively to identify binding constants between host and guest type interactions[5, 95]. Using TAMRA modified Extended DNA (Figure 76), the modification was monitored by fluorescence emission as a function of Tus concentration, after which, a binding model was applied. In order to abstract binding information, the relative concentrations between host and guest was required to be within the approximate binding constant range. Using significantly higher concentrations could potentially result in saturation of the equilibrium between host and guest, and other potential unwanted effects, such as aggregation or intermolecular interactions. However, at the other extreme, for very tight binding and hence very low concentrations ranges, experimental limitation may arise from instrumental sensitivity (especially for CD and UV-Vis application, relatively high concentrations are required for sufficient signal-to-noise).

```
T4   5′ GCT GCC AGC TCC GAA TAA GTA TGT TGT AAC TAA AG4 GGA 3′
     3′ CGA CGG TCG AGG CTT ATT CAT ACA ACA TTG ATT TCA CCT 5′
```

Figure 76 TAMRA Extended DNA sequence applied for Tus-DNA fluorescence titration analysis, TAMRA modification shown in red, protein-DNA contact region shown in yellow, Kamada sequence shown underlined.

The literature binding constant (K_D) value for *TerB* binding to Tus is 0.5 nM ± 0.2 (see above)[94]. When using FAM and TAMRA fluorophores, fluorescence and fluorescence polarisation instruments are typically sensitive to measure changes approaching this low concentration range (high extinction coefficient and quantum yield). Conditions were optimised to obtain complexation information, but also attuned to the fluorophore in question. Tus was titrated into TAMRA modified Extended DNA (T^4), the extent of complexation was monitored by observing the change in fluorescence at 580 nm, as shown in Figure 77 (α). Comparing the increased fluorescence from 0 to 1.8 equivalents of Tus added, yielded a substantial increase in fluorescence emission by

approximately 30 % (normalised by integrating the 0 and 1.8 equivalents curves and dividing appropriately). The fluorescence increase in TAMRA emission is related to changes in the fluorophores local environment. After Tus addition and subsequent binding, the fluorophore potentially no longer has a full undisrupted random orientation, i.e. before Tus complexation, TAMRA is attached to the DNA *via* a relatively long and flexible hexyl linker, which can rotate relatively freely around a cone shape, perpendicular to the DNA duplex (see introduction, section 1.1). After Tus binding, this random orientation experienced by TAMRA is perturbed, potentially, the cone shape volume of fluorophore occupation is reduced, and hence the fluorophore is held in a 'fixed' position, thus, giving rise to an increased fluorescence.

Figure 77 (β) displays the normalised fluorescence emission of TAMRA versus the equivalents of Tus added. The data was normalised by taking the 0 and 1.8 equivalents of Tus added, as 0 and 100 % of Tus bound respectively. Fitting the curve to a 1 to 1 binding model gave an approximate K_D of 74 nM ± 20. The tight binding constant obtained is in approximate agreement with the literature value. The result suggests the Extended DNA sequence has an effect on binding affinity, in comparison to the *TerB* DNA sequence, as used by Neylon *et al*[94]. This change, however, could be attributed to the modification itself, the *TerB* DNA sequence used by Neylon *et al.*, contained a Biotin modification at the non-permissive terminus, whilst the extended DNA sequence used above contained a TAMRA modification located within close proximity to the direct binding region.

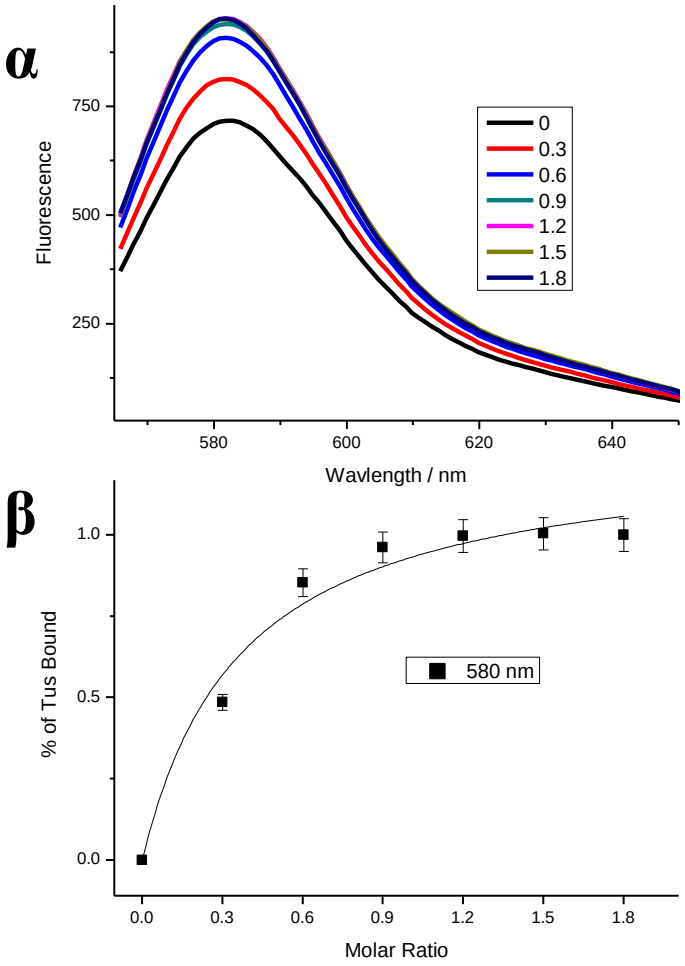


Figure 77 Tus titration with T^4 in duplex form, fluorescence emission spectra (α) and normalised fluorescence change at 580 nm (β). The curve shown is a model fit for a single binding process with a K_D of 74 nM ± 20, DNA concentration at 1 μM in Tus buffer, exciting at 560 nm, scanning at 600 nm per minute, 1 cm quartz cell, 500 PMT voltage.

.

4.5 Determining the solution phase structure of Tus and its complexes

The solution phase structure of Tus and its Extended DNA complex were analysed to complement the subsequent FRET and exciton coupling molecular ruler experiments. Analytical gel filtration was employed to identify the hydrodynamic radius of Tus and its Extended DNA complex (see section 4.6.1). Small angle neutron scattering was applied to identify a high resolution solution phase shape and structure of

Tus (see section 4.6.2). Finally, CD analysis of the protein and DNA regions was conducted to identify the key protein and DNA regions involved in binding (see section 4.6.3).

4.5.1 Analytical gel filtration

Analytical gel filtration (AGF) analysis was applied to determine the hydrodynamic radius of Tus and its Extended DNA complex. The hydrodynamic radius (R_H) is the sum of the effective radius of the molecule in question, and the solvent shell associated to this molecule. The R_H can be identified using light scattering techniques (for example dynamic light scattering), and applying the Stokes-Einstein equation[96]. However, the R_H can be identified with respect to known standards. By passing the standards through a size exclusion column, then comparing the elution volumes of the standards with the unknown sample; taking into account the void volumes (known as the column distribution coefficient, k_{ev}), the R_H can be identified. Size exclusion chromatography is a non-contact purification technique, small pores in the silica or polystyrene column enable small molecules to enter and leave. Larger molecules, conversely, cannot fit inside these pores, and pass through the column in a shorter time, to elute first. The R_H of native Tus (without the histidine tag) has already been established by Hill and co-workers[7]. Hill established a 23.2 ± 1.5 Å radius for Tus. To identify hydrodynamic radius of the Tus-Extended DNA complex, AGF of the protein in the absence of the DNA was performed first, followed by complex analysis. This method was selected to prevent complications arising from overlapping UV absorption peaks.

The R_H of Tus was determined with respect to the known standards: ferritin, ovalbumin, aldolase and ribonuclease, all in Tus buffer[96, 97, 98, 99, 100]. The largest protein, ferritin (400 kDa) elutes first, and the smallest, ribonuclease (1.4 kDa) last. The standards elution volumes were plotted versus their known molecular weight Figure 78). The elution volume of Tus (61.4 mL) was plotted with respect to the known standards, to yield a molecular weight of 37 kDa. This approximate molecular mass is in good agreement with the literature, and mass spectra value obtained experimentally. Finally,

the R_H of Tus was determined by plotting the K_{av} (column distribution coefficient) versus R_H of the known standards, then comparing the Tus K_{av} value to the standards. The result yielded a R_H value of 34 ± 2.5 Å. This value is in approximate agreement with the result established by Hill and co-workers. The discrepancy, however, could be attributed to experimental differences related to the standards and AGF column applied, or potentially, this could also be the effect of the histidine tag and sortase recognition sequence, which may increase the R_H of Tus in solution.

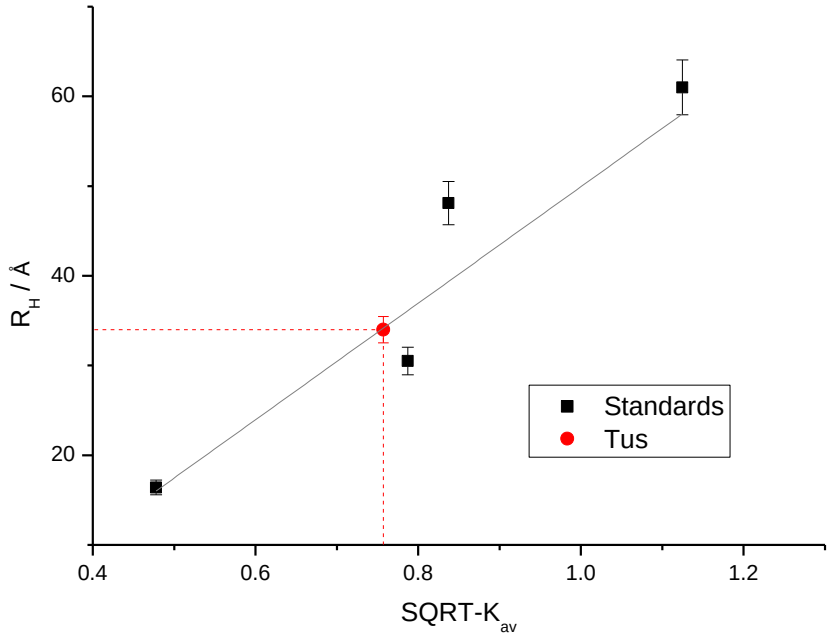


Figure 78 Hydrodynamic radius versus square root - K_{av} for Tus (red) and known standards (black) in Tus buffer, 9 µM injections, flow rate 0.7 mL per minute, using S75 size exclusion column monitoring UV absorption at 280 nm.

The uncomplexed structure of Tus was expected to be smaller than the complexed version, since Tus is a DNA binding protein, for it to locate its recognition sequence efficiently, Tus is expected to be flexible. Once the complex is formed, the structure is expected to be more stable, rigid and larger than the unbound structure, to confirm this hypothesis, AGF of the complex was performed. Analysing the elution volumes of the standards, an elution volume of approximately 58 mL was expected for the complex (Note, AGF of the complex was also an opportunity to optimise

purification of the complex as required for SAXS synchrotron studies, see 4.6.2). As a control, the Extended DNA in duplex form was passed through the column to elute a very broad peak at approximately 55 mL (Figure 79). This elution volume corresponded to a molecular weight of approximately 100 kDa. The molecular weight of the duplex is approximately 24 kDa, therefore, an the DNA only was expected to elute after the Tus only elution volume (> 61.4 mL). Either multiple duplexes are interacting, for example self-assembly or aggregation, or the standards being proteins, are not a good comparison to determine the molecular weight and R_H information of DNA. For instance, the cylindrical shape of the duplex DNA, compared to the typically globular shape of proteins, could explain the deviation in elution volumes with anomalous molecular weight.

The Tus-Extended DNA complex species should have a larger R_H than the individual Tus and Extended DNA components, hence a lower elution volume than either was expected. The molecular weight of the complex is approximately 60 kDa, and should therefore elute at around 58 mL. Equimolar quantities of Tus and Extended DNA were mixed and equilibrated for 30 minutes. The solution was passed through the S75 size exclusion column in Tus buffer to yield the chromatogram in Figure 79. Multiple peaks are observed arising from the complex and DNA only entities. The Tus only peak has disappeared, indicating Tus is in a complex state. The expected peak at 58 mL is observed for the complex. However, at least two other peaks at around 48 and 55 mL are present, potentially related to the anomalous DNA behaviour stated above. Since baseline separation of the peaks was not obtained, identification of the peaks responsible for each component was challenging, i.e. PAGE analysis would also be complicated.

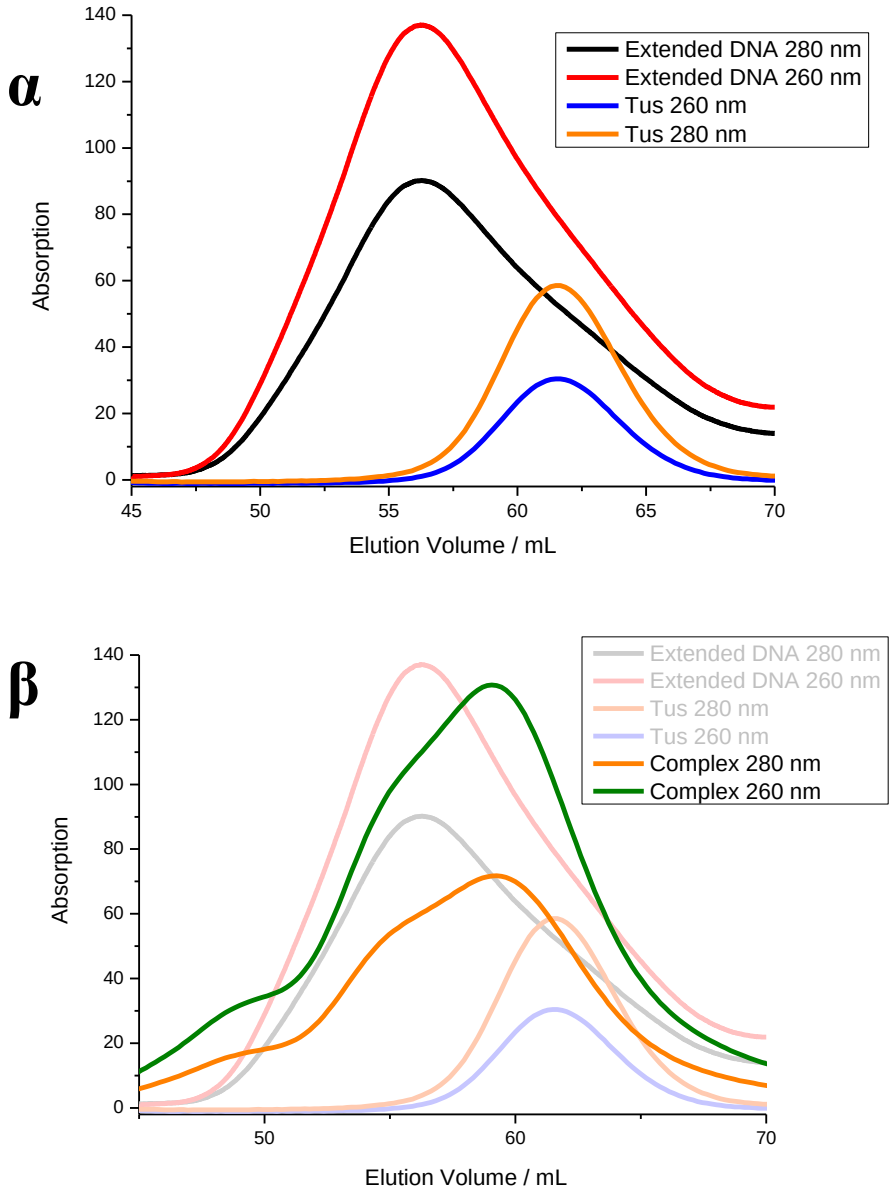


Figure 79 AGF chromatogram of controls DNA 9 µM and Tus only 18 µM (α), and the Tus-DNA complex 9 µM (β), flow rate 0.7 mL per minute, using S75 size exclusion column monitoring UV absorption at 280 nm and 260 nm.

4.5.2 SAXS and SANS analysis of Tus

The uncomplexed Tus protein was analysed by small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) in an attempt to identify its solution phase structure. Small angle scattering techniques (SAS) can determine the shape and size of biomolecules within the 1-25 nm range (see introduction, section 1.6.4). Firstly, the small angle scattering of the Tus-*Ter* complex was calculated using the X-ray crystal structure to obtain scattering intensity versus distance (Q), and distance distribution graphs of the complex, as shown in the appendix (see section 8.28). The crystal structure of the complex yielded a distance distribution maximum around 30 Å, which gave rise to a radius of gyration (R_G) of 21.4 Å. Potentially, the experimental Tus uncomplexed SAXS analysis was expected to yield a distance distribution similar to the calculated version, since the DNA of the crystal structure does not extended beyond the protein exterior.

Firstly, the scattering intensity of the sample had to be identified. Tus was scanned at 100 μM using the Diamond Ltd light source, the instrument yielded intense scattering. However, the distance distribution gave an approximate maximum around 500 Å, over an order of magnitude higher than the calculated version. This result suggests extensive aggregation of the protein, deriving from the high sample concentration. The experiment was repeated scanning Tus at 30 μM to yield a R_G of 21.0 ± 0.4 Å, which is in good agreement with the calculated graph. Figure 80 displays the Tus and normalised crystal structure (normalised crystal structure to Tus intensity) SAXS spectra. The experimental spectra exhibited high errors below 0.05 nm^{-1}. This result indicates instrumental limitations in the ability to resolve the Tus structure at very low distances.

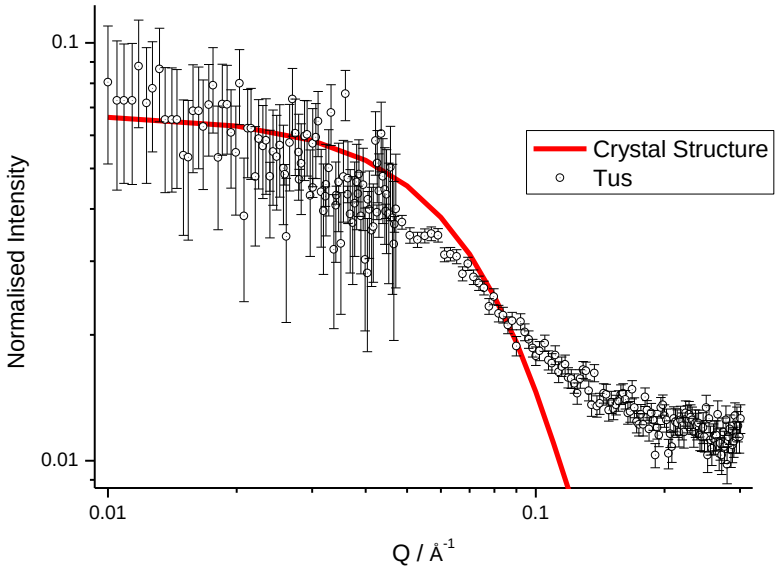


Figure 80 SANS spectra of Tus (black circles) and Tus-*TerB* X-ray crystal structure (red line), at 100 μM in Tus buffer.

4.5.3 CD spectroscopic analysis of the Tus-DNA complex

Circular dichroism (CD) is a spectroscopic technique able to give insight into binding and structural information between proteins and DNA[12-14]. CD is closely related to UV-Vis spectroscopy, a CD instrument measures the absorption of left and right circularly polarised light, the difference between the two absorptions yields the final CD spectra (see introduction 1.6.3). Beamline CD instruments utilise powerful synchrotron light sources to expose greater sensitivity (signal-to-noise) at lower wavelengths unavailable in bench top instruments. Another big advantage for using synchrotron CD instruments is related to the background saturation effect. For accurate protein or DNA analysis, matching the solvent conditions found in the native system is an essential prerequisite. Bench top instruments are less powerful than beamline instruments, therefore require greater path length cells, or higher sample concentrations to obtain the same CD resolution. For proteins and DNA, material is always at a minimum. With higher concentrations, there is also an increased possibility of unwanted aggregation or light scattering interactions. The buffer of the system may also cause problems, typically salt scattering (< 230 nm for chloride anions) overlaps with protein and DNA CD. Using a longer path length cell can quickly lead to saturation of the CD detector (PMT voltage > 650 volts), from which no valid CD results can be made. Using a beamline instrument allows the user to use lower sample concentrations, in combination with shorter path length cells, and therefore enabling a greater window for experimentation, in particular for titrations. The CD light source at Diamond Ltd was used to study Tus and its complexes, including: Tus CD Beamline stability, Tus-DNA CD complexes, melting studies of Tus-DNA CD complexes and CD titrations of Tus-DNA complexes. These experiments were conducted to identify the protein and DNA regions which change upon complexation.

4.5.3.1 Beamline Tus stability

Tus is a natural protein consisting of a liner chain of amino acids, which fold up to form the active protein structure, ready for DNA binding (see section 1.10). Tus is

held together by weak interactions from hydrogen bonds and hydrophobic interactions. When applying a light source from synchrotron radiation to study biological interactions, caution was paid to the sensitive nature of the sample, and take action to prevent sample degradation, for example oxidation and bond cleavage. To help preserve the natural integrity of the sample, the addition of a reducing agent (for example DTT), and filters to reduce the intensity of the light source, was applied. Firstly, Tus was repeatedly scanned 80 times, from 300-200 nm to ensure beamline stability (Figure 81). Plotting the cross-section at 220 nm revealed extensive degradation of Tus. Tus degraded quickly from its native state, to a damaged and degraded state within 10 scans. This result displays the relative vulnerability of Tus and the power of the synchrotron light. Exposure also resulted in visible precipitation of Tus. Reducing the beamline slit width from 0.5 mm to 0.28 mm, reduced the intensity of the light. Repeating the experiment with the new slit width, yielded stable analysis of Tus even after 80 scans (see appendix section 8.29), after which, reliable and reproducible results were obtained.

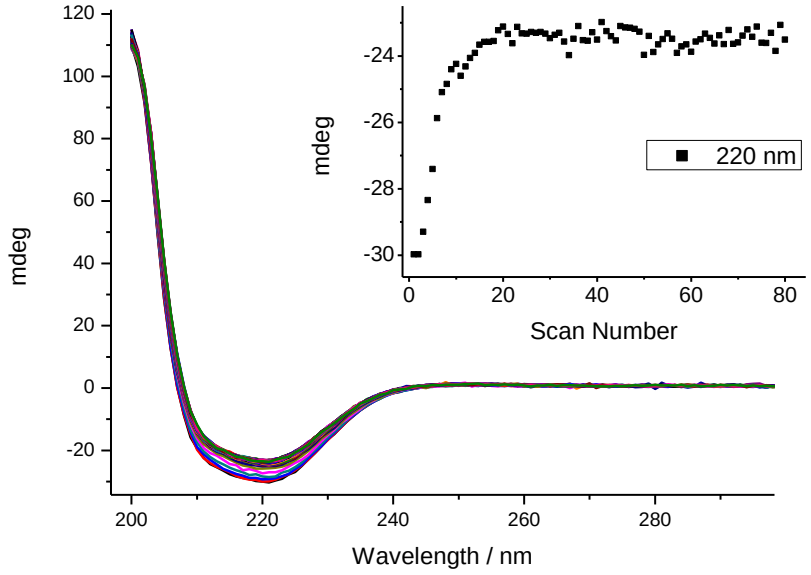


Figure 81 Tus CD spectroscopy beamline stability test, 4 μM Tus and 220 nm cross-section (inset) in Tus buffer, 0.2 cm cell, 0.5 mm slit width, scanning at 2 seconds per nm.

4.5.3.2 High resolution Tus analysis

Figure 82 displays the high resolution beamline CD analysis of Tus. The sample was scanned 40 times at 0.5 nm per second. To estimate the secondary structure of Tus from the CD curve, the data was fitted using the CD Spectra Deconvolution software. Fitting between 200 and 285 nm, the estimated secondary structure region gave rise to: 33.6 % α-helix, 16.9 & β-turn and 32 % random coil. These values are largely in agreement with those calculated previously by Hill and co-workers[64]. With the CD structure of Tus identified and confirmed, Tus's complex was analysed by CD.

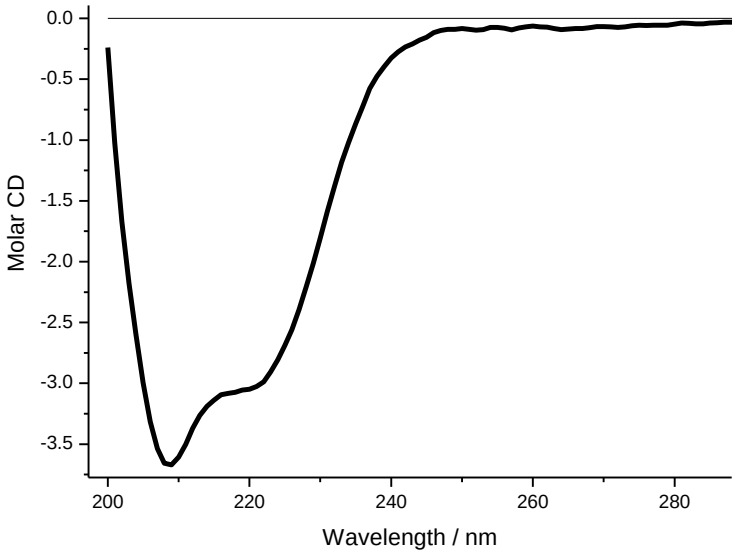


Figure 82 Beamline CD analysis of Tus in Tus buffer, 7.1 μM in 0.1 mm cell, 0.28 mm slit width, scanning at 0.5 nm per second.

4.5.3.3 Tus-Extended DNA 1 : 1 analysis

CD analysis of the Tus-Extended DNA complex was expected to reveal structural changes upon complexation. Scanning each entity separately, followed by complex form analysis, the comparison between simulated and experimental spectra was enabled. The experimental components were identified to give: Tus only (orange curve), DNA only (red curve) and the complex (blue curve) entities in Figure 83 (α). Tus CD analysis gave a strong negative curvature around 220 nm, which corresponded to the protein secondary structure region, a result of peptide linkages and folding CD absorptions. The Extended DNA displayed typical β-DNA CD curvature. Combining the two in equimolar ratios yielded the combined spectra. Potentially, the protein and DNA components change upon complexation, to identify the changes, further analysis was required.

Firstly, the complex spectra (blue curve) was compared to the sum of the Tus only and DNA only spectra; identified as the simulated combined curve (pink curve), as shown in Figure 83 (β). The majority of the CD difference is located in the region below 225 nm. Both the protein and the DNA have a CD absorption in this region, suggesting that potentially both the protein and the DNA change upon complexation. Comparing the Tus only curve to the experimental complex curve, minus the DNA only component (black curve); gives rise to a change in CD below 225 nm, as shown in Figure 83 (γ). Conducting the reverse, comparing the DNA only curve to the experimental complex, minus the Tus only component (grey curve), also gave rise to CD change below 225 nm, as shown in Figure 83 (δ). Finally, subtracting the sum of the Tus only and DNA only curves, to the experimental combined curve, revealed changes between the simulated and experimental curves (green curve); the majority of the CD change is again, located below 225 nm, as shown in Figure 83 (ε). Like above, the change in this CD region cannot be attributed to one component; since both entities have a CD absorption at 225 nm. CD and fluorescence titrations were conducted to further characterise the binding event, (see below).

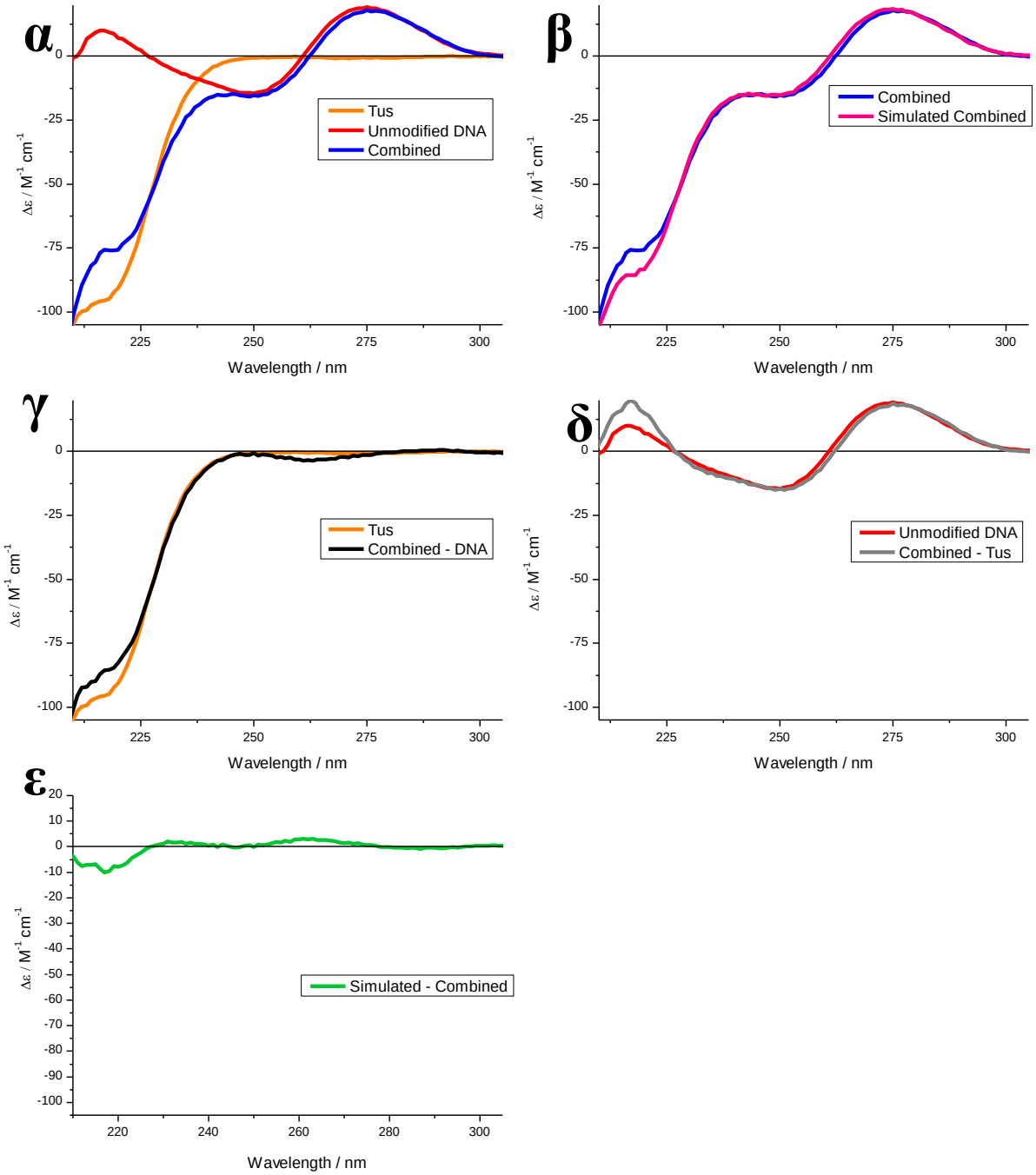


Figure 83 CD spectroscopy of Tus, Unmodified Extended DNA and combined in 1 : 1 ratio (α), combined and calculated combined (β), Tus and combined - DNA (γ), Unmodified Extended DNA and combined - Tus (δ) and simulated - combined (ε), Tus 18 μM, Extended DNA 18 μM, Combined 9 μM, in Tus buffer, 2 seconds per nm, 0.1 mm cell.

4.5.3.4 Tus-DNA complex melting analysis

CD Melting analysis of the Tus-DNA complex was conducted, in an attempt to identify information about the thermal stability of Tus, and various Tus-DNA complexes. The complexes were monitored at 222 nm, the region giving rise to significant protein CD. Each of the melting curves were compared to the protein only and DNA only controls, as shown in Figure 84. CD melting analysis of Tus revealed the structure is stable from 10-35 °C (red curve). Further heating resulted in a sharp irreversible unfolding of the protein. Integration of the first heating curve yielded a melting temperature (T_m) of approximately 45 °C. Theoretically, the Tus-DNA complex was expected to form a more stable structure than the Tus only; hence, a higher T_m was expected.

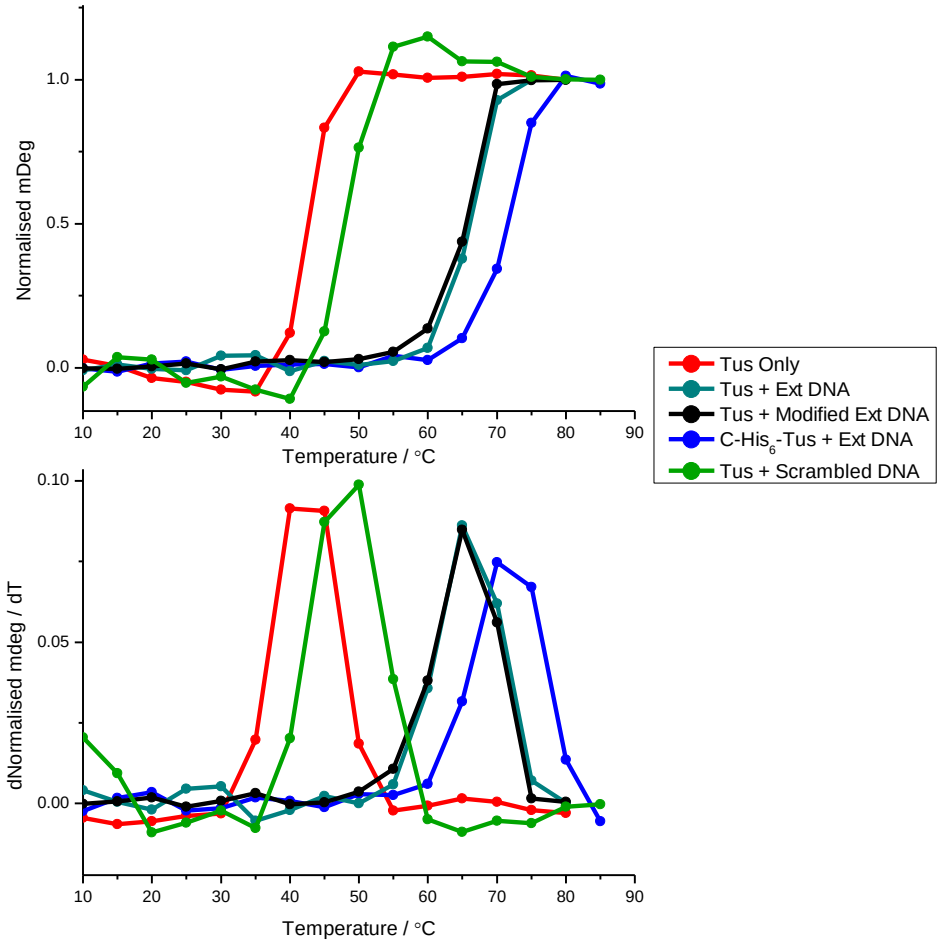


Figure 84 CD beamline spectroscopy melting analysis of Tus combinations, Tus to DNA molar ratio 1 : 1.6, Tus 1 μM, in Tus buffer, heating at 1 °C per minute, collecting data at 5 °C intervals in Tus buffer, 4 mm cell, scanning at 2 seconds per nm, 0.5 nm slit width.

The DNA only CD melting analysis did not give rise to a melting transition at 222 nm (see appendix section 8.30); simplifying interpretation of the results. With the addition of Tus to the Extended DNA, with the DNA in a 1.6 molar ratio excess (to ensure all protein present is in the complexed form, magenta curve), the complex gave rise to a T_m of approximately 65 °C. The formation of the complex resulted in significant structural stabilisation of approximately 25 °C, compared to the Tus only T_m. However, like Tus only, once the complex had melted, the unfolding of the protein was irreversible. As a control, a scrambled version of the Extended DNA sequence was applied, which contained the same base-pair content, but different DNA sequence (green curve). Melting analysis of the Tus-Scrambled DNA complex generated a T_m of approximately 50 °C, approximately 15 °C below the Tus-Extended DNA complex. This result highlights the proteins affinity for DNA containing the ter sequence. These results correlate well with the reduced binding affinity of non-specific DNA; in comparison to the Biacore results obtained by Neylon *et al.*[94].

The Tus-DNA complex with FAM and TAMRA Extended DNA was analysed to identify the effect of these modifications (black curve). Differentiation of the melting curve gave rise to a T_m of approximately 65 °C. Evidently, the modifications have a negligible effect on complexes stability, further verifying previous modified DNA complex results. CD melting analysis of carboxy labelled histidine tagged Tus with Extended DNA was conducted to identify the effect of the histidine tag upon heating (blue curve). Differentiation of the melting curve gave rise to a T_m of approximately 70 °C, indicating the histidine tag exhibits a small thermodynamic destabilising effect on the complex at the amine terminus.

4.5.3.5 Tus-DNA titration analysis

CD titrations with the Tus-DNA complex were conducted to identify information about the binding event; the binding constant and which entities undergo structural change upon binding. The following section describes the CD titrations conducted at Diamond Ltd, all of which were conducted on the beamline, unless stated.

Experimental planning was essential for quality CD results; all spectra were required to be within the PMT voltage window to avoid saturation of the detector. Hence, when conducting the titration, the starting concentrations were kept to a minimum. However, resolution was required to be sufficient to obtain meaningful data, i.e. sufficient signal-to-noise. Problems encountered when conducting a Tus-DNA CD titration were related to the small extinction coefficient of the protein between 280 and approximately 230 nm (lower wavelengths leads to quicker saturation of the detector, due to high protein and DNA absorptions). To obtain sufficient signal-to-noise, the concentration was increased. However, the DNA has an extinction coefficient an order of magnitude higher than the protein, after a few additions of DNA, saturation of the detector would of quickly been achieved (PMT above 650 volts). Before experimentation, the upper and lower limits of the PMT voltage, and hence absorption was estimated before, to ensure they were within the PMT voltage window. When conducting the titration, it was essential sufficient mixing of reagents was conducted; hence mixing using relevant and accurate pipettes, for example positive pressure pipettes, ± 0.05 μL. Once the substrate was added and mixed, sufficient time was allowed to pass to ensure the complex had reached equilibrium before scanning.

The Extended DNA (substrate) was titrated into Tus (host), to establish the binding event and further clarify the native PAGE and CD 1 : 1 results from above, as shown in Figure 85 (α). The spectra shows Tus only (black line), with increasing DNA molar ratio. The more DNA added, the more prevalent the DNA CD curvature from approximately 300-240 nm. In an attempt to identify a binding transition form the titration curves, the software CD Titration v1.6 (developed by Dr. Daniel Myatt at Diamond Light Source Ltd) was applied. The CD difference titration between Tus and the Extended DNA (calculated using the Tus only and DNA only spectra, not shown), displays no plateau for the expected 1 : 1 complex above 1 equivalent of DNA added, as

shown in Figure 85 (β). The curvature indicates incomplete saturation of the protein even after 1.4 equivalents of DNA; completely unexpected from pervious results. Potentially, the primary binding event could be masked by other non-specific binding events as visualised by CD at this concentration range. However, this is unlikely since the primary binding event of Tus to *TerB* DNA is very tight (0.5 nM). Any non-specific binding would also have to be very strong, which was not observed in the Biacore results by Neylon *et al.*[94]. The linearity of the curve could be related to the concentration of the reagents; if the binding constant is in the 0.5 nM region, then the titration would have to start close to the expected binding constant to obtain binding information. However, even if this was the case, an expected plateau upon saturation Tus should be observed[95]. Another possibility for no binding plateau could be related to aggregation of Tus at higher concentrations. Tus in the nM range may behave differently to Tus in the micromolar range, i.e. self-assembly and / or aggregation. The titration was repeated, and the curvature confirmed. To rule out DNA sequence effects of the Extended DNA sequence, the experiment was repeated using the *TerB* DNA sequence, to replicate the results obtained from Neylon *et al.* in their Biacore binding assay (see introduction section 1.10).

Deconvolution of the Tus-*TerB* titration also yielded no plateau, and hence no dominant binding event, even within 2.5 equivalents of substrate (see appendix section 8.30). In conclusion, it appears CD is not appropriate to study the Tus-DNA binding mechanism; comparing the Neylon results, the binding constants were established using surface plasmon resonance (SPR), whereby DNA was tethered to a surface, to which carboxy tagged histidine Tus was passed through. Potentially, it is wrong to compare the two techniques, SPR is on an interface, whilst the CD performed above, is in solution; both are different phases from which different interactions can be observed, for example a host tethered to a surface, is a two dimension interaction, where non-specific interactions are less likely. Whereas, in the solution phase other interactions may become more prevalent due to three dimensional interactions. Fluorescence spectroscopy was applied to identify the binding event, which allowed the host-guest system to be studied at lower concentrations, potentially minimising unwanted non-specific interactions.

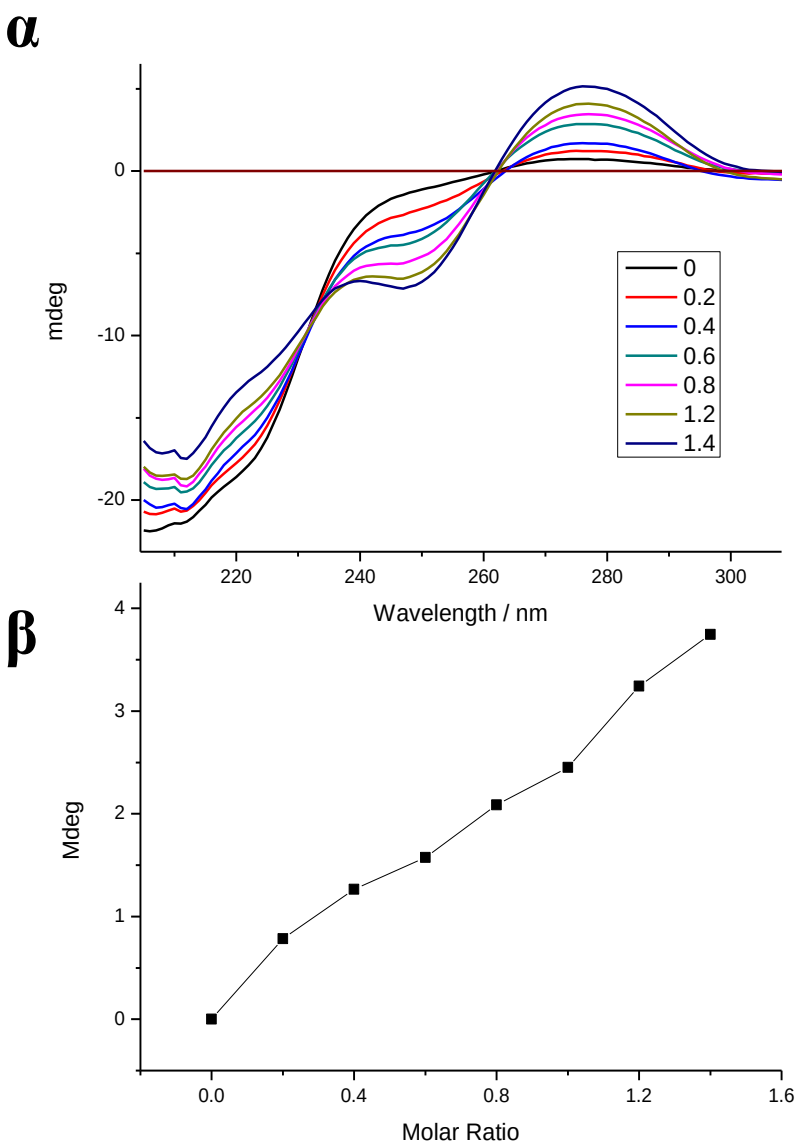


Figure 85 CD spectroscopy of Tus at 1.7 μM (host) with increasing Extended DNA (guest) titration (α), and CD difference spectra versus molar ratio (β), in Tus buffer, 2 seconds per nm, 4 mm cell.

4.6 Molecular ruler analysis of Tus-DNA complexes

Three different molecular rulers were applied to analyse different regions of the Tus-Extended DNA complex. A FAM and TAMRA molecular ruler was applied to investigate potential DNA flanking regions wrapping around the protein. Zinc and 2H porphyrin Extended DNA was applied to investigate the non-permissive region of the Extended DNA upon protein complexation. Finally, fluorescein and 2H porphyrin FRET pairs were applied to study complex formation. However, instead of a DNA based molecular ruler, fluorescein was attached to the protein to conduct FRET between labelled protein and DNA.

4.6.1 FAM and TAMRA FRET analysis

FAM and TAMRA FRET pairs were applied to identify potential DNA wrapping around Tus. If the Extended 45-mer duplex DNA does wrap around the protein, this will result in a large DNA conformational change. DNA wrapping should therefore give rise to a large change in FRET when FRET pairs are positioned either side of the binding domain. Calibration of the FAM and TAMRA FRET pairs Extended DNA sequence was achieved in chapter 3. FRET pair combinations with F^2 and F^3 displayed typical FRET behaviour; as predicated by the Förster model. F^1 combinations were omitted due to quenching effects related to the donor's terminal positioning, resulting in single-strand quenching, and subsequent non-FRET type behaviour. FRET experimentation of the FAM and TAMRA Extended sequence with Tus was conducted to identify binding information of the complex, and any protein induced changes on the DNA conformation. Tus binds to the Extended DNA sequence at the *TerB* 16 base-pair region. Modifications were replaced for T's outside this strong protein-DNA contact region. Modifications have been shown previously not to affect protein binding, due to the direct 'clamping' nature of binding. Fluorescence titration analysis of TAMRA labelled Extended DNA displayed the relatively tight binding to Tus is maintained (K_D = 74 nM ± 20), which led to steady-state FRET experimentation being conducted. Firstly, the various DNA combinations were measured and annealed together, then Tus

in 1 to 1 molar ratio was added; the protein-DNA was left to equilibrate at room temperature for 15 minutes, after which UV-Vis and fluorescence spectroscopy was performed.

Figure 86 displays the fluorescence emission of the eight F^2 and F^3 combinations. The donor emission at around 515 nm yields a high fluorescence intensity for combinations: $F^2.T^3$, $F^3.T^1$ and $F^3.T^2$, indicating low FRET, and therefore large FAM and TAMRA distance separations. However, to accurately identify the peaks emission, the donor and acceptors peaks were deconvoluted using the Clegg method (see introduction, section 1.6.2).

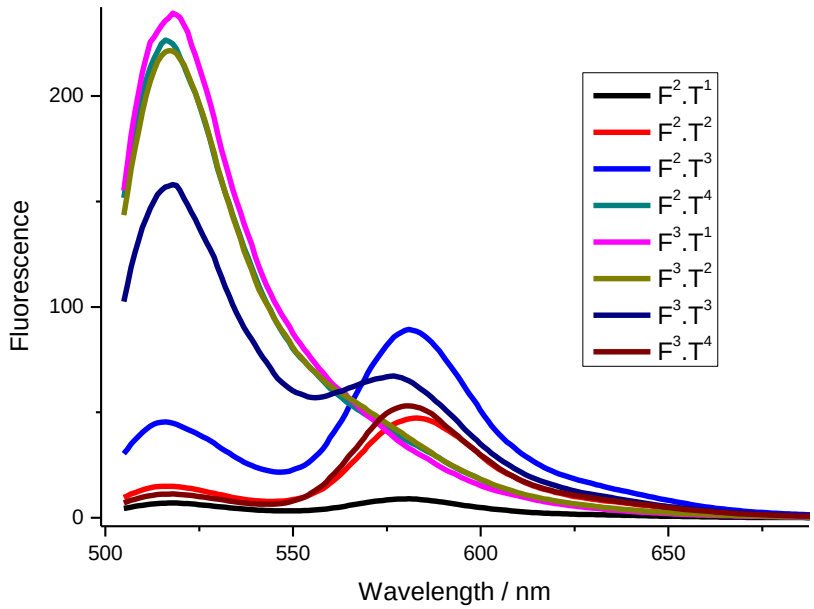


Figure 86 Fluorescence emission of FAM and TAMRA extended DNA combinations with Tus 1 : 1, excitation at 495 nm, Tus and DNA combinations at 1 μM, in Tus buffer, 1 cm fluorescence cell, scanning 600 nm per minute, 500 PMT voltage.

The deconvoluted donor values from above were normalised using the E_{FRET} equation (see section 1.6.2), from which the E_{FRET} values were plotted and compared to the values before and after the addition of Tus, as shown in Figure 87. The graph shows at distances less than the Förster distance (R_0, < 54 Å), the E_{FRET} efficiency

increases in comparison to the values obtained without Tus, suggesting the donor-acceptor distances decrease. At distances greater than the R_0 value, the E_{FRET} efficiency decreases, i.e. the donor-acceptor distances increases. These results suggest anomalous behaviour other than FRET potentially occurring; how can more E_{FRET} occur with Tus, when the FRET pairs before the addition of Tus (small FRET distances), are supposedly at a maximum? The results suggest the model is not valid when comparing the two systems, which is potentially explained by additional interactions between the protein and the fluorescent moieties.

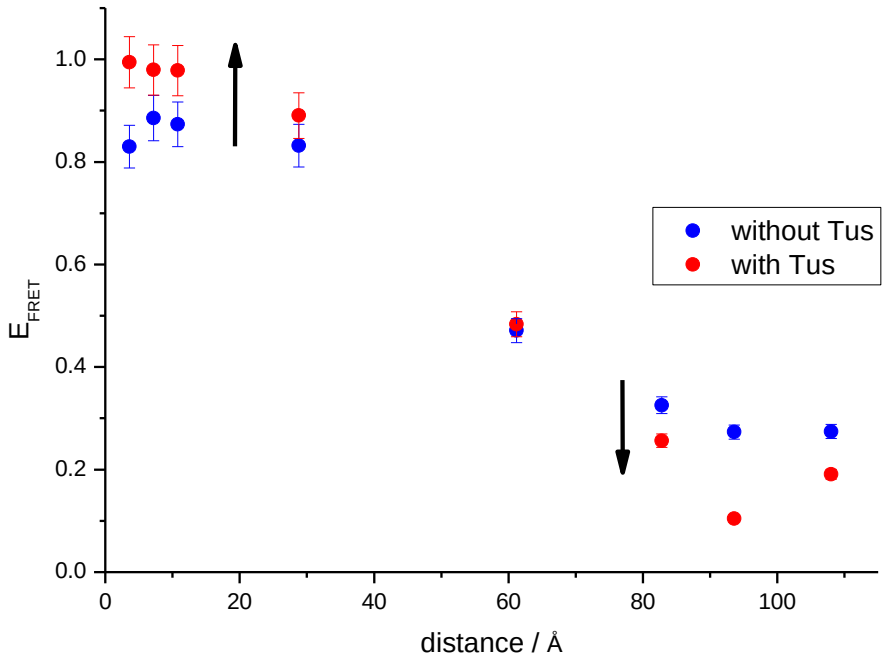


Figure 87 E_{FRET} versus Distance for FAM and TAMRA FRET pairs combinations F^2 and F^3 with (Red) and without (Blue) 1 equivalent of Tus.

To identify the anomalous behaviour reported above, a fluorescence titration with increasing Tus concentration to FAM and TAMRA modified Extended DNA was conducted. With increasing Tus concentration and subsequent complex formation, any significant change in DNA upon binding was expected to result in a change of the donor-acceptor distance and hence donor and acceptor emission, to potentially reveal a

binding isosbestic point[15]. The fluorescence emission of F^1 with T^3 Extended DNA (Figure 88), when exciting at 495 nm yields the fluorescence emission in Figure 89 (black line, left spectra). The peak heights correspond to approximately 55 % donor and 45 % acceptor emission (calculated by deconvoluting the peaks and integrating the area under the curve), suggesting approximately 45 % FRET efficiency; hence donor-acceptor are approximately 58 Å apart. Addition of Tus in 0.3 molar ratio changes the emission intensity of the TAMRA moiety, Figure 89, which is in agreement with previous results (see section 4.2), FAM emission remains largely unchanged. After complete complexation of the DNA, quantities > 0.9 equivalents of Tus added, the TAMRA emission increased by approximately 22 %. This change however is not resulting from changes in the FRET pair distances, i.e. changes in the FRET efficiency, instead these changes are the result of complexation effects affecting TAMRA's local environment, and therefore no binding isosbestic point was exhibited.

Interestingly the local induced change on TAMRA upon Tus binding gives a 22 % increase in fluorescence arising from FRET (Figure 89), whilst direct excitation at 560 nm gives rise to a 30 % increase (see section 4.2). It is interesting to note that these values are not the same, potentially indicating another anomalous effect upon binding when donor and acceptor are within the FRET distance range. The fluorescence titration was repeated using FRET pairs positioned either side of the binding domain, using combinations. Again no isosbestic point was present between the FAM and TAMRA emission peaks. However, change was observed in the TAMRA emission peaks, which again was attributed to protein binding. Overall, these results show that the DNA does not undergo significant change in DNA conformation, such as DNA flanking region wrapping to increase complex affinity. If wrapping did occur, then corresponding change would be observed in donor-acceptor emission, even with the protein induced binding effects on TAMRA.

```
T3 Ext  5′ GCT GCC AGC TCC GAA 3AA GTA TGT TGT AAC TAA AGT GGA 3′
F1 Ext  3′ CG1 CGG TCG AGG CTT ATT CAT ACA ACA TTG ATT TCA CCT 5′
```

Figure 88 FAM + TAMRA Extended DNA sequence applied for Tus-DNA fluorescence titration analysis, TAMRA modification shown in red, FAM modification shown in green, protein-DNA contact region shown in yellow, Kamada sequence shown underlined (top).

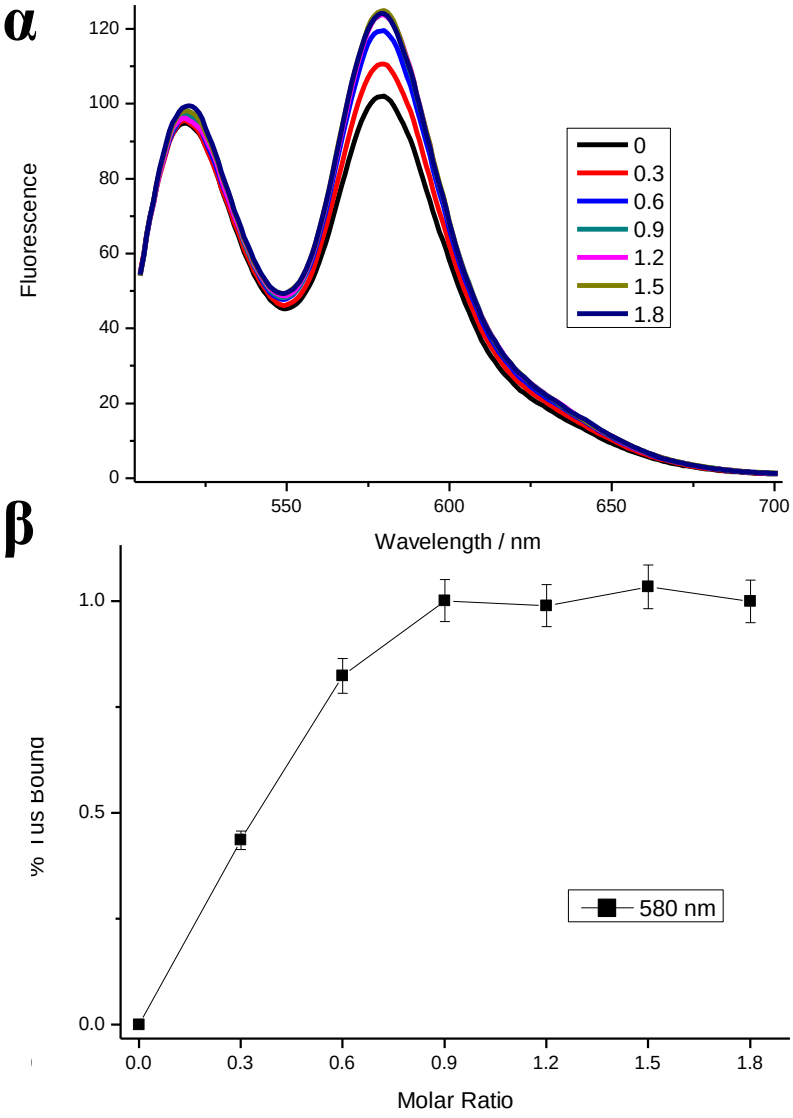


Figure 89 Tus titration to FAM + TAMRA modified DNA, fluorescence emission spectra (left) exciting at 560 nm at 1 μM in Tus buffer, and normalised fluorescence change at 580 nm (right), scanning at 600 nm per minute, 500 PMT voltage (bottom).

4.6.2 Porphyrin exciton coupling analysis

Zinc metalled and 2H porphyrins have been used previously for FRET, CD and SAXS experimentation (see chapter 2). Exciton coupling between porphyrin labelled DNA has previously shown to be distance and angle dependant. This system was applied to study the Tus-*Ter* complex, to potentially identify subtle changes in the non-permissive DNA region upon complex formation. The porphyrins were synthesised and incorporated into the Extended DNA sequence in Figure 90. Both porphyrins were located in the non-permissive region which had been identified to show anomalous DNA behaviour (chapter 3.2.5). Therefore, a porphyrin DNA CD molecular ruler was applied to observe any protein induced change in excitonic coupling, and therefore any non-permissive region conformational changes.

Non-permissive end **Permissive end**

W² 5′ GCA GCC AGC 2CC GAA TAA GTA TGT TGT AAC TAA AGT GGA 3′

E¹ 3′ CGT CGG 1CG AGG CTT ATT CAT ACA ACA TTG ATT TCA CCT 5′

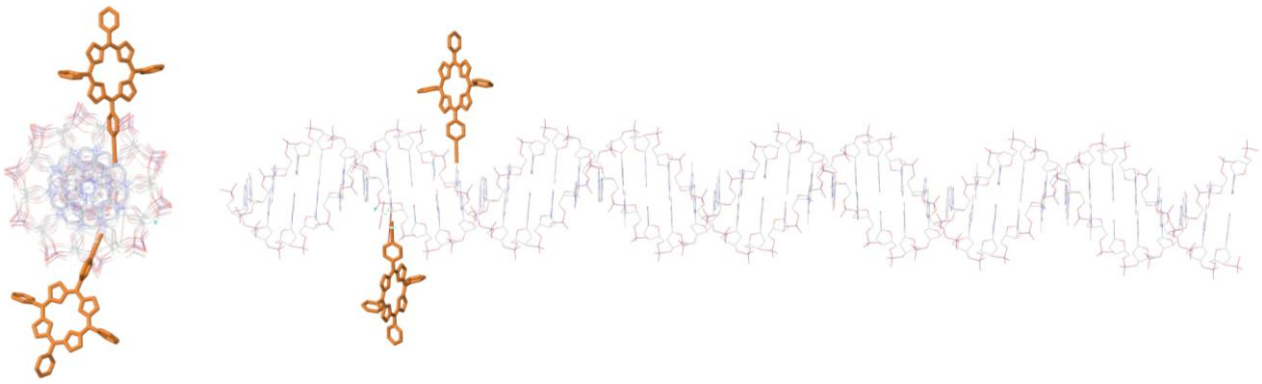


Figure 90 2H-2H porphyrin Extended DNA sequence applied for complex analysis (top). Macromodel top and side images of $W^1.E^1$ porphyrin labelled Extended DNA, porphyrin-porphyrin dihedreal angle approximately 133°, separated by approximately 10 Å (bottom).

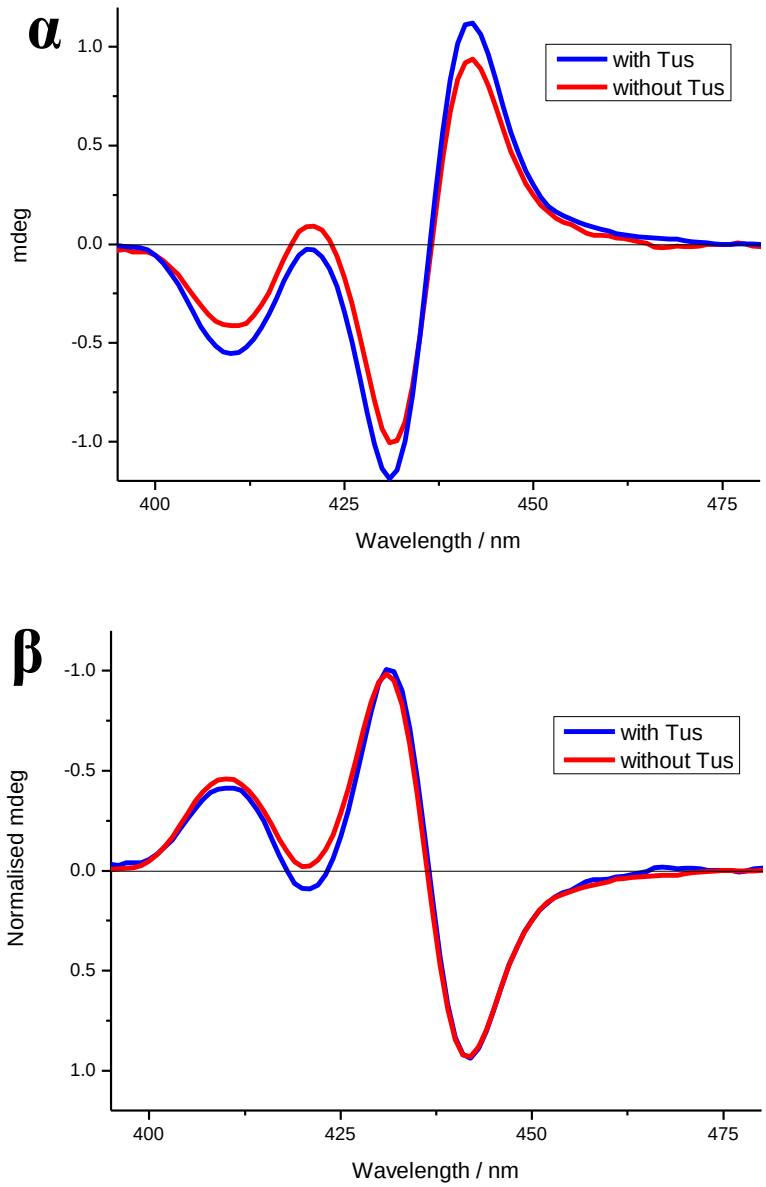


Figure 91 CD Porphyrin region of $W^1.E^1$ porphyrin labelled Extended DNA (red) with 1 equivalent of Tus (blue), experimental spectra (α) and normalised spectra at 440 nm (β). Chirascan CD instrument, sample concentrations at 41.5 μM, scanning range 500-200 nm, 2 seconds per nm, 2 mm path length cell in Tus buffer.

The porphyrin strands were synthesised, purified and annealed using protocols established in chapter 2, after which CD of the Soret band region was performed. Firstly, the porphyrins were analysed in duplex form only, then in complex form with 1 equivalent of Tus (the DNA and protein was equilibrated for 30 minutes prior to

scanning). The experiment was conducted at a relatively high concentration to obtain sufficient porphyrin signal-to-noise ratio. The CD porphyrin Soret band region in Figure 91 (α) displays a + / - / - tri-signet Cotton pattern. The intense bisignet peaks at 440 and 430 nm correspond to porphyrin-porphyrin exciton coupling interactions. The additional negative peak at 420-390 nm has been described as intermolecular π-π porphyrin interactions by Berova *et al.*[101], expected for porphyrin DNA at these relatively high concentrations. After the addition of 1 equivalent of Tus, the porphyrin region displayed subtle changes in intensity and undergoes a small peak shift. These changes are clarified in the normalised CD spectra, as shown in Figure 92 (β), normalisation at the positive 440 nm peak yielded a slight red-shift of the negative 430 nm peak, and reduced intensity of the negative 420 nm peak. These changes in CD of the porphyrin exciton coupling indicate a reduction in porphyrin aggregation as Tus binds. However, substantial changes are not observed in the main Soret band region, indicating the porphyrins and therefore the non-permissive region of the DNA does not undergo significant conformational protein induced changes. Therefore it is apparent that the hypothesised G-quadruplex formation of the non-permissive region is derived from the high sample concentration, and representative of Tus-Extended DNA in solution, i.e. the Extended DNA does not adopt a pre-formed G-quadruplex structure to aid complex affinity.

4.6.3 F-Tus and porphyrin DNA FRET analysis

FRET between fluorescein and porphyrins have been applied before by Brown *et al.*[77]. Applying excitation at the peak absorption maximum of FAM (donor), yielded FRET to the 2H-porphyrin Q bands (acceptor) since the FRET prerequisites: spectral overlap of donor emission and acceptor absorption, free rotation of the donor-acceptor dipoles and high quantum yield of the donor (0.4) were satisfied. Hence, efficient FRET between FAM labelled Tus (F-Tus) with 2H porphyrin DNA was expected. The exact positioning of the fluorescein moiety in FAM-Tus was likely to vary from that shown in Figure 92. The moiety could potentially fray out from the bulk of the protein to occupy a large volume, or favour a hydrophobic pocket within Tus itself, since

fluorescein is aromatic. The image shows the group extended linearly from the protein, using the Tus-*Ter* crystal structure from Kamada *et a.* [67].

The literature Förster distance (R_0) for the FRET pairs is 45 Å[77]. From the design stages of the experiment, this R_0 distance was sufficient to measure distance changes between FAM-Tus and the labelled 2H porphyrin DNA, Figure 91 shows F-Tus with 2H porphyrin Extended DNA. The donor and acceptor are separated by approximately 60 Å using the x-ray crystal structure from Kamada *et al.*[21].

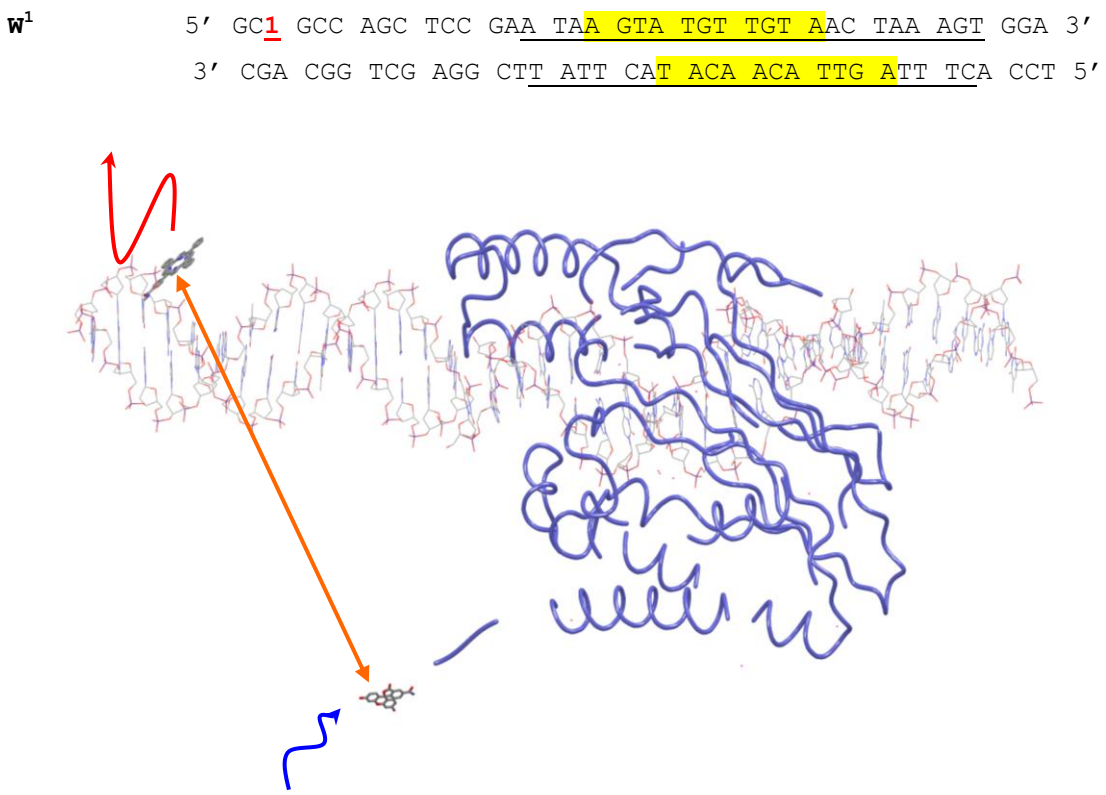


Figure 92 Fluorescein-Tus (blue ribbon) with W^1 2H Porphyrin Extended DNA, chromophores are separated by approximately 60 Å as indicated by orange double headed arrow, Tus structure modified from Kamada *et al.* [21].

.7.3.1 Fluorescein-Tus biosynthesis

The molecular ruler does not specifically have to be assigned to DNA; the molecular ruler probes can be attached to the protein itself, from which additional protein-DNA distances can be obtained. Neylon *et al.* have used sortase to modify and label the carboxy terminus of proteins[91, 92]. Sortase can also attach modifications to the amine terminus of proteins. The mechanism works by sortase's affinity for a specific carboxy terminus amino acid sequence: R_1-LPXTGXX-COOH (where X is any amino acid residue). Sortase contains a key cysteine-184 amino acid which is activated by proton transfer from histidine-120, it is this sulphur atom from cysteine-184 which binds to the carboxy terminus, cleaving the GXX-COOH sequence (see Figure 93), forming an R_1-LPXT-sortase intermediate. This intermediate is susceptible to nucleophilic attack by an addition-elimination reaction of the newly formed thiol-ester bond. If an amine is present, then it will react to form a new amide bond with displacement of the sortase enzyme. The reaction has the potential to form a hydrolysed product with a water molecule in the final nucleophilic addition-elimination reaction, rather than the desired amine. If the reagents are left for extended time then increased quantities of the hydrolysed product will be obtained[91, 92].

The R groups for the R_1-LPXTGXX-COOH binding sequence and the amine nucleophile can be proteins or small organic molecules, it does not matter which is which. The versatility of modification opens up an enormous potential for functionalised molecular biology.

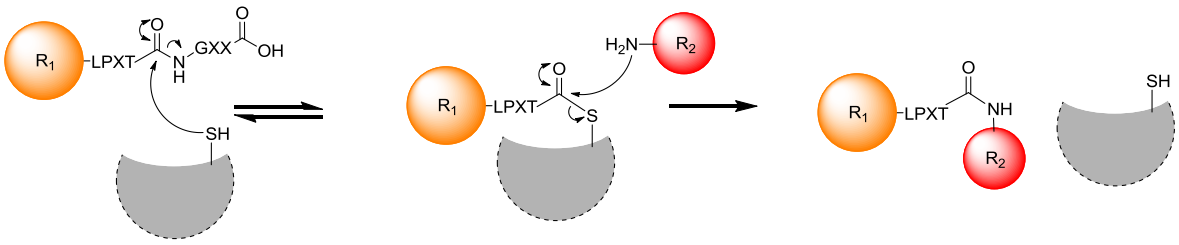


Figure 93 Sortase mediated reaction of carboxy terminus of LPXTGXX-COOH (R_1) with nucleophilic amine (R_2).

Neylon *et al.* attached fluorescein-lysine diglycine (see appendix section 8.4 for chemical structure) to the carboxy terminus of Tus using the sortase enzyme as described above, where R_1 is Tus and R_2 is fluorescein-lysine diglycine. Tus contained the LPXTGXX-COOH sequence with an additional histidine tag at the carboxy-terminus. Upon formation of the sortase intermediate, the histidine tag and GXX sequence was eliminated. Sortase labelling was applied to the Tus-DNA system for molecular ruler application *via* FRET, with fluorescein labelled Tus and 2H porphyrin DNA. Adding modified lysine-diglycine to the protein's carboxy or amine terminus enables extended functionality of the protein itself. The reaction proceeds *via* a Sortase mediated reaction with displacement of the histidine tag. Attaching a fluorophore to Tus in combination with fluorescently modified DNA enables further FRET distances / binding information as compared to the DNA only approach. Commercially available fluorescein-lysine-diglycine was attached to His_6-Tus *via* a sortase mediated reaction. Sortase binds to the LPETTGG-His_6 sequence by displacement of the GG-His_6 sequence. The free nucleophilic amine of fluorescein-lysine-diglycine binds to the carboxy terminus of Tus whereby cleaving sortase. The reagents; Tus, sortase and fluorescein-lysine-diglycine were simply mixed at room temperature overnight (see appendix section 8.5 for protocol). The product was purified by nickel affinity chromatography. The fluorescein-Tus (F-Tus) product has a lower affinity for nickel than the starting material, hence F-Tus elutes with low imidazole concentration. The product was further purified by size exclusion chromatography, as shown in Figure 94.

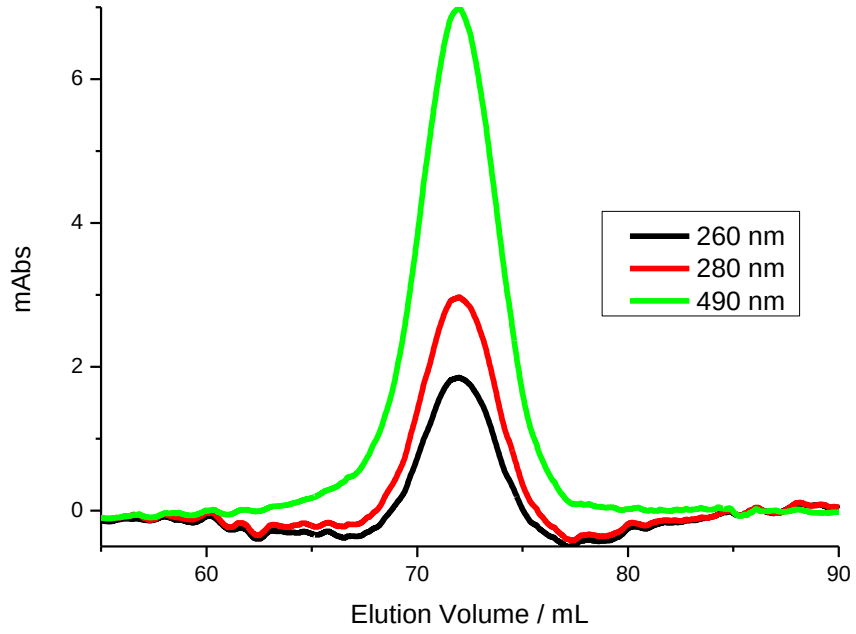


Figure 94 Size exclusion chromatogram of pure F-Tus in Tus buffer, 10 nanomol injection, monitoring the UV absorption at 260, 280 and 490 nm, using S75 size exclusion column, 1 mL per minute flow rate.

4.7.3.2 FRET analysis

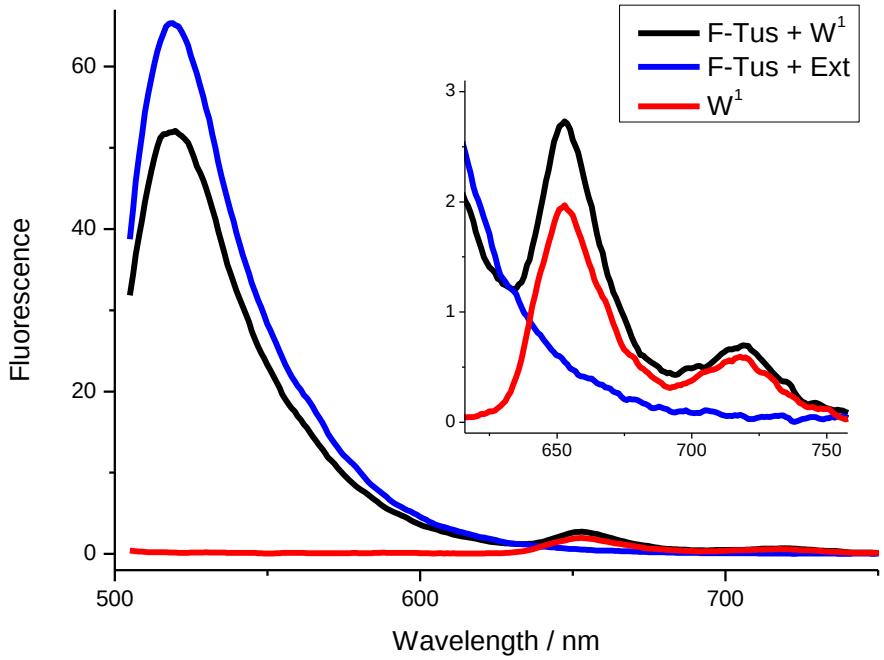


Figure 95 FAM labelled Tus (F-Tus) with 2H porphyrin DNA (W^1), and control with unlabelled DNA (Ext) at 2 μM in Tus buffer, inset zoomed porphyrin emission region, excitation 495 nm, scanning at 600 nm per minute, 500 PMT voltage.

Figure 95 displays the control donor only fluorescence emission of F-Tus with the unlabelled Extended duplex DNA when exciting at 495 nm (1 to 1 molar ratio). After complexation of F-Tus with W^1 (donor with acceptor), the emission decreases by approximately 19 % indicating FRET. Monitoring the 2H porphyrin emission region does not yield substantial increases in acceptor emission, as expected from FRET when considering the 19 % decrease of the donor emission. Potentially, the majority of the emission is the result of direct excitation of the porphyrin absorption at 495 nm, as evidenced in the porphyrin only spectra. The results suggest potential quenching of the fluorescein emission upon the addition of 2H porphyrin moiety, rather than quenching upon addition of DNA only, as verified by the control with unmodified DNA. There is also the possibility of direct interaction between the donor and acceptor leading to quenching, however for this to occur the FRET pairs are required to come into close

proximity (< 10 Å) which is unlikely given the simulated models in Figure 92 suggesting approximately 60 Å distance separation.

Deconvolution of the F-Tus emission and applying the E_{FRET} equation (see section 1.6.2), an E_{FRET} value of approximately 0.15 was obtained. Plotting the efficiency and applying the literature Förster distance gave rise to results in good agreement with the simulated distance from above (Figure 96).

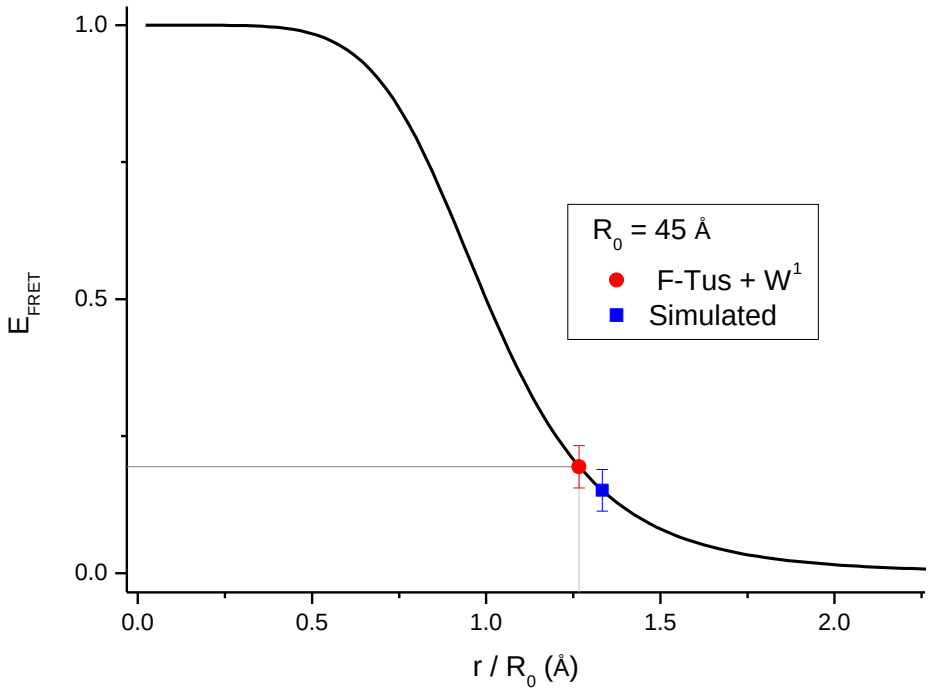


Figure 96 E_{FRET} versus Distance for experimental fluorescein labelled Tus with 2H porphyrin modified DNA (red square) and simulated distance separation (blue square) using R_0 from [77].

4.8 Summary

Firstly, the Tus biosynthesis was explained, followed by purification and analysis (section 4.2). Tus has been shown to form a protein-DNA complex with an elongated version of the *TerB* binding sequence. Native PAGE analysis of the unlabelled DNA complex showed Tus binding activity was maintained (section 4..3). The results were subsequently compared to FAM and TAMRA labelled Extended DNA (section 4.4); blocking either the permissive or non-permissive region of the binding sequence did not reduce the binding affinity, suggesting the Tus binding mechanism is either a direct binding, or more likely is a non-specific sliding mechanism. Fluorescence titrations yielded fluorescence change in the TAMRA emission upon the addition of Tus (section 4.5). Fitting a 1 to 1 binding model resulted in a binding constant in approximate agreement with the literature value for Tus-*TerB* complexation.

AGF and CD analysis was applied to determine the solution phase structure of the Tus-Extended DNA complex (section 4.6). The hydrodynamic radius of Tus was determined using AGF with respect to known standards. The result was in good agreement with the literature value and the X-ray crystal structure. However, AGF analysis of the Extended DNA complex was inconclusive; the complex appeared unstable on the column which lead to streaking and anomalous behaviour. Synchrotron CD studies of Tus had to be optimised to prevent protein degradation arising form the intense beamline light source. CD 1 : 1 studies of the complex showed the majority of the change in the CD was located below 215 nm. Due to overlapping protein and DNA CD signals, the change could not conclusively be attributed to either. However, most change was expected to arise from the protein when considering the crystal structure and the DNA's limited deviation from typical helical arrangement. CD melting studies of the various complexes displayed the expected behaviour; Tus melted at 45 °C, once melted, unfolding was irreversible. The various complexes yielded melting temperatures up to 20 °C higher, even with modifications. The control experiment with scrambled DNA resulted in non-specific binding, to yield a melting temperature above the Tus only melt, but 15 °C below any of the specific DNA combinations. CD titration analysis did not reveal the expected 1 to 1 binding curves, even after extensive

optimisation. This result was potentially attributed to the high concentrations required analysis.

FAM and TAMRA FRET pairs were applied to analyse potentially DNA wrapping of the Extended DNA complex (section 4.7). FRET FAM and TAMRA molecular ruler yielded substantial increases in the TAMRA emission upon protein binding. However, this increase was not the result of changes in the FRET pair distance, due to the lack of an isosbestic point. The result suggests the DNA does not undergo significant changes upon protein binding, i.e. the DNA flanking regions do not wrap around the protein. Smaller changes may take place however, which are not detectable using FAM and TAMRA FRET pairs due to their relatively large and flexible linkers.

Appling the porphyrin CD molecular ruler to the non-permissive flanking region did reveal subtle changes upon complexation. The probes were attached to the non-permissive side of the duplex to investigate the anomalous behaviour observed in this region from previous CD studies. The porphyrins gave rise to a small change in exciton coupling, suggesting small displacement of the DNA strands. Applying FRET pair's fluorescein and 2H porphyrin attached to the protein and the DNA respectively, gave rise to donor emission quenching upon binding. The result did not clearly giving rise to the increased acceptor emission expected. However, this quenching only came from the addition of the donor in the presence of the acceptor, validated by experimentation with relevant controls, indicating non-fluorescent FRET. Comparison of the donor quenching gave rise to FRET distances in approximate agreement with those calculated using the X-ray crystal structure from Kamada *et al.*.

Adjustable Strap

5.1 Introduction

The adjustable strap (AS) is a 45-mer DNA single-strand sequence which is comprised of 80 % self-complementary bases, and 20 % non-self-complementary bases. The AS strand was designed to form four different hairpin conformations and one duplex conformation. Each conformation is formed upon addition of the appropriate complementary strand, followed by subsequent annealing. Attaching FRET pairs to the extremes of the AS sequence enables multiple donor-acceptor distances, and therefore multiple FRET efficiencies to be obtained.

The AS utilises DNA's ability to form hairpin conformations; hairpin structures form from self-complementary sequences of DNA single-strands. The required minimum length for hairpin formation is three bases, as found in nature's tetraloop[102]. The stability of the hairpin loop is dependant on the length of the helix region; the greater the helix region, the more stable the hairpin. Incorporation of a non-self complementary terminus enables hairpins to unwind spontaneously upon the addition of a complementary strand. When the hairpin unwinds, energy is released; the greater the loop length, the greater the energy released. Note, however, the larger loop sizes, the more prevalent intermolecular hairpin interactions become, such as kissing complexes[103]. Turberfield and coworkers selectively utilised hairpins as autonomous molecular motors with the addition of a non-self complementary region (sticky terminus)[104]. The addition of the complementary strand, leads to initial sticky terminus binding, followed by unwinding of the hairpin, to eventually form a full duplex; energy is given out by the system to generate the thermodynamically more stable duplex. The energy released is approximately 1.4 kcal mol^{-1} per single-strand base in the hairpin region at 20 °C[105].

Brown and co-workers designed Hybeacons® which contain two fluorophores attached to a DNA strand which anneals to form a hairpin. The hairpin brings the two fluorophores into close proximity, of which undergo fluorescence quenching. The hairpin sequence selectively binds to its complementary strand when added *in situ*. Once added, the hairpin unwinds to form a full duplex with the release of energy. The comparatively rigidified duplex arrangement leaves the fluorophores positioned apart

preventing quenching, and hence giving rise to a contrastingly increase in fluorescence[106, 107].

Firstly, the AS sequence design and modification positions are identified (section 5.2), after which the expected modification distances were calculated. Two FRET pairs were applied to analyse the different AS conformations, established FRET pairs FAM and TAMRA and zinc and 2H porphyrins. FAM and TAMRA moieties were applied first to calibrate the AS sequence. The five donor-acceptor distances were calculated, and the expected E_{FRET} intensities plotted (section 5.3). Experimental analysis of the AS sequence began by monitoring the hairpin formation by varying the annealing rate (section 5.4). With the annealing rate optimised for maximum hairpin formation, fluorescence melting analysis was conducted on the five combinations (section 5.5). The donor and acceptor emission was monitored to observe how the FRET intensity varies upon conformation melting. Finally, steady-state fluorescence analysis was performed to identify the donor emission intensities (section 5.6). The donor emission peaks were deconvoluted, from which the E_{FRET} values were identified and plotted and compared to the calculated values. Following FAM and TAMRA analysis, the zinc and 2H porphyrin FRET pairs were applied to the AS sequence (section 5.7). Firstly, synthetic details are described (section 5.8), followed by calculated E_{FRET} values (section 5.9). After which steady-state fluorescence experimentation was performed and analysed (section 5.10).

5.2 AS sequence

The AS is a 45-mer DNA single-strand containing a 4 base 1-2 DNA poly-repeat unit, see Figure 97. The repeating sequence is either ATTA or TAAT sequence. The actual sequence selected could be any number of variants, including 3 or 5 base repeats. However, 3 repeats were not selected due to decreased hairpin stabilisation. G's and hence C's were not used due to potential G-quadruplex or i-motif formation respectively. The AS also contains a 13 base non-repeating section called the sticky terminus. The sticky terminus region is required as a 'foothold' for complementary strand addition.

1 - 2 - 1 - 2 - 1 - 2 - 1 - 2 - Sticky End

5' AAAT ATTA TAAT ATTA TAAT ATTA TAAT ATTA GCD GTA ACT AAA G 3'

Figure 97 AS regions (top), 1 and 2 self-complementary bases, where 1 = TAAT, 2 = ATTA, sticky end = non-self-complementary region. AS sequence, D = Donor A = Acceptor.

If FRET pairs are attached to the primary DNA sequence at either terminus as shown in Figure 98, when the primary strand is in full duplex format, i.e. the complementary strand added contains the full complement for the entire primary strand, the FRET pairs at the terminal ends should be as far apart as possible, and hence conduct minimal FRET ($S^1.C^1$ in Figure 98). At the other extreme, when the sticky-end only complement only is added, the 1,2 poly-repeat should form the largest hairpin structure, yielding FRET pairs in close-as-possible proximity, and so maximum FRET was expected ($S^1.G^5$ in Figure 98). If other complementary strands are added each containing the sticky end complement and a section of the 1,2 repeat unit, then up to 5 different AS configurations are possible, and hence 5 different FRET distances.

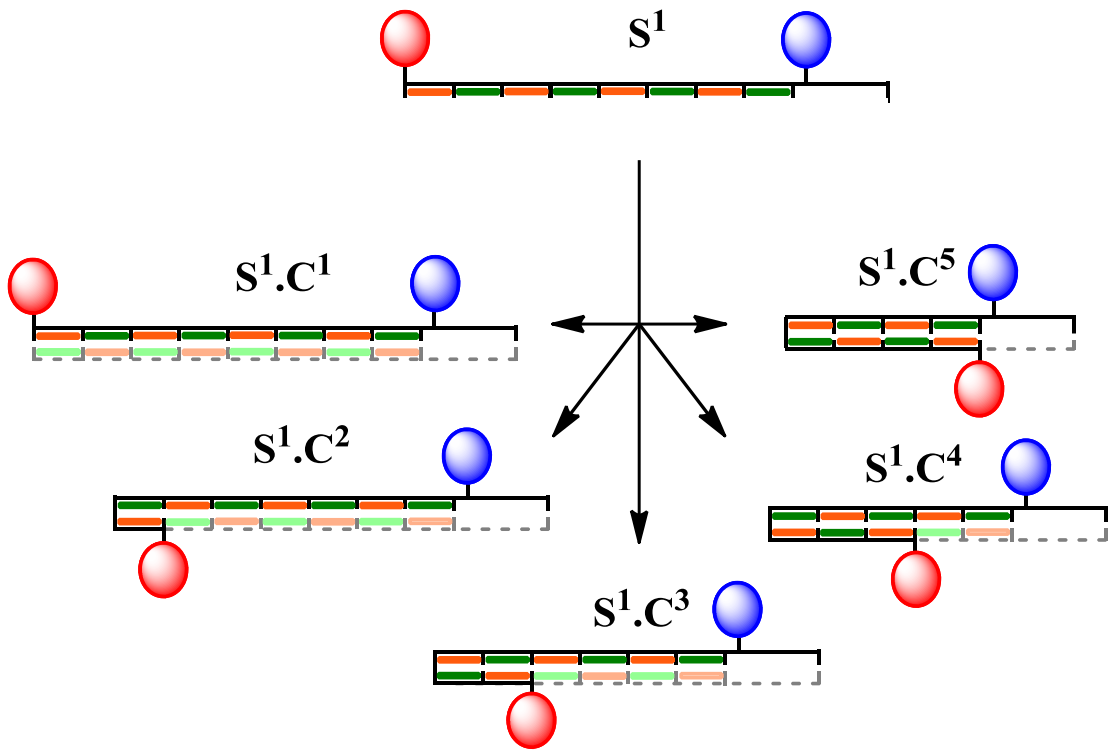


Figure 98 AS schematic with varying complementary strand, blue sphere = Donor, red sphere = Acceptor, orange and green rectangles = self-complementary region, hollow region = sticky end, grey colours = complementary strands.

Attention was paid to the total length of the primary strand, the 45-mer chosen was approaching the limit for doubly modified DNA synthesis, longer and final yields could be untenable, however much shorter and too few hairpins are formed, leading to insufficient FRET distances. Another possibility to increase potential FRET distances was to reduce the 1,2-poly repeat unit, for example 3 bases could have been used instead of 4. This format was considered in the design stage, however was omitted due to decreased stability of the smaller hairpins compared to 4 base-pair overlap regions.

The AS and full complement ($S^1.C^1$) combination was designed to anneal and form a DNA duplex with no hairpin region; a 45-mer, consisting of the highest T_m (mfold calculated T_m approximately 60 °C)[87]. Minimal FRET efficiency for the combination was expected, i.e. maximum donor emission in comparison to the other FRET combinations. The $S^1.C^2$ combination was designed to anneal and form the smallest hairpin conformation. Due to only approximately 4 bases in the hairpin

sequence, the $S^1.C^2$ hairpin was expected to be the least stable[87]. The sticky region of the AS sequence enables the C^2 complement to anneal and form a duplex covering 82 % of the AS strand. The $S^1.C^3$ combination was designed to form a larger hairpin conformation deriving from 16 self-complementary bases. The duplex region of the $S^1.C^3$ combination should cover approximately 64 % of the AS strand. The $S^1.C^5$ combination was designed to form a hairpin conformation deriving from 24 self-complementary bases. The duplex region should cover approximately 47 % of the AS strand. The $S^1.C^5$ complement was designed to form the maximum hairpin conformation deriving from 32 self-complementary bases. The combination should form a duplex covering approximately 29 % of the AS strand. The $S^1.C^6$ combination enables direct comparison of the single-stand quenching effect on the acceptor chromophore when compared to the S^1 only strand.

If the AS only (S^1) sequence is heated and annealed slowly (1 °C / min), an intramolecular hairpin with maximum helix formation is expected, leaving the sticky end, a 13-mer, to fray around in a single-stranded fashion. The hairpin region will comprise of about 14 base-pairs which should have an approximate T_m of 53 °C (calculated from mfold). During the formation of the hairpin, the annealing rate is important; too fast, then multiple hairpin species can be obtained, too slow and the potential for homodimerisation is increased (note, intermolecular complex formation, also increases with increasing concentration)[103]. Below the T_m for the S^1, the FRET pairs should be in close proximity, and hence maximum FRET efficiency was expected, analogous to the $S^1.C^5$ combination. One must note the partial single-stranded environment of the acceptor positioning; this effect was expected to complicate analysis with the other FRET combinations. The single-stranded environment of the AS only combination was therefore expected to lead to quenching of the acceptor's fluorescence.

Table 6

AS combination	BP separation of donor and acceptor	Donor-Acceptor distance (3.6 Å model)
$S^1.C^1$	33	122.4
$S^1.C^2$	26	97.2
$S^1.C^3$	18	68.4
$S^1.C^4$	10	39.6
$S^1.C^5$	2	10.8

The distance of the FRET pair modification sites between donor and acceptor combinations was calculated using the 3.6 Å distance separation (see introduction, section 1.6.2)[3]. Table 6 shows the calculated distances displaying sequential change in distance of FRET pairs, resulting from the 4 base elongation of the complementary strand ($S^1.C^5 \rightarrow S^1.C^2$). The $S^1.C^1$ combination deviates from this incremental change as a result of the positioning of the donor moiety, which gives rise to an additional 7 bases-pair separation. The 3.6 Å model does not allow for the helical curvature of the duplex for application during FRET studies, however this effect was not expected to be large for fluorophores attached to DNA *via* long flexible linkers, as used by Lilley and co-workers[3].

5.3 Calculated E_{FRET} vs. donor-acceptor distances

FAM and TAMRA as FRET pairs have been used previously for steady-state FRET measurement (chapter 3), and by many other groups for monitoring FRET on DNA[7-8]. Since the AS covers a relatively large distance, suitable FRET pairs are required to test the full AS combinations. The approximate Förster distance (R_0) for FAM and TAMRA FRET pairs (approximately 54 Å) is therefore ideally suited for AS analysis. Using this Förster distance, the $S^1.C^1$ combination was expected to yield no FRET (approximately 122 Å), whilst the $S^1.C^5$ combination maximum FRET. For the

other AS combinations, FRET was expected to show an approximate sigmoidal curvature between $S^1.C^5$ and $S^1.C^1$ combinations, as shown in Figure 99.

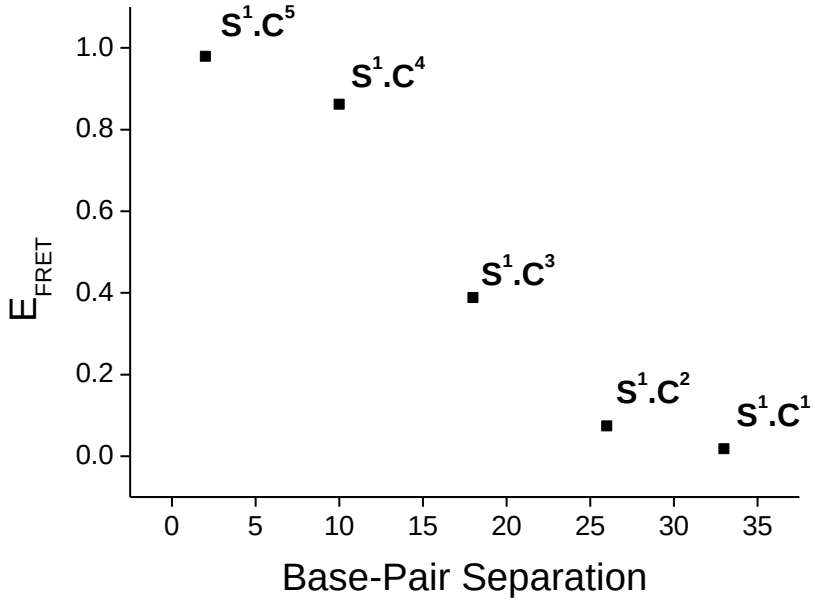


Figure 99 Calculated AS E_{FRET} versus Base-pair separation for FAM and TAMRA FRET pairs.

Firstly, the annealing rates were analysed to identify the extent of intermolecular versus intramolecular interactions, and hence duplex versus hairpin formation. Mfold calculations were performed on S^1 to calculate the expected T_m and thermodynamic stability of the hairpin only conformation[87]. The calculated values were compared to experimental values obtained by UV-Vis and CD melting studies. Experimentation was required to identify these annealing limitations (see fluorescence melting analysis) after which steady-state FRET experimentation was performed and analysed.

5.4 Variable annealing rates

When annealing the different AS combinations, the annealing rates have to be considered; annealing quickly (>> 1 °C / min) can lead to excessive mis-matched duplex fragments and bulges, which will promote the kinetically most stable conformation. Conversely, annealing slowly (<< 1 °C / min) may promote inter-strand complexation; the AS contains large repeating units of ATTA and TAAT segments which will be vulnerable to both quick and slow annealing complications. Variable annealing rate fluorescence analysis was performed to identify the extent of unwanted confirmations, by monitoring the fluorescence emission of FAM and TAMRA modifications as a function of temperature.

Annealing at 5 °C per min for the $S^1.C^5$ combination, below its T_m (53 °C), high donor emission is observed, as shown in Figure 100. This high intensity translates to low FRET efficiency, which corresponds to the fluorophores being held far apart, hence the high donor emission. This result is not expected for the maximum hairpin conformation, which was expected to have fluorophores positioned in close proximity and to undergo maximum FRET. When annealing at rates of 1 °C per min, the curve displays substantially less donor fluorescence below the T_m, and hence high levels of FRET, indicating the fluorophores are in close proximity. Annealing at rates of 0.5 °C per min gives negligible change in emission compared to 1 °C per min, suggesting little homodimerisation when comparing the 1 and 0.5 °C per min annealing rates. These results indicated subsequent annealing at rates of 1-0.5 °C per min were suitable for AS hairpin formation.

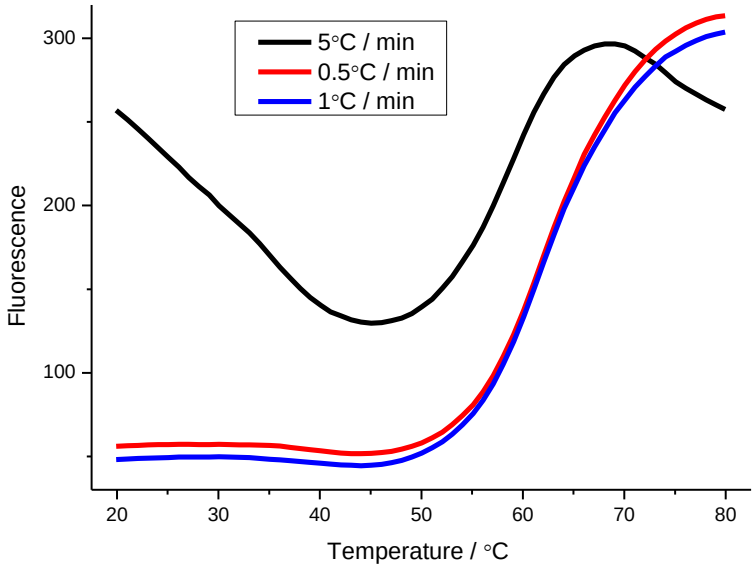


Figure 100 Fluorescence variable melting profile of $S^1.C^5$ FAM and TAMRA version, 0.3 μM in 0.1 M sodium phosphate buffer, excitation at 495 nm, emission at 515 nm, 1 cm quartz cell, 500 PMT voltage.

5.5 AS melting analysis

UV-Vis, CD and fluorescence melting analysis was conducted to monitor the extent of the five AS combinations helix and duplex formation. These three techniques enabled the DNA and fluorophores environment to be determined upon heating.

5.5.1 AS DNA melting analysis

UV and CD melting studies on this DNA absorption region enabled the properties of the DNA structure to be identified. UV melting analysis was conducted to identify the DNA hairpin and duplex character of the AS only combination (S^1). The peak height absorption at 260 nm was monitored as a function of temperature to obtain the melting temperature. Firstly, S^1 was monitored at 260 nm to give (see appendix section 8.32). The UV melting profile displayed typical DNA melting curvature for a duplex and hairpin[10, 104, 106, 108]. Differentiating the melting curve yielded a T_m of 57.4 °C. The extent of hystersis between melting and annealing was negligible,

indicating similar mechanistic pathways for both transitions. The hyporchromicity between the annealed and denatured states gave rise to a 15 % change in UV absorption. Figure 101 displays the UV melting analysis of the AS full duplex ($S^1.C^1$) and the AS full hairpin combinations ($S^1.C^5$). The $S^1.C^1$ combination UV melt differs from the S^1 combination from above; the well defined melting transition is not as obvious with the C^1 complementary strand present. This result indicates different conformations in the sticky terminus region, i.e. the result of single-strand and double-strand effects. The $S^1.C^5$ combination gave rise to a sharp melting temperature of 61.2 °C. CD melting was conducted to further identify these changes since CD is more susceptible to conformational DNA changes.

CD melting profiles for the $S^1.C^1$ and $S^1.C^5$ combinations display clear melting transitions at 227 nm. Differentiation of the $S^1.C^5$ spectra gave rise to a T_m of 61.4 °C, whilst the $S^1.C^1$ resulted in a T_m of 51.3 °C. These differences in T_m are mirrored in the general trends of the CD curvature, for the $S^1.C^5$ combination, the melt gives a sharp change above approximately 50 °C. The $S^1.C^1$ combination, however, exhibits a gradual melting transition above about 30 °C. Comparing the two combinations full hairpin and full duplex, resulted in different melting profiles, corresponding to two different conformations as expected. To further identify these effects, fluorescence analysis of the fluorophores was conducted.

5.5.2 Donor fluorescence melting analysis

Fluorescence melting analysis of the FAM and TAMRA modified AS combinations enabled visualisation of the fluorophore's environment and DNA conformation. Both the donor and acceptor emission, and hence FRET efficiencies were monitored as a function of temperature[10, 104, 106, 108]. Heating and annealing the AS combinations was expected to be a competition between two competing states: duplex and hairpin formation, arising from intramolecular and intermolecular interactions respectively. Fluorescence melting studies was conducted to reveal the conformational state over a temperature range for each of the AS combinations. Taking S^1 labelled with FAM and TAMRA; the donor and acceptor proximities were expected

to be as close as possible, due to the formation of the largest hairpin, and was therefore expected to conduct minimum donor emission/maximum FRET. The $S^1.C^1$ combination was expected to form a full duplex with donor and acceptor on opposite sides. FRET for this combination was expected to be minimal, i.e. maximum donor emission/minimal FRET emission. For the other AS combinations, the donor emission was expected to be in-between these extremes in the respective order: $S^1.C^1 > S^1.C^2 > S^1.C^3 > S^1.C^4 > S^1.C^5$, as reflected in the length of the hairpin moiety.

Observing the fluorescence melting profile in Figure 101, the FAM emission at 515 nm, at 20 °C, the fluorescence emission for the $S^1.C^1$ combination (grey line) is the largest of all the AS combinations, indicating maximum donor emission and hence minimum FRET. The other AS combinations also exhibit donor emission intensities in the order as expected, as in agreement with the design of the structure. For full FRET distance analysis, deconvolution of the donor and acceptor peaks was required.

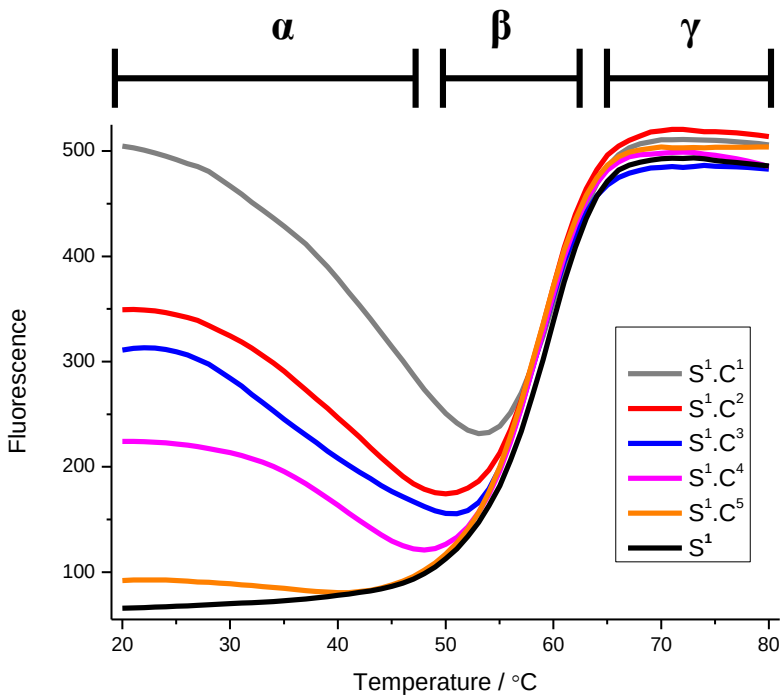


Figure 101 Fluorescence melting profile of the AS combinations donor emission, at 0.3 μM in 0.1 M sodium phosphate buffer, excitation at 495 nm, emission at 510 nm, annealing at 1 °C / min, 1 cm quartz cell, 500 PMT voltage. Melting profile split into three melting regions: α, β and γ corresponding to duplex versus adjustable strap, adjustable strap and single-strand melting regions respectively.

The S^1 strand exhibited a subtle positive change in fluorescence in the α region (between 20 and 45 °C), whereas the other combinations resulted in a negative change. The subtle increase in fluorescence suggests a small proportion of hairpin melting, i.e. donor and acceptor are becoming separated by a greater distance. Further heating unwinds the full hairpin yielding typical sigmoidal curvature for normal melting of duplex DNA in the β region. Differentiation of the curve yields a 60.0 °C T_m. At temperatures approximately higher than the T_m, above 65 °C, in the γ region, the donor fluorescence plateaus to an approximate fluorescence intensity of 500. This linearity suggests donor and acceptor are as far apart as possible, and that only one conformation is adopted at these elevated temperatures, indicating the S^1 combination is in complete single-strand mode. The other AS combinations also exhibit similar fluorescence intensities above 60.0 °C, demonstrating complete melting of all conformations. These intensities of the melted combinations further verifies the annealed $S^1.C^1$ combination at 20 °C, in which donor and acceptor are outside the E_{FRET} distance, i.e. no FRET, indicative of the full duplex conformation as expected.

The α region for the AS combinations gives rise to steady negative changes in fluorescence, except for the S^1 combination which does not contain an intermolecular complementary strand. These negative changes in this region are the result of a competition between hairpin and duplex formation. Towards the higher temperature range of the α region, the largest hairpin formation based on S^1's conformation begins to dominate. Hence, intramolecular interactions dominating over intermolecular interactions, giving rise to the negative change fluorescence. Conversely, towards the lower end of the α region, the desired hairpins and duplex conformations dominate. The α region is very informative, it suggests the intermolecular and intramolecular winding and unwinding of the duplex and hairpins are reversible. However, towards 20 °C, the desired AS conformations form, even though they are thermodynamically less stable than the intramolecular full hairpin conformation. The β and γ regions for the other combinations are very similar to S^1's profile, indicating no conformational deviations in these temperature ranges.

For the $S^1.C^1$ fluorescence melt at temperatures above 20 °C, the donor emission decreases steadily. This change indicates the donor and acceptor are coming closer together. After coming to a fluorescence emission minimum at approximately 53 °C,

the $S^1.C^1$ emission follows typical curvature for the S^1 combination, potentially evidence for intramolecular stabilisation overriding intermolecular stabilisation in temperature the β region. The $S^1.C^1$ T_m is interesting since the expected melting temperature for a 45 mer duplex is expected to be approximately 80 °C, however, 80 °C is expected for random DNA, the AS is A-T rich which reduces the expected melting temperature to approximately 60 °C, calculated using DNA melting temperature calculators. A-T base-pairs are inherently weaker than G-C base pairs, A-T base-pairs contain 2 hydrogen bonds compared to 3 from G-C base-pairs, leading to A-T rich regions melting at lower temperatures [10, 87], hence explaining the relatively low melting temperature for the AS.

5.5.3 Acceptor fluorescence melting analysis

Exciting the acceptor directly at 557 nm and monitoring the emission at 580 nm, displays similar trends to the donor melting from above, as shown in Figure 102. At temperatures greater than 20 °C, intramolecular interactions begin to dominate in the α region. Excitation at 557 nm and monitoring at 580 nm enabled direct observation of the acceptor (TAMRA) fluorophore environment, i.e. no emission arising from FRET. One would therefore expect all combinations to have the same florescence intensities and curvature since the acceptor receives no energy from the donor *via* FRET. Potentially, other interactions can affect the fluorescence below the melting temperature, such as exciton coupling and / or electron transfer. The coupling effect is a result of the relatively close proximity of the donor and acceptor for some of the combinations; excitonic coupling interactions may interfere with donor absorption and emission spectrums, and hence effect the acceptor emission as a function of temperature. For the S^1 strand, the donor and acceptor are in close proximity, excitation at 557 nm only excites the acceptor moiety since the donor does not absorb at 557 nm, and hence emission is only from direct excitation of the acceptor. Due to the donor's close proximity, this could potentially result in excitonic coupling interactions perturbing the fluorescence emission of the acceptor shown above. The $S^1.C^5$ combination has the same curvature above 45 °C, but below this temperature the acceptor fluorescence

increases as a direct result of the additional C^6 complement, placing the FRET pairs in different environments compared to the C^1 only combination. This curvature is in agreement with the donor emission melting curvature from above.

The general fluorescence emission of TAMRA attached to DNA in water decreases upon heating, as observed in the single-strand temperature range in the γ region for all the AS combinations. The S^1 strand forms a stable hairpin in α region, after which a fast melting transition occurs, corresponding to a melting temperature of 57.5 °C. The other AS combinations all have different starting emission intensities, all however, display strong negative changes in fluorescence as the duplexes and hairpin conformations compete in α region, in agreement with the changes observed in the donor melting profile.

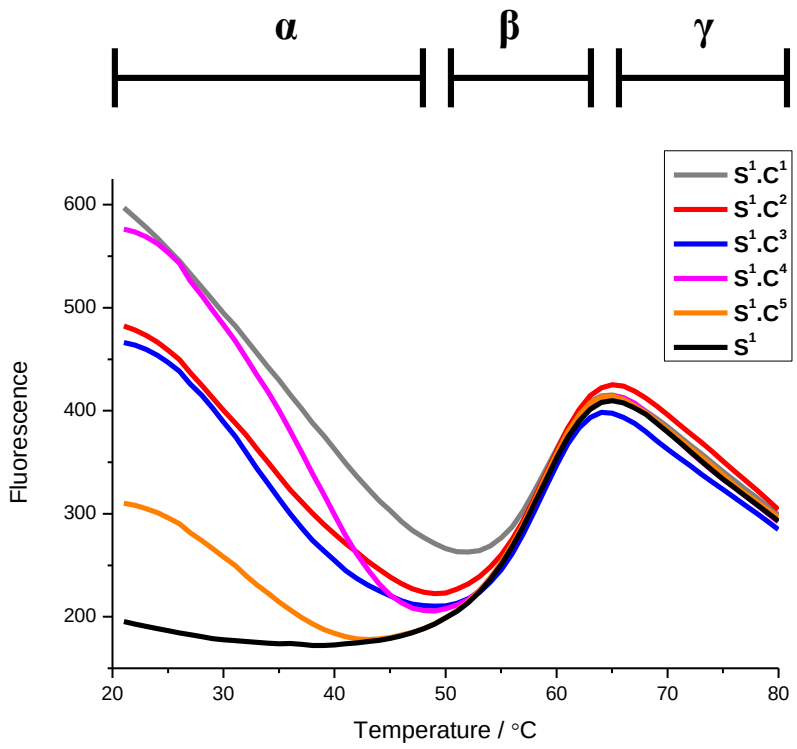


Figure 102 Fluorescence melting profile of the AS combinations acceptor emission at 0.3 μM in 0.1 M sodium phosphate buffer, excitation at 557 nm and emission at 580 nm, annealing at 1 °C / min, 1 cm quartz cell, 500 PMT voltage. Melting profile split into three melting regions: α, β and γ corresponding to duplex versus adjustable strap, adjustable strap and single-strand melting regions respectively.

5.6 FAM and TAMRA steady-state FRET

The steady-state donor emission spectrum for the AS combinations displays emission intensities in agreement to those calculated, as shown in Figure 103. For the longest fluorophore distance separation, i.e. no FRET, the $S^1.C^1$ combination, the donor emission is greatest, hence FRET is minimal. For the smallest fluorophore distance separation, the $S^1.C^5$ combination, the donor emission is minimal, and hence FRET is the greatest. However, to fully identify the E_{FRET} values the donor and acceptor peaks must be deconvoluted due to the overlapping donor and acceptor emission peaks at around 515 nm. Deconvolution of the peaks was conducted according to methods established by Clegg *et al.* (see section 1.6.2).

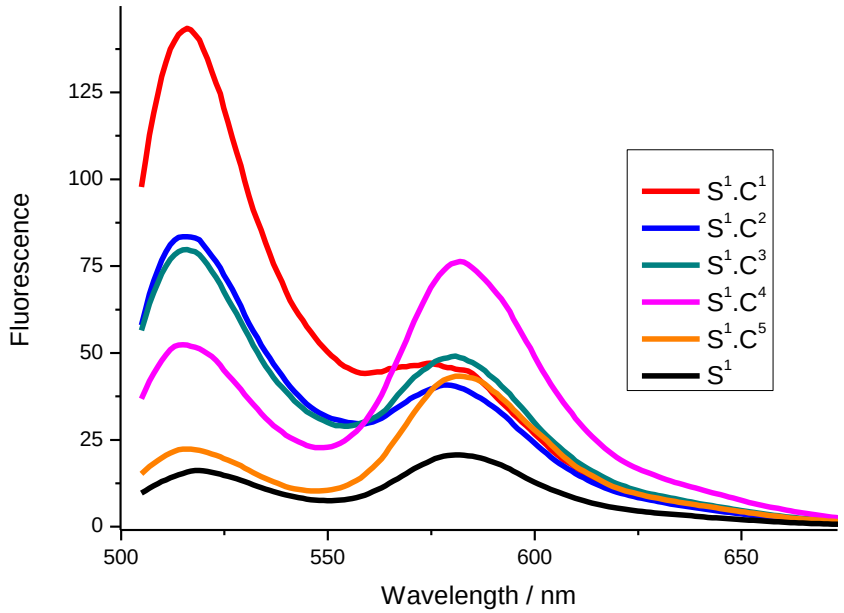


Figure 103 Steady-state fluorescence emission spectrum for AS combinations, 1 μM in 0.1 M sodium phosphate buffer, excitation at 495 nm, 1 cm cell, 500 PMT voltage.

The deconvoluted fluorescence emission of the donor in the five AS combinations, S^1 with C^1, C^2, C^3, C^4 and C^5, the general emission trend is in good correlation with the expected: $S^1.C^1 > S^1.C^2 > S^1.C^3 > S^1.C^4 > S^1.C^5$ (see appendix section 8.33). The AS combinations were normalised with respect to $S^1.C^1$ combination (since this combination has been verified to conduct no FRET from fluorescence

melting analysis). The distance separation of the fluorophores (as calculated by base-pair separation using the 3.6 Å approximation) was plotted versus the E_{FRET} term in Figure 104. The data points are generally in agreement with the calculated values, with the exception of $S^1.C^2$ combination. Observing the $S^1.C^2$ steady-state fluorescence emission, the intensity appears to be reduced from what was expected when comparing the other combinations. The $S^1.C^2$ deviation could potentially be indicative of structural deviation from the desired conformation. The $S^1.C^2$ combination was designed to form the smallest hairpin, comprised of approximately four bases, and therefore was also expected to be the weakest. The effect of the donor modification for the $S^1.C^2$ combination has to be noted; the TAMRA fluorophore may of resulted in increased destabilisation of the surrounding base-pair interactions, related to the modifications terminal positioning, consequently leading to increased terminal fraying of the DNA strands. For the $S^1.C^2$ combination this will result in decreased desired hairpin formation.

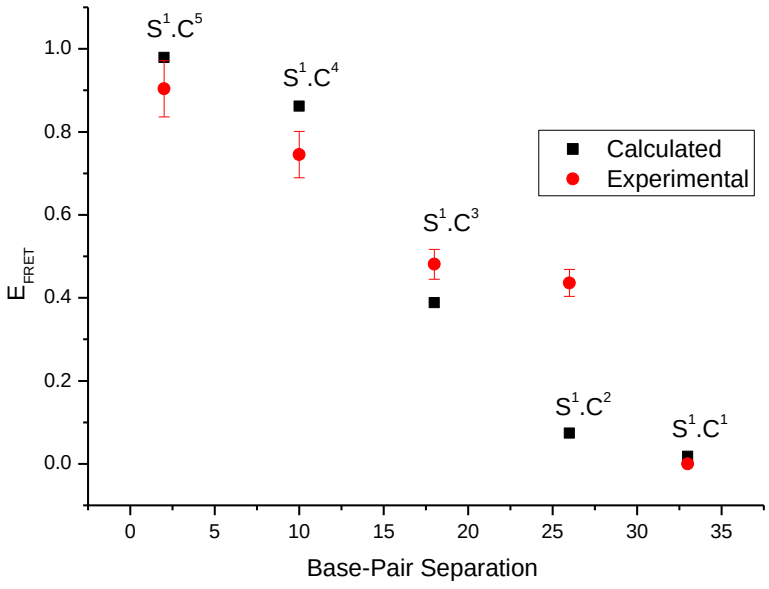


Figure 104 E_{FRET} versus Distance graph for AS FRET combinations from steady-state data obtained from above.

The fluorescence emission intensity of the $S^1.C^2$ combination suggests the fluorophores and hence the hairpin, forms a similar conformation to the $S^1.C^3$ combination, which translates to the formation of a 2 x 1-2 repeat hairpin, instead of a 1 x 1-2 repeat, leading to the formation of a more stable hairpin, as shown in Figure 105. Consequently, the unmodified complement will form either a small 1 x 1-2 repeat hairpin or remain in single-stranded form.

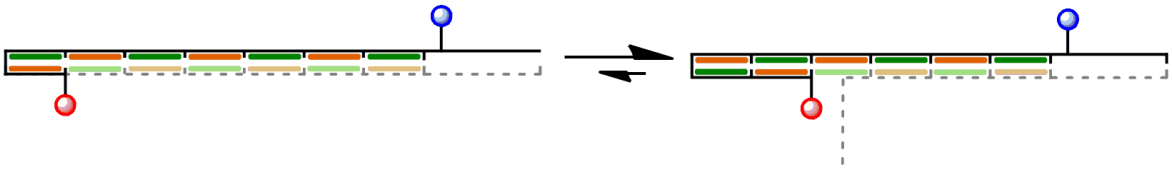


Figure 105 Schematic representation of the potential $S^1.C^2$ combination structural conformation in equilibrium between two states based on fluorescence data obtained from above.

Comparing the fluorescence emission of the S^1 and $S^1.C^5$ AS combinations enabled direct observation of the surrounding nucleobase quenching effect in the fluorophores. The peaks from Figure 103 were deconvoluted using a Lorentzian deconvolution, as shown in Figure 106. For the S^1 strand, the FAM moiety is in a single-stranded environment, whilst for the $S^1.C^5$ combination, the FAM moiety is attached to DNA in a duplex environment. The single-strand format can lead to single-strand-FAM interactions, and hence donor emission quenching[109, 110]. This effect is outlined in the peak height red-shift of donor emission from 515 nm ($S^1.C^5$) to 517 nm (S^1). Comparing the two peak integrals, the S^1 combination is 8 % less intense than the $S^1.C^5$ combination. The red-shift and 8 % decrease in fluorescence are indicative of quenching effects.

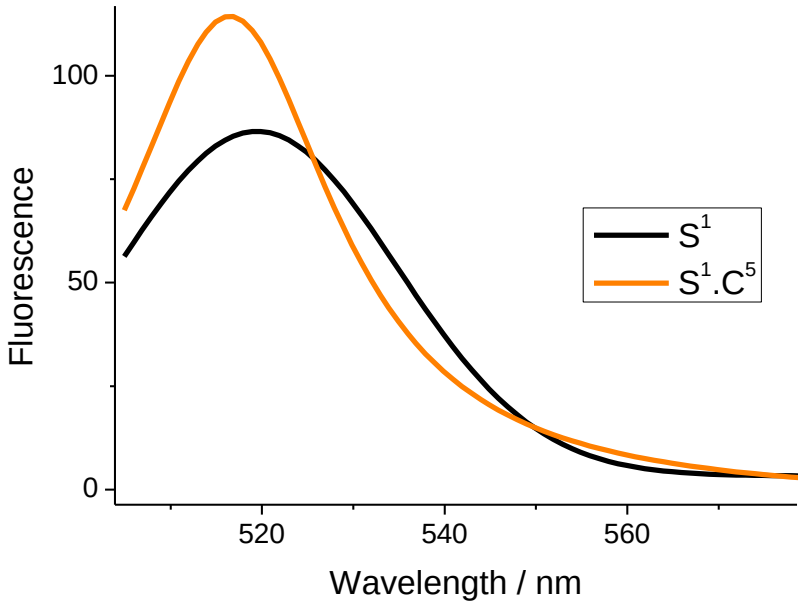


Figure 106 S^1 and $S^1.C^5$ Lorentzian deconvoluted fluorescence of FAM emission spectrums, exciting at 495 nm at 1 μM, in 0.1 M sodium phosphate buffer, 1 cm path length, 500 PMT voltage.

5.7 Summary

The adjustable strap (AS) concept was introduced (section 5.1 and 5.2). Two different functionalised versions of the AS have been synthesised and analysed, a FAM and TAMRA and a zinc and 2H porphyrin versions. FAM and TAMRA are ideally suited for FRET analysis of the AS sequence due to the optimum Förster distance (R_0), which covers nearly the entire length of the full AS sequence. FAM and TAMRA modifications enabled thermodynamic information about the different AS combinations to be identified (section 5.3 and 5.4). Fluorescence melting analysis of the combinations yielded melting curves in the expected order of fluorescence intensity for each combination (section 5.5). This analysis shows the AS design works, four hairpin conformations and one duplex conformation was generated. The calculated steady-state FRET values were largely in agreement with experimental values obtained (section 5.6), with the exception of the $S^1.C^2$ combination. The $S^1.C^2$ combination was reluctant to form the smallest (up to) 8 base hairpin, instead this combination mimics the $S^1.C^3$ combination, to generate a 16 base hairpin. Steady-state fluorescence analysis enabled some fundamental questions to be asked and answered directly, for example, how does a single-strand environment effect a fluorophore. By comparing the fluorescence emission of the S^1 and $S^1.C^5$ combinations this question was answered. For S^1, the donor (FAM) is in a single-strand environment, whilst in the $S^1.C^5$ format, the donor moiety is in a double strand environment. This environmental change resulted in a 2 nm blue-shift and an 8 % increase in fluorescence emission.

An iodinated or metallated porphyrin version of the AS would be ideal for SAXS studies, to potentially identify the metal-to-metal scattering, analogous to FRET results from above. The major obstacle for small-angle scattering application is the relatively high concentrations required (typically > 100 μM), leading to problems associated when annealing, for example, inter-strand complexation. This could be overcome by annealing as dilute solutions, followed by gentle concentration using spin-filtration techniques.

Terpyridine DNA arrays†

† This chapter was published in the journal *Molecules*, by Burns and co-workers[111]

6.1 Introduction

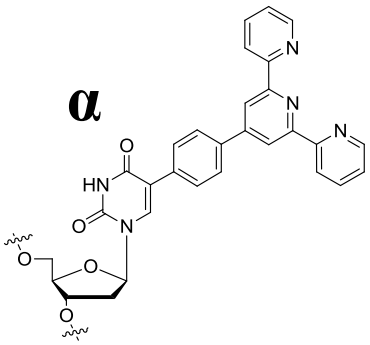
A terpyridine monomer was synthesised and incorporated into DNA in an attempt to generate highly ordered two and three dimensional DNA arrays. Functionalised DNA nanotubes were created using orthogonal hydrogen bonding and metal complexation interactions. Short terpyridine labelled DNA strands where designed to self-assemble into long duplexes by the use of overlapping DNA regions. Adding weakly terpyridine binding metals such as zinc and nickel, induced the formation of inter-strand metal complexes to generate DNA tubular arrays 50-200 nm wide and 2-50 nm high.

Seeman[112, 113, 114], Turberfield[115, 116, 117, 118], LaBean[119, 120], Rothemund[121, 122], Yan[123, 124], Mirkin[125, 126] and others[127, 128, 129, 130, 131, 132, 133] have utilised DNA to generate novel self-assembled nanostructures. Structures are now being developed to carry out specific tasks, to behave like nanomachines. To obtain a better understanding and to utilise these higher-order properties to design new functional materials based on biologically derived templates, reliable construction methods are required. DNA is an ideal scaffold from which supra-molecular structures can be built; DNA forms highly predictability, specific and reliability structures[57, 134, 135, 136, 137]. Controlled formation of nanotubes from polymeric DNA duplexes containing a terpyridine-deoxyuridine modification at specific positions is presented. To minimise DNA duplex disruption, the terpyridine modifications were incorporated into the major groove of B-type DNA. This system makes use of orthogonal binding modes through selective metal recognition together with Watson-Crick base pairing of 'sticky ends' of the DNA. Firstly, the terpyridine monomer synthesis is discussed, including the terpyridine formation mechanism (section 6.2), with the remaining sections displaying terpyridine array analysis, including UV-Vis, CD, TEM and AFM (section 6.3 and section 6.3).

6.2 Terpyridine monomer and DNA sequences

Terpyridine or phenanthroline modified DNA are useful maintain DNA duplex stability through metal complexation between two complementary DNA strands[138, 139, 140], or for the formation of specific supramolecular DNA structures[141, 142]. A bisNTA-Ni(II)-His_6 complex was recently used to create DNA arrays on surfaces[143]. In our system, the terpyridine was linked to the deoxyuridine nucleoside *via* a short rigid linker, which allows precise positioning of the metal complex, see Figure 107 (α). The modification is placed at every tenth position in both DNA single strands, this will lead to a modification at every fifth position in the annealed duplex, see Figure 107 (β and γ). Therefore, the terpyridines are evenly placed along the helical DNA, at approximately half and full helical turns, to maintain orthogonal terpyridine and DNA alignments.

α



β

```
2     5' CAT TGT GAG GCA TAC GTA TC 3'
3     5' CCT CAC AAT GGA TAC GTA TG 3'

2t    5' CAX TGT GAG GCA XAC GTA TC 3'
3t    5' CCX CAC AAT GGA XAC GTA TG 3'
```

γ

```
5' -- CAX TGT GAG GCA XAC GTA TCC ATT GTG AGG -- 3'
3' -- GTA ACA CTC CGT ATG CAX AGG TAA CAC XCC -- 5'
```

Figure 107 Terpyridine deoxyuridine modification (α), DNA sequences (β) where X = T for DNA strands: 2 and 3, and X = terpyridine deoxyuridine modification, and annealed poly-DNA strands (γ, grey represents repeat sequences).

6.3 Terpyridine monomer synthesis

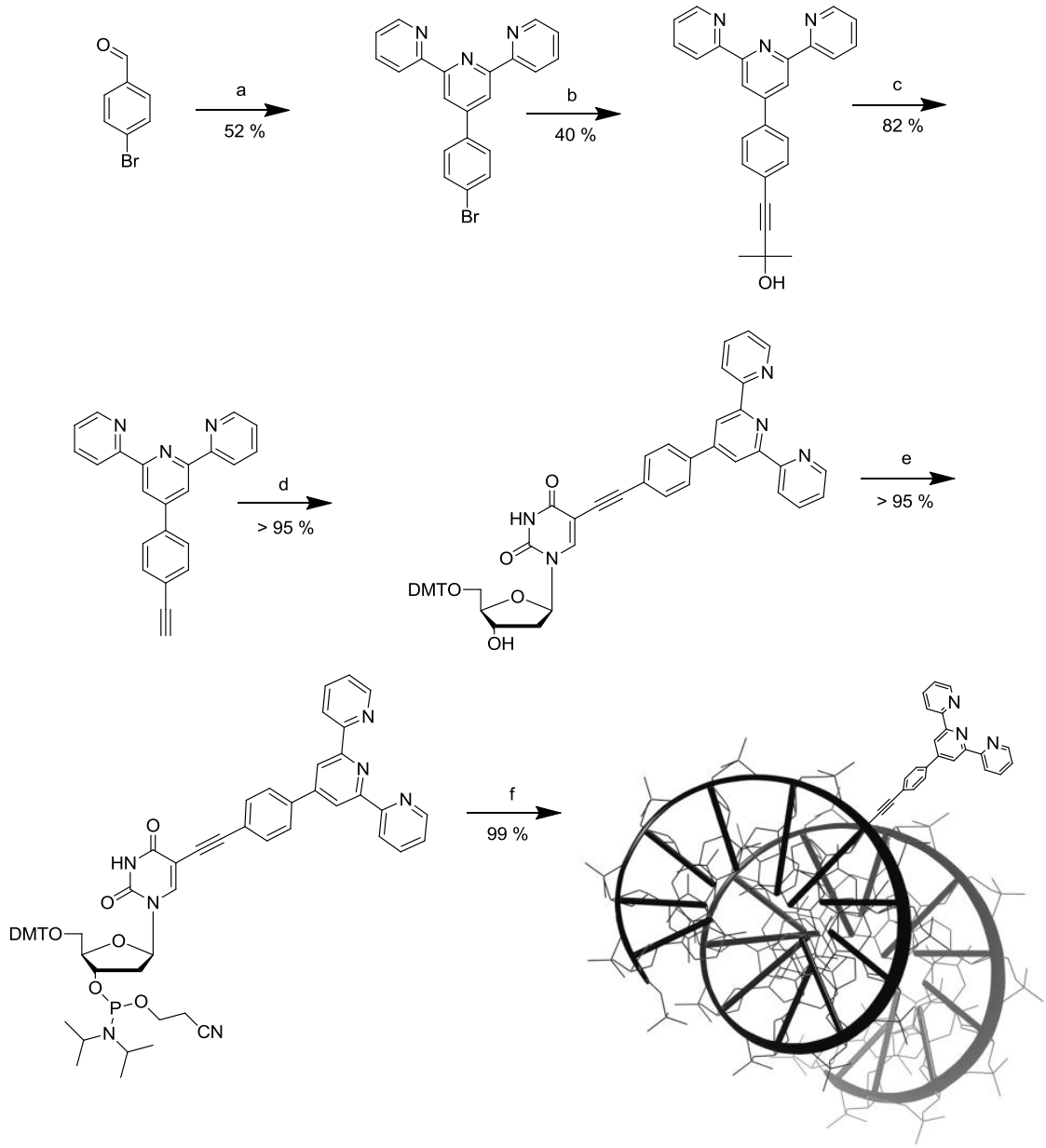


Figure 108 Terpyridine phenyl acetylene DNA synthesis [87].

a KOH, NH_3, 2-Acetyl-pyridine, Ethanol

b Hydroxy-dimethyl-acetylene, CuI, PPh_3, DMF, $Pd(PPh_3)_4$, 90 °C, DME/H_2O

c NaOMe, Toluene, 90 °C

d DMT-iodo-dU, CuI, PPh_3, DMF, $Pd(PPh_3)_4$, 90 °C, DME/H_2O

e DCM, DIPEA, CEP-Cl

f Solid phase DNA synthesis

The terpyridine monomer was synthesised using modified protocols established from Rühl and co-workers, as shown in Figure 108 (see appendix section for synthetic protocol, sections 8.9-8.12)[88].

6.3.1 Terpyridine DNA synthesis - Step a

Bromo-phenyl terpyridine was synthesised using ammonia, potassium hydroxide, 2-acetyl-pyridine and hydroxy-dimethyl-acetylene benzaldehyde (reaction a). The reaction proceeds *via* a complex mechanism under equilibrium; ammonia converts the ketone from 2-acetyl-pyridine to an enamine *via* an amine tautomerisation, with the loss of a water molecule, Figure 109. The primary amine formed repeats this enamine formation with a second 2-acetyl-pyridine molecule. The di-enamie complex undergoes an electrophilic addition to the aldehyde, leading to the evolution of another water molecule. Finally the loss of two protons fully aromatises and flattens the three linked pyridines, producing the terpyridine structure. The terpyridine precipitates in the reaction vessel forming a thick substance. The product was purified by washing and re-crystallisation in hot ethanol.

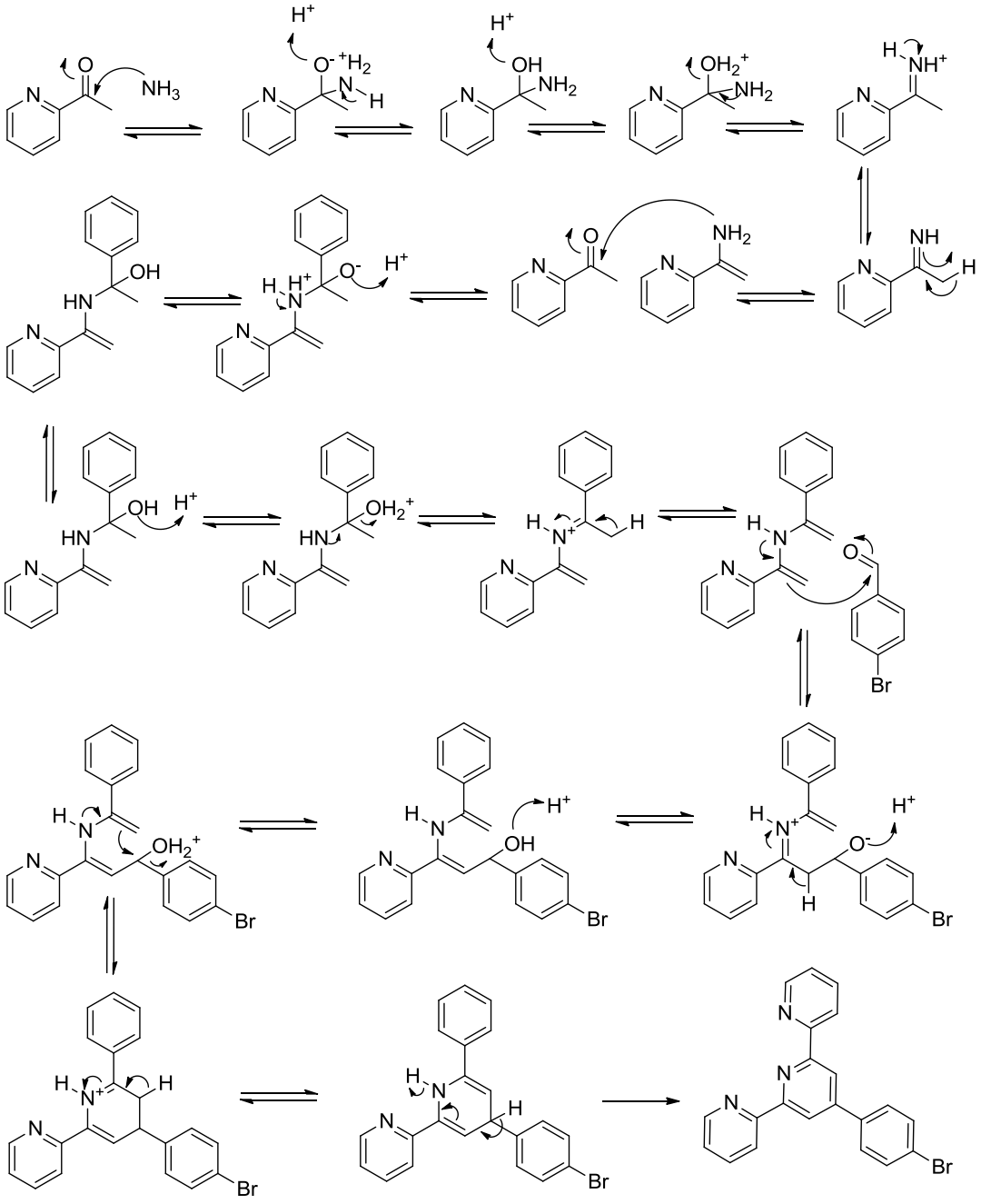


Figure 109 Terpyridine mechanism, from 2-acetyl-pyridine, ammonia, potassium hydroxide in ethanol.

6.3.2 Terpyridine DNA synthesis - Step b

Hydroxy-dimethyl acetylene terpyridine was synthesised using the well-established Sonogashira palladium catalysed cross-coupling reaction (reaction b). Bromo-phenyl terpyridine was cross-coupled to hydroxy-dimethyl acetylene at 90 °C under N_2 and in the dark, in a water / ethylene glycol solvent, the solvent ratio was measured as accurately as possible for complete solvation of all reagents at high temperatures. The acetylene reagent was added in 0.1 equivalents over 4 hours to prevent saturation of the triethylamine catalyst.

6.3.3 Terpyridine DNA synthesis - Step c

Deprotection of the acetylene protecting group was achieved using excess NaOH in dry toluene, under N_2 whilst heating to reflux (reaction c). The reaction went to completion in 30 minutes, as monitored by TLC. Terpyridines stain brown with ferric chloride *via* formation of a iron-terpyridine complex[88]. The product was purified with column chromatography with silica gel using ethyl acetate pre-washed with ammonia (prevents extensive streaking as a result of strong silica-terpyridine interactions).

6.3.4 Terpyridine DNA synthesis - Step d

The Rühl *et al.* protocol for the terpyridine monomer Sonogashira coupling was followed multiple times; however, this method produced extremely poor yields (less than 2 %), most probably due to extensive terpyridine-palladium, terpyridine-copper or terpyridine-iodine complexes[88]. These complexes distort the relative reagent concentration required for efficient and desired cross-coupling reactions. In conclusion, the same conditions used for the first Sonogashira coupling were applied from above. Whereby heating to reflux prevented significant terpyridine-metal complexation and resulted in to near-quantitative yields. The solution was washed with EDTA solution at

90 °C to prevent terpyridine-metal complexation upon cooling. The monomer was purified using silica chromatography in ethyl acetate pre-washed with ammonia.

6.3.5 Terpyridine DNA synthesis - Step e

Terpyridine DNA was synthesised and purified by Dr. Thomas Rühl. DNA was conducted on a 1 µmol scale using solid phase supports. The relative terpyridine modified deoxyuridine phosphoramidite concentration and coupling times were doubled to allow for the reduced coupling efficiencies of the modified nucleobases. The terpyridine modified DNA was purified using HPLC. Figure 110 displays the UV-Vis and fluorescence spectra of terpyridine modified DNA (exciting at 330 nm).

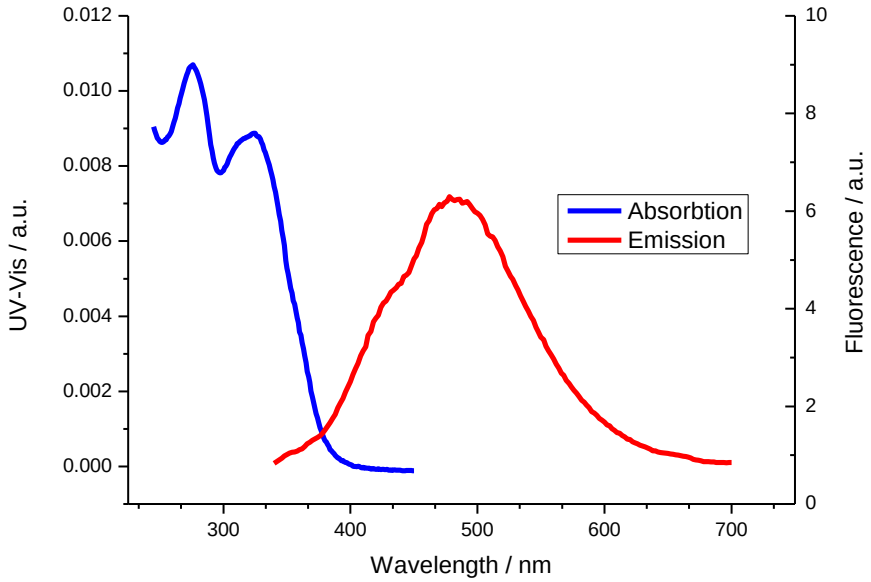


Figure 110 UV-Vis and fluorescence spectra of duplex form terpyridine phenyl acetylene Reduced DNA, absorption peaks at 260 and 330 nm (blue line, 260 nm coincides with DNA peak). Excitation at 330 nm yields the fluorescence emission at 495 nm (red line).

6.4 Spectroscopic analysis of terpyridine arrays

The terpyridine labelled DNA strands 2t and 3t display the characteristic terpyridine absorbance at around 330 nm, λ_{max} at 325 nm (Figure 111 (α)). The thermal denaturing of the unmodified DNA duplex 2.3 gives rise to a T_m of 45.4 ºC; however, the curvature of the lower temperature region indicates cooperative denaturing due to the overlapping DNA interactions (Figure 111 (β)). Surprisingly, the terpyridine modifications have a stabilising effect on the DNA duplex, rather than an expected destabilising effect; 2t.3t shows a T_m of 56.3 ºC. This T_m is also closer to the T_m calculated for the full complementary sequence without overlapping regions (61.3 ºC calculated using mfold)[87]. For the 2t.3t combination, the cooperativity is even more pronounced, and a 2 ºC hystersis indicates different kinetics upon melting and annealing the labelled strands. The extent of hyperchromicity for the 2t.3t combination is significantly lower than for 2.3 combination. The nature of the terpyridine effect is unclear; the CD spectra of both the labelled and unlabelled DNA combinations are very similar in shape (Figure 111 (γ)), therefore B-type DNA structure at ambient temperatures can be assumed for both versions[42]. The terpyridine moiety gives rise to a very weak and broad induced CD signal. The intensities of the terpyridine DNA 2t.3t peaks are lower than those of the unmodified DNA, in particular for the peak at 244 nm, therefore, the terpyridine unit does not have an significant effect on the local structure of the DNA.

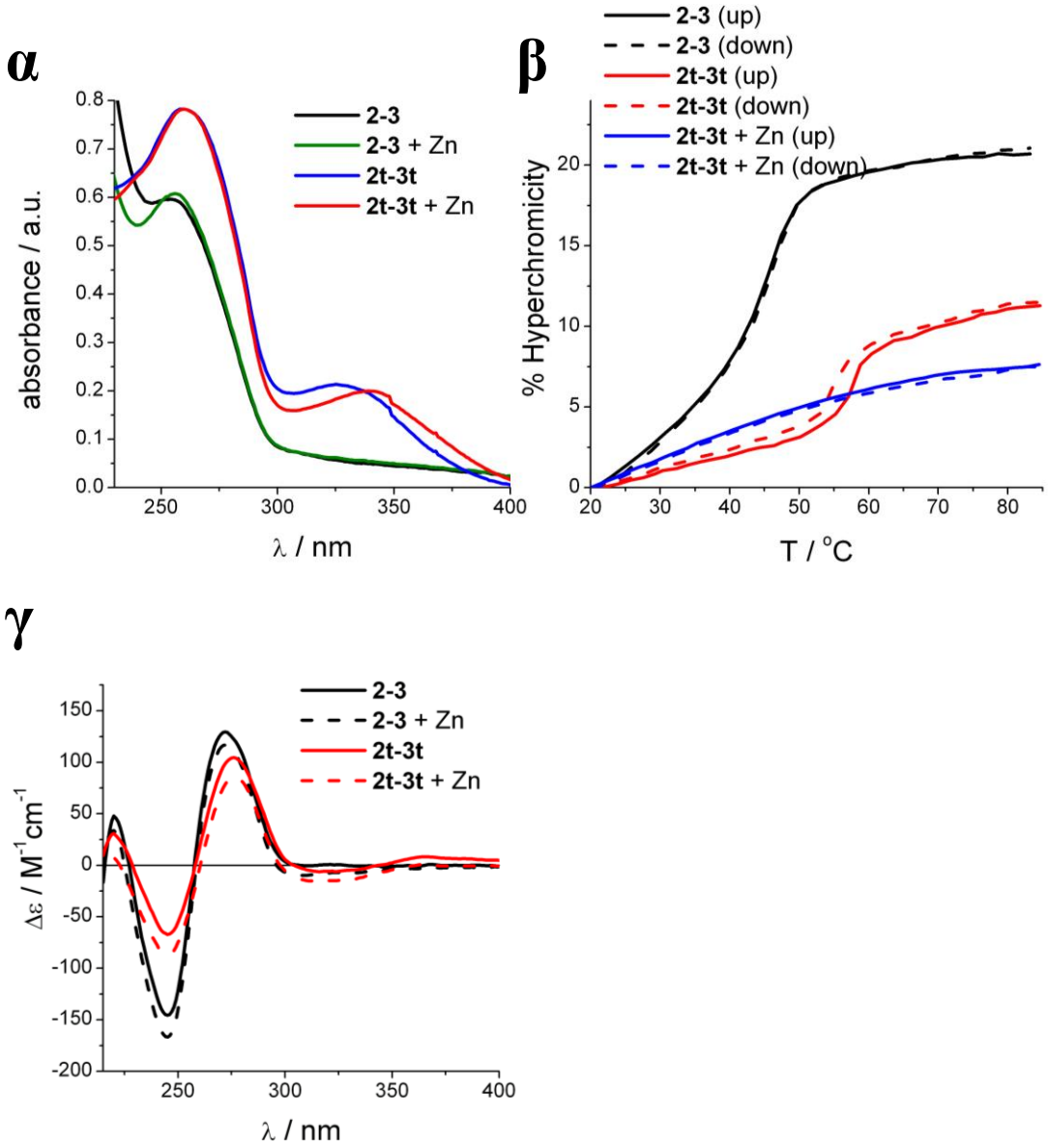


Figure 111 UV-Vis spectra of 2.3 and 2t.3t in the presence an absence of zinc(II) (α), thermal denaturing of the DNA strands 2.3, 2t.3t and of 2t.3t in the presence of 0.5 eq zinc(II) (β), CD spectra of 2.3 and 2t.3t, both in the presence and absence of zinc(II) (γ). Conditions for the measurements: 2 μM DNA in 100 mM NaCl, 50 mM phosphate buffer pH 7.0, spectra taken from Burns *et al.*[111].

If 0.5 equivalents of $ZnCl_2$ per terpyridine moiety are added to solution, the melting behaviour of the terpyridine DNA 2t.3t changes significantly (Figure 111 (α and β)); the UV-Vis spectrum of 2t.3t shows a distinctive change in the terpyridine absorbance region, a bathochromic shift of 24 nm to 339 nm was observed. The pronounced melting curve transition of 2t.3t is absent with the addition of $ZnCl_2$, only a

steady increase in absorbance was observed. These results are indicative of the formation of intermolecular terpyridine metal complexes, to induce the formation of larger arrays through non-covalent connections of the DNA strands. This effect is likely to lead to a highly cooperative denaturing process, the T_m of which is outside the experimental window (> 85 °C). The change in the CD spectrum is almost completely suppressed; therefore, it can be assumed that all terpyridine units form zinc *bis*-terpyridine complexes. Interestingly, both 2.3 and 2t.3t combinations show similar changes in the DNA region of the CD spectrum; the addition of zinc reduces the peak at 275 nm by 10 % for 2.3 and 18 % for 2t.3t, whilst the peak at 245 nm increases by 15 % for 2.3 and 33 % for 2t.3t. Therefore, the addition of metal cation does affect the overall structure of the duplexes, however, this effect is more pronounced for the terpyridine labelled DNA strands.

6.5 TEM and AFM analysis

Transmission Electron Microscopy (TEM, conducted in conjunction with Dr. Dorota Bartczak) and Atomic Force Microscopy (AFM, conducted in conjunction with Dr. Jurgita Zekonyte, see appendix section 8.34) techniques were employed to investigate the arrays formation via metal complexation. The terpyridine labelled and unlabelled DNA samples were slowly annealed overnight in the presence or absence of metal ions. For TEM analyses, the samples were dried on a copper grid and stained using uranyl acetate and lead citrate. Both the 2.3 and 2t.3t with out zinc display relatively unordered organic matter (Figure 122); this does not change for 2.3 upon addition of $ZnCl_2$. Therefore, the interactions observed in the CD spectra do not have a long range effect on the structure. However, the terpyridine DNA duplexes 2t.3t with $ZnCl_2$ exhibit a contrasting image. Complexation of the terpyridine groups with the cationic metal ions leads to a long range into fibre like structures. Higher magnification of 2t.3t with zinc shows that the fibres seem to consist of arrays of small globular structures ranging from 2 to 5 nm.

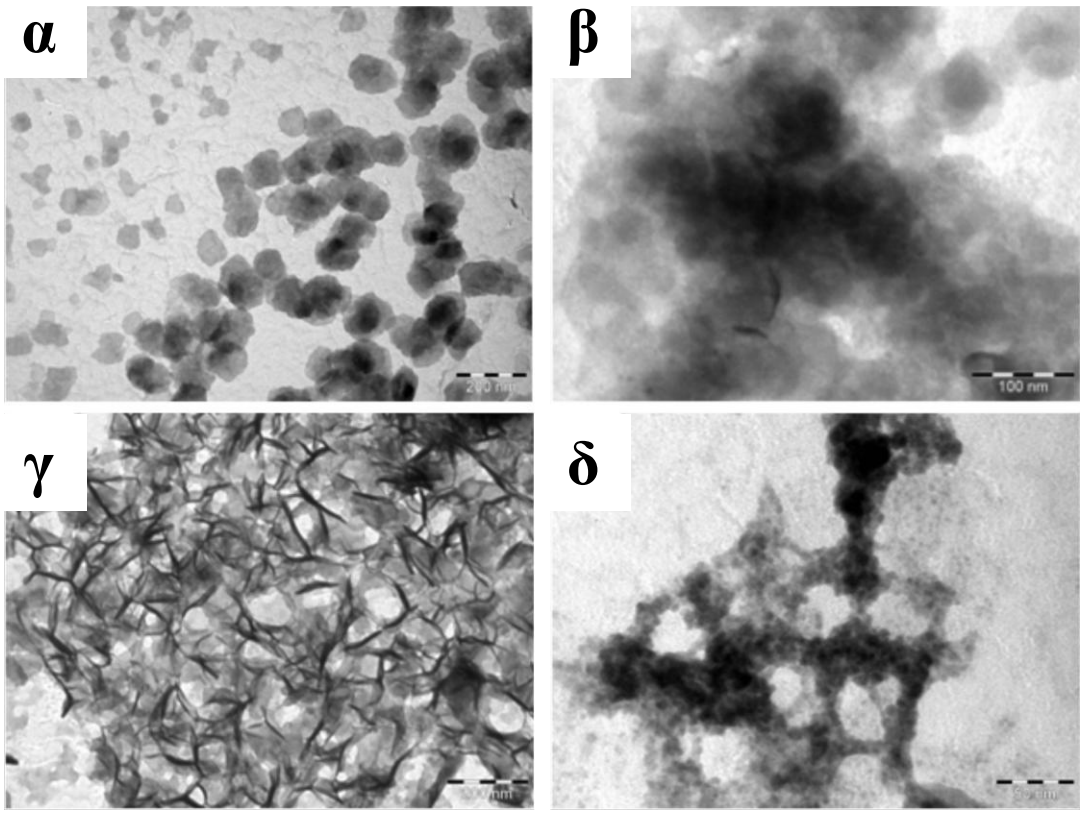


Figure 112 TEM images of the DNA arrays obtained after annealing and staining, in the presence or absence of $ZnCl_2$ 2.3 (α, scale bar: 200 nm) 2t.3t (β, scale bar: 100 nm) 2t.3t with $ZnCl_2$ (γ scale bar: 200 nm), 2t.3t with $ZnCl_2$ zoomed in (δ scale bar: 50 nm), images taken from Burns *et al.*[111].

AFM was applied to investigate the terpyridine DNA array formation in greater detail and in the solution phase (see appendix section 8.34). The samples in buffer were deposited onto a freshly cleaved mica surface for two hours before analysis was performed. In contrast to was expected, but in agreement with the TEM pictures in Figure 112, both the terpyridine labelled and unlabelled DNA systems do not display long strands as a result of the overlapping DNA sticky ends. Rather the DNA forms globular assemblies of variable size, which range from about 10 to 100 nm. This result indicates that the overlapping DNA regions are not stable enough to assemble larger arrays on their own. The AFM pictures of the 2t.3t with $ZnCl_2$ reveals tubular structures of 2-5 μM length. The height and width of these assemblies varies between 30-50 nm and 100-300 nm, respectively. Analysing the terpyridine DNA structure using Maestro (Figure 113) shows that the terpyridine units are not aligned at 180°

(when looking down the helix); the angle between the terpyridines is actually around 135°.

6.6 Summary

Terpyridine labelled DNA can selectively form polymeric tubular arrays using orthogonal binding modes; overlapping DNA strands and metal *bis*-terpyridine complexes. The terpyridines were attached to the DNA using a rigid acetylene linker, the terpyridine unit can react with cationic metals to form non-covalent interstrand DNA linkages. The arrays generated form long tubular assemblies rather than planar arrays. The terpyridine moiety is attached rigidly to the DNA, which seems to guide the 3D structure towards the observed tubular arrays to represent the thermodynamically most favourable arrangement, obtained after slow annealing, as shown schematically in Figure 113.

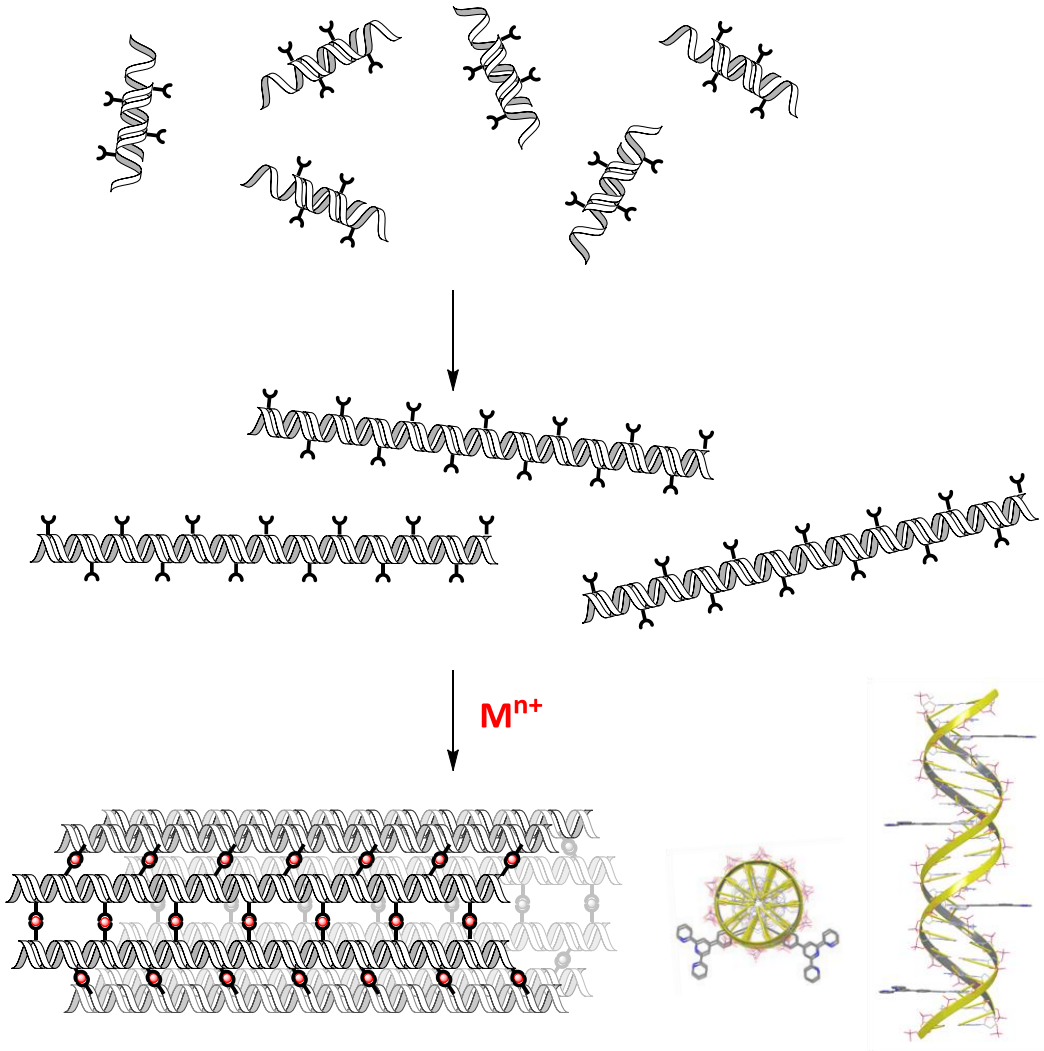


Figure 113 Putative representation of the annealing of long DNA strands through sticky-end recognition and formation of tubular assemblies by metal complexation, images taken from Burns *et al.*[111].

Conclusions and Outlook

7.1 Conclusions

A highly accurate and sensitive porphyrin molecular ruler was generated. Porphyrin monomers were synthesised and incorporated into DNA, their molecular ruler properties were calibrated and applied to analyse a protein-DNA complex. High precision was achieved by rigidly attaching the porphyrin moieties to DNA via an acetylene linker. The porphyrin molecular ruler was calibrated on a 21-mer *TerB* DNA sequence; a zinc porphyrin moiety was incorporated into one position on one strand of the calibration sequence, and seven 2H porphyrins were incorporated into seven positions on seven complementary strands, in total giving rise to seven different zinc and 2H porphyrin combinations. Steady-state fluorescence and circular dichroism spectroscopic analysis showed the different labelled DNA combinations exhibit different characteristics depending on the distance and angles between the intramolecular porphyrins. Individual base-pair resolution was observed for both techniques. The FRET experimental results were compared to two theoretical models, one of which accounted for a rigid linker, and the other did not. The experimental results were in good agreement to the rigid model.

After calibration, the ruler was applied to investigate a protein-DNA complex, the Tus-*Ter* complex. The aim was to identify information about the binding mechanism and any protein induced conformational change on the DNA molecular ruler. As a control experiment, a FAM and TAMRA molecular ruler was synthesised in parallel. This control assay enabled comparison of the rigid and flexible linker effect on the molecular ruler. Twelve FAM and TAMRA combinations were incorporated into seven elongated *Ter* DNA sequences. Steady-state fluorescence spectroscopic analysis showed the averaged κ^2 model for predicting the FRET efficiency was accurate over longer distances. A porphyrin CD ruler was applied to study the non-permissive DNA region of the Tus-*Ter* complex. For protein binding, the non-permissive region, even though there is no direct contact, is highly sequence dependant, indicating potential protein-non-permissive DNA interaction. Soret band analysis of the porphyrins showed the non-permissive DNA region did not undergo significant protein induced conformational change, indicating the non-permissive region maintains its conformation. Due to the relatively low distance dependency of the zinc and 2H porphyrin molecular

ruler, a FAM and TAMRA molecular ruler was applied to identify potential DNA-protein wrapping to further increase binding affinity. Steady-state fluorescence analysis showed the flanking regions of the main binding site does not wrap around the protein to a detectable extent.

A new DNA based molecular device was designed and developed; the adjustable strap (AS). The AS is a 45-mer DNA single-strand which contains a poly-repeating eight base sequence and a non-repeating region. Four hairpins and one duplex conformation can be formed selectively from the AS sequence. The desired conformation was selectively generated upon addition of the appropriate complementary strand. FAM and TAMRA FRET pairs were incorporated into the extremes of AS sequence and used to characterise the AS conformations. UV-Vis and fluorescence melting analysis showed the desired conformations were generated for all of the AS conformations, except for the smallest hairpin conformation.

Finally, terpyridine moieties were incorporated into a polymer DNA duplex. The polymer was generated using overlapping DNA sequences, which annealed to form a duplex polymer in one dimension. With the incorporation of terpyridines at ten base intervals, and applying zinc chloride, an intermolecular zinc bis-terpyridine complex was generated. Multiple complexes formed to produce higher ordered structures; tubular arrays were generated ranging in diameter from 50-200 nm wide and 2-20 nm high.

7.2 Outlook

The use of porphyrins as molecular ruler probes should continue; individual base-pair resolution was achieved, a high target for molecular ruler resolution. Two versatile types of molecular ruler analysis was possible using bench top instruments; after the porphyrin molecular ruler was prepared, UV-Vis, steady-state fluorescence and CD analysis was conducted, without even changing the cell. Meaning all this molecular ruler information was obtained within minutes. This analysis can even co-inside with the additional information gathered in the DNA and protein UV-Vis and CD regions, to ultimately provide a vast wealth of complementary information very quickly. To fully characterise the porphyrin zinc to 2H energy transfer process in detail, time-resolved and single-molecule fluorescence analysis should be employed. This can be compared to literature values where experimentation was conducted in supramolecular arrays in organic solvents. To expand the CD molecular ruler aspects of the porphyrin molecular ruler, the exciton coupling effect should be modelled and calculated. This will enable distances and angles to be extracted in an analogous approach to the FRET method.

The porphyrin molecular ruler also enabled the identification of duplex DNA fraying; by placing a moiety towards the end of the duplex, the porphyrin clearly exhibited different spectroscopic properties, as observed in UV-Vis, fluorescence and CD spectroscopy. The porphyrins can therefore be used to identify single-strand DNA environments as well as duplex environments. To further develop the FRET porphyrin ruler, using FRET pairs with greater spectroscopic differences will ease FRET analysis (and also increase the Förster distance provided there is greater spectral overlap). This will give a greater spread of FRET results and therefore distances covered, for example, applying rigid FRET pairs: terpyridine and 2H porphyrin; both are rigidly attached to DNA via the acetylene bond, they have excellent spectral overlap, and the quantum yield of the terpyridine is high. Therefore, terpyridine and 2H porphyrin FRET pairs look like ideal candidates to further the highly accurate molecular ruler field.

The porphyrins being aromatic and having large and flat π systems can undergo unwanted intermolecular π-π interactions at high concentrations. For example, SAXS analysis of the zinc porphyrin molecular ruler, and MALDI-ToF attempts resulted in porphyrin aggregation. To overcome this concentration dependency, potentially

increasing the number of lone pairs and therefore polar groups on the porphyrin entity will aid water solubility. For instance, introducing carboxylic acids groups to the phenyl positions will disrupt π stacking interactions, but also co-ordinate with water molecules to increase solubility.

The adjustable strap molecular device can be investigated further; other interesting experiments may include competition experiments; for example using a preformed $S^1.C^5$ combination, and adding excess C^5 complement, the result should give a large change in the fluorescence emission of both donor and acceptor upon conformational change. Another experiment could involve changing the 'sticky' terminus of the AS sequence to, for example, the *TerB* 21-mer. Addition of Tus (or F-Tus) in solution will attach to its binding region of this modified AS sequence, which will enable further binding, functionalisation and energy transfer mechanisms to be observed.

The field of DNA supramolecular structures (DNA origami) is growing rapidly; introducing more functional groups to the structures will increase their applications and diversity. The terpyridine DNA arrays can be expanded and applied to generate new structures. Different metals can be applied to generate different metal *bis*-terpyridine complexes. These arrays could even be incorporated with the porphyrin molecular ruler to create a molecular wire embedded on a terpyridine array surface.

References

8.1 References

1. Stryer, L.; Haugland, R.P.: *Proceedings of the National Academy of Sciences* **1967**, 58 (2), 719-726.
2. Lewis, F.D.; Zhang, L.; Liu, X.; Zuo, X.; Tiede, D. M.; Long, H.; Schatz, G. C.: *Journal of the American Chemical Society* **2005**, 127 (41), 14445-14453.
3. Iqbal, A.; Arslan, S.; Okumus, B.; Wilson, T. J.; Giraud, G.; Norman, D. G.; Ha, T.; Lilley, D.M.J.: *Proceedings of the National Academy of Sciences* **2008**, 105 (32), 11176-11181.
4. Sönnichsen, C.; Reinhard, B.M..; Liphardt, J.; Alivisatos, P.A.: *Nature biotechnology*, **2005**, 23, 741-745.
5. Lakowicz, J.R.: *Principles of Fluorescence Spectroscopy*, Springer **2006.**
6. Kamada, K.; Horiuchi, T.; Ohsumi, K.; Shimamoto, N.; Morikawa, K.: *Nature* **1996**, 383 (6601), 598-603.
7. Dragan, A.I.; Privalov, P.L.: *Methods in Enzymology*, Academic Press: **2008**, 450, 185-199.
8. Jin, C.; Marsden, I.; Chen, X.; Liao, X..: *Journal of molecular biology* **1999**, 289, 683-90.
9. Maye, M.M.; Kumara, M. T.; Nykypanchuk, D.; Sherman, W.B.; Gang, O.: *Nature nanotechnology* **2010**, 5, 116-20.
10. Williams. D.M: *Nucleic Acids in Chemistry and Biology*, Royal Society of Chemistry, **2006**.
11. Rothemund, P.W.K.: *Nature* **2006**, 440, 297-302.
12. Boïjesson, K.; Preus, S.; El-Sagheer, A.H.; Brown, T.; Albinsson, B.; Wilhelmsson, L.M.: *Journal of the American Chemical Society* **2009**, 131 (12), 4288-4293.
13. Hurley, D.J.; Tor, Y.: *Journal of the American Chemical Society* **2002**, 124, 13231-41.
14. Stryer, L.; Berg, J.M.; Tymoczko, J.L.: *Biochemistry: International Edition*. W. H. Freeman **2006**.
15. Crick, F.; Watson, J.: *Nature*, **1953**, 171, 737.

16. Clayden J.; Greeves N.; Warren S.; Wothers P.: *Organic Chemistry*, Oxford **2001**.
17. Atkins, P.; Paula. J.D.: *Physical Chemistry*, Oxford University Press, 9th edition, **2009**.
18. Förster, T.: *Naturwissenschaften*, **1946**, 33 (6), 166-175.
19. Förster, T.: *Annalen der Physik*, **1948**, 437 (1-2), 55-75.
20. Darby, R. A.J.; Sollogoub, M.; McKeen, C.; Brown, L.; Risitano, A.; Brown, N.; Barton, C.; Brown, T.; Fox, K. R.: *Nucleic Acids Research*, **2002**, 30 (9), e39.
21. Wang, K.; Tang, Z.; Yang, C. J.; Kim, Y.; Fang, X.; Li, W.; Wu, Y.; Medley, C. D.; Cao, Z.; Li, J.; Colon, P.; Lin, H.; Tan, W.: *Angewandte Chemie International Edition*, **2009**, 48 (5): 856-870.
22. Clegg, R.M.; Murchie, A.I.; Zechel, A; Carlberg, C.; Diekmann, S.; Lilley, D.M.: *Biochemistry* **1992**, 31, 4846-56.
23 Clegg, R.M.; Murchie, A.I.; Zechel, A; Lilley, D. M.: *Proceedings of the National Academy of Sciences*, **1993**, 90, 2994-8.
24. Clegg, R.M.; Murchie, A.I.; Lilley, D.M.: *Biophysical journal* **1994**, 66, 99-109.
25. Lewis, F. D.; Zhang, L.; Zuo, X.: *Journal of the American Chemical Society* **2005**, 127 (28), 10002-10003.
26. Van der Meer, B.W.: *Reviews in Molecular Biotechnology* **2002**, 82 (3), 181-196.
27. Clegg, R.M.: *Methods in Enzymology*, Academic Press, **1992**; 211, 353-388.
28. Kolstermeier. D.; Millar. D.P.: *Biopolymers*, **2002**, 61, 159-179.
29. Kuznetsov, S.V.; Sugimura, S.; Vivas, P.; Crothers, D.M.; Ansari, A.: *Proceedings of the National Academy of Sciences*, **2006**, 103, 49, 18515-18520.
30. Roy, R.; Hohung. S.; Ha. T.: *Nature Methods*, **2008**, 18, 16-26
31. Gell. C.; Brockwell, D.; Smith, A.: *Handbook of Single Molecule Fluorescence Spectroscopy*, Oxford University Press, **2006**.
32. Wozniak, A.K.; Schröder, G. F.; Grubmüller, H.; Seidel, C.A.M.; Oesterhelt, F. *Proceedings of the National Academy of Sciences*, **2008**, 105, 18337-42.
33. Rodger, A.; Norden, B., *Circular Dichroism and Linear Dichroism*. Oxford University Press, **1997**.

34. Berova, N.; Nakanishi, K.; Woody, R.W., *Circular Dichroism Principles and Applications,* 2nd edition, **2000**.
35. Mammana, A.; Pescitelli, G.; Asakawa, T.; Jockusch, S.; Petrovic, A.G.; Monaco, R.R.; Purrello, R.; Turro, N.J.; Nakanishi, K.; Ellestad, G.A.; Balaz, M.; Berova, N.: *Chemistry - A European Journal*. **2009**, 15 (44), 11853-11866.
36. Anderson, H.L.: *Inorganic Chemistry* **1994**, 33 (5), 972-981.
37. Harada, N.; Nakanishi, K.: *Circular Dichroic Spectroscopy: exciton coupling in organic stereochemistry*. University Science Books, **1983**.
38. Neylon, C.: *European Biophysics Journal* **2008**, 37 (5), 531-541.
39. Lipfert, J.; Doniach, S.: *Annual Reviews of Biophysics*. **2007**, 36, 307-27.
40. Brown, T.; Brown, D.J.S.: *Methods in Enzymology*, 211, **1992**, Pages 20–35.
41. Pearson, W.H.; Berry, D.A.; Stoy, P.; Jung, K.; Sercel, A.D.: *Methods*, ***2005***, 7114-7122.
42. Fendt, L.A.; Bouamaied, I.; Thöni, S.; Amiot, N.; Stulz, E.: *Journal of the American Chemical Society*, **2007**, 129, 15319–15329.
43 Barbaric, J.; Wagenknecht, H.A.: *Organic and Biomolecular Chemistry*, **2006**, 4, 2088–2090.
44 Mayer-Enthart, E.; Wagenknecht, H.A.: *Angewandte Chemie International Edition*, **2006**, 45, 3372–3375.
45 Vrabel, M.; Horakova, P.; Pivonkova, H.; Kalachova, L.; Cernocka, H.; Cahova, H.; Pohl, R.; Sebest, P.; Havran, L.; Hocek; M.; Fojta, M.: *Chemistry - A European Journal*, **2009**, 15, 1144–1154.
46 Khan, S.I.; Beilstein, A.E.; Grinstaff, M.W.: *Inorganic Chemistry*, **1999**, 38, 418–419.
47 Kottysch, T.; Ahlborn, C.; Brotzel, F.; Richert, C.: *Chemistry - A European Journal*, **2004**, 10, 4017–4028.
48 Hobbs, F. W.: *Journal of Organic Chemistry*, **1989**, 54, 3420–3422.
49. Hocek, M.: *European Journal of Organic Chemistry*, **2003**, 245–254.
50 Jager, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M.: *Journal of the American Chemical Society*, **2005**, 127, 15071–15082.

51 Seela, F.; Feiling, E.; Gross, J.; Hillenkamp, F.; Ramzaeva, N.; Rosemeyer, H.; Zulauf, M.; *Journal of Biotechnology*, **2001**, 86, 269–279.

52. Bandy, T.J.; Brewer, A.; Burns, J.R.; Marth, G.; Nguyen, T.; Stulz, E.: *Chemical Society reviews* **2011**, *40*, 138-48.

53. Gouterman, M.; Wagnière, G.H.; Snyder, L.C.: *Journal of Molecular Spectroscopy* **1963**, 11 (1-6), 108-127.

54. Caughey, W. S.; Smythe, G. A.; O'Keeffe, D. H.; Maskasky, J. E.; Smith, M. I.: *Journal of Biological Chemistry* **1975**, 250 (19), 7602-7622.

55. Gouterman, M.: *The Porphyrins*, Academic Press **1978**, 111.

56. Kadish K.; Smith K.; Guilard R., *The Porphyrins Handbook*. Academic Press **2000**.

57. Nguyen, T.; Brewer, A.; Stulz, E.: *Angewandte Chemie International Edition* **2009**, 48 (11), 1974-1977.

58. Asano, M.S.; Yamashita, K.; Kitabayashi, M.; Kusama, K.; Kagota, D.; Sugiura, K.: *Physical Chemistry Chemical Physics* **2011**, 13 (28), 12712-12715.

59. Otsuki, J.; Iwasaki, K.; Nakano, Y.; Itou, M.; Araki, Y.; Ito, O.: *Chemistry – A European Journal* **2004**, 10 (14), 3461-3466.

60. Bouamaied, I.; Stulz, E.: *Synlett*, **2004**, 2004 (09), 1579-1583.

61. Prathapan, S.; Johnson, T. E.; Lindsey, J.S.: *Biopolymers* **1993**, 7519-7520.

62. Rempel, U., Maltzan, B.; Borczyskowski, C.: *Journal of Luminescence*, 48-49, **1991**, 415-418.

63. Yang, S.I.; Seth, J.; Balasubramanian, T.; Kim, D.; Lindsey, J.S.; Holten, D.; Bocian, D.F.: *Journal of the American Chemical Society* **1999**, *121*, 4008-4018.

64. Hill, T.M.: *Annual review of microbiology* **1992**, *46*, 603-33.

65. Neylon, C.; Kralicek, A.V.; Hill, T.M.; Dixon, N.E. *Society* **2005**, 69, 501-526.

66 Mulcair, M.D.; Schaeffer, P.M.; Oakley, A.J.; Cross, H.F.; Neylon, C.; Hill, T.M.; Dixon, N.E. *Cell* **2006**, 125, 1309-1319.

67. Kamada, K.; Horiuchi, T.; Ohsumi, K.; Shimamoto, N.; Morikawa, K.: *Nature* **1996**, 383 (6601), 598-603.

68. Wang, H.; Lu, M.; Tang, M.-shong; Houten, B. Van; Ross, J.B.A.; Weinfeld, M.; Le, X.C. *Proceedings of the National Academy of Sciences* **2009**, *106*, 12849-54.

69. Rivetti, C.; Guthold, M.; Bustamante, C. *The EMBO journal* **1999**, *18*, 4464-4475.
70. Saecker, R.M.; Record, M.T.: *Current opinion in structural biology*, 12, (3), **2002**, 311-319
71. Geier, G.R.; Callinan, J.B.; Rao, P.D.; Lindsey, J.S., *Journal of Porphyrins and Phthalocyanines* **2001**, 5 (12), 810-823.
72. Geier III, G.R.; Riggs, J.A.; Lindsey, J.S., *Journal of Porphyrins and Phthalocyanines* **2001**, 5 (9), 681-690.
73. Lindsey, J.S.; Wagner, R.W., *The Journal of Organic Chemistry* **1989**, *54* (4), 828-836.
74. Geier III, G.R.; Lindsey, J.S.: *Journal of the Chemical Society, Perkin Transactions 2* **2001**, 5, 677-686.
75. Wagner, R. W.; Johnson, T.E.; Li, F.; Lindsey, J.S.: *The Journal of Organic Chemistry* **1995,** *60* (16), 5266-5273.
76. Mammana, A.; Asakawa, T.; Bitsch-Jensen, K.; Wolfe, A.; Chaturantabut, S.; Otani, Y.; Li, X.; Li, Z.; Nakanishi, K.; Balaz, M.; Ellestad, G. A.; Berova, N.: *Bioorganic & Medicinal Chemistry* **2008,** *16* (13), 6544-6551
77. Börjesson, K.; Tumpane, J.; Ljungdahl, T.; Wilhelmsson, L.M.; Nordén, B.; Brown, T.; Mårtensson, J.; Albinsson, B.: *Journal of the American Chemical Society* **2009,** *131* (8), 2831-2839.
78. Nordhoff, E.; Cramer, R.; Karas, M.; Hillenkamp, F.; Kirpekar, F.; Kristiansen, K.; Roepstorff, P.: *Nucleic Acids Research* **1993,** *21* (15), 3347-3357.
79. Tang, W.; Zhu, L.; Smith, L. M.: *Analytical Chemistry* **1997,** *69* (3), 302-312.
80. Brewer, A.; Siligardi, G.; Neylon, C.; Stulz, E.: *Organic & Biomolecular Chemistry* **2011**, 9, 777-82.
81. Stephenson, A.W.I.; Partridge, A.C.; Filichev, V.V.: *Chemistry - A European Journal*, **2011**, 17, 6227-6238.
82. Moreira, B.G.; You, Y.; Behlke, M.A; Owczarzy, R.: *Biochemical and biophysical research communications* **2005**, 327, 473-84.
83. Gray, D.M.; Ratliff, R.L.; Vaughan, M.R.: *Methods in Enzymology*, **1992**, 211, 389-406.

84. Nguyen, T.; Brewer, A.; Stulz, E.: *Angewandte Chemie International Edition*, **2009**, 48, 1974-7.
85. Matile, S.; Berova, N.; Nakanishi, K.; Fleischhauer, J.; Woody, R.W.: *Journal of the American Chemical Society* **1996**, 118, 5198-5206.
86. Seidel, C.A.M.; Schulz, A.; Sauer, M.H.M.: *The Journal of Physical Chemistry* **1996**, 100, 5541-5553.
87. Zuker, M.: *Nucleic Acids Research* **2003**, 31, 3406-3415.
88. Rühl, T.; Stulz, E.: *Supramolecular Chemistry*, **2010**, 22, 103.
89. Suntharalingam, K.; White, A.J.P.; Vilar, R.: *Inorganic chemistry* **2009**, *48*, 9427-35.
90. Green, S.J.; Lubrich, D.; Turberfield, A.J.: *Biophysical journal* **2006**, 91, 2966-75.
91. Chan, L.; Cross, H.F.; She, J.K.; Cavalli, G.; Martins, H.F.P.; Neylon, C.: *PloS one* **2007**, *2*, e1164.
92. Popp, M.W.L.; Ploegh, H.L.: *Angewandte Chemie International Edition*, **2011**, *50*, 5024-32.
93. Sambrook, J.; Russell, D.: *The Condensed Protocols from Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press **2006**.
94. Neylon, C.; Brown, S.E.; Kralicek, A.V.; Miles, C.S.; Love, C.A.; Dixon, N.E.: *Biochemistry* **2000**, 39, 11989-119899.
95. Schneider, H. J.; Yatsimirsky, A.: *Principles and Methods in Supramolecular Chemistry*. Wiley **2000**.
96. Cutler, P.: *Methods in Molecular Biology, Protein Purification Protocols*. Humana Press **2004**.
97. Grubisic, Z.; Rempp, P.; Benoit, H.: *Journal of Polymer Science Part B: Polymer Physics* **1996,** 34 10, 1707-1713.
98. Wang, Y.; Teraoka, I.; Hansen, F.Y.; Peters, G.N.H.; Hassager, O.: *Macromolecules* **2010,** 43 3, 1651-1659.
99. Boulanger, P.; le Maire, M.; Bonhivers, M, Dubois, S.; Desmadril, M.; Letellier, L.; Biochemistry, **1996**, 35, 14216-14224.
100. Janson, J.C.; Ryden, L., *Protein Purification*. Wiley **1998**.

101. Stephenson, A.W.I.; Partridge, A.C.; Filichev, V.V., *Chemistry (Weinheim an der Bergstrasse, Germany)* **2011**, 17, 6227-38.
102. Woese, C.R.; Winker, S.; Gutell, R.R.: *Proceedings of the National Academy of Sciences* **1990,** *87* 21, 8467-8471.
103. Green, S.J.; Lubrich, D.; Turberfield, A.J.: *Biophysical Journal*, **2006,** *91* 8, 2966-2975.
104. Fegan, A.; Shirude, P.S.; Ying, L.; Balasubramanian, S.: *Chemical Communications*, **2010,** *46* (6), 946-948.
105. SantaLucia, J.: *Proceedings of the National Academy of Sciences*, **1998,** *95* (4), 1460-1465.
106. Whitcombe, D.; Theaker, J.; Guy, S.P.; Brown, T.; Little, S.: *Nature Biotechology*, **1999,** *17* (8), 804-807.
107. Gale, N.; French, D. J.; Howard, R. L.; McDowell, D. G.; Debenham, P. G.; Brown, T.: *Organic & Biomolecular Chemistry* **2008,** *6* (24), 4553-4559.
108. Rachwal, P. A.; Fox, K. R.: *Methods* **2007,** *43* (4), 291-301.
109. Morrison, L. E.; Stols, L. M., *Biochemistry* **1993,** *32* (12), 3095-3104.
110. Parkhurst, K. M.; Parkhurst, L. J., *Biochemistry* **1995,** *34* (1), 285-292.
111. Burns, J. R.; Zekonyte, J.; Siligardi, G.; Hussain, R.; Stulz, E.: *Molecules*, **2011**, 16, (6), 4912-4922.
112. Chen, J.H.; Seeman, N.C.: *Nature* **1991**, (350), 631-633.
113. Endo, M.; Seeman, N.C.; Majima, T.: *Angewandte Chemie International Edition.* **2005**, 44, 6074-6077.
114. Seeman, N.C.: *Nano Letters.* **2010**, 10, 1971-1978.
115. Goodman, R.P.; Berry, R.M.; Turberfield, A.J.: *Chemical Communications*. **2004**, 1372-1373.
116. Erben, C.M.; Goodman, R.P.; Turberfield, A.J.: *Angewandte Chemie International Edition* **2006**, 45, 7414-7417.
117. Green, S.J.; Lubrich, D.; Turberfield,: *Biophysics Journal*. **2006**, 91, 2966-2975.
118. Bath, J.; Turberfield, A.J.: *Nature Nanotechnology*. **2007**, 2, 275-284.
119. LaBean, T.: *Nature Materials*. **2006**, 5, 767-768.
120. Park, S.H.; Pistol, C.; Ahn, S.J.; Reif, J.H.; Lebeck, A.R.; Dwyer, C.; LaBean, T.H.: *Angewandte Chemie International Edition*. **2006**, 45, 735-739.

121. Rothemund, P.W.K.: *Nature*. **2006**, 440, 297-302.
122. Kershner, R.J.; Bozano, L.D.; Micheel, C.M.; Hung, A.M.; Fornof, A.R.; Cha, J.N.; Rettner, C.T.; Bersani, M.; Frommer, J.; Rothemund, P.W.K.; Wallraff, G.M.: *Nature Nanotechnology*. **2009**, 4, 557-561.
123. Ke, Y.G.; Sharma, J.; Liu, M.H.; Jahn, K.; Liu, Y.; Yan, H.: *Nanotechnology Letters*. **2009**, 9, 2445-2447.
124. Ke, Y.G.; Lindsay, S.; Chang, Y.; Liu, Y.; Yan, H.: *Science*. **2008**, 319, 180-183.
125. Park, S.-J.; Lazarides, A.A.; Storhoff, J.J.; Pesce, L.; Mirkin, C.A.: *Journal of Physical Chemistry B*. **2004**, 108, 12375-12380.
126. Zhang, H.; Li, Z.; Mirkin, C.A.: *Advanced Materials*. **2002**, 14, 1472-1274.
127. Yang, H.; McLaughlin, C.K.; Aldaye, F.A.; Hamblin, G.D.; Rys, A.Z.; Rouiller, I.; Sleiman, H.F.: *Nature Chemistry*. **2009**, 1, 390-396.
128. Lo, P.K.; Karam, P.; Aldaye, F.A.; McLaughlin, C.K.; Hamblin, G.D.; Cosa, G.; Sleiman, H.F.: *Nature Chemistry*. **2010**, 2, 319-328.
129. Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H.: *Journal of the American Chemical Society*. **2010**, 132, 1592-1597.
130. Shen, W.; Zhong, H.; Neff, D.; Norton, M.L.: *Journal of the American Chemical Society*. **2009**, 131, 6660-6661.
131. Sinensky, A.K.; Belcher, A.M.: *Nature Nanotechnology*. **2007**, 2, 653-659.
132. Heilemann, M.; Kasper, R.; Tinnefeld, P.; Sauer, M.: *Journal of the American Chemical Society*. **2006**, 128, 16864-16875.
133. Tumpane, J.; Sandin, P.; Kumar, R.; Powers, V.E.C.; Lundberg, E.P.; Gale, N.; Baglioni, P.; Lehn, J.M.; Albinsson, B.; Lincoln, P.; Wilhelmsson, L.M.; Brown, T.; Norden, B.: *Chemical Physical Letters*. **2007**, 440, 125-129.
134. Malinovskii, V.L.; Wenger, D.; Häner, R.: *Chemical Society Reviews*, **2010**, 39, 410-422.
135. Weisbrod, S.H.: *Chemical Communications*, **2008**, 5675-5685.
136. Schwartz, E.; Le Gac, S.; Cornelissen, J.; Nolte, Roeland, .J.M.; Rowan, A.E.: *Chemical Society Reviews*. **2010**, 39, 1576-1599.
137. Wengel, J.: *Organic and Biomolecular Chemistry*, **2004**, (2), 277-280.

138. Kalek, M.; Madsen, A.S.; Wengel, J.: *Journal of the American Chemical Society*, **2007**, 129, 9392-9400.
139. Karlsen, K.K.; Jensen, T.B.; Wengel, J.: *Journal of Organic Chemistry*, **2009**, 74, 8838-8841.
140. Yang, H.; Rys, A.Z.; McLaughlin, C.K.; Sleiman, H.F.: *Angewandte Chemie International Edition*, **2009**, 48, 9919-9923.
141. Ghosh, S.; Pignot-Paintrand, I.; Dumy, P.; Defrancq, E.: *Organic and Biomolecular Chemistry*, **2009**, 7, 2729-2737.
142. Yang, H.; Sleiman, H.F.: *Angewandte Chemie International Edition*. **2008**, 47, 2443-2446.
143. Mitchell, N.; Ebner, A.; Hinterdorfer, P.; Tampe, R.; Howorka, S.: *Small* **2010**, 6, 1732-1735.

Appendix

8.1 Experimental

Reagents and solvents:

All reagents and solvents were from commercially available sources (Fisher Scientific, Sigma-Aldrich, Apollo Scientific, Glen Research, Link Technology, Berry & Associates and ATD Bio Ltd) unless stated.

Column chromatography:

Column chromatography was performed using 60 Å silica gel unless stated (Fluka).

Thin Layer Chromatography:

Aluminium sheets, 0.2 mm silica gel 60 Å F_{254} (Fluka), imaged using short and long range UV light (254 nm and 365 nm) and stained using anisaldehyde where relevant. For DMT containing reagents and products, the group was visualised by exposing to concentrated HCl vapour.

NMR:

^{1}H and ^{13}C NMR obtained using a 300 and 400 MHz Bruker DPX-300 and DPX-400 spectrometers, spectrums referenced to undeuterated solvent peaks (Aldrich).

ESI Micro-ToF and MALDI-ToF:

Samples analysed by Julie Herman at the University of Southampton.

UV-Vis absorption spectroscopy:

UV-Vis spectroscopy was performed using a Varian Cary 300 Bio UV-Vis Spectrophotometer in a 1 cm quartz cells (Fisher), duplex and hairpin DNA was annealed and denatured at 1 °C / min using the instruments peltier device unless stated.

Fluorescence emission spectroscopy:

Fluorescence analysis was performed using a Cary Eclipse Fluorescence Spectrophotometer in 1 cm quartz cells (Fisher) and 500 PMT voltage unless stated.

Duplex and hairpin DNA was annealed and denatured at 1 °C / min using the Varian's peltier device.

Circular dichroism spectroscopy:
Spectra were collected using B23 beamline at Diamond Light Source using a module B end-station with 0.5 mm slit equivalent to 1.2 nm bandwidth, 1 s integration time, 38 nm / min scan speed at a wavelength range 190-400 nm, or collected using an Applied Photophysics Chriscan bench-top instrument.

TEM:
TEM analysis was performed together with Dr. Dorota Bartczak using a Hitachi H7000 TEM, operating at a bias voltage of 75 kV. 10 μL of the 2 μM annealed DNA was loaded on to carbon film 400 mesh copper (50) TEM grids (Agar Scientific), left to dry and stained using lead citrate stain and uranyl acetate 2 % aqueous solution.

AFM:
AFM topographical images were acquired together with Dr. Jurgita Zekonyte, in buffer at room temperature with magnetically driven dynamic force microscopy (MAC Mode III, 5500 Scanning Probe Microscopy from Agilent Technologies, US), and magnetically coated Type VI MAC levers (Agilent Technologies, US). The nominal spring constant of MAC lever was 0.2 N / m, resonance frequency 22 kHz, and the lateral scan frequency was 2 Hz. PicoView 1.6 and PicoImage (Agilent Technologies, US) software were used for data acquisition and image analysis, respectively. For AFM imaging, DNA samples were deposited onto freshly cleaved mica substrate attached to the commercial fluid cell. The mica discs (14 mm, Agar Scientific) were pre-treated with $NiCl_2$ solution according to established methods to which the terpyridine arrays were loaded; AFM imaging was performed after 1 hr deposition time.

High pressure liquid chromatography:
HPLC was performed using a Varian 920-LC HPLC, with a Waters Xbridge OST C_{18} 2.5 μM, 4.6 x 50 mm column, monitoring the UV-Vis absorption at 200-800 nm.

Table 7 Typical HPCL buffers and gradients for native DNA

TEAA (100 mM), CH_3CN (1 %) percentage	**MeOH percentage**	**Time (mins)**
99	1	0→1
80	20	1→17
20	80	18→24
99	1	25→33

Table 8 Typical HPLC buffers and gradients for porphyrin DNA

HFIP (38.1 mM), TEA (8.4 mM)	**MeOH percentage**	**Time (mins)**
88	12	0→1
10	90	2→30
10	90	31→38
88	12	39→45

Size exclusion chromatography:

SEC was performed using a Varian 920-LC HPLC, with a Superdex 75 10/300 GL column (GE Healthcare) monitoring the UV-Vis absorption at 200-800 nm.

Analytical gel filtration:

AGF was performed using the same FPLC and Superdex 75 column as in SEC. The standards ferritin, ovalbumin, aldolase and ribonuclease (GE healthcare) were passed though using exactly the same conditions as the unknown samples of Tus, DNA and its complexes.

SDS Poly acrylamide gel electrophoresis:

12 % SDS poly acrylamide gel electrophoresis (PAGE) was conducted using protocols from: *Molecular Cloning, A Laboratory Manual* by Sambrook and Russell (2001, third edition CSHL Press). The 12 % SDS gel solution was comprised of acrylamide (50 %, 6 mL), water (4.9 mL), Tris (pH 8.8, 1.5 M, 3.8 mL) and SDS (10 %, 150 μL). The solution was mixed, after which ammonium persulphate (10 %, 150 μL) and TEMED (6 μL) was added and the gel quickly loaded. Water (1 mL) was added to the top of the gel, and the solution gentle rocked 3 times, after 1 hour the water was removed and the gel stack prepared. The 5 % SDS gel stack solution was comprised of acrylamide (50 %, 500 μL), water (2.1 mL), Tris (pH 8.8, 1.5 M, 380 μL) and SDS (10 %, 30 μL). The solution was mixed, after which ammonium persulphate (10 %, 30 μL) and TEMED (3 μL) was added and the gel stack quickly loaded. After 1 hour the stacking plate was removed and the gel prepared for electrophoresis. The running buffer (1 L) comprised of tris base (3.03 g), g;lycine (14.4 g) and SDS (1 g) was added. The wells were washed with running buffer, after which the lanes were loaded with the relevant samples. Wide range markers (Sigma Alrich) were run in the extreme lanes. The gels were run at 140 volts for 1 hour, however, the progression of the gel was monitored visually by observing position the markers. After 1 hour, the gel was stopped and the gel extracted. SDS was removed from the gel by washing in deionised water for 1 hour. After which the gel was stained by washing with coomassie solution (Sigma Aldrich) for 30 minutes. Excess coomassie was removed by washing with deionised water for 6 hours.

Native Poly acrylamide gel electrophoresis:

12 % native PAGE was conducted using protocols from *Molecular Cloning, A Laboratory Manual* by Sambrook and Russell (2001, third edition CSHL Press). 5X TBE was made by combining tris base (26.5 g), boric acid (13.75 g) and EDTA (0.5 M, pH 8, 10 mL) to water (490 mL). The 12 % native gel solution was comprised of acrylamide (50 %, 4 mL), water (3.9 mL), 5X TBE (2 mL)., the solution was mixed, after which ammonium persulphate (10 %, 70 μL) and TEMED (7 μL) was added and the gel quickly loaded. Water (1 mL) was added to the top of the gel, and the solution gently rocked 3 times, after 1 hour the water was removed and the gel stack prepared. The 5 % native gel stack solution was comprised of acrylamide (50 %, 1.7 mL), water

(6.2 mL), 5X TBE (2 mL). The solution was mixed, after which ammonium persulphate (10 %, 70 μL) and TEMED (6 μL) was added and the gel stack quickly loaded. After 1 hour the stacking plate was removed and the gel prepared for electrophoresis. 1X TBE (1 L) was used as a running buffer. The wells were washed with running buffer, after which the lanes were loaded with the relevant samples. The gels were run at 140 volts for 1 hour. After which, the current was stopped and the gel extracted. TBE buffer was removed from the gel by washing in water for 1 hour. After which the gel was imaged to observe any fluorescently labelled entities. DNA was imaged using ethidium bromide solution (Sigma Aldrich) by carefully washing for 30 minutes. Excess ethidium bromide was washed with deionised water for 2 hours, after which the DNA was imaged. Finally, the protein entities were visualised by staining using coomassie solution (Sigma Aldrich) for 30 minutes. Excess coomassie was removed by washing with water for 6 hours.

Solid phase DNA synthesis:

Solid phase DNA synthesis was conducted using an Expedite automated solid phase DNA synthesiser, 1 μmol scale, using 1 μmol columns with 500 Å pores. All reagents used were DNA grade, degassed under nitrogen or argon. Various DNA sequences have been synthesised ranging in length from 14- to 45-mers. The instrument was cleaned regularly and all lines washed with DNA grade degassed ammonia / acetonitrile solution (1 : 1) prior to use. The degassed acetonitrile used was of DNA grade purity and contained less than 30 ppm H_2O. The protected monomers and reagents were commercially available (Proligo), the monomers were made up to 0.11 M in acetonitrile. Coupling efficiency was automatically monitored by the UV absorption reading of the cleaved dimethoxy trityl cation. The efficiency was monitored to ensure efficient base coupling in the 3' → 5' direction. The final DMT group was cleaved using the acid wash protocol (unless required for fluorous affinity for porphyrin labelled DNA, in which case the final DMT deprotection was not conducted).

The protecting groups and resin of the DNA synthesised was deprotected and cleaved respectively by heating to 55 °C in ammonia using 1.5 mL volume screw-top vials. After 6 hours the transparent solution was cooled to room temperature, and the ammonia was removed with care under reduced pressure using a Eppendorf

concentrator 5301. The DNA was dissolved in deionised H_2O (0.5 mL, sometimes requiring high temperatures and excess H_2O), to which size exclusion chromatography was performed using a Nap-25 column. The column was preconditioned with deionised water, passing though 5 mL with an approximate 1 mL flow per minute under gravity. The DNA was loaded onto the column; to which the first 1.3 mL was collected, whereby removing small impurities associated with deprotection. The concentration was confirmed by UV analysis, after which HPLC purification was performed.

The DNA was purified using by HPLC, and C60 reverse phase column column. All HPLC flows for unmodified DNA started with 3 % HPLC grade acetonitrile and 97 % 50 mM TEAA buffer (the water used to make the TEAA buffer was distilled and deionised prior to use). Acetonitrile was added to the DNA solution to match the starting HPLC flow rates, after which the solution was filtered using a 0.45 μm spin filters. The conditioned DNA sample was transferred to a 1 mL HPLC loading vial to which HPLC analysis (loaded using the autosampler), followed by HPLC purification was performed (loaded by manual injection to a 1 mL sample loop). For HPLC analysis; small quantities of sample, typically about 1 nmol was injected and ran using a standard DNA HPLC solvent gradient. The purity was monitored and 260 nm for unmodified DNA using a UV-Vis detector. The elution volume and resolution was monitored, to which the gradient was modified at/near the main product elution volume. Typically the gradient around the elution volume was reduced. The analytical injection was repeated to ensure maximum resolution was obtained, following this the remainder of the sample was injected using the manual injection loop.

Conducting semi-preparative HPLC on large quantities of DNA, typically injecting approximately 500 nmol's in one run can potentially result in effects not observed during analytical HPLC, for example, aggregation, saturation of column and/or detector. When injecting high quantities, no aggregates / particulates should be present, often requiring efficient injection from initial NAP and spin filtration (if extended periods of time between steps occurred, additional spin filtration was conducted). Fractions were collected using a fraction collector, the delay volume from UV-Vis detector to fraction collector was calibrated using standards (approximately 330 μL), and further verified by independent UV-Vis analysis to the collected fractions. Once injected, the chromatogram was monitored at 295 nm, due to saturation of the

detector at 260 nm. The fractions were collected in 1 mL fractions; the purest fractions were combined and concentrated under vacuum overnight. The dry samples were redissolved in deionised water and concentrated again to remove residual DNA-TEAA association. Molecular mass confirmation of the synthesised and purified DNA was confirmed using either MALDI-ToF or ESI micro-ToF analysis. Samples submitted required no further desalting due to the absence of salt throughout purification. For ESI micro-ToF analysis, temperatures above 55 °C were required to ionise the DNA to break DNA-TEAA interactions.

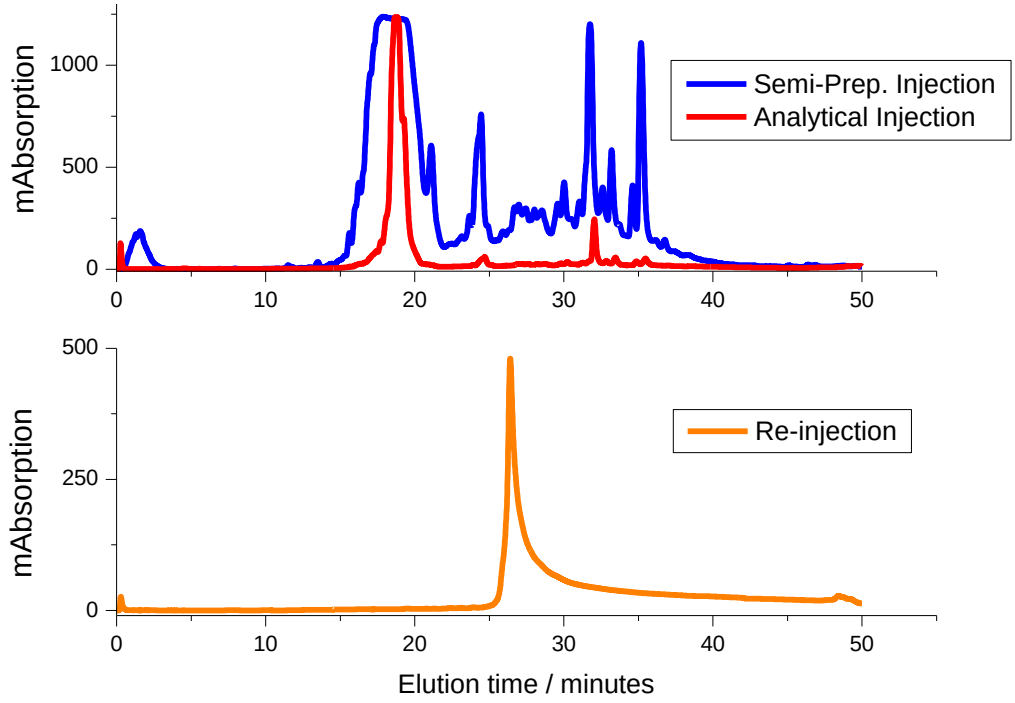


Figure 114 HPLC Extended DNA sequence chromatogram analytical injection (10 µL, red line), semi-preparative injection (990 µL, blue line) and re-injection of purified DNA for confirmation (50 µL, orange line). Note re-injection gradient reduced compared to initial HPLC traces above to ensure complete purity.

DNA hybridisation:

Duplex DNA was annealed either using an Eppendorf thermomixer, by heating to 85 °C for 5 minutes, followed by cooling to 15 °C over 3 hours, or the peltier devices of the

UV-Vis and fluorescence spectrophotometers by cooling at 1 °C per minute from 85 °C to 15 °C.

Fluorous affinity:

Porphyrin DNA was purified using Fluoro-PAK columns by incorporating a fluorous DMT group to the terminal base of the DNA (Berry & Associates). Their protocols were followed to elute pure porphyrin labelled DNA. The elution was monitored by UV-Vis absorption spectroscopy, purity was confirmed by analytical HPLC analysis.

NAP:

DNA was purified and desalted using NAP-10 and NAP-25 columns (GE Healthcare). Firstly 5 mL of deionised water was passed through the column under gravity, after which the DNA was loaded onto the column in the smallest volume of liquid possible, and water was added to elute the desalted DNA within 1 mL for NAP-10 columns and with 2 mL for NAP-25 columns. The elution was monitored by UV-Vis absorption spectroscopy.

Tus buffer:

250 mM KCl, 50 mM Tris-HCl, 1 mM EDTA, 1 mM DTT pH 7.4.

Adjustable Strap DNA:

The FAM and TAMRA Adjustable Strap DNA sequence (S^1) and the complementary strands were synthesised and purified by ATDBio Ltd. Synthesis was conducted on a 0.2 μmol scale.

8.2 Tus-*Ter* crystal structure DNA sequence

```
5'  TTAGTTACAACATACT  3'
3'   ATCAATGTTGTATGAT 5'
```

8.3 Tus (amino-His^{6}) sequence

MARYDLVDRLNTTFRQMEQELAIFAAHLEQHKLLVARVFSLPEVKKEDEHNPLNRIEVKQHLGNDAQSLA
LRHFRHLFIQQQSENRSSKAAVRLPGVLCYQVDNLSQAALVSHIQHINKLKTTFEHIVTVESELPTAARF
EWVHRHLPGLITLNAYRTLTVLHDPATLRFGWANKHIIKNLHRDEVLAQLEKSLKSPRSVAPWTREEWQR
KLEREYQDIAALPQNAKLKIKRPVKVQPIARVWYKGDQKQVQHACPTPLIALINRDNGAGVPDVGELLNY
DADNVQHRYKPQAQPLRLIIPRLHLYVADGRGLELPETGGHHHHHH

8.4 Lycine-di-glycine chemical structure

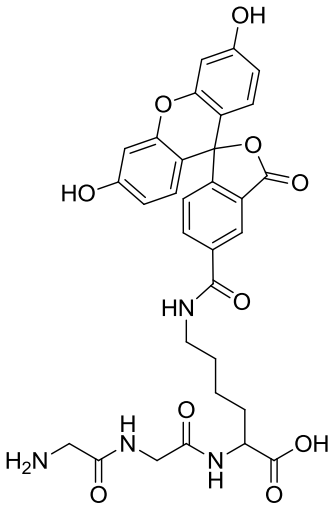


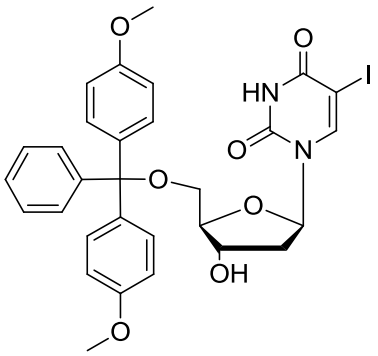
Figure 115 Chemical structured of lycine-di-glycine (supplied from Dr. Cameron Neylon).

8.5 Fluorescein-Tus

Sortase (5 uL, 10 μM), Calcilum chloride (14 mg) and fluorescein-di-glycie (39 uL, 8 μM) was added to Tus (1 mL, 28 μM), the solution was left overnight at room

temperature. After 16 hours, the product was purified by nickel affinity chromatography. The F-Tus fractions were combined, after which the imidazole was removed by dialysis (MWCO <12'00-14'000, Spectrum Lab. inc.). The fluorescent solution was concentrated using a spin filter (Millipore biomax,10 K NMWL membrane), after which glycerol (10 %) was added for storage at -80 °C.

8.6 Iodo-deoxyuridine monomer



1

Deoxyuridine (5 g, 14 mmol, 1 eq.) was added to DMF (3 mL) under nitrogen. Over 6 hours DMT-Cl (4.78 g, 14 mmol, 1 eq.) was added, after completion, the solvent was removed under vacuum by co-evaporation with toluene and methanol, dried overnight under ultra high vacuum. Column chromatography was performed (eluent = DCM / MeOH 20:1 TEA, R_f = 0.4), the pure product was dried overnight under ultra high vacuum to yield a white solid.

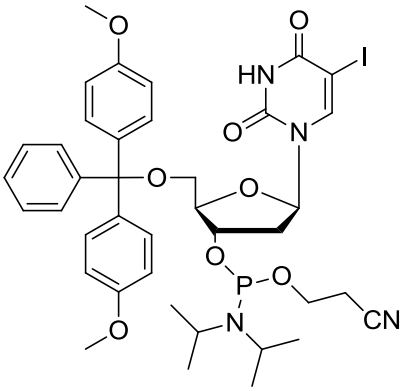
Mass 5.0 g Yield 54 %

^{1}H NMR (400MHz, Chloroform-d) δ = 8.17 (s, 1 H), 7.44 (d, *J* = 7.5 Hz, 2 H), 7.39 - 7.20 (m, 7 H), 6.87 (d, *J* = 8.8 Hz, 4 H), 6.35 (dd, *J* = 5.9, 7.7 Hz, 1 H), 4.61 - 4.53 (m, 1 H), 4.15 (q, *J* = 7.1 Hz, 2 H), 3.81 (s, 6 H), 3.47 - 3.33 (m, 2 H), 3.02 - 2.87 (m, 1 H), 2.60 - 2.49 (m, 1 H), 2.31 (t, *J* = 7.5 Hz, 1 H), 2.20 (s, 2 H), 2.07 (s, 2 H), 1.99 - 1.82 (m, 1 H), 1.29 (t, *J* = 7.2 Hz, 2 H)

^{13}C NMR (101MHz, Chloroform-d) δ = 171.2, 160.1, 158.6, 150.1, 144.3, 144.3, 135.4, 135.3, 130.1, 130.0, 128.1, 128.0, 127.0, 113.4, 113.4, 87.0, 86.6, 85.7, 77.3, 76.7, 72.4, 68.7, 63.5, 60.4, 55.2, 41.4, 30.9, 21.0, 14.1

ESI$^-$: Calculated 656.5 Da, Found 655.2 Da $[M\text{-}H^+]^-$

8.7 Iodo-deoxyuridine phosphoramidite monomer



2

To oven dried glassware under N_2, **1** (100 mg, 0.15 mmol, 1 eq), CEP-Cl (99 μL, 106 mg, 0.45 mmol, 3 eq.) and DIPEA (94 μL, 70 mg, 0.61 mmol, 4 eq.) was added to anhydrous DCM (2 mL). The solution was stirred for 3 hours, the reaction was monitored by TLC (DCM / MeOH 10 : 1, R_f = 0.4). After the reaction had gone to completion, the product was precipitated in degassed hexane (2 x 5 mL). The solvent was removed by filtration, washed with more hexane (2 x 1 mL), and dried under ultra high vacuum for 20 minutes. The product was dissolved in DCM / CH_3CN (10 : 1 Total volume 2 mL) and transferred to the DNA synthesiser.

Mass 123 mg Yield > 95 %

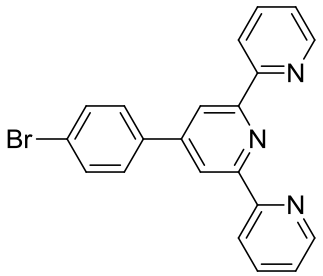
ESI⁻: Calculated 856.7 Da, Found 856.6 Da $[M\text{-}H^+]^-$

8.8 Iodine DNA analysis

Table 9 Iodine DNA analysis

DNA	MW (Da)	ε	Nanomoles	ESI⁻ Found (Da)
Y^1	6604.229	224000	383.4821	6602.78
Y^2	6604.229	224000	415.625	6186.89
Y^3	6604.229	224000	392.8571	6602.78
Y^4	6604.229	224000	397.3214	6602.78
Y^5	6604.229	224000	397.3214	6602.78
Y^6	6604.229	224000	397.3214	6602.78
Y^7	6604.229	224000	397.3214	6602.78
X^1	6716.129	204100	450.2695	6050

8.9 4'-(4-Bromophenyl)-2,2':6',2''-terpyridine



3

Under N_2, ethanol (250 mL), 2-acetylpyridine (7 g, 58 mmol, 4 eq.) and bromo-benzaldehyde (2.68 g, 14.5 mmol, 1 eq.) was added. The solution was left to stir for 20 minutes, after which, ammonia (6 mL, 33 %) and potassium hydroxide (12.2 g, 217 mmol, 15 eq.) was added. Additional ammonia (3 mL, 33 %) was added to solution after 4 hours. The solution changed colour from green to thick white precipitate overnight. The product was filtered with cold ethanol (4 x 50 mL), then re-crystallised in hot ethanol (2 x 50 mL). The product was dried overnight under ultra high vacuum.

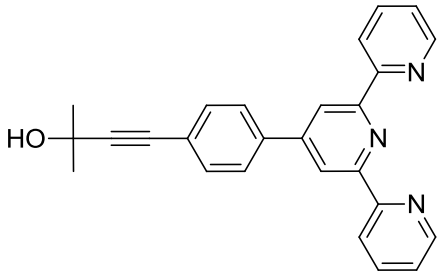
Mass 2.9 g Yield 52 %

^{1}H NMR (400MHz, Chloroform-d) δ = 8.75 - 8.65 (m, 6 H), 7.89 (m, 2 H), 7.78 (d, *J* = 8.5 Hz, 1 H), 7.65 (d, *J* = 8.5 Hz, 2 H), 7.37 (m, 2H).
^{13}C NMR (101MHz, , Chloroform-d) δ = 156.1, 149.1, 149.0, 137.4, 136.9, 132.1, 128.9, 123.9, 123.5, 121.4, 118.5.

ESI^{+}: Calculated 388.3 Da, Found 390.1 Da $[M+H^{+}]^{+}$

MALDI-ToF: Calculated 387.04 Da, Found 388.03 Da $[M+H^{+}]^{+}$

8.10 4-(4-([2,2':6',2''-Terpyridin]-4'-yl)phenyl)-2-methylbut-3-yn-2-ol



4

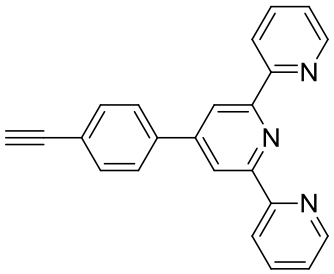
3 (1.64 g, 4.2 mmol), potassium carbonate (1.78 g, 12.8 mmol), copper iodide (8.2 mg, 0.045 mmol), triphenyl phosphine (0.16 g, 0.6 mmol) and palladium tetrakis-triphenylphosphine (52 mg, 0.045 mmol) was added to dimethoxy ethane (18.6 mL) and water (8.7 mL). The solution was sonicated for 5 minutes, and then heated under nitrogen at 80 °C. After 20 minutes, 2-metyl-but-3-yn-2-ol (0.65 g, 7.7 mmol) was added in 3 portions over 4 hours. The reaction was monitored by TLC (eluent EA, R_f = streaky, product intense blue emission under short UV light, stains brown in ferric chloride), after 6 hours the solution was cooled to room temperature, washed and extracted in EA (5 x 50 mL). The solvent removed under vacuum, followed by crystallisation in DCM / MeOH (10 : 1 mL) over 72 hours at room temperature. The green-white crystals were washed with petrol ether (2 x 50 mL), then redissolved in DCM, to which column chromatography was performed (eluent DCM / MeOH 50 : 1, product R_f value approximately 0.2 streaky). The pure product was collected, the solvent removed and dried overnight under ultra high vacuum.

Mass 490 mg Yield 40 %

^{1}H NMR (300MHz, Chloroform-d) δ = 8.78 - 8.63 (m, 6 H), 7.95 - 7.79 (m, 4 H), 7.55 (d, *J* = 8.4 Hz, 2 H), 7.36 (ddd, *J* = 0.9, 4.8, 7.4 Hz, 2 H), 1.66 (s, 6 H)
^{13}C NMR (75MHz, Chloroform-d) δ = 220.1, 156.1, 156.0, 149.3, 149.1, 138.1, 136.9, 132.2, 127.1, 123.9, 123.5, 121.4, 118.7, 95.2, 81.8, 77.4, 76.6, 65.6, 31.5

MALDI-ToF: Calculated 391.4 Da, Found 392.4 Da $[M+H^+]^+$

8.11 4'-(4-Ethynylphenyl)-2,2':6',2''-terpyridine



5

4 (0.16 g, 0.047 mmol, 1 eq.) was added to sodium methoxide (0.46 g, 1.1 mmol, 30 eq.) in dry toluene (50 mL) at 130 °C under nitrogen overnight. The reaction was monitored by TLC (eluent petrol ether / ethyl acetate 1 : 1). Upon completion the solvent was removed under vacuum, followed by an acid and basic wash (hydrochloric acid / water 1 : 1, sodium hydrocarbonate (saturated), the product extracted in EA (3 x 50 mL), the solvent removed under vacuum, the pure product was dried under ultra high vacuum for 72 hours.

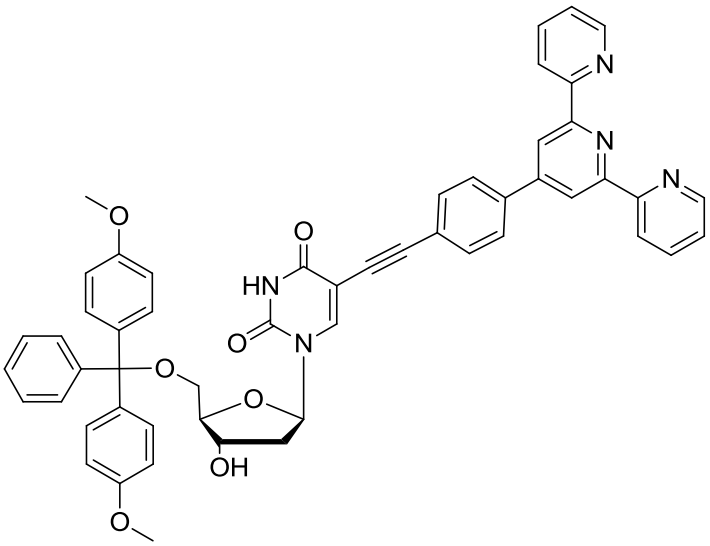
Mass 0.112 g Yield 82 %

^{1}H NMR (400MHz, Chloroform-d) δ = 8.65 - 8.58 (m, 4 H), 8.55 (d, *J* = 8.0 Hz, 2 H), 7.82 - 7.71 (m, 4 H), 7.53 (d, *J* = 8.0 Hz, 2 H), 7.27 - 7.20 (m, 2 H), 3.10 (s, 1 H)
^{13}C NMR (101MHz, Chloroform-d) δ = 156.0, 149.2, 149.1, 138.7, 136.8, 132.6, 127.2, 123.8, 122.7, 121.3, 118.6, 83.3, 78.5, 30.8

ESI+: Calculated 388.3 Da, Found 390.1 Da $[M+H]^+$

MALDI-ToF: Calculated 387.04 Da, Found 388.03 Da $[M+H^+]^+$

8.12 Terpyridine-deoxyuridine monomer



6

Under nitrogen, **1** (300 mg), triphenlyphosphine (14 mg), copper iodide (7.8 mg), potassium carbonate (157 mg) and palladium tetrakis triphenyl phosphine (15.8 mg) was added to ethylene glycol (2.5 mL) with water (1 mL). The solution was stirred at 80 °C for 20 minutes to which **5** (228 mg) was added. The reaction was monitored by TLC (eluent EA, R_f 0.1, product stays alone baseline, stains brown with ferric chloride, turns orange upon acid exposure). After 16 hours, EDTA (2M, 50 μL) was added and the solution was cooled to room temperature. The product was washed water (50 x 2 mL) then extracted in EA (50 mL). The product was purified by column chromatography (eluent EA), the pure product was collected, concentrated, and dried overnight under ultra high vacuum.

^{1}H NMR (400MHz, Chloroform-d) δ = 8.68 - 8.51 (m, 6 H), 8.16 (s, 1 H), 7.79 (dt, *J* = 1.8, 7.7 Hz, 2 H), 7.62 (d, *J* = 8.5 Hz, 2 H), 7.43 - 7.02 (m, 15 H), 6.71 (dd, *J* = 3.0, 9.0 Hz, 4 H), 6.37 - 6.22 (m, 1 H), 4.54 - 4.47 (m, 1 H), 4.08 (d, *J* = 2.5 Hz, 1 H), 3.60 (d, *J* = 3.5 Hz, 6 H), 3.43 - 3.20 (m, 2 H), 2.53 - 2.40 (m, 1 H), 2.31 - 2.14 (m, 1 H)
^{13}C NMR (101MHz, Chloroform-d) δ = 207.0, 180.8, 161.6, 158.6, 156.1, 155.9, 149.4, 149.3, 149.1, 144.4, 142.4, 138.0, 136.9, 135.5, 132.2, 132.0, 129.9, 129.9, 128.0, 127.9,

127.0, 126.8, 123.9, 123.3, 121.4, 118.6, 113.3, 100.4, 93.4, 87.1, 86.8, 85.9, 81.5, 77.3, 77.2, 76.7, 72.2, 63.5, 55.1, 45.7, 41.7, 32.2, 31.5, 30.9, 22.6, 14.1, 10.7

ESI$^-$: Calculated 861.9 Da, Found 860.6 Da $[M-H^+]^-$

Hi-Resolution ESI$^+$: Calculated 861.9 Da, Found 560.192 Da $[M-DMT^-+H^+]^-$

UV-vis:

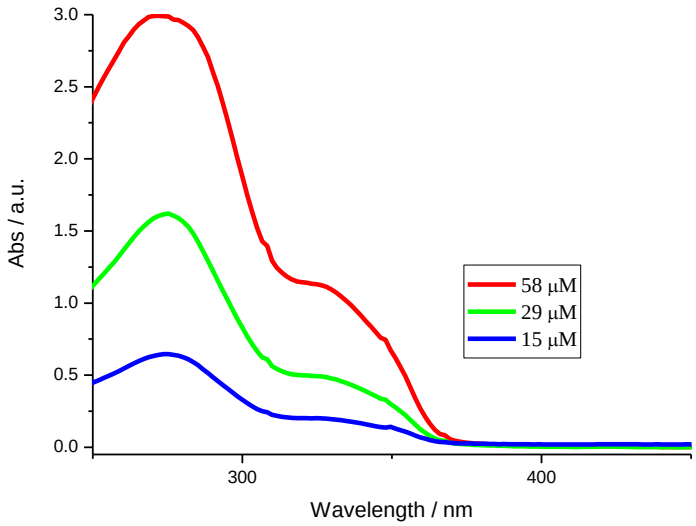


Figure 116 UV-Vis analysis of DMT-terpyridine-deoxyuridine at 58 µM (red line), 29 µM (green line) and 15 µM (blue line) in ethanol, 1 cm quartz path length, scanning at 600 nm per minute.

Table 10 Extinction coefficient values of DMT-terpyridine-deoxyuridine in methanol.

µM	ε_{260nm} **(Abs)**	ε_{330nm} **(Abs)**	ε_{424nm} **(Abs)**
58	48448, log 4.7 (2.81)	18876, log 4.3 (1.1)	345, log 2.6 (0.02)
29	47241, log 4.7 (1.37_	16483, log 4.2 (0.48)	345, log 2.6 (0.02)
15	36667, log 4.6 (0.55)	12980, log 4.1 (0.19)	345, log 2.6 (0.02)
Mean	44119, log 4.6	16113, log 4.2	345, log 2.6

Fluorescence:

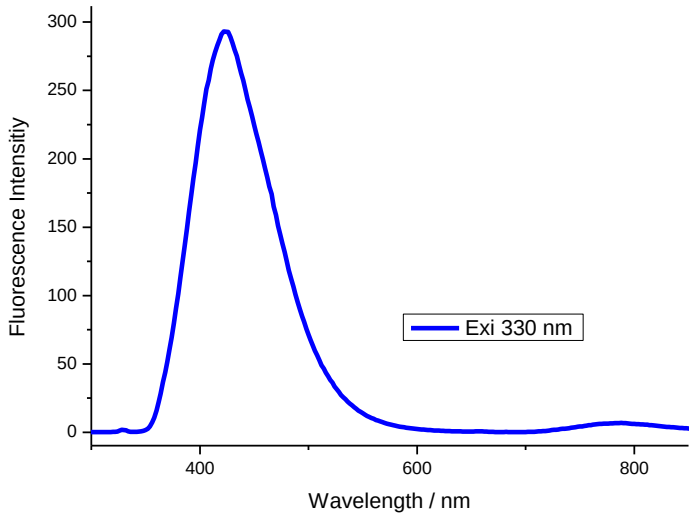


Figure 117 Fluorescence analysis of **6**, excitation at 330 nm, 10 µM in ethanol. 1 cm path length cell, 500 PMT voltage, scanning at 600 nm per minute.

Circular dichroism:

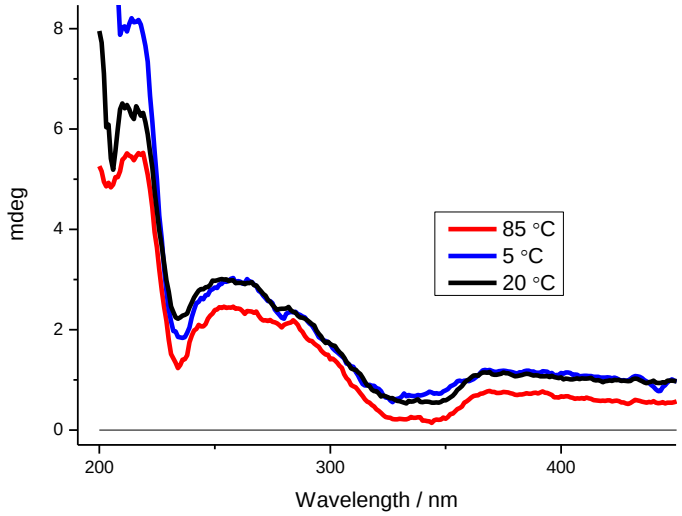
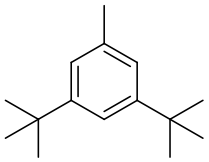


Figure 118 Circular Dichroism analysis of DMT-deoxyuridine terpyridine, 10 µM in ethanol, 1 cm path length cell, scanning at 2 seconds per nm.

8.13 1,3-Di-tert-butyl-5-methylbenzene



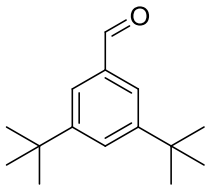
7

Toluene (81.6 mL, 0.77 mol, 1 eq.) was added to 2-chloromethylpropane (160 mL, 1.5 mol, 2 eq.) and purged under nitrogen for 20 minutes. Aluminum trichloride (3.96 g, 0.03 mol) was added in 8 portions over 8 hours. The solution was left to stir overnight, after which, water (40 mL) was added to quench the reaction. The solution was taken up in diethyl ether (3 x 50 mL), the solvent removed and the product purified by distillation. The pure product crystallised at 4 °C after 72 hours.

Mass 26.7 g Yield 17 %

H^1 NMR (300 MHz, Acetone) δ ppm 7.1 (s, 1 H) 6.9 (s, 2 H) 2.2 (s, 3 H) 1.2 (s, 18 H)
^{13}C NMR (101MHz, Chloroform-d) δ = 152.2, 138.4, 124.9, 120.9, 36.3, 33.1, 23.5

8.14 3,5-Di-tert-butylbenzaldehyde



8

7 (10.4 g, 51 mmol, 1 eq.) was added to carbon tetra chloride (90 mL) and degassed under nitrogen for 20 minutes. NBS (11.8 g, 66 mmol, 1.3 eq.), AIBN (4.17 g, 25 mmol, 0.5 eq.) was added to the solution and stirred at reflux for 3 hours. The solution was left to cool and the product purified by distillation.

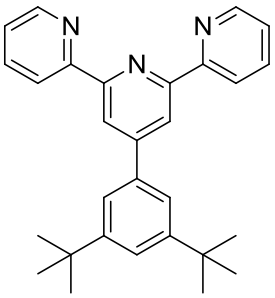
Crude Yield 71.9 %

Under nitrogen, 3,5-ditert-butyl-bromo-benzaldehyde (7 g, 24.7 mmol, 1 eq.) was added to MeOH / H_2O (1 : 1, 200 mL). After 15 minutes, HMTA (34.6 g, 0.24 mol, 10 eq.) was added, stirred for 4 hours (the solution turned dark red). Concentrated HCl (30 mL) was added to quench the reaction, stirred for 30 minutes, then left to cool to room temperature. The solution was extracted in DCM (50 mL), the solvent removed and the product recyrstallised from hot ethanol / water.

Mass 1.8 g Yield 33 %

^{1}H NMR (300 MHz, Chloroform-d) δ = 9.9 (s, 1 H) 7.6 - 7.6 (m, 3 H) 1.3 (s, 18 H)
^{13}C NMR (101MHz, Chloroform-d) δ = 193.1, 151.8, 136.2, 128.8, 124.1, 34.9, 31.3

8.15 4'-(3,5-Di-tert-butylphenyl)-2,2':6',2''-terpyridine



9

2-Acetylpryidine (1.5 g, 12.4 mmol, 2 eq.), **8** (1.35 g, 6.1 mmol, 1 eq.), acetamide (7.3 g, 0.37 mol, 30 eq.) and ammonium acetate (4.7 g, 0.18 mol, 15 eq.) was added and left to stir at 160 ^{o}C for 2 hours. Sodium hydroxide (20 g in 40 mL H_20) was added dropwise to the solution at 100 ^{o}C. The solution was heated to 120 ^{o}C for 2 hours with no stirring, which formed a black layer. The solution was cooled to room temperature, the black layer was extracted with DCM, washed with water and the solvent removed. Column chromatography was performed (eluent hexane → hexane / diethyl ether 1 : 1).

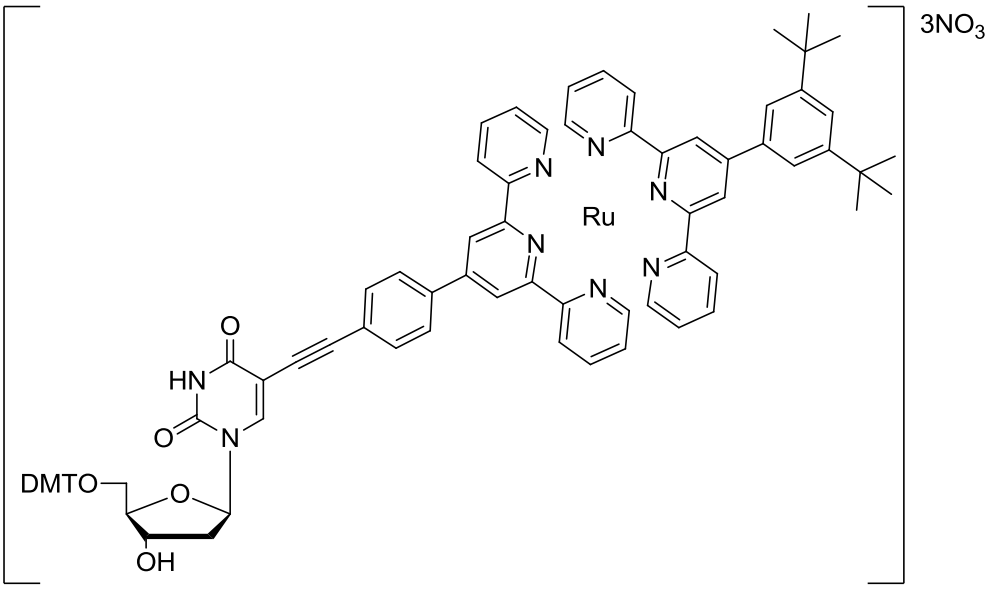
Mass 0.61 g Yield 21 %

ESI^{+}: Calculated 421.6 Da, Found 422 Da $[M+H]^{+}$

MALDI-ToF: Calculated 421.6 Found 420.9 Da $[M-H^{+}]^{-}$

^{1}H NMR (300 MHz, Chloroform-d) δ = 8.8 (d, *J* = 4.7 Hz, 2 H) 8.7 (d, 2 H) 8.7 (d, *J* = 8.0 Hz, 2 H) 7.9 (td, *J* = 7.7, 1.7 Hz, 2 H) 7.7 (d, *J* = 1.7 Hz, 2 H) 7.6 (t, 1 H) 7.4 (ddd, *J* = 7.4, 4.8, 1.0 Hz, 2 H) 1.5 (s, 18 H)
^{13}C NMR (101MHz, Chloroform-d) δ = 156.4, 155.7, 151.9, 151.3, 149.1, 138.3, 136.7, 123.6, 123.0, 121.7, 121.4, 119.4, 35.0, 31.5

8.16 Ruthenium bis-terpyridine deoxyuridine monomer



10

9 (292 mg, 0.69 mmol, 1 eq.), $RuCl_3.XH_20$ (130 mg, 0.62 mmol, 0.9 eq.) was added to ethanol (95 mL) and left to stir at reflux for 2 hours under N_2. The solution was then cooled to room temperature and the product filtered off and washed with cold ethanol (2 x 25 mL) to obtain a crude yield of 62 %. 4-Ethylmorpholine (24 mg, 2.1 mmol, 3.6 eq.) was added to methanol (10 mL) under nitrogen. Ruthenium 3,5-di-tert-butyl phenyl terpyridine (37 mg, 0.06 mmol, 1 eq) and **6** (50 mg, 0.06 mmol, 1 eq) was added, the solution heated to reflux, after 2 hours water (1 mL) was added. The solution was heated for a further 15 hours, the reaction was monitored by TLC (eluent acetone / saturated potassium nitrate 20 : 1, R_f = 0.4). The reaction was left to cool to room temperature and the solvent evaporated. The product was then purified by column chromatography (eluent acetone / saturated potassium nitrate 20 : 1), and the solvent removed, to which excess saturated potassium hexafluorophosphate was added (15 mL). The solvent removed and dried under ultra high vacuum.

Mass 56 mg Yield 58 %

Hi-resolution ESI^+: Calculated 689.2373 Da Found 689.2371 Da $[M+H-2PF_6]^{2+}$

Table 11 Extinction coefficient values of DMT-terpyridine-deoxyuridine in methanol.

µM	ε_{260nm} **(Abs)**	ε_{284nm} **(Abs)**	ε_{308nm} **(Abs)**	ε_{491nm} **(Abs)**
0.3	20000, log 4.3(0.6)	43456, log 4.64 (1.3)	46667, log 4.7, (1.4)	20600, log 4.3 (0.62)
0.15	23333, log 4.4 (0.35)	49333, log 4.7 (0.74)	52000, log 4.7 (0.78)	23333, log 4.4 (0.35)
0.0125	16000, log 4.2 (0.2)	36000, log 4.5 (0.42)	35200, log 4.5 (0.44)	16000, log 4.2 (0.2)
Mean	19778, log 4.3 (0.38)	39596, log 4.6 (0.82)	44622, log 4.6 (0.87)	19977. log 4.3 (0.39)

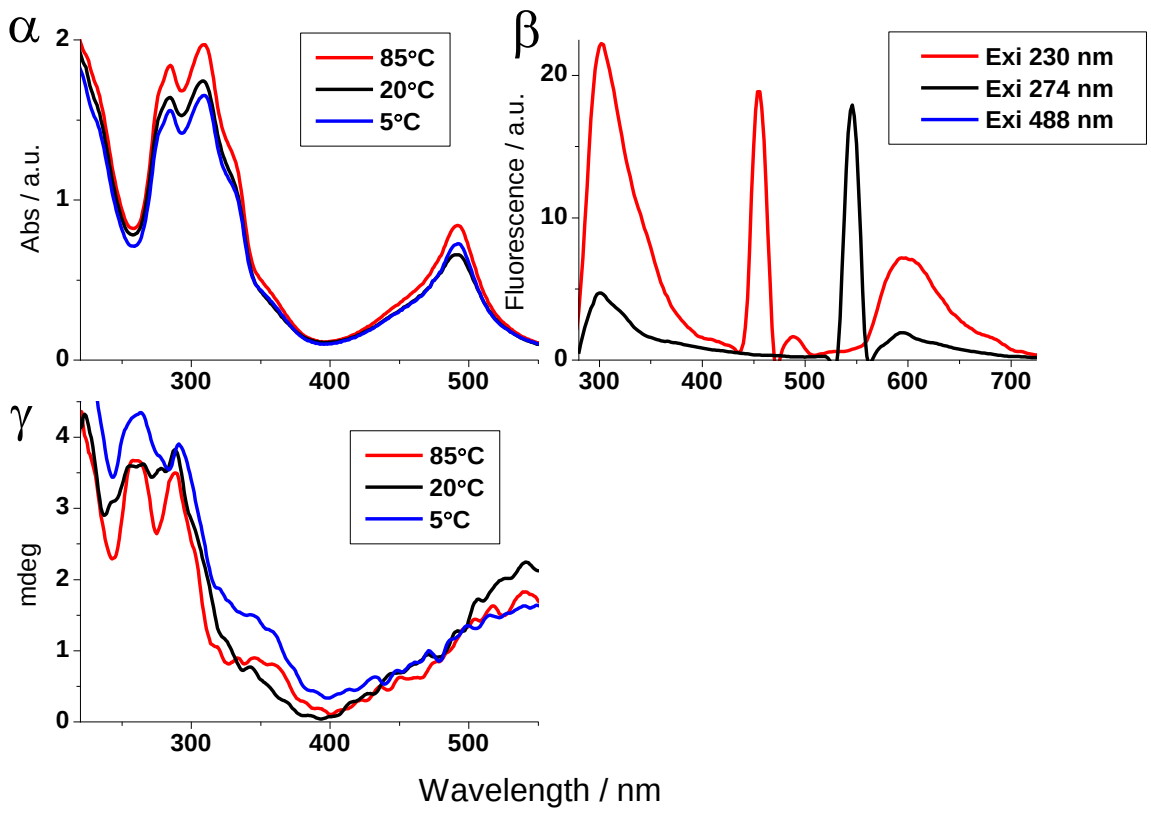
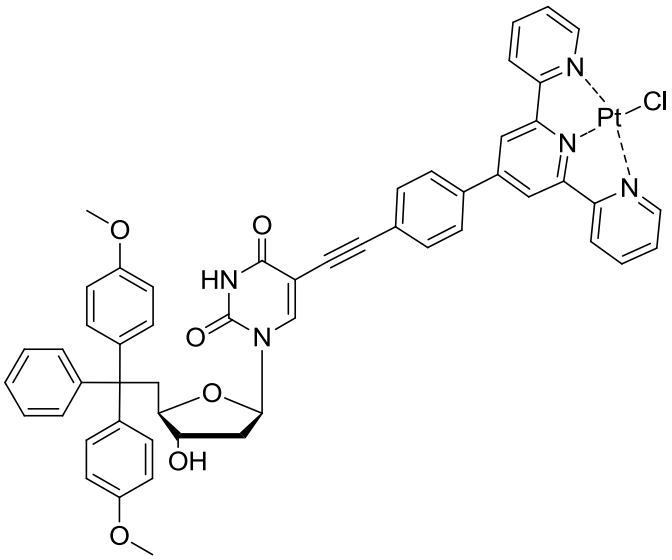


Figure 119 Ruthenium terpyridine monomer (**10**) spectroscopic analysis at 10 μM in ethanol; UV-Vis (α), CD (γ) and fluorescence analysis with varying excitation wavelength (β, harmonic overtones at 460 and 548 nm, a direct result of the excitation wavelength applied), UV-Vis and fluorescence scanning at 600 nm per minute, fluorescence at 500 PMT voltage and CD scanning at 2 seconds per nm.

UV-Vis spectroscopic analysis of the ruthenium *bis*-terpyridine monomer displays absorption bands at around 200, 250-350 and 495 nm, as shown in Figure 37. The absorption below 350 nm is from *bis*-terpyridine absorption, whilst the broad peak at 495 nm derives form ruthenium absorption, all of which slightly increase with higher temperatures. Ruthenium *bis*-terpyridine monomer exhibits limited circular dichroism (CD), which also changes negligibly with increasing temperatures. Excitation at the absorption bands yields low fluorescence, indicating ruthenium quenching properties of the surrounding terpyridines.

8.17 Platinum terpyridine deoxyuridine monomer



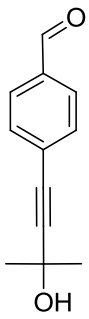
11

4 (46 mg, 53 mmol, 1 eq.) and K_2PtCl_4 (22.2 mg, 53 mmol, 1 eq.) was added to DMSO (2 mL), the solution was stirred at 50 degrees for 15 hours. The reaction was monitored by TLC (eluent EA, R_f 0.1, product appears yellow) The solution was cooled to room temperature, then precipitated by adding the solution dropwise to acetone (100 mL). The solid was washed with diethylether (2 x 50 mL), methanol (2 x 50 mL) and water (2 x 50 mL). The product was dissolved in DMSO (5 mL), to which a saturated solution of KPF_6 (2 mL) was added. The precipitate was washed with diethylether (2 x 50 mL), methanol (2 x 50 mL) and water (2x 50 mL). The product was dissolved in pyridine (5 mL), the solvent removed under ultra high vacuum overnight.

Mass 50 mg Yield 85 %

ESI^+: Calculated 1092.5 Da, Found 1092.8 Da

8.18 4-(3-Hydroxy-3-methylbut-1-yn-1-yl)benzaldehyde



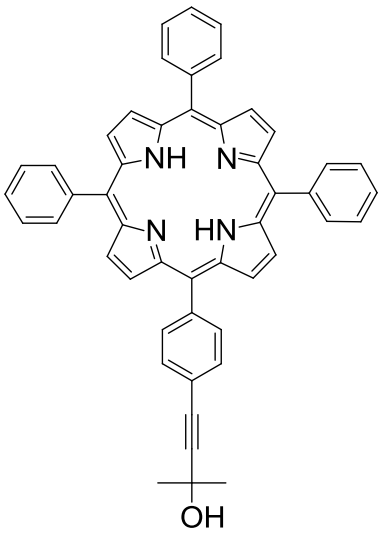
12

p-Bromobenzaldehyde (6.0 g, 32 mmol, 1 eq.), potassium carbonate (20.58 g, 148 mmol, 4.6 eq.), copper iodide (0.55 g, 2.9 mmol, 0.09 eq.), palladium-carbon (0.34 g, , 1.2 mmol, 10 %) and triphenyl phosphine (1.26 g, 4.8 mmol, 0.15 eq.) was added to dimethoxyethane / water (1 : 1, 120 mL) at room temperature and stirred for 30 minutes under nitrogen. 2-Methyl-but-3-yn-2ol (14.7 mL, 150 mmol, 4.7 eq.) was added via syringe, the mixture heated to 80 °C and stirred for 3 hours, the reaction was monitored by TLC (eluent diethyl ether / ethyl acetate 1:1, R_f = 0.77). Upon completion, Pd-C was removed by filtration over celite, taken up in ethyl acetate (100 mL), washed with saturated potassium carbonate in water (50 mL) and water (2 x 50 mL). The solution was dried with magnesium sulphate, filtered, and the solvent removed. The residue was taken-up in dichloromethane and column chromatography was performed. The solvent was removed to give a dark yellow oily product.

Mass 3.65 g Yield 60 %

^{1}H NMR (400MHz ,CHLOROFORM-d) δ = 9.96 (s, 1 H), 7.84 - 7.69 (m, 2 H), 7.51 (d, *J* = 8.1 Hz, 2 H), 2.72 (s, 1 H), 2.92 - 2.62 (m, 1 H), 1.62 (s, 6 H)
^{13}C NMR (101MHz ,CHLOROFORM-d) δ = 191.5, 135.3, 132.0, 129.2, 98.0, 81.1, 65.4, 31.2.

8.19 5-Hydroxy-^{t}propylethinyl-10,15,20-triphenyl-2H-porphyrin



13

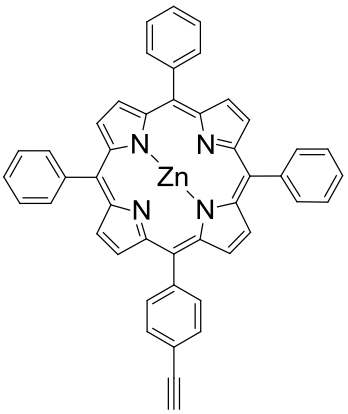
Benzaldehyde (3.4 g, 0.32 mmol, 6 eq.), **12** (1 g, 0.053 mmol, 1 eq.) and pyrrole (2.04 g, 0.32 mmol, 6 eq) was added to degassed chloroform (400 mL) for 20 minutes. Boron trifluoride ethyl etherate (0.41 g, 0.28 mmol, 0.9 eq.) was added, stirred for 1 hour under nitrogen (colour change observed from brown to dark brown / black). 2,3-Dichloro-5-6-dicyano-*p*-benzochinone (4.3 g, 1.91 mmol 6 eq.) was added to quench the reaction and stirred over night, the reaction was monitored by TLC (eluent dichloromethane / methanol 80:1, R_f 0.65). The solvent was removed and column chromatography was performed (eluent dichloromethane / methanol 80 : 1).

Mass 0.67 g Yield 16 %

^{1}H NMR (400MHz ,CHLOROFORM-d) δ = 9.05 - 8.96 (m, 8 H), 8.38 - 8.32 (m, 6 H), 8.30 (d, *J* = 8.0 Hz, 2 H), 7.96 - 7.91 (m, 2 H), 7.88 - 7.77 (m, 9 H), 2.37 (br. s., 2 H), 1.89 (s, 6 H), -2.56 (br. s, 2 H)
^{13}C NMR (101MHz ,CHLOROFORM-d) δ = 133.4, 128.9, 126.6, 125.6, 30.5

MALDI-ToF: Calculated 696.8 Da, Found 696.7 Da $[M+H^{+}]^{+}$

8.20 Acetylene-phenyl-10,15,20-triphenyl-zinc-porphyrin



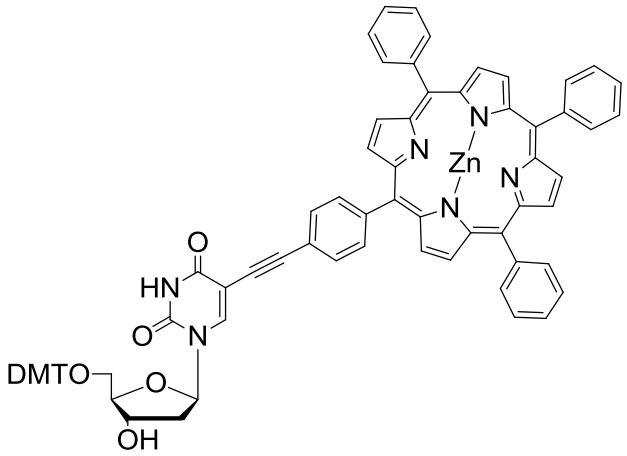
14

13 (0.21 g, 0.28 mmol 1 eq.), sodium methoxide (0.45 g, 8.4 mmol, 30 eq.) and toluene (40 mL) were stirred under nitrogen for 4 hours at 110 °C, the reaction was monitored by TLC (eluent dichloromethane, R_f = 0.2), the solvent removed under vacuum. The product was dissolved in DCM (100 mL), washed, extracted in DCM (3 x 30 mL), after which column chromatography was performed. The pure product was dried overnight under ultra high vacuum to give **14** as a violet powder.

Mass 0.18 g, Yield 95 %

^{1}H NMR (400MHz ,CHLOROFORM-d) δ = 9.04 - 8.95 (m, 8 H), 8.24 (d, *J* = 8.0 Hz, 2 H), 8.30 - 8.20 (m, 9 H), 7.91 (d, *J* = 8.0 Hz, 2 H), 7.84 - 7.74 (m, 9 H), 3.30 (s, 1 H)
^{13}C NMR (101MHz ,CHLOROFORM-d) δ = 134.9, 132.5, 130.8, 127.0, 31.3

8.21 Zinc-porphyrin DMT deoxyuridine monomer



15

14 (0.15 g, 0.021 mmol 1 eq.), triethylamine (200 µl), copper iodide (8.2 mg, 0.43 mmol, 20 %) and DMT iodo-deoxyuridine (0.37 g, 0.32 mmol, 1.3 eq.) was added to DMF (3 mL) and stirred for 10 minutes under nitrogen in the dark. Palladium tetrakis triphenylphosphine (0.14 g, 0.132 mmol, 3 eq.) was added, the reaction was monitored by TLC (eluent dichloromethane / methanol 10 : 1, R_f 0.4). After 30 hours the palladium was removed by filtration over celite, the crude product was taken up in diethyl ether (70 mL) and washed with water (5 x 50 mL), dried with magnesium sulphate and the solvent removed. Column chromatography was performed (eluent dichloromethane / methanol / triethylamine 25:1:0.5 + 10 % silica H), the solvent was removed to give a purple solid.

Mass 0.229 g , Yield 87 %

H^1 NMR (300 MHz, *CHLOROFORM-d*) δ ppm 9.0 (d, 2 H) 8.9 (s, 4 H) 8.9 (d, *J* = 4.6 Hz, 2 H) 8.2 (d, 6 H) 8.0 (d, *J* = 8.1 Hz, 2 H) 7.9 (s, 1 H) 7.7 (d, *J* = 5.8 Hz, 9 H) 7.1 - 7.5 (m, 13 H) 6.8 (d, *J* = 8.4 Hz, 2 H) 6.4 (t, 1 H) 4.1 - 4.2 (m, 1 H) 3.8 (s, 1 H) 3.7 (s, 6 H) 3.3 - 3.4 (m, 1 H) 3.0 - 3.1 (m, 1 H) 2.8 (s, 1 H) 2.7 (s, 1 H) 2.1 - 2.1 (m, 1 H) 1.9 - 2.0 (m, 1 H)

MALDI-ToF: Calculated 1230.71 Da, Found 1230.7 Da $[M\text{-}H^{+}]^{-}$

UV-vis:

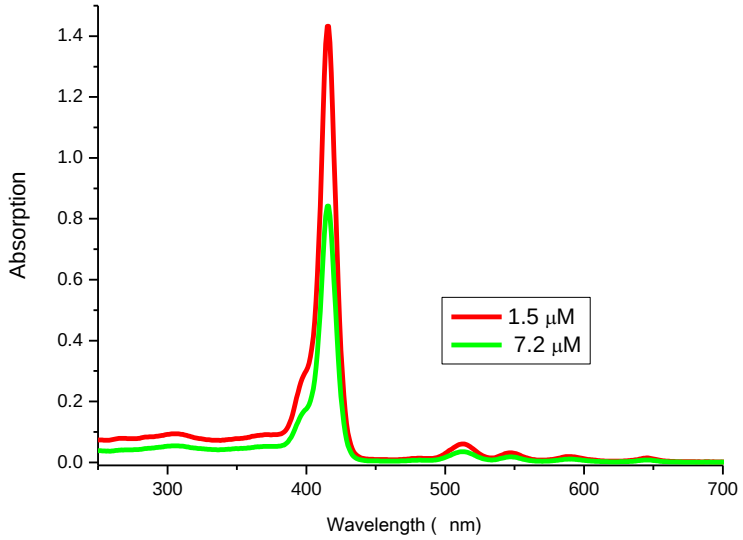


Figure 120 Fluorescence analysis of DMT-deoxyuridine 2H porphyrin, excitation at 330 nm, in ethanol, scanning at 600 nm per minute using a 1cm quartz cell.

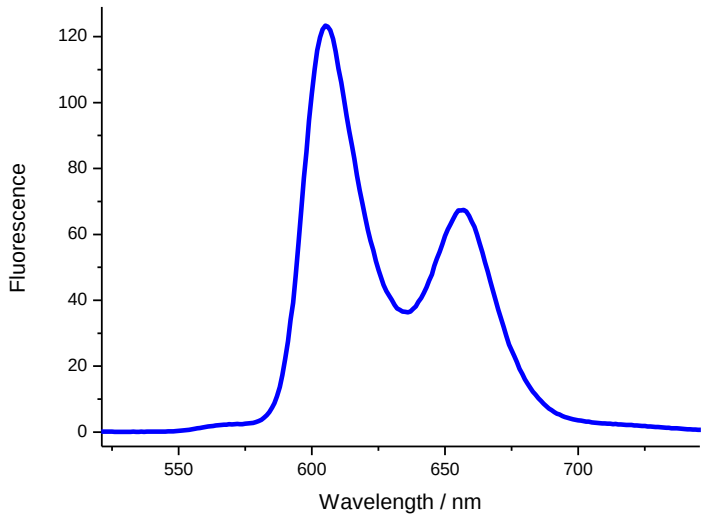
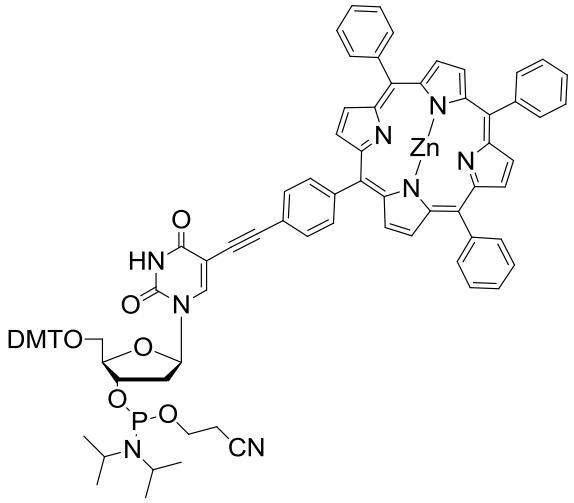


Figure 121 Fluorescence analysis of DMT-deoxyuridine 2H porphyrin, excitation at 330 nm, 1.5 μM in ethanol, 1 cm path length cell, 500 PMT voltage, scanning at 600 nm per minute.

8.22 Zinc-porphyrin DMT deoxyuridine phosphoramidite monomer



16

15 (100 mg, 0.081 mmol, 1 eq.), DIPEA (55 µl, 0.32 mmol, 4 eq.) and CEP-Cl (54 µl, 0.24 mmol, 3 eq.) was added to degassed DCM (3 mL). The solution was stirred in the dark under N_2 for 2 hours. After the reaction had gone to completion as indicated by TLC (eluent DCM / MeOH 1 : 1, R_f = 0.6), the product was precipitated in degassed hexane (2 x 5 mL). The solvent was removed by filtration, washed with more hexane (2 x 1 mL), and dried under ultra high vacuum for 20 minutes. The product was dissolved in DCM / CH_3CN (10 : 1, 2 mL) and transferred to the DNA synthesiser.

Table 12 Porphyrin DNA yields

Porphyrin DNA	MW (Da)	ε	Nanomoles
H^1	7114.2	224000	373.1649295
H^2	7114.2	224000	380.7958849
H^3	7114.2	224000	302.127792
H^4	7114.2	224000	396.9872137
H^5	7114.2	224000	365.899556
H^6	7114.2	224000	371.3364922
H^7	7114.2	224000	365.0519781
Z^1	6976.2	204100	373.1649295

8.23 Tus Expression

Luria broth solution was prepared (5 g NaCl, 5 g Tryptone, 2.5 g yeast extract per 0.5 L) and autoclaved (heating to 80 °C). To Luria broth solution (20 mL) under sterile environment, Tus cell stock phcs151 (< 1 μL) was added, solution was stirred overnight at 37 °C at 200 rpm. The solution was transferred to a conical flask (5 L), Luria broth (980 mL) and ampicillin (1 mL) was added, the solution was stirred at 37 °C at 200 rpm. The sample was monitored by UV-Vis spectroscopy at 580-620 nm, after 3 hours, when the absorption reached 0.4 a.u, IPTG (1 M, 1 mL) was added, and the solution stirred for a further 4 hours. After which, the solution was centrifuged for 20 minutes at 4 °C, at 20'000 rpm for 20 minutes. The pellets were collected, weighed (~ 65 g), and dissolved in lysis buffer (50 mM Tris pH 6.8, 1 mM EDTA, 10 % w / v Sucrose, 50 mM NaCl, 0.2 mg per mL lysozyme). The solution was frozen at - 80 °C, and then thawed to room temperature; this process was repeated 3 times. The solution was centrifuged for 20 minutes, at 4 °C at 20'000 rpm, the supernatant was collected, the NaCl concentration was increased to 1 M. The solution was stirred on ice for 1 hour, after which ammonium sulphate (0.07 g / mL) was added, and stirred for a further 20 minutes. The solution was centrifuged (18000 rpm, 20 min. 4 °C), the supernatent collected, amonium suplhate concentration was increased (0.29 g / mL, obtained 175 mL), stirred for 1.5 hours under ice. Centrifuged again, the solid was collected and dissolved in loading buffer (50 mM Tris, 50 mM NaCl, pH 7.4). The protein was purified by nickel affinity chromatography, and eluted using eluting buffer (50 mM Tris pH 7.4, 50 mM NaCl, 500 M Imidazole). Concentrated (100 → 10 mL) using milipore 0.4 micron filter paper, and purified using HPLC gel filtration (S200 column, buffer: 0.25 M KCl, 50 mM Tris-HCl, 1 mM EDTA, 1 mM DTT pH 7.4). To the pure protein, glycerol (10 % v / v) was added, the solution then concentrated (1 mg / mL) and aliquoted, and stored at - 80 °C.

See synthesis and biosynthesis chapter 2 for Tus analysis.

8.24 UV-Vis melting analysis

TerB:

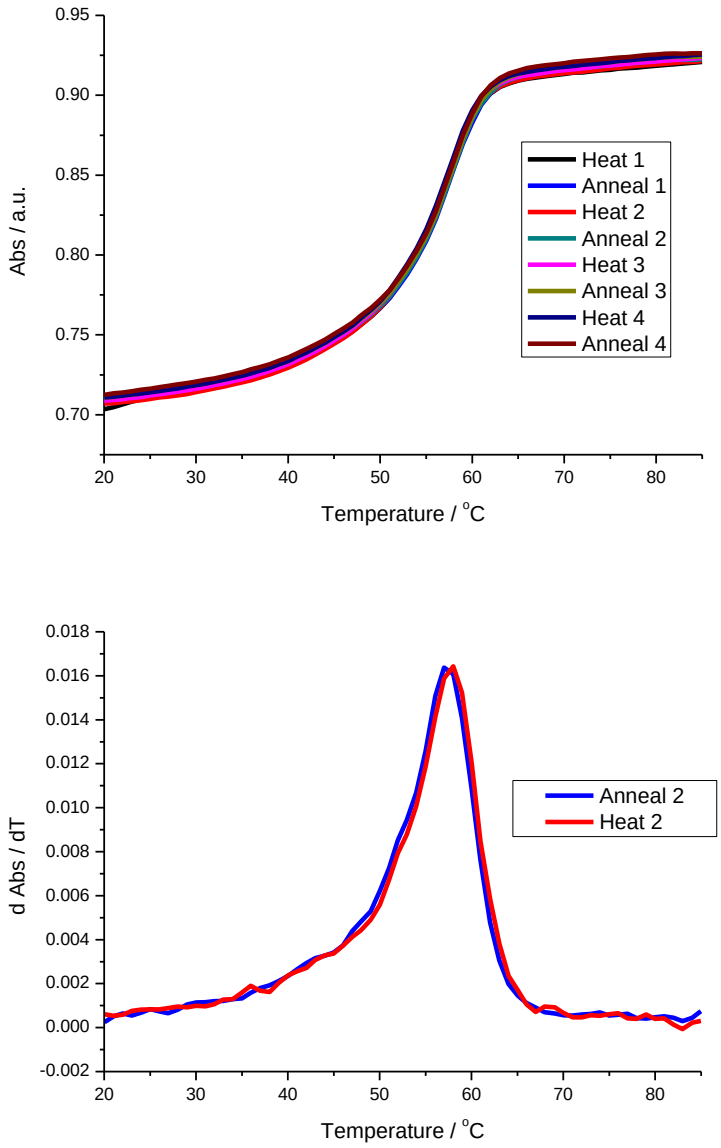


Figure 122 UV melt at 260 nm of *TerB* duplex at 1 μM, in Tus buffer, 1 cm path length quartz cell, heating and annealing a 1 °C per minute.

Extended:

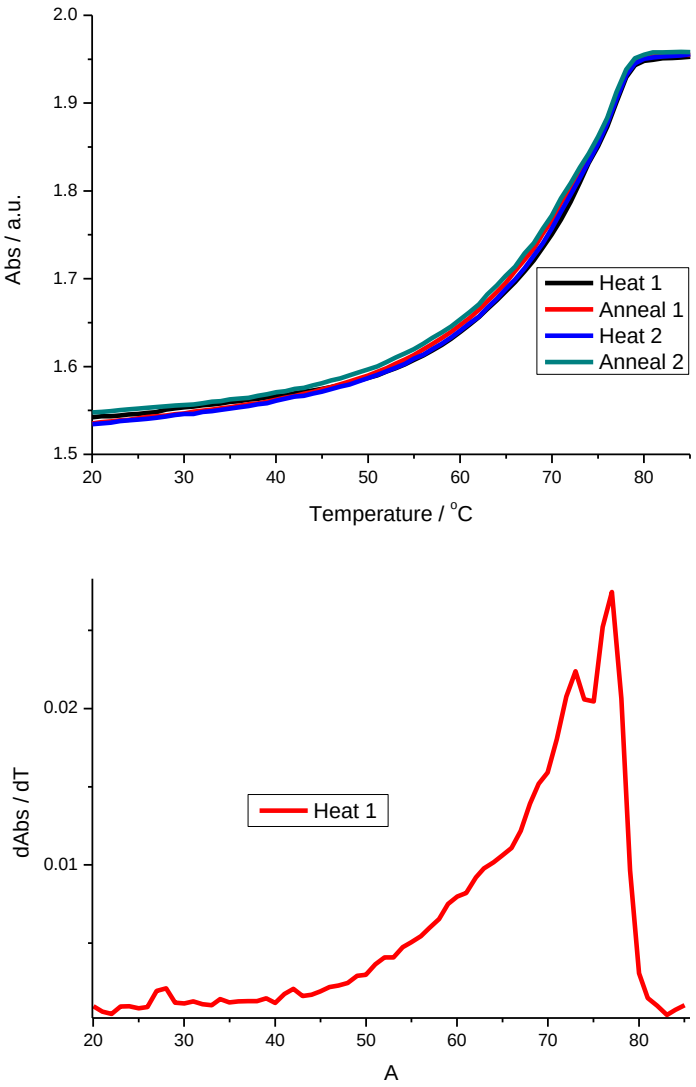


Figure 123 UV melt at 260 nm of Extended duplex at 1 μM, in Tus buffer, 1 cm path length quartz cell, heating and annealing a 1 °C per minute.

Permissive:

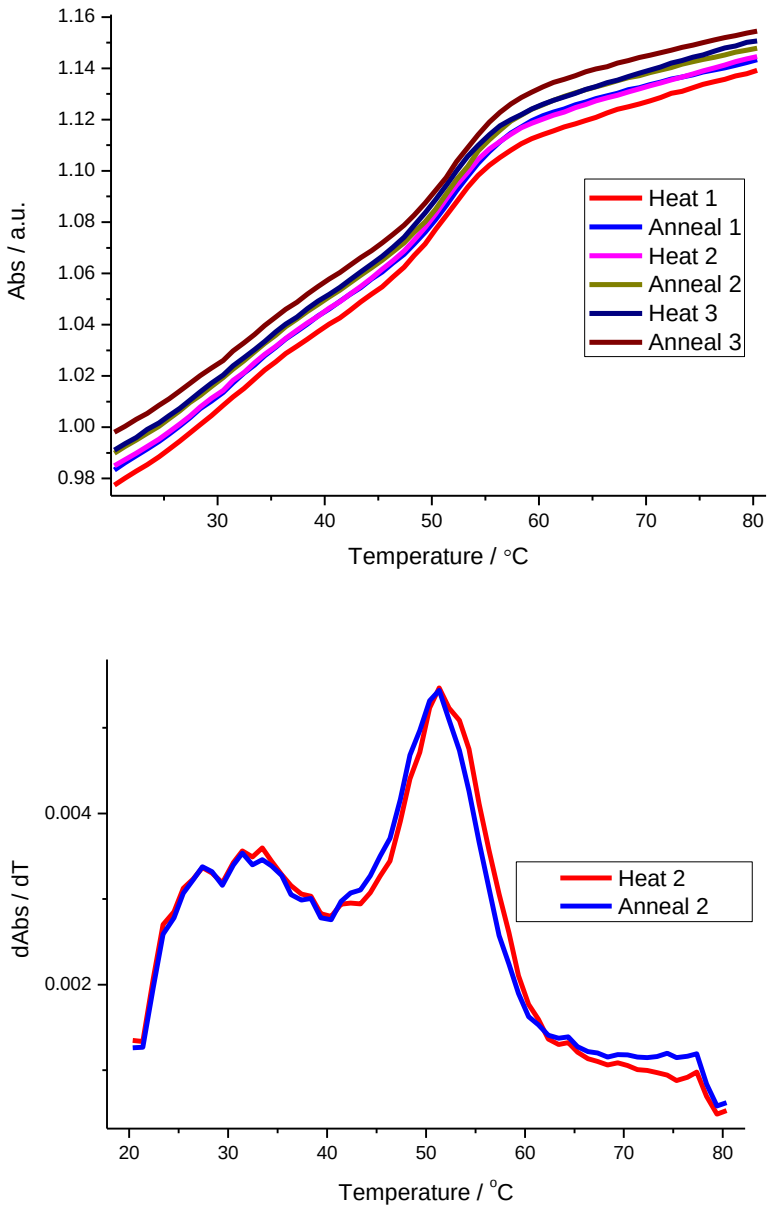


Figure 124 UV melt at 260 nm of permissive duplex at 1 μM, in Tus buffer, 1 cm path length quartz cell, heating and annealing a 1 °C per minute.

Non-permissive:

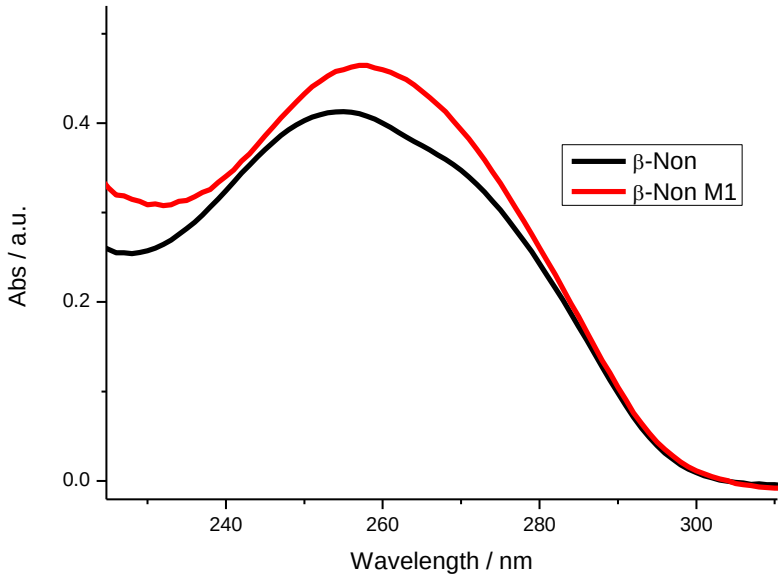


Figure 125 UV-Vis spectra of β-Non permissive strand (black) and the β-Non permissive mutant (red), approximately 5 μM, in Tus buffer, 1 cm path length quartz cell, heating and annealing a 1 °C per minute.

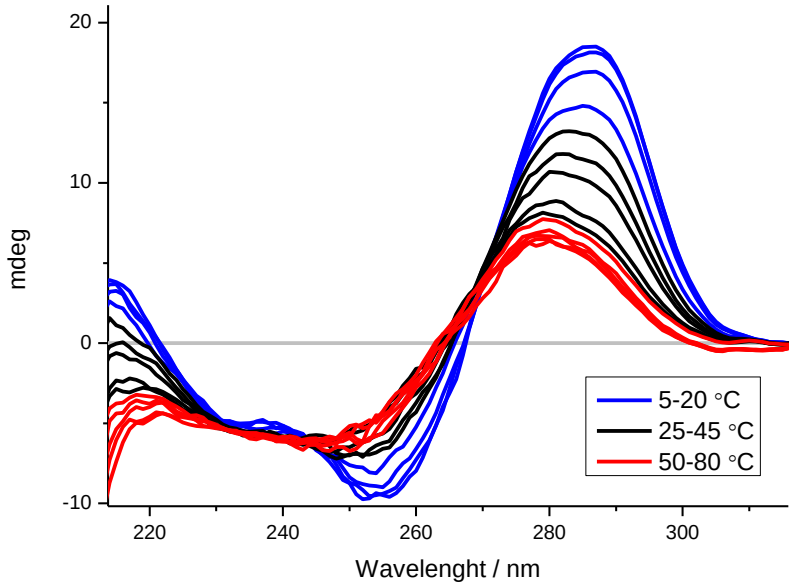


Figure 126 CD melt of β-Non permissive strand at 15 μM, in Tus buffer, 1 cm path length cell, heating at 5 degree intervals, scanning a 2 seconds per nm.

Non-permissive M1:

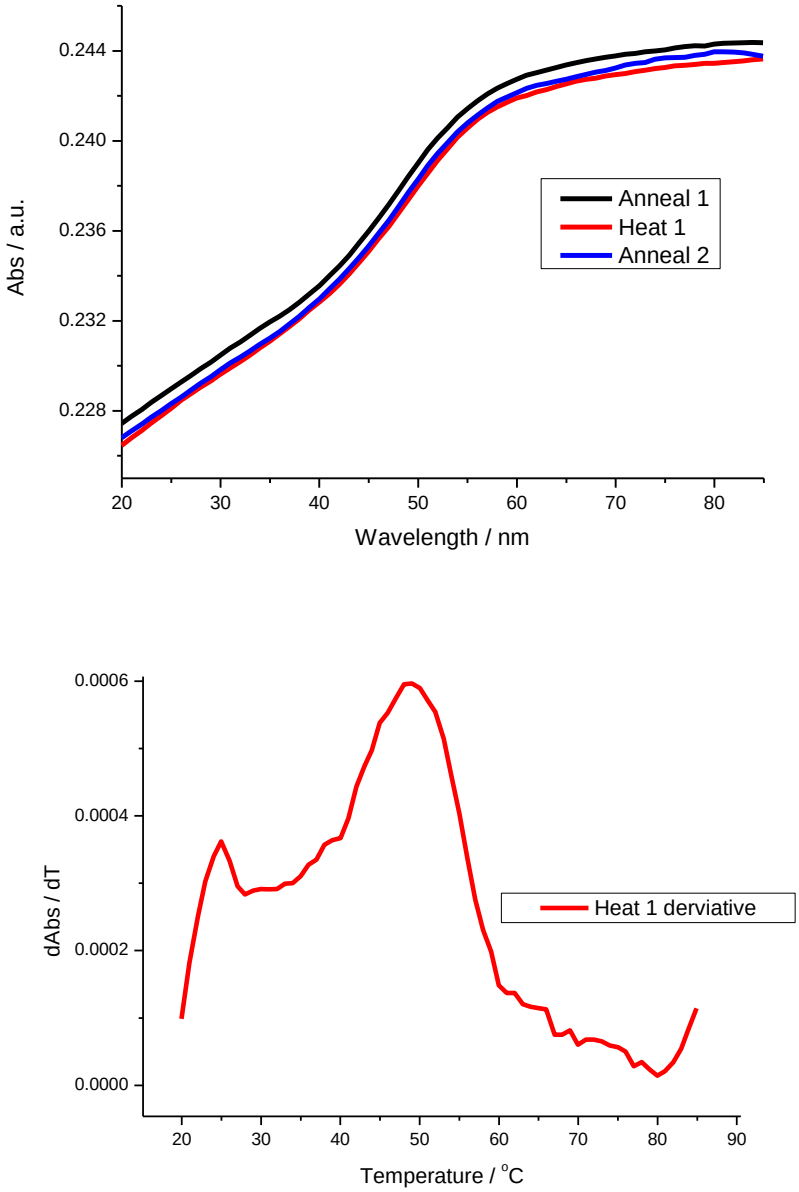


Figure 127 UV melt at 260 nm of permissive mutant 1 duplex strand at 1 μM, in Tus buffer, 1 cm path length quartz cell, heating and annealing a 1 °C per minute.

8.25 FAM and TAMRA Quantum Yield, ε and Spectral Overlap:

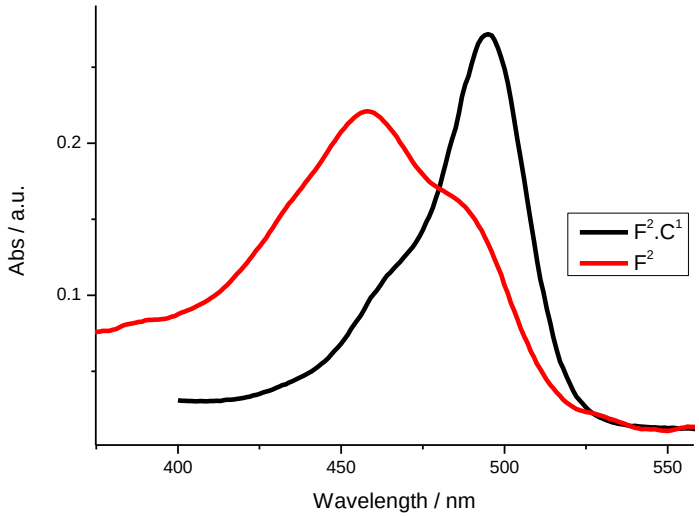


Figure 128 UV-Vis spectra of duplex (black) and single-strand (red) F^2, in Tus buffer, 1 cm path length quartz cell, scanning at 600 nm per minute.

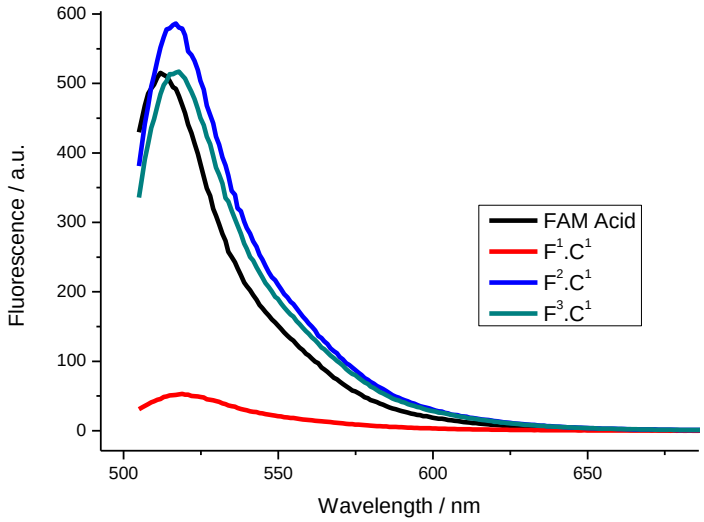


Figure 129 Fluorescence of duplex F^1, F^2, F^3 and FAM acid, in Tus buffer, 1 cm path length quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

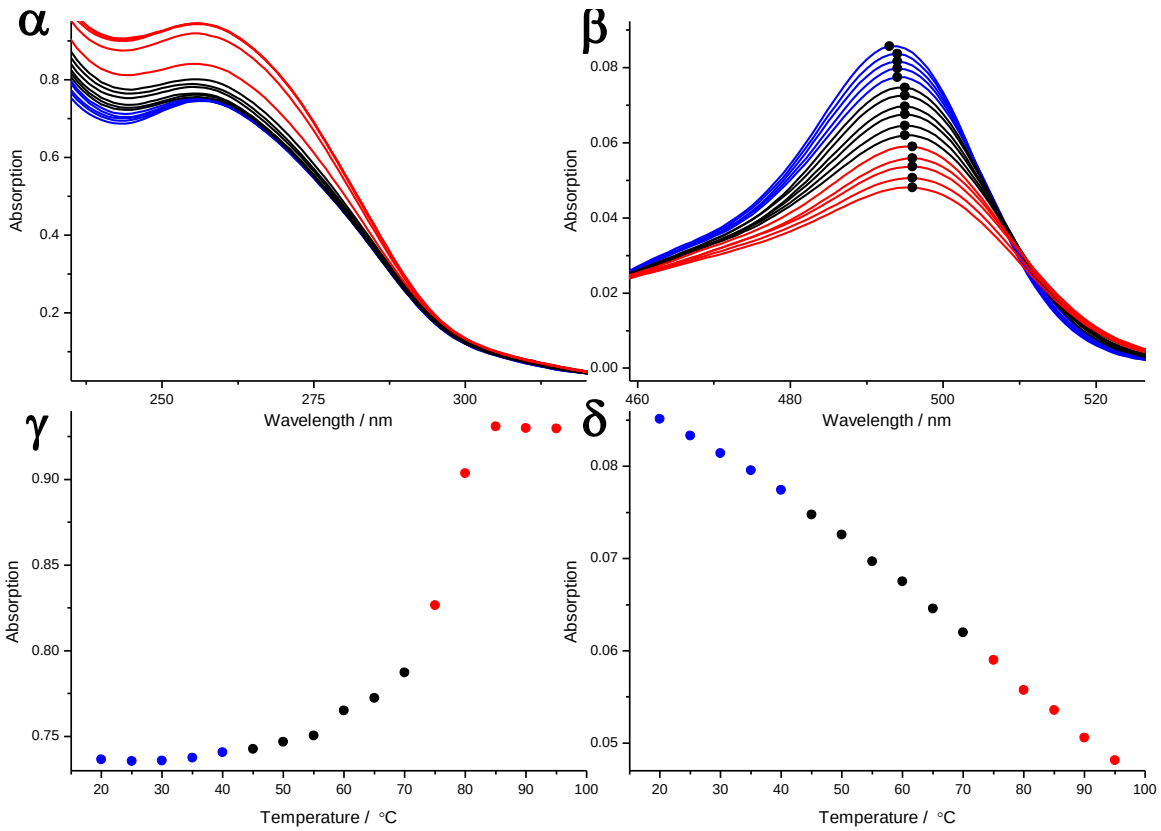


Figure 130 UV-Vis melting spectrums of F^3 Extended DNA with unmodified complement; DNA absorption (α), FAM absorption (β) and melting cross-sections at 260 nm (γ) and 495 nm (δ), at 2 μM, melting at 5 °C intervals, scanning at 600 nm per minute, 1 cm quartz cell, blue lines 10-40 °C, black lines 45-70 °C and red lines 75-95 °C, scanning at 600 nm per minute, heating at 5 °C intervals.

8.26 FAM and TAMRA calculated EFRET vs. donor-acceptor distance:

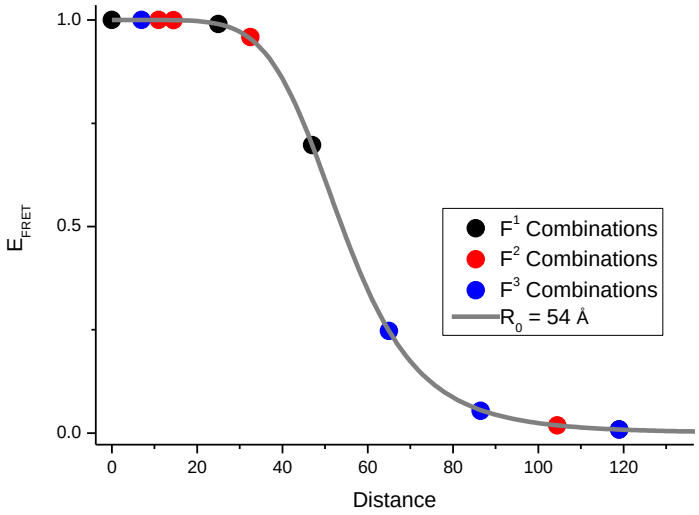


Fig. 131 Calculated E_{FRET} versus Distance for FAM and TAMRA combinations, calculated using the 3.6 Å model as used by Lilley *et al.*[3].

8.27 Porphyrin DNA with Tus native PAGE analysis

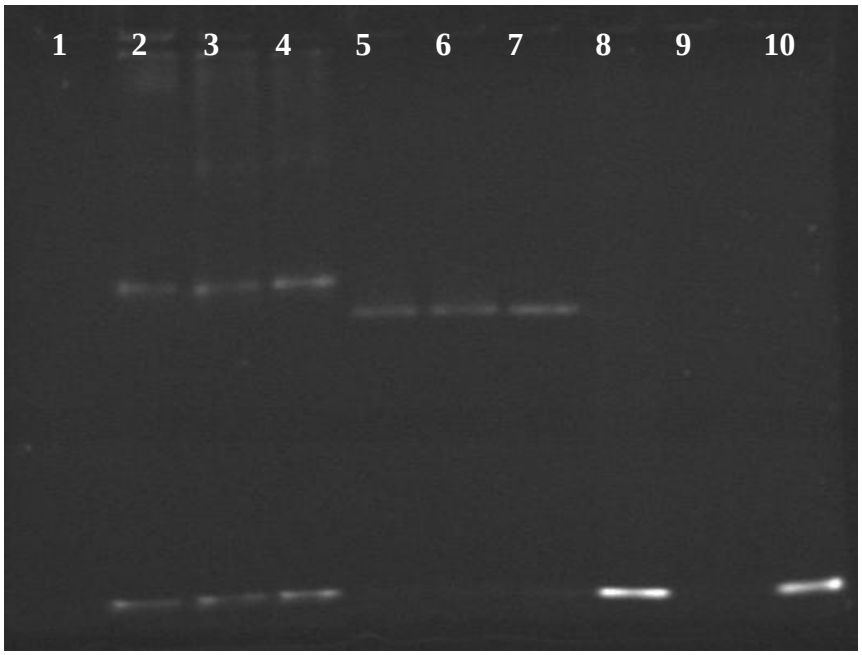


Figure 133 $Z^1.P^2$ Native PAGE analysis with varying Tus molar ratio, ethidium bromide stain, imaged under UV light. Lane 1 = Tus 400 nM only, lane 8 = $Z^1.P^1$ only, lane 2 = Tus 400 nM with $Z^1.P^1$ 400 nM, lane 3 = Tus 400 nM with $Z^1.P^1$ 800 nM, lane 4 = Tus 400 nM with $Z^1.P^1$ 1600 nM.

8.28 Tus-*Ter* X-ray crystal simulated SAS graphs

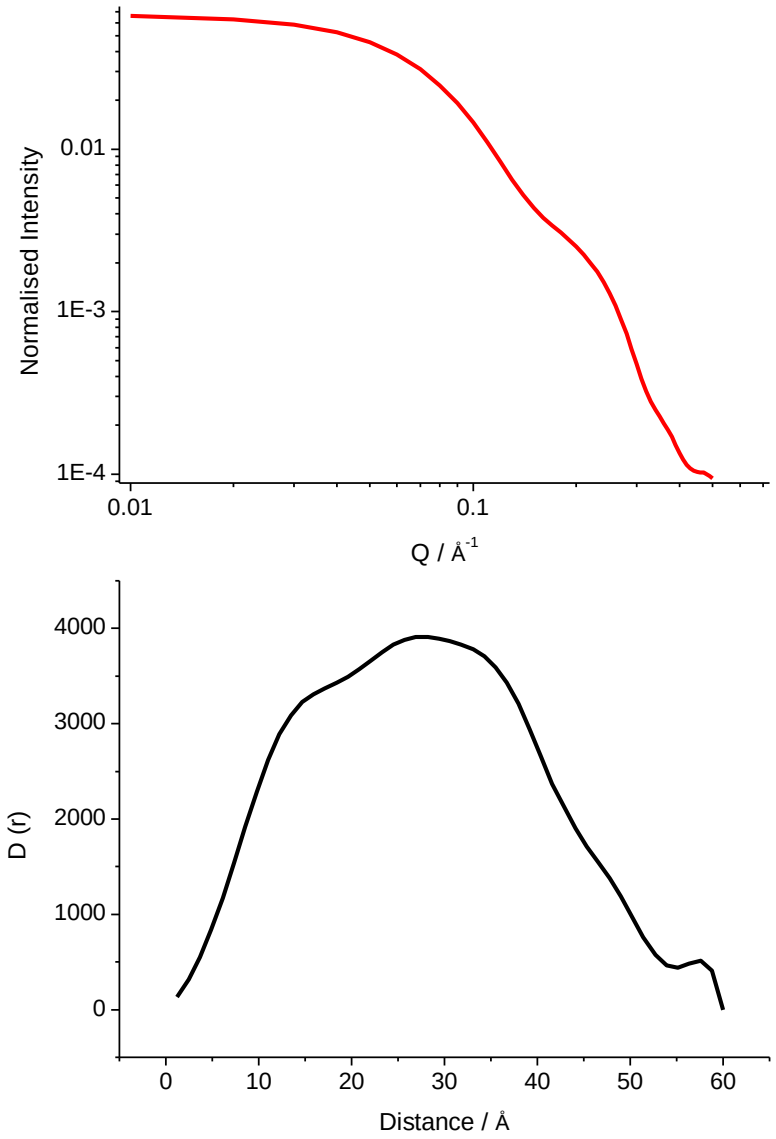


Figure 134 Tus-*Ter* X-ray crystal structure scattering intensity versus distance (top), and distance distribution (bottom) graphs, calculated by Dr. Luke Clifton using Crysol software.

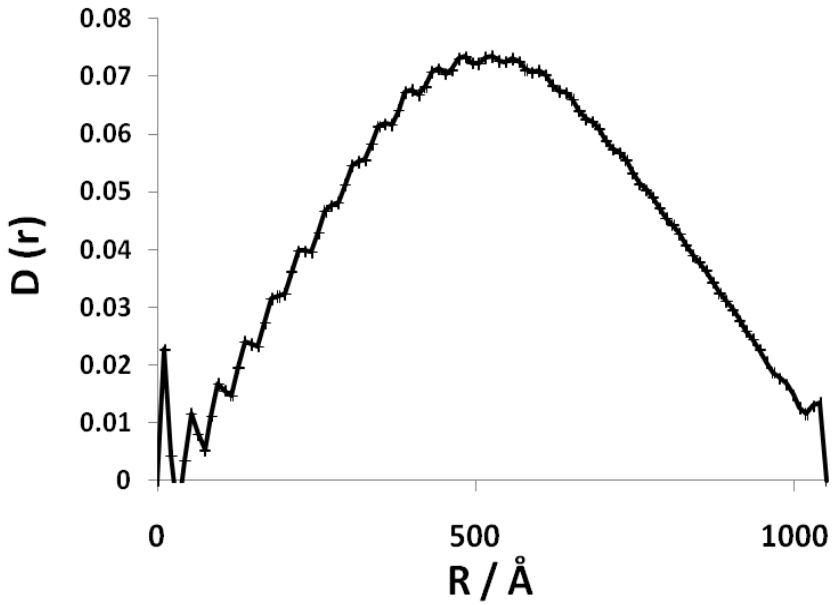


Figure 135 Tus SAXS distance distribution graph, 100 μM, in Tus buffer.

8.29 CD beamline graphs

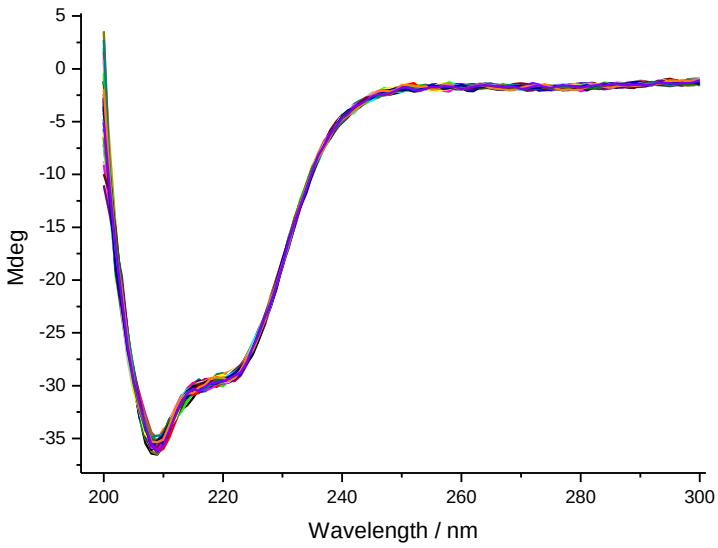


Figure 136 Tus CD beamline stability test, 7 μM, 0.28 mm slit width, Tus buffer, 1 mm quartz cell, scanning at 2 seconds per nm.

8.30 CD titrations

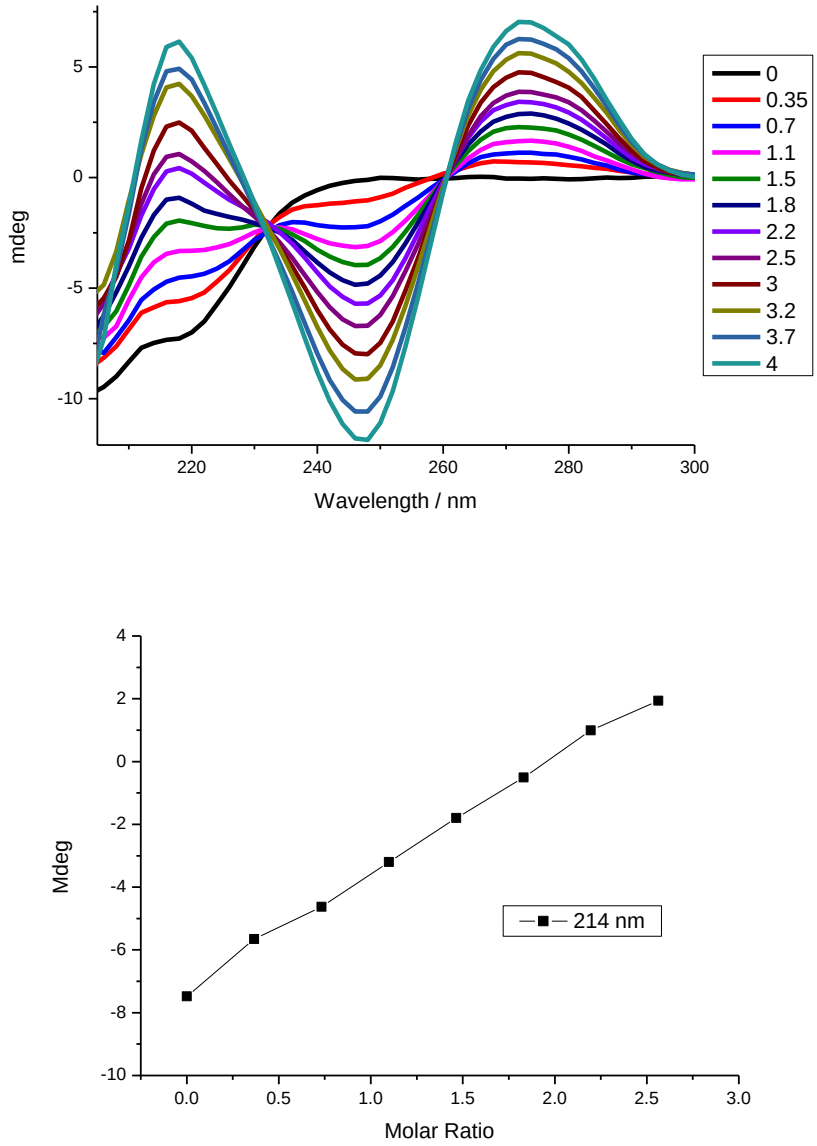


Figure 137 Tus-*TerB* CD titration, Tus 1.4 μM in Tus buffer with increasing TerB molar ration, scanning at 2 seconds per nm, 0.1 mm path length quartz cell.

8.31 AS sequences

Figure 138 The five AS hairpin combinations with complementary DNA sequences and the AS only sequence. AS sequence shown in black, hairpin region shown as black line, complementary sequence shown in grey, **D** = donor, **A** = acceptor.

8.32 AS melting curves

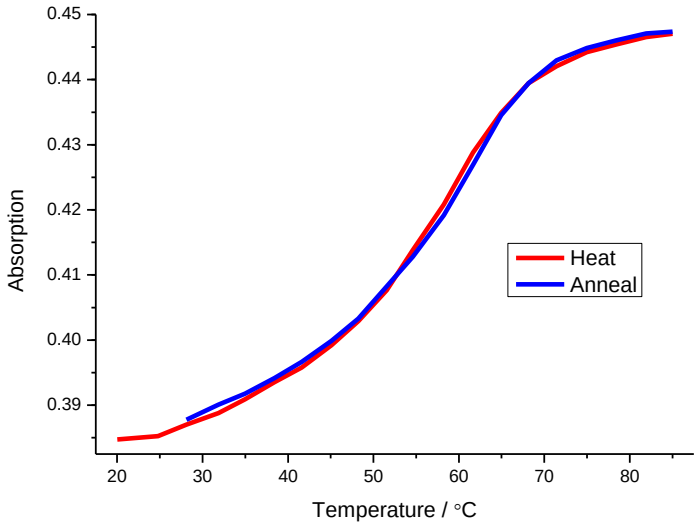


Figure 139 UV melting profile of S^1 only, at 1 μM in 0.1 M sodium phosphate buffer, heating and annealing at 1 °C per minute, 1 cm quartz cell.

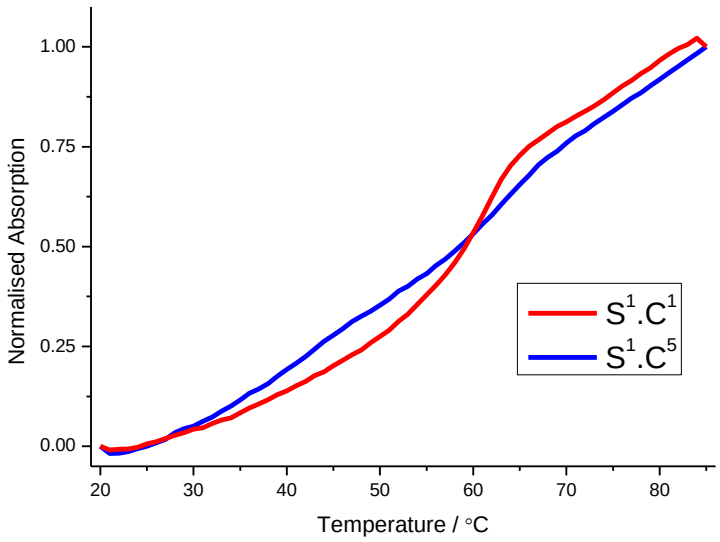


Figure 140 UV normalised melt spectrum of $S^1.C^1$ and $S^1.C^5$, at 1 μM in 0.1 M sodium phosphate buffer. Heating and annealing at 1 °C per minute, 1 cm quartz cell.

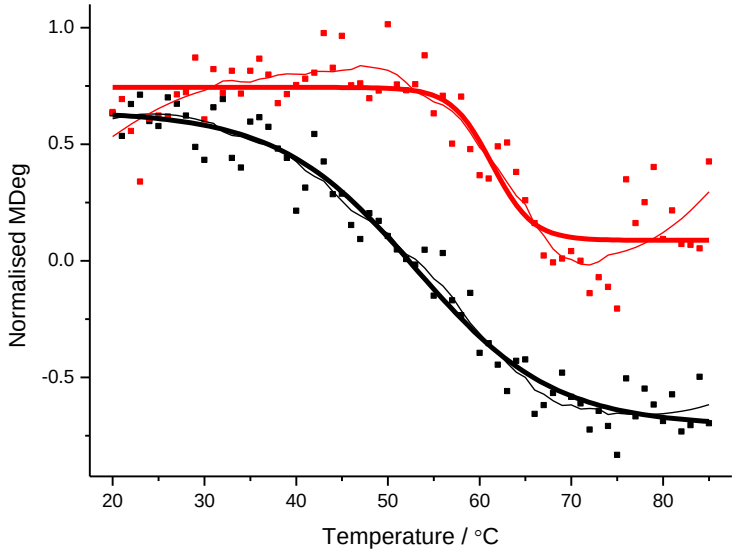


Figure 141 CD melting profile of $S^1.C^1$ (red) and $S^1.C^5$ (black) combinations, melting at 1 °C per minute, at 1 μM, 1 cm quartz cell. Raw CD data (dots), smoothed (thin lines) and boltzman fit (thick lines).

8.33 Deconvoluted S^1 AS combinations fluorescence peaks

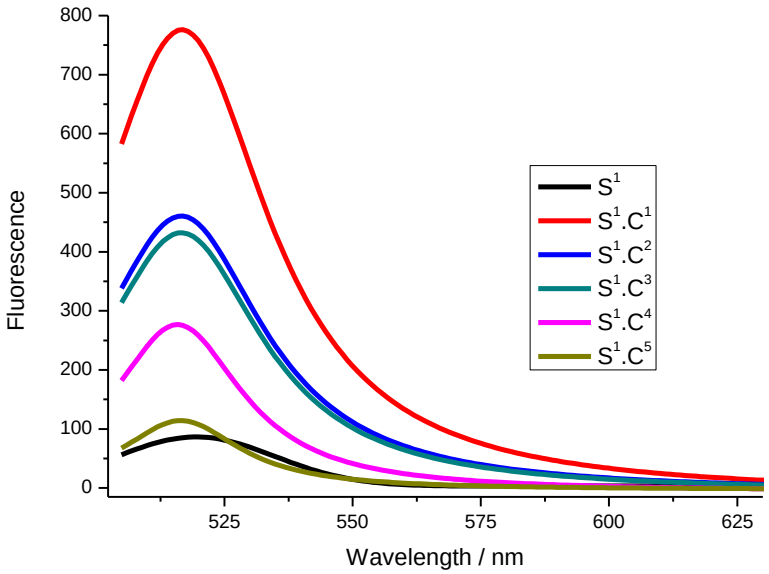


Figure 142 S^1 AS combinations, FAM deconvoluted fluorescence emission peaks, 1 μM in 0.1 M sodium phosphate buffer, excitation at 495 nm, scanning at 600 nm per minute, 500 PMT voltage.

8.34 Terpyridine arrays AFM analysis

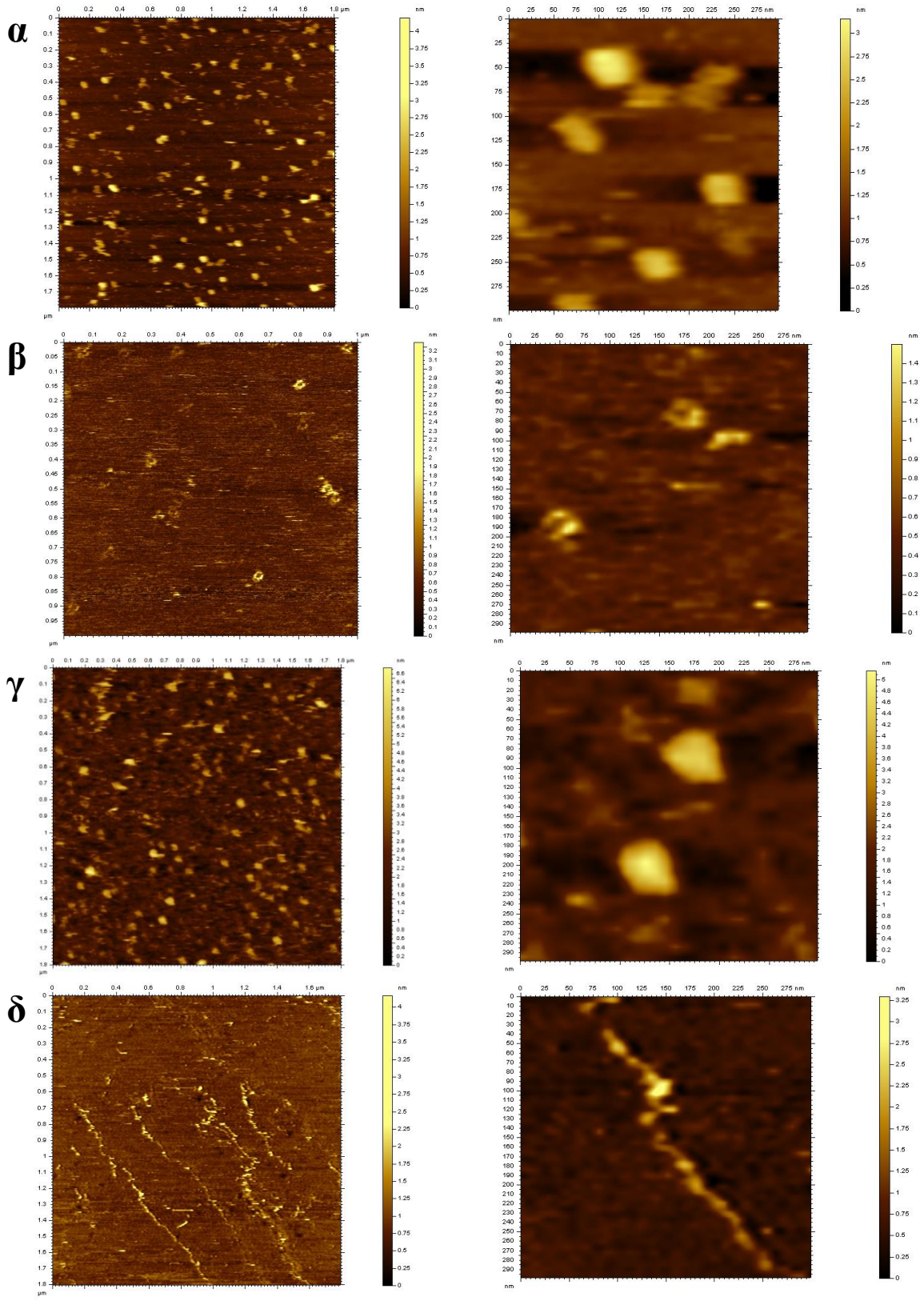


Figure 143 AFM images of the DNA systems. (α) 2.3. (bβ) 2.3 + Zn(II). (γ) 2t.3t. (δ) 2t.3t + Zn(II). Scan size is 1.8 μm (left) and 300 nm (right) for all pictures except (b), which is 2 μm and 300 nm, images taken from Burns *et al.*[111].