Construction of DNA-Based Photonic Wire Assemblies by Programmable Polyamides

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Stephen James Gray MChem (Leicester)

Department of Chemistry

University of Leicester

July 2011

Abstract

Construction of DNA-based Photonic Wire Assemblies by Programmable Polyamides

Stephen James Gray MChem (Leicester)

A key problem in nanotechnology is the integration of individual components into larger networks capable of more complex processes. DNA based photonic wires are a promising solution as they have been shown to transmit light energy over 10 nm distances, but are limited by their problematic assembly and reliance on fluorophore labelled DNA. This thesis describes efforts to construct an improved photonic wire using functionalised DNA binding small molecules as proof of principle for a 'mix and match' approach to nanotechnology which delivers individual components to a specific site on DNA.

Polyamides have been shown to bind to DNA with very high affinity and specificity which together with their modular nature makes them an ideal 'delivery system'. To combine this with the versatility and efficiency of copper catalysed click chemistry, novel internally functionalised alkyne polyamides were synthesised using both solution and solid phase chemistry. A general route to produce these internally modified polyamides was developed and the synthesis of the standard polyamide building blocks was improved. Test click reactions on alkyne polyamide fragments showed up to 92% conversion, but the same reactions failed on the full length polyamides and previously reported modification methods were used to create a fluorophore labelled polyamide.

A coumarin based fluorophore was selected to allow direct substitution into proven photonic wires, but when the DNA binding affinity of this polyamide was tested, it was found that only weak binding was observed with 1.5 equivalents of polyamide. Upon construction, the improved photonic wire transported energy over a distance of 6 nm with an overall efficiency of 9% which was attributed to the poor DNA affinity. This poor performance makes it difficult to assess the general potential for this 'mix and match' approach, but the non-applicability of click chemistry and improvements in the synthesis will inform future designs.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Glenn Burley for providing me with the opportunity to work on this project and never running out of ideas when things didn't go to plan. I'd also like to thank our collaborator Prof. Clive Bagshaw for use of his labs fluorescence spectrometer and other equipment as well as all his useful advice. My second supervisor, Prof. Paul Cullis has also been a great source of friendly advice and information, particularly in the short months before Glenn's arrival and since his departure for pastures new. Many people have passed through the doors since I began working in the Burley group and I would like to thank all of the project students who have contributed to this work as well as my fellow PhD students who have often provided a fresh perspective. In particular I would like to thank Andrew Perrett and Helen Lewis for aiding me with the synthesis of the DNA strands and Dr. Wu Su who carried out the fluorescence measurements and provided key advice during the solid phase synthesis of the polyamides. This project would not have been possible without the help and technical support of the many instrument specialists I have worked with: Dr. Gerry Griffith (NMR), Dr. Graham Eaton and Dr. Emma Harry (ESI mass spectrometry), and Dr. Sharad Mistry and Shairbanu Ibrahim (MALDI mass spectrometry). I would also like to thank Mick Lee for his help and the rest of the organic lab, past and present for providing a stimulating and friendly working environment. I could not have made it through this project without the support of my family and friends, particularly Malcolm Spence and Jenny Hartley who helped me preserve my sanity during the most trying times. Finally I would like to thank Prof. Eric Hope for introducing me to this project and for providing such a warm welcome when I first came to this department as an undergraduate nine years ago.

List of Abbreviations

BIM	Tris((1*H*-benzo[d]imidazol-2-yl)methyl)amine
Boc	Tert-butyl carbonate
bp	base pairs
BTC	bis(trichloromethyl) carbonate
DABCO	1,4-Diazabicyclo[2.2.2]octane
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
DIEA	Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic Acid
ESI	Electrospray ionization
EtOAc	Ethyl acetate
FRET	Fluorescence or Förster Resonance Energy Transfer
HATU	*O*-(7-Azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate
HBTU	*O*-Benzotriazole-*N*,*N*,*N*’,*N*’-tetramethyluronium hexafluorophosphate
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	1-Hydroxybenzotriazole
MALDI	Matrix Assisted Laser Desorption Ionization
nm	nanometres
NMP	*N*-Methylpyrrolidone

PCR	Polymerase Chain Reaction
TCEP.HCl	Tris(2-carboxyethyl)phosphine hydrochloride
TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TPW	Trifluoroacetic acid (92.5% v/v), Phenol (2.5% w/v), Water (5% v/v)
TRIS.HCl	Tris(2-carboxyethyl)phosphine hydrochloride

Contents

Chapter 1 DNA nanotechnology

1.1 The need for new nanotechnology

Research in the area of nanotechnology has become increasingly important due to the ubiquitous use of electronics in modern devices creating a consumer as well as scientific demand for cheaper, smaller and smarter microprocessors for new applications and improving existing technology. The processors powering the latest generation of computers already contain discrete features as little as 32 nm apart with this distance expected to drop to 10 nm by 2022[1] but the technology required to meet this projection has only been developed for the next few years and significant challenges are expected. Nanotechnology is a broader area than just electronics and covers all areas working on the nanometre scale where quantum mechanical effects begin to alter the properties of objects[2] and other deviations can be observed in the behaviour of materials compared to larger samples, attracting more fundamental research.

In particular, research into these novel properties has led to the development of a fundamentally different way of thinking which avoids the problems of 'sculpting' and controlling objects on this scale associated with traditional 'top-down' manufacturing technology by instead constructing devices from simple molecular building blocks[3] (Figure 1). A limited number of disparate building blocks have already been created that show the merit of this 'bottom-up' approach, but the lack of a common 'language' to govern their behaviour and interactions limits their practical applications. In order for real progress to be made a common, a versatile and robust system is needed to ensure reliable and scalable control of assembly, communication and positioning within a

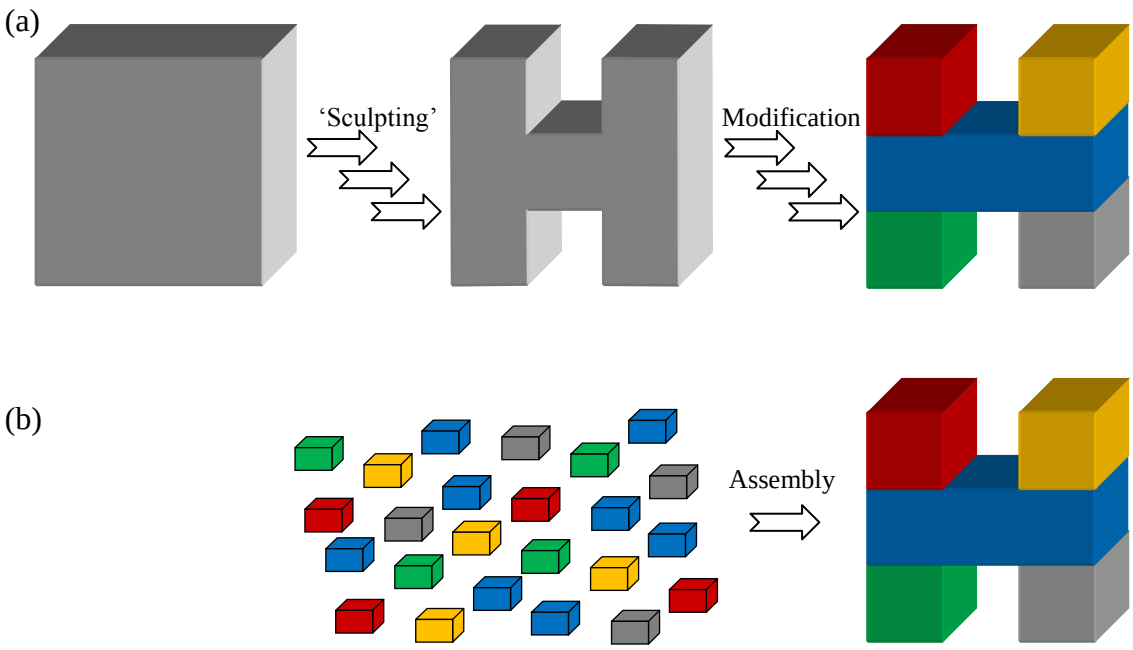


Figure 1 *- Creation of a 'H' motif using the* ***(a)*** *top-down and* ***(b)*** *bottom-up approaches*

larger framework in order to recreate the complex networks of components modern devices require.

Developing a system capable of achieving these objectives is a difficult challenge but this is only one of the barriers that must be overcome before this new 'bottom-up' approach to nanotechnology can fully mature. Pioneering work by Rothemund[4] and Seeman[5] has shown how DNA can be manipulated to produce a programmable scaffold that is rich in information and dynamic 3D networks have also been demonstrated,[6] but DNA alone is too limited to produce the degree of complexity required. Previous work concentrated on addressing this issue has focussed on proteins[7-11] and single stranded regions of DNA to 'address' sites on the scaffold, but these involve a significant synthetic challenge and are ill-suited for larger networks. DNA binding small molecules offer an alternative method to address a 'payload' to specific locations on a scaffold with nanoscale precision which can be used as a general mechanism for a wide variety of materials and target sequences. This capability is important for relatively small functionalities such as fluorophores which interact with light and provide a way of

connecting the nanometre scale assemblies with the macrosized world in which they must operate, another of the key challenges associated with the 'bottom-up' approach. As the energy of visible light and UV photons is significantly above the thermal background, it is able to provide a strong driving force for changes within the building blocks, making it far superior to alternative methods of interacting with the building blocks such as the addition of small control molecules or electrochemistry which are also much slower to respond to changes when manipulating components.

These principles can be applied to design a 'toolbox' of mix and match components where any building block can be positioned at any position on a DNA scaffold. Significant progress has recently been made in the monofunctionalisation of potential nanometre scale building blocks such as nanoparticles[12, 13] and quantum dots[14] with DNA, but first reliable methods of controlling the attachment of smaller molecules must be established. In addition to this, specific connections analogous to the wires in conventional electronic circuits also need to be formed between the individual components to allow them to interact and transfer signals. This work aims to address this problem by creating a light based nanometre scale wire that demonstrates that these design principles are generally applicable in the construction of full 'bottom-up' nanotechnology devices capable of surpassing existing technology.

1.2 Lithography and the top-down approach

The underlying methods used today for the mass production of microprocessors have remained essentially unchanged since their development in the 1960's,[15] where the small scale features are produced on a semiconductor surface using light based lithography. The capabilities of this lithographic process are dependent on controlling the pattern of the modification through the mask. Lithography is typical of 'top-down'

techniques in that it is the challenge of controlling large scale objects such as templates with sufficient precision that limits the size of the features that can be produced.[16] The process consists of several cycles of modification where complexity is gradually built up through repeated operations to create the finished devices. Firstly, the bare surface is coated with a photolabile 'mask' (Figure 2a, b) which acts as a protective layer to prevent modification of the covered areas. Next, a template is used to pattern a beam of light focused onto the surface (Figure 2c), selectively removing the protective mask and exposing the areas where modification is required (Figure 2d). This can take the form of etching away material (Figure 2e) or involve the introduction of trace impurities that alter the electrical properties of the exposed region before the remaining mask is removed chemically (Figure 2f) and the cycle can begin again to introduce other modifications.

The limiting process for the conventional lithographic process is translating a design from the template into a modification on the surface which has become increasingly difficult as the desired size of each feature is reduced. A key factor in this

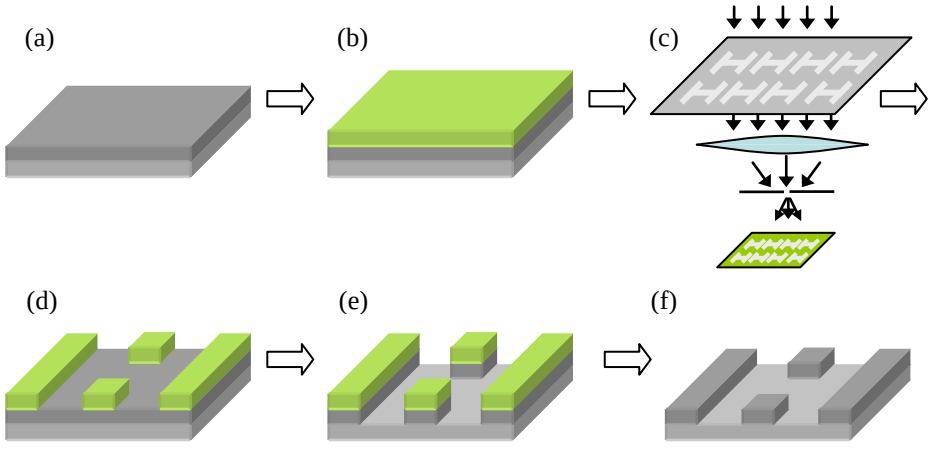


Figure 2 *- Schematic representation of the lithographic approach to circuit formation:* ***(a)*** *an extremely pure and flat sheet of silicon is created;* ***(b)*** *this is coated with a protective masking layer (green);* ***(c)*** *a template (grey) is used to create a pattern of light which is reduced in size and focused on the surface;* ***(d)*** *the light selectively removes the masking layer;* ***(e)*** *the newly exposed surface is modified as required;* ***(f)*** *any remaining mask is removed*

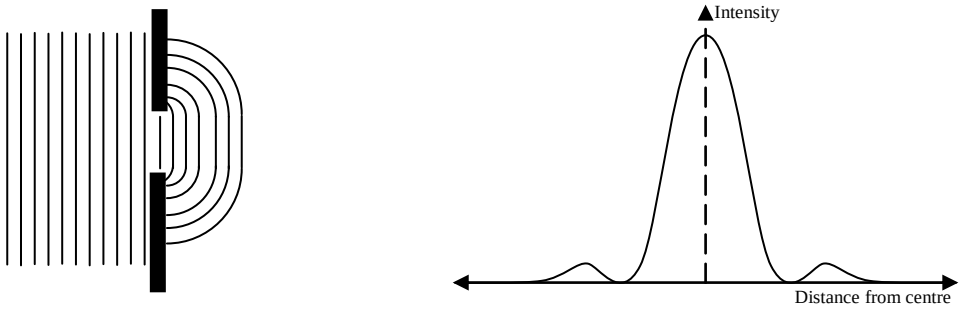


Figure 3 *- Schematic illustrating the diffraction of monochromatic light as it passes through a slit of comparable size to its wavelength and the intensity distribution observed*

is that as the size of the gaps in the template approach the wavelength of light being used, diffraction around the edges creates a spread of light (Figure 3) rather than a single point and this imposes a minimum width restriction on the features that can be produced. Although it is an inevitable part of light based lithography, the effects of diffraction are wavelength dependent which has led recent trends towards shorter wavelengths of light[16] to further reduce the minimum size feature that can be produced and fulfil the long term demand for smaller devices. This process means that wavelength used for current microprocessors manufacture has moved far beyond the visible and ultraviolet regions of the electromagnetic spectrum into the extreme ultraviolet region (120 – 10 nm). The energy of photons in this region (1000 – 12,000 $kJmol^{-1}$) is much greater than dissociation energy of chemical bonds within the masking layer with the excess energy dissipated in the form of photoelectrons and secondary electrons which radiate out from the site of absorption still containing enough energy to cause removal of the mask, broadening the size of the feature being produced.

These lithographic processes allow the introduction of a single modification onto a surface, but it is the combination of different materials that creates the functional devices required for microprocessors. Modern microprocessors are based on adjacent regions containing one of just two behaviour modifying impurities which are introduced onto the surface of the silicon, but the complex patterns required need several cycles of

lithography to produce and as the scale of the features decreases consistently modifying these regions is increasingly problematic. The lithographic approach to microelectronics has endured due to the simplicity of the technique and ease of automation, but a fundamental shift from these methods may soon be required as the level of complexity demanded by modern devices approaches the fundamental limits of existing technology which are increasingly difficult to overcome.[16-18] The miniaturising nature of the top-down approach means that there are limited opportunities for discovering the new properties or alternative materials that might answer some of these issues, but the alternative bottom-up approach is more flexible and provides a better framework for novel designs.

1.3 The bottom-up approach as an alternative

The top-down approach to nanotechnology[3, 19, 20] is analogous to the process of sculpture where the detail of the item being formed is 'carved' out of a larger, simpler starting material and then 'decorated' with the required functionality (Figure 1a). In contrast to the 'top-down' approach, the 'bottom-up' approach is analogous to constructing the item out of smaller modular construction blocks where the 'decoration' has already been applied and parts can easily be exchanged to modify the design (Figure 1b). This ability to substitute in novel building blocks represents a significant advantage for the integration of new materials and investigating the interactions within assemblies of building blocks that may lead to the discovery of novel properties. In addition, the convergent nature of the assembly process and the use of common building blocks simplify the process of mass production and allow a wide range of final products to be created from the same source materials. Finally, the size of the products that can be created is not limited by the level of control over the 'carving' as with top-down

techniques, but by the dimensions of the building blocks which can be as small as single molecules.

Possibly the greatest challenges for bottom-up methodology are those of establishing the interactions between the individual parts that make up the assembly, controlling the dimensions of this assembly and identifying and correcting errors. Some degree of self-assembly is essential for all bottom-up techniques but is most prominent in the field of supramolecular chemistry which began with the creation of macromolecules by Lehn[22] which recognised specific metal ions and has since been extended to produce larger structures which can selectively self-assemble from a mixture of starting materials[23] (Figure 4). The central role of metal ions within these structures also provides a mechanism to control the shape of the assemblies dynamically and has been demonstrated by reducing metal ions in-situ[24] to interconvert between two distinct structures. In addition to this metal based approach, simpler combinations of organic molecules have also been developed[25] which self-assemble in specific ratios to

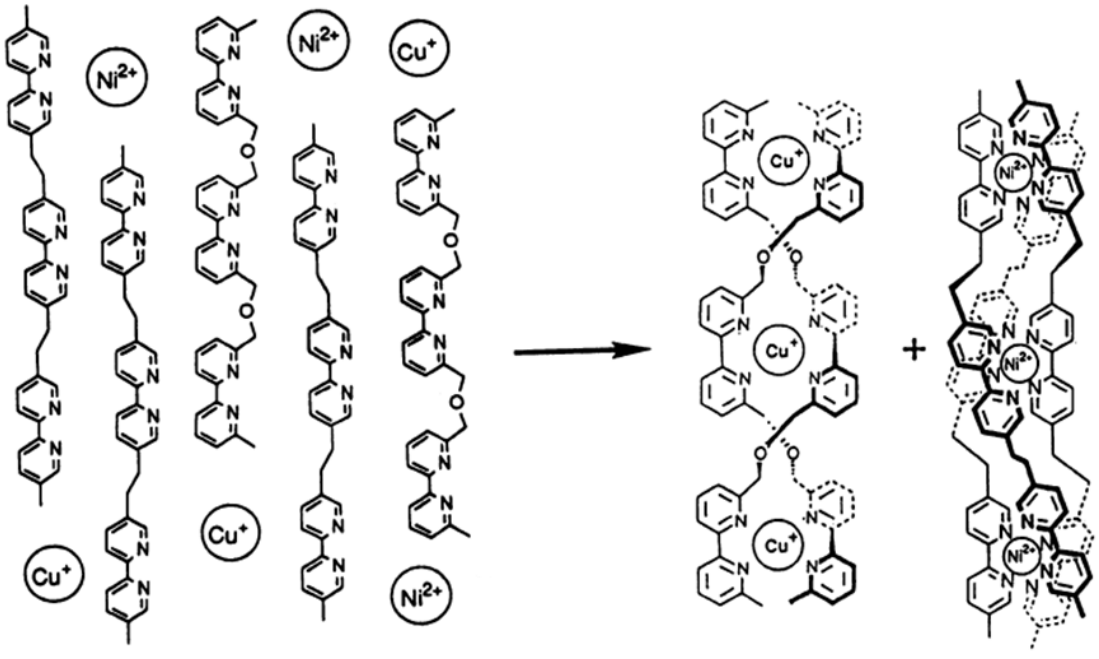


Figure 4 *– Self-assembly of supramolecular structures based on metal ligand preferences*[21]

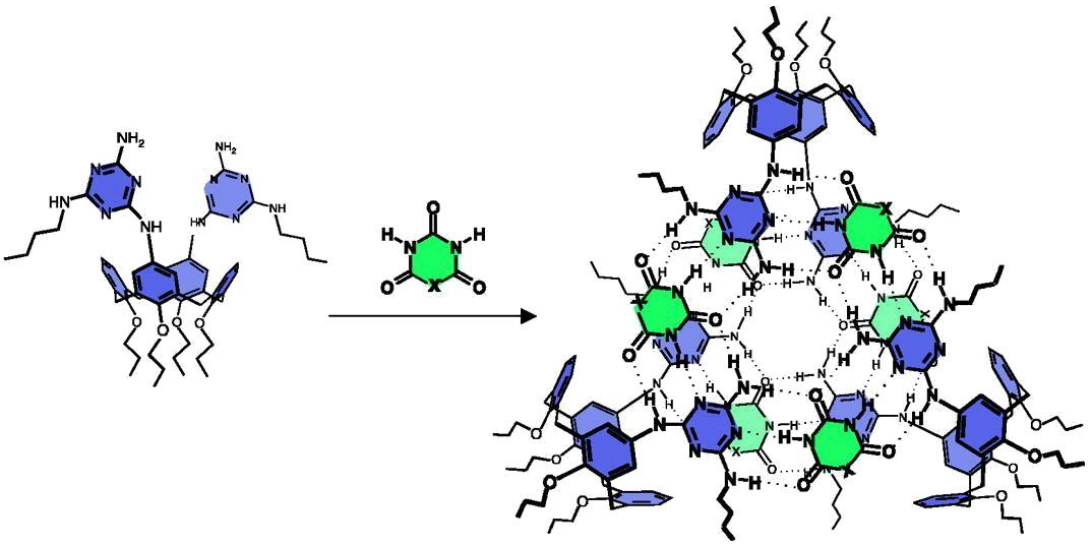


Figure 5 *– Self-assembly of organic molecules into three dimensional structures*[25]

create well defined 3D structures (Figure 5). Selective assembly is achieved by designing each molecule with a pattern of hydrogen bond donors and acceptors that allow it form a network of specific interactions with neighbouring molecules. This network determines the overall size and shape of the assembled structure and allows a wide variety of discrete structures to be formed. The usefulness of both organic and metal templated structures are ultimately limited by the lack of variety and specificity in the recognition processes used, precluding their use in more complex systems where several different assemblies are required. In addition to directing the association process for the individual parts, an ideal self-assembly system would also provide a means of directing the interactions between the assembled building blocks and the formation of more complex groupings, something which none of these artificial schemes have yet achieved.

Rather than designing a novel system from scratch in this way, existing biological recognition systems have been adapted such as the streptavidin-biotin interaction,[26, 27] DNA aptamers,[28, 29] DNA duplex formation and protein recognition of DNA sequences.[30-33] DNA plays a prominent part in many of these methods because it has an easily manipulated recognition system that is intrinsic to its structure, but the

diverse range of 3D shapes and interactions formed by natural proteins has also attracted much attention.[33, 34] The 20 different natural amino acids[35] that can be used to make up proteins might be thought to form an ideal method to introduce a unique sequence of bases as an 'address' to identify a particular component for further modification or assembly into a larger network, but the problems of 'reading' this pattern and more significantly of designing a novel protein with a specific functionality or to adopt a particular shape currently make this impractical. In contrast, DNA is capable of recognising corresponding strands without any modification and whilst being more limited than proteins it is far easier to design a structure and be able to predict the outcome.

The complexity of using biological building blocks has led to interest in creating simpler molecule sized components, particularly the idea of forming mechanical devices that respond to environmental triggers. Rotaxanes are typical of these types of molecules (Figure 6).[36] Rotaxanes consist of a central 'rod' which contain multiple 'stages' between which a movable 'torus' treaded is positioned using specific triggers such as pH or the presence of metal ions.[37-39] The mechanical movement produced in rotaxanes has potential applications such as creating artificial 'muscles' which contract in response to stimuli[37, 39-41] or controlling the release of drugs.[42] Rotaxanes have also been used as a simple two state memory[43] and a more complicated example used to perform very simple calculations,[44] but in all of these applications each molecule acts independently. An effective mechanism for interaction is important for allowing multiple components to be linked together to achieve more difficult tasks in the same

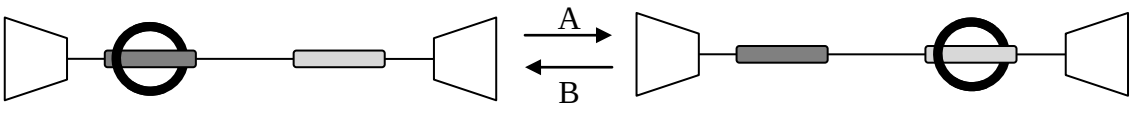
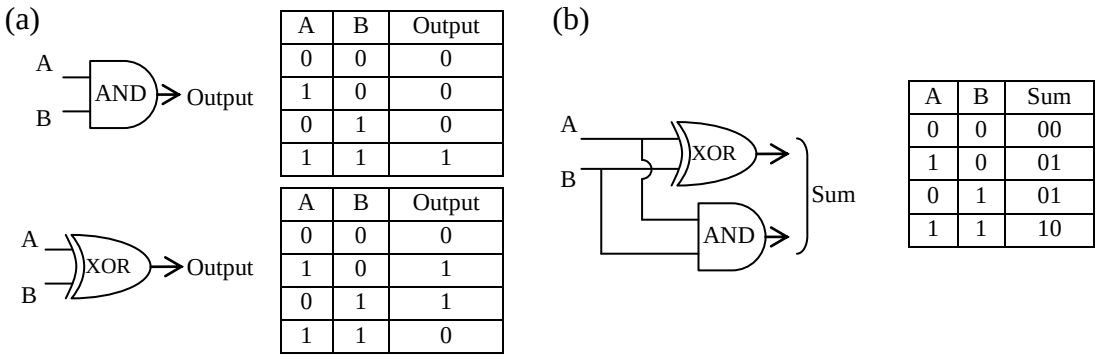


Figure 6 *– Schematic representation of a two stage rotaxane*



A	B	Output
0	0	0
1	0	0
0	1	0
1	1	1

A	B	Output
0	0	0
1	0	1
0	1	1
1	1	0

A	B	Sum
0	0	00
1	0	01
0	1	01
1	1	10

Figure 7*- Combining basic functions to create more complex responses: (a) symbolic representations and input responses of AND and XOR (eXclusive OR) logical functions; (b) combining these functions to create a device capable of adding inputs A and B*

way as modern microprocessors. Although a simple sensor which converts one kind of signal into another might be considered the simplest kind of component or device, these have limited utility beyond this function whilst components such as AND gates, which produce an output only when both inputs A and B are provided (Figure 7a), are able to act as the basic units for almost all microprocessors because they can be combined to form networks capable of 'useful' functions such as counting or adding inputs (Figure 7b). Two key requirements for achieving this are identifying orthogonally triggers to act as external inputs and controlling the path of signals between components. Simple environmental triggers such as the presence of a particular metal ion or pH that work well for discrete components[37-39] are impractical for use in more complex environments due to the limited number of compatible triggers and relatively slow rate at which they can be altered. More promising are the examples of 'molecular machines' which respond to irradiation with specific wavelengths of light (Figure 8).[45-48] Although these

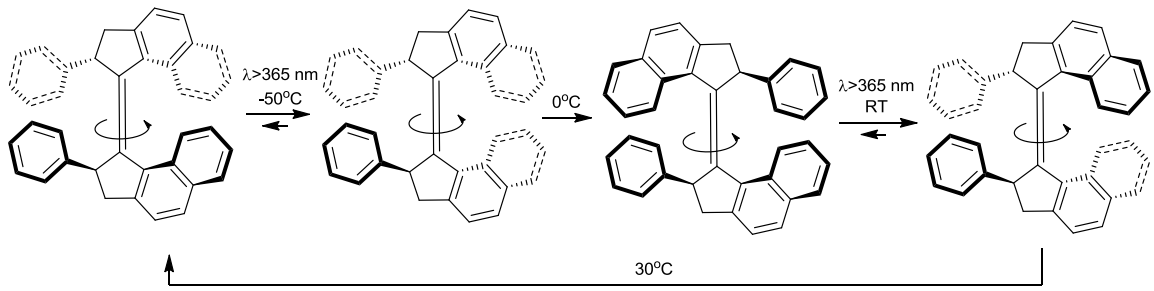


Figure 8 *- A example of a light driven 'molecular motor'*[45]

building blocks are designed to create a physical movement when triggered rather than a more easily read output signal, they illustrate how light can be attractive mechanism for interacting with the nanometre scale world.

This highlights another key problem with creating addressable nanometre scale objects which is the limited information that can be obtained by a macroscale observer many orders of magnitude larger in order to characterise the product and allow errors to be detected. Atomic resolution is available using atomic force microscopy (AFM)[49] and high resolution transmission electron microscopy (HRTEM)[50, 51] but these techniques are relatively slow and an impractical method of 'reading' the output of a device outside of the research lab. Highly sensitive light detection systems are now routinely used to study the behaviour of single molecules attached to biological targets,[52-54] allowing direct observation of device responses and as the interaction normally takes the form of a specific absorption by the molecule, the response can be made selective for particular wavelengths. Characterisation remains a difficult challenge for the bottom-up approach, but light based processes are a rapid and simple method of gaining an insight into the dynamic structure of nanometre scale assemblies.

A key challenge remaining to be addressed is integrating the discrete building blocks into a network capable of replicating the functions available using top-down methodology. In order to address this, a reliable method of locating individual parts which also allows selective connections to be made is required, a significant challenge given the lack of distinguishing features these molecules typically exhibit. An ideal system would be modular, allow the integration of a wide variety of features and include a fixed scaffold on which the active components can be positioned and connected. This capability would also be a significant advantage for research into the interactions between individual building blocks and nanometre scale materials which is

currently limited by the lack of reliable methods to position the materials of interest. Although a challenging proposition, this is a vital part of moving beyond forming discrete components in the laboratory and towards developing fully functioning devices and investigating novel properties.

1.4 DNA Structural Nanotechnology

1.4.1 DNA Primary and Secondary Structure

The highly selective recognition properties of DNA that allow it to play an important role in many bottom-up techniques are a product of its underlying structure, from which all of the properties of a particular strand arise. DNA is a linear copolymer

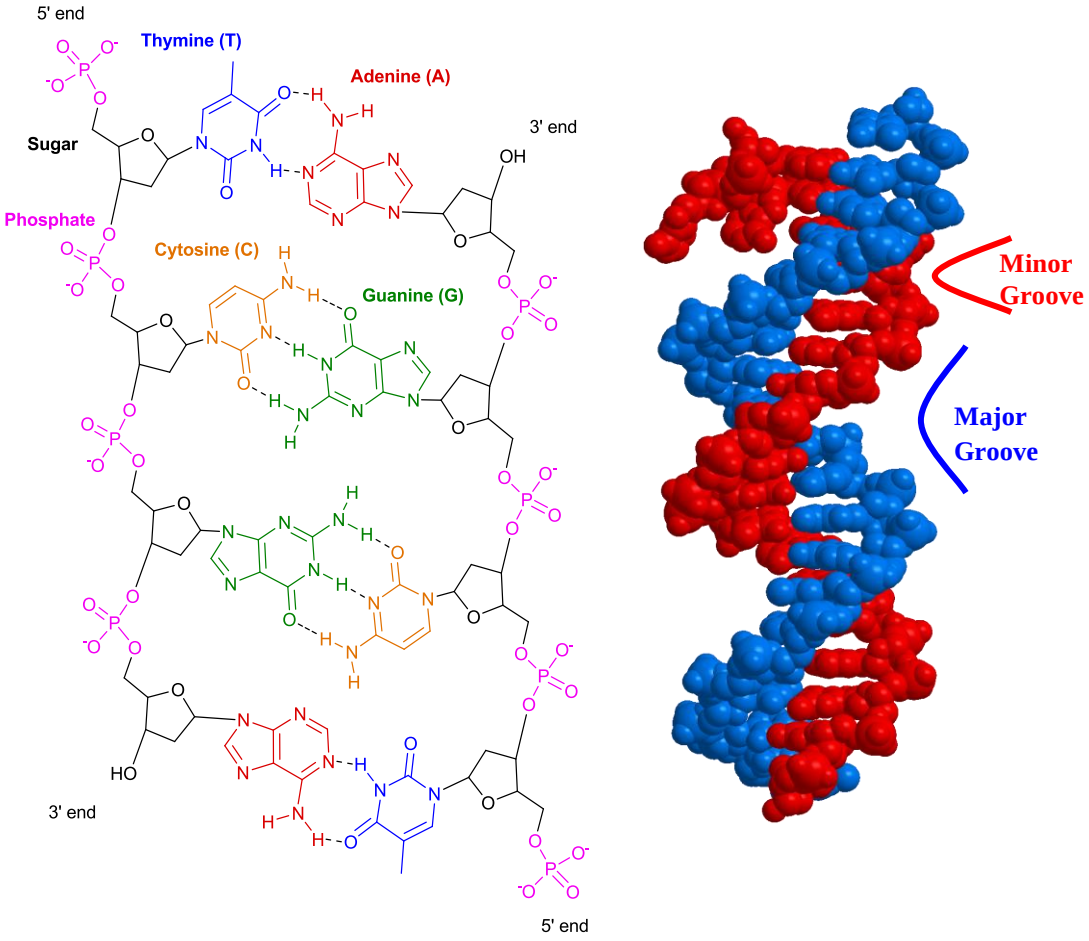
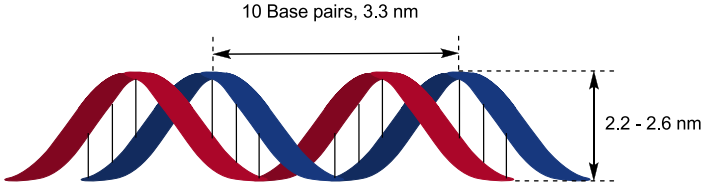


Figure 9 *- The base pairing and secondary structure of DNA*

consisting of alternating phosphate groups with ribose sugars that have an aromatic base attached capable of hydrogen bonding. When guanine (G) bases are paired against cytosine (C) and thymine (T) with adenine (A), non-covalent interactions create stable base pairs which position the strands into a double helix[55] (Figure 9). The structure of the double helix in solution normally contains both a major and minor groove with the hydrophobic bases stacked in a central 'column' and the negatively charged phosphate groups positioned on the outside. Although significant variations on this basic structure have been identified with altered proportions for the grooves and a reversed helical twist,[56] in general the only observable variation for medium and long DNA strands is a slight narrowing of the minor groove in A and T rich regions.[57]

The stability of a DNA duplex is strongly influenced by the sequence as the three hydrogen bonds of the G-C base pair produce a stronger interaction than the two formed by the A-T base pair. As a result G-C rich duplexes dissociate at higher temperatures and undergo the partial thermal dissociation ('DNA breathing') less frequently than A:T rich strands. For a typical DNA strand with a random sequence these variations average out and allow a general model of the double helix to be constructed (Figure 10) with each turn of the helix containing 10 base pairs over a distance of 3.3 nm and a diameter of 2.2-2.6 nm. Although this model essentially describes a rigid 'rod', the bases within the double helix retain enough freedom of movement to prevent any correlation between the direction of the start and end of any strand longer than 120-150 base pairs (the persistence length of DNA, approximately



	Width	Depth
Minor groove	6.0 Å	8.2 Å
Major groove	11.6 Å	8.5 Å

Figure 10 *- Typical dimensions for a B-DNA strand with a random sequence*[58]

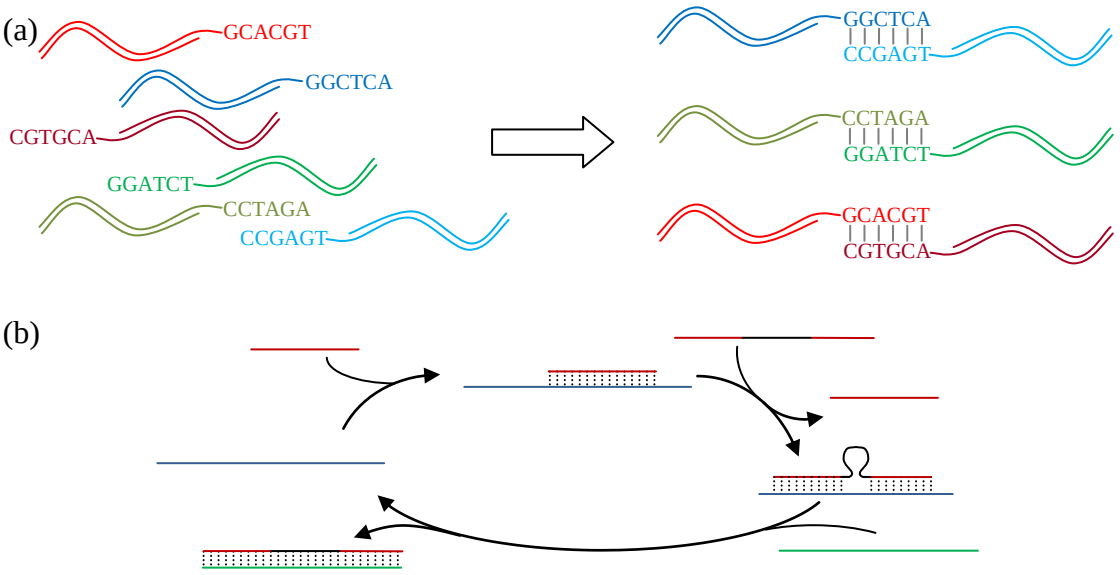


Figure 11 *– Applications of single stranded DNA 'sticky ends' –* ***(a)*** *Specific assembly of designed pairs of strands from a mixture;* ***(b)*** *Schematic of a strand displacement cycle where shorter strands are displaced and a longer, more stable duplex formed*

40-50 nm[59]). Beyond this limit, double stranded DNA has a random structure, but the rod model can be used for much shorter duplexes and 'braced' which allows them to be used as a 'molecular ruler' to produce well defined separations on the nanometre scale.

The formation of a duplex forms the basis of most DNA recognition systems as the temperature at which the two strands dissociate is sensitive to the number of correctly paired bases, allowing the kinetics and thermodynamics of this process to be carefully controlled through modifications to the sequence. This process is typically exploited using overhanging single stranded regions known as 'sticky ends'. The sequence of this region can be used as unique 'address' that allows strands to be designed with a complementary sequence to form specific associations within a complex mixture (Figure 11a). Through careful design of 'fuel' DNA strands which displace weaker binding short sequences and then be removed, complex cycles can be created (Figure 11b) which have formed the basis of simple 'molecular machines'[60] capable of moving along a DNA strand[61] or complex logical functions.[62] This advanced level of control is readily accessible through manipulation of the sequence making DNA

a key part of the nanotechnology 'toolbox', particularly as the attachment of non-natural components to the end of the strands extends these capabilities to a wide variety of materials and allows individual building blocks to be positioned at the base pair level.[10-12, 63-71]

1.4.2 Synthesis of DNA

Methods for the artificial production of DNA vary according the number of bases required. Short strands are typically produced using solid phase methodology with DMT protected phosphoramidite derivatives of the bases working from the 3' end of the strand.[72, 73] The exocyclic amines of the bases are also protected by a variety of base labile groups to prevent side reactions and a β-cyanoethyl group attached to the phosphorous chain to improve the solubility of the phosphoramidite. The cycle (Figure 12) begins with the acidic deprotection of the DMT group to expose the reacting 5' hydroxyl group. The new base is then introduced using an excess of the appropriate phosphoramidite mixed with a coupling agent such as tetrazole to form an activated tetrazole phosphoramidite that reacts with the freshly exposed hydroxyl group. Solid

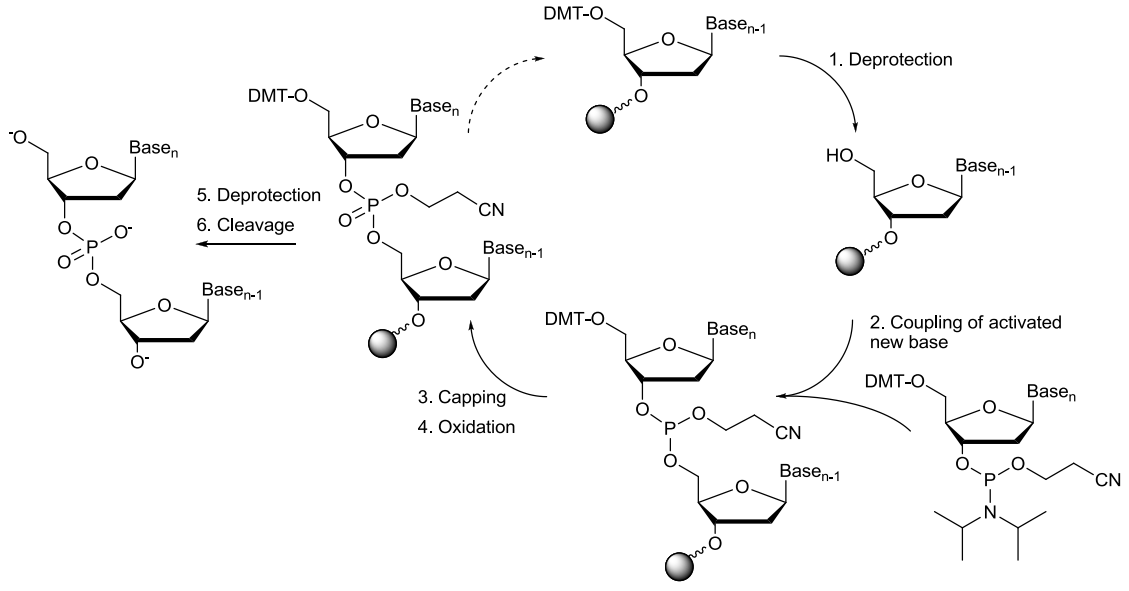


Figure 12 *- DNA solid phase synthesis cycle using phosphoramidite chemistry*

phase methodology allows easy removal of this excess which is needed to maximise the yield of each step and any unreacted groups are then capped using acetic anhydride to prevent the formation of truncated sequences with missing bases. Finally, the still reactive phosphite linkage is oxidised up to the phosphate triester using iodine and water.

The dinucleotide produced can then be cycled again until the desired sequence of DNA has been produced when the complete strand is deprotected and cleaved from the solid support using a concentrated ammonium hydroxide solution. The number of bases that can be introduced in this way varies depending on the solid support used, but the efficiency of the coupling reaction typically limits the strands produced in this way to a maximum of 150 bases.[73] Solid phase DNA synthesis is relatively simple to implement and is a quick way to produce small quantities of a several strands, but consumes large quantities of reagents and is ill-suited for creating extended stretches of DNA. Where these longer strands of DNA are required, an alternative approach using the biological machinery of DNA replication is used.

The polymerase chain reaction (PCR)[74, 75] uses a bacterial DNA polymerase that binds to double stranded regions of DNA where it moves along the DNA from the 5' to the 3' end, extending the complementary strand in any single stranded regions. This allows a long 'template' strand of DNA to be replicated by synthesising short strands

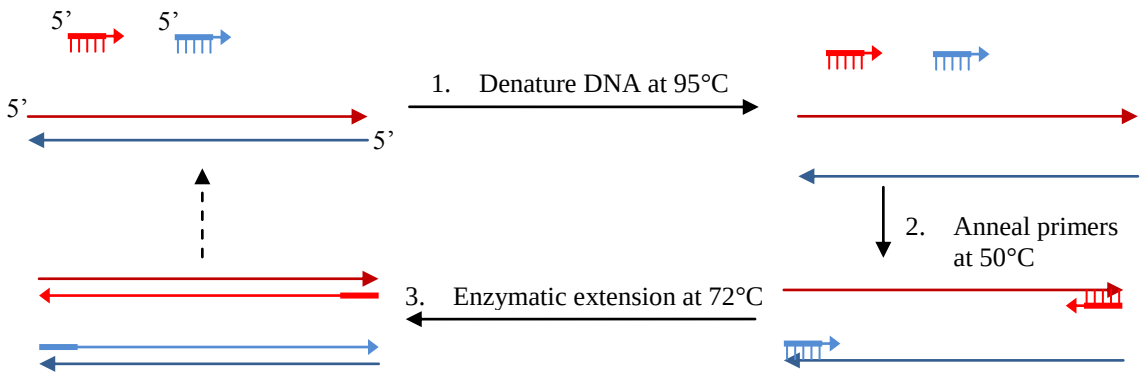


Figure 13 *- Polymerase chain reaction (PCR) cycle*

known as ‘primers’ that bind at the ends of the desired sequence to act as the starting points (Figure 13). After dissociating the template DNA strands by heating to 95°C, the mixture of primer strands and template DNA is cooled to 50°C to allow the primers to bind, then reheated to 72°C to dissociate any weakly bound (i.e. non-matching) DNA strands and provide the optimum conditions for the enzyme to operate. The enzyme uses the naturally produced base triphosphates which are also added as the raw materials to replicate the DNA and was specially chosen to withstand these relatively harsh conditions which would denature most polymerases.[75] After the DNA has been replicated, the matching strands produced are used as the template strands for the next cycle, allowing exponential growth in the number of strands introduced and the amplification of sequences up to 42,000 bases long[76] starting from a very small amount of template DNA. PCR is ideal for creating long stretches of DNA and once the template strands and primers have been produced can rapidly produce relatively large quantities of DNA, making it a key method for genetic research. The technique is not without drawbacks though as it is ill-suited to replicating short strands and optimising the conditions for a particular sequence can present significant challenges. In addition, as PCR exploits the biological machinery of the cell to recognise the template strands and insert the matching bases it also retains the substrate specificity of these proteins, which normally acts to preserve the accuracy of the replicated strand but also limits the alterations to the basic structure of DNA that can be tolerated.

1.4.3 Modified DNA bases

One of simplest ways of directly adapting DNA for nanotechnology applications is by introducing modified bases during the synthesis.[77] Direct substitution of the standard bases positions the modification inside the double helix (Figure 14a) which

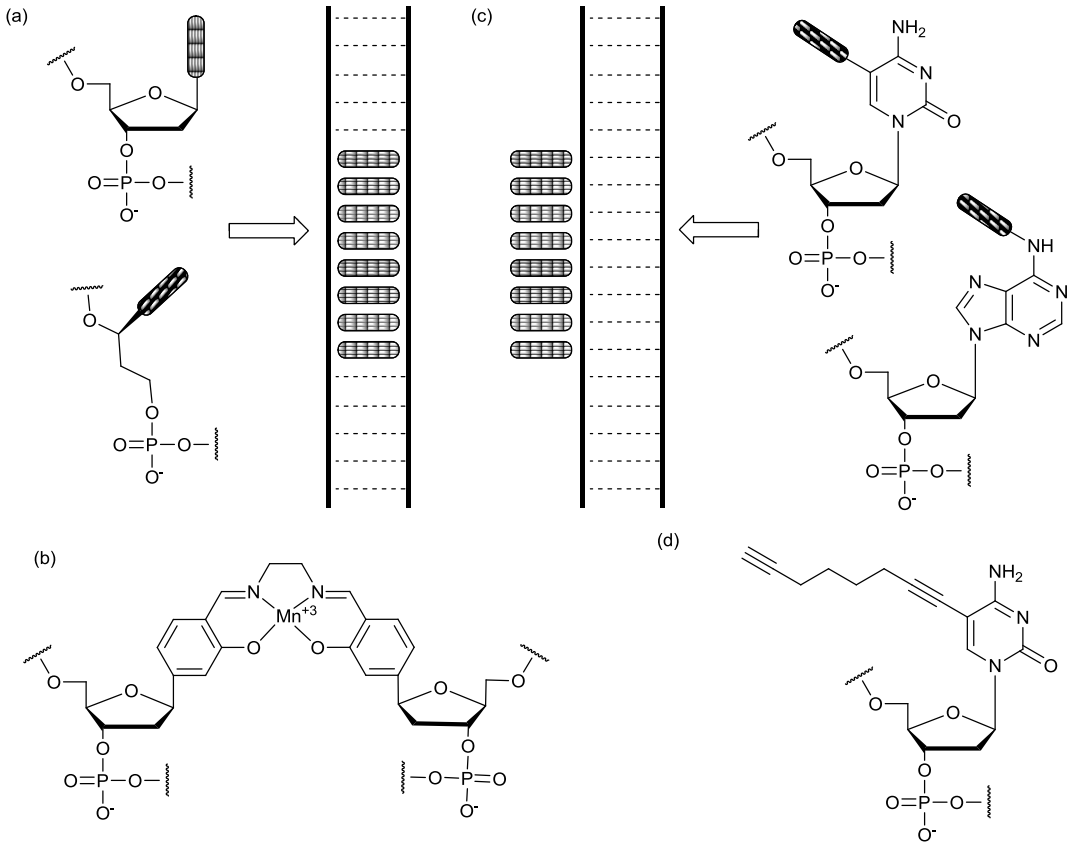


Figure 14 *- Methods of including modified bases into DNA strands: positioning(a) and an example of internal modification (b);*[79] *positioning(c) and an example of external modification (d)*[80]

has been used to create unnatural pairs of bases and to create a string of metal ions by substituting with two ligand-like halves among other applications (Figure 14b).[78] In contrast, by adding functionality to the natural bases these modifications can be positioned within the major groove of the DNA which retains the core structure of the DNA and opens up new possibilities (Figure 14c, d). The regular arrangement of the double helix means that each base of DNA provides a potential attachment site at a precise 0.4 nm intervals along what can be considered as a rigid 'rod' for short distances.

Using an existing base as the basis of a modification also has the advantage that the modified position retains its recognition properties, allowing better control of its assembly and proximity to other modified sites. However, all of these modified bases represent a significant synthetic challenge not only to make, but also to introduce into

DNA strands. Incorporation into short DNA strands can be achieved after optimisation with reduced synthetic yields, but selectively modifying long strands is more difficult. Firstly, the conditions used for the PCR must be optimised to allow the enzyme to accept the larger modified bases,[81] a process usually achieved by reducing the specificity of the enzyme which can cause errors to be introduced into the DNA replication. Despite this, complete substitution of a DNA strand with modified bases has been achieved[82] and it is possible to produce statistical mixtures (e.g. 5% substitution) by using a mixture of modified and natural base triphosphates during the amplification. Unlike solid phase oligonucleotide synthesis, it is impossible to selectively introduce a modification at a single position using PCR without the use of modified primers, making it difficult to create long strands of DNA with specific internal modifications. Methods may eventually be developed to overcome these problems, but the extended chemical synthesis inherent in creating modified bases is unavoidable and requires significant commitments in the form of timescale and complexity.

1.4.4 Two Dimensional DNA structures

Although modified bases are a simple way to create a unique site within a DNA strand, the information contained within the sequence of bases has the potential to act as an 'address' if it could be accessed. Just 10 sets of DNA base pairs can be assembled into over 500,000 different arrangements compressed into a helix just 3.3 nm long. Unfortunately this incredible density of information is difficult to exploit, but the use of overhanging single stranded regions of DNA allows some of this information to be accessed. These 'sticky-ends' (Figure 11a) expose the sequence information and allow separate strands to be joined together in a programmable way using the complementary

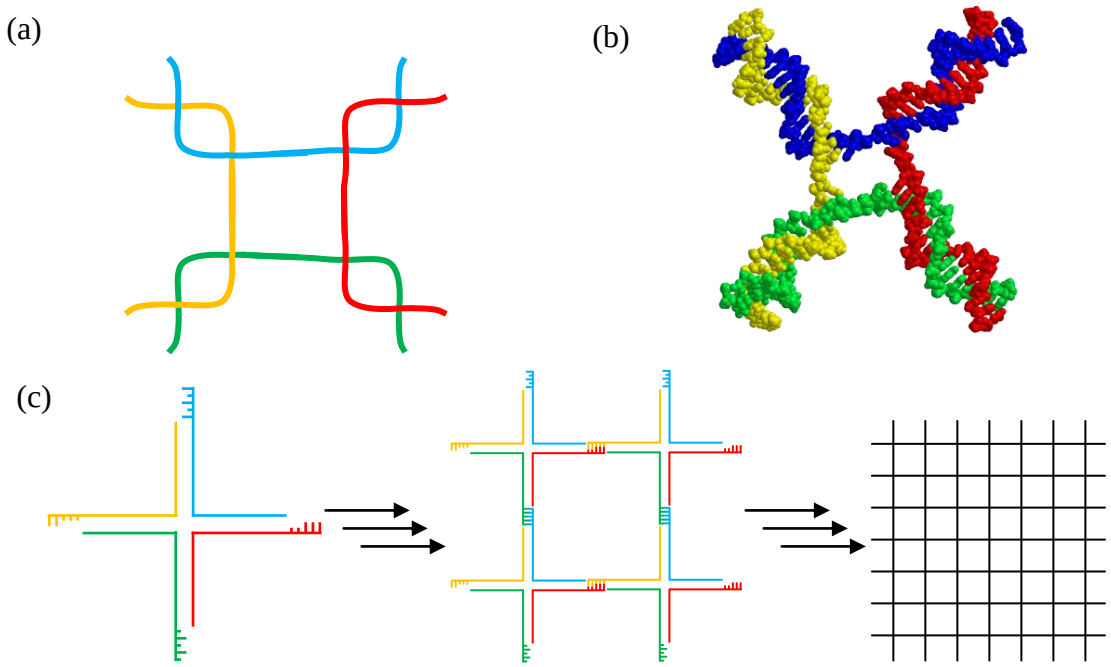


Figure 15 *- Forming shapes out of DNA –* ***(a), (b)*** *Schematic representation and crystal structure of a 4 way DNA junction (PDB: 1FLO[83]);* ***(c)*** *self assembly of a large DNA grid from a single building block created using 4 DNA strands*

sequence. In addition, by designing a sequence with several complementary regions, the same piece of DNA can form duplexes with multiple strands.

Combining strands in this way allows junctions between the duplexes to be formed and with that comes the ability to create complex shapes using DNA rather than linear structures (Figure 15a, b). In this way, four short strands of DNA designed with two complementary regions and an 'address' overhang can be combined to create an extended grid pattern many times larger than the length of the original strands (Figure 15c).[84] These types of structures are possible because DNA behaves both as a rigid rod for distances less than 40 nm[59] and as a flexible 'string' within single stranded regions and over longer distances. Regular patterns such as this have been used to create one and two dimensional arrays of materials such as gold nanoparticles,[67, 85] quantum dots[68] and fullerenes[86] as well as patterning proteins and creating a metallised grid.[11] The lack of distinct sequences or structures makes these types of structure a limited material as the basis for more complex network. The level of complexity that can be introduced is

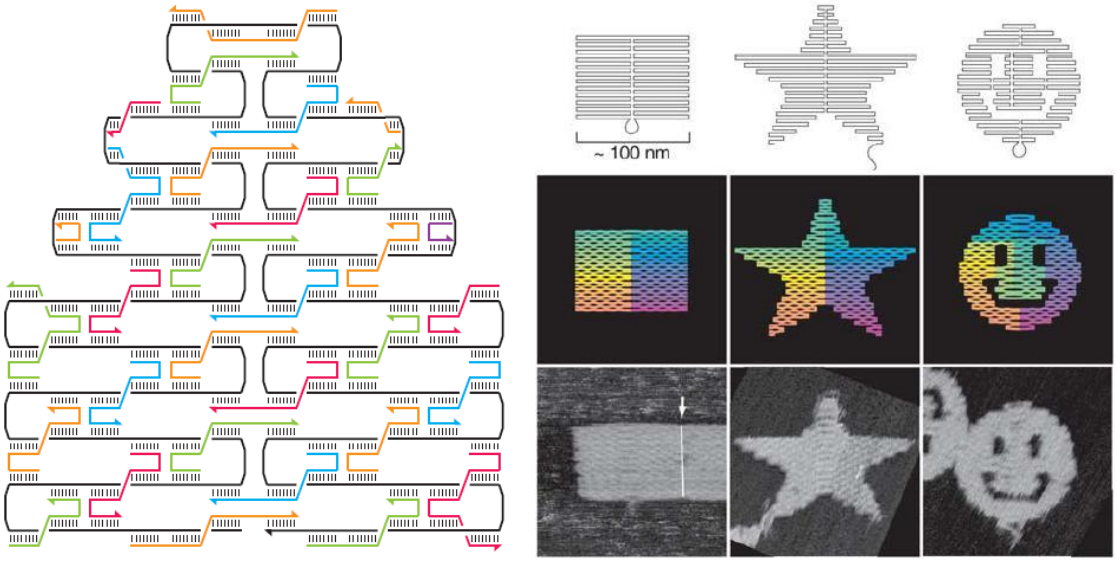


Figure 16 *- DNA origami*[4] *-* ***(a)*** *The network of staple strands underpinning DNA origami structures;* ***(b)*** *Structures formed using the same long DNA strand, top and middle: route of long DNA strand, strand direction red to purple, bottom: AFM images of structures produced*

also limited by the length of DNA strands that can be synthesised, restricting this type approach to DNA construction to simple repetitive motifs.

A promising alternative for more complex structures is the 'DNA origami' technique which uses a very long single strand of viral DNA running throughout the entire structure with a number of short 'staple' strands to fix the shape of the molecule (Figure 16a).[4] As the complementary regions of the staple strands determine which regions of the viral strand are positioned together, they also determine the final shape of the structure and this modularity allows the same viral strand to be used to create a variety of structures (Figure 16b). This approach also has the advantage that each section of the structure has a unique sequence, allowing staple strands modified using 'sticky ends',[66, 87] biotin[8] and DNA aptamers[7] to act as a unique 'address' for the attachment of materials such as quantum dots,[68] nanoparticles[67] and proteins[7, 8] to a specific location or as part of a more regular array. Selective functionalisation can also be achieved using multiple 'addresses',[9] offering a mechanism to assemble devices using multiple components selectively positioned onto a well defined DNA scaffold. However, in order to achieve this goal, in addition to designing a new DNA scaffold for

each new device, an orthogonal recognition method for targeting each 'address' would be required for each unique functionality being introduced. This problem can be avoided by functionalising each modified staple strand before assembly, but this may interfere with the assembly of the structure and imposes more restrictions on the materials and connections that can be made. These examples illustrate the potential applications that can be achieved by controlling the assembly of the existing components into simple two dimensional arrangements as with existing technology, but DNA structures are capable of creating much more complex three dimensional networks which allow many more connections to be made.

1.4.5 Three Dimensional DNA structures

In addition to these flat structures, the same principles can also be applied to create three dimensional DNA objects. Structures ranging from simple tetrahedra[88] through to more complex polyhedra[90] can be produced using short strands (Figure 17a), whilst larger structures can be created using the DNA origami approach including the formation of an entirely enclosed box with a dynamically controllable lid[6] and 'solid' structures formed by multiple helices.[89, 91] The ability to construct a truly three dimensional scaffold on a nanometre scale would represent a fundamental advance over the top-down approach which in minimising the lithographic process retains its two dimensional architecture. Cryo-electron microscopy has shown these designs are

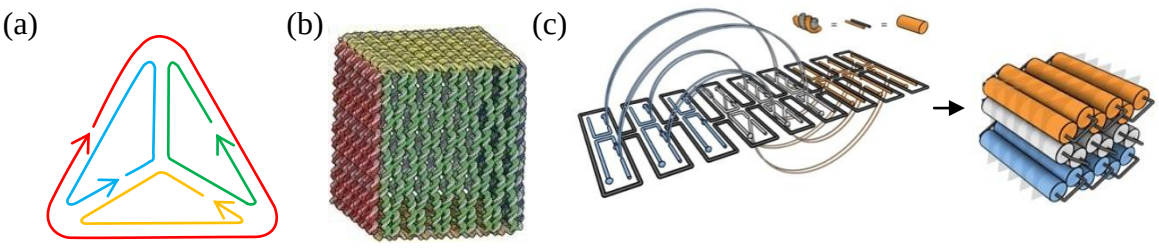


Figure 17 *– Three dimensional DNA structures:* ***(a)*** *a DNA tetrahedron constructed from four DNA strands;*[88] ***(b)*** *DNA origami box with a controllable lid;*[6] ***(c)*** *linking of staple strands to create a solid bundle of helices using DNA origami*[89]

accurately translated into their respective structures,[6, 92] potentially opening up new areas for innovation which are currently inaccessible.

1.5 Using DNA to construct photonic wires

Bridging the gulf between the nanometre scale and the macroscale user is key to creating practical nanotechnology devices. Even if a device can be assembled in high yield and is able to perform complex functions, it will be fundamentally limited if there is not a simple, rapid and selective method of supplying the information to be processed and recording the resulting output of the device. The photonic approach to this problem has many advantages over alternative methods of interaction such as small molecules, pH and short DNA strands. Light based interactions are extremely rapid, allow the injection of significant amounts of energy into a system and interact in a selective and tuneable manner with the organic molecules that form the basis of many existing building blocks. The chromophores and fluorophores that mediate interactions with the devices are small and a wide variety of attachment chemistry has already been developed, allowing easy incorporation into building blocks that are not already light based. In addition, studies of fluorescently labelled biological systems has led to the development of highly sensitive detection techniques capable of studying single molecules[52] which would allow the return of the processed information to the macroscale domain.

Photonic interactions provide a powerful mechanism to interact with nanometre scale devices, but well defined positioning of the components is essential in order to maximise its potential. The ability of DNA to anchor functional materials with 0.4 nm resolution and to provide a larger framework onto which building blocks can be assembled makes it ideal for this role. This precise positioning coupled with the

complex three dimensional shapes that can be created and the wide variety of materials that can be attached to DNA also allows more fundamental aspects of photonic interactions between individual components to be investigated systematically.

1.5.1 Fluorescence and Simple FRET

Many organic molecules absorb light in the ultraviolet or visible regions of the electromagnetic spectrum where the energy the photons corresponds to electronic transitions between molecular orbitals (e.g. $\pi \rightarrow \pi^*$).[93] In most cases the energy absorbed is dissipated as heat in the form of increased vibrational and rotational motion of the solvent but in some cases part of the energy is rapidly released as a photon of longer wavelength, a process known as fluorescence (Figure 18a). For most molecules,

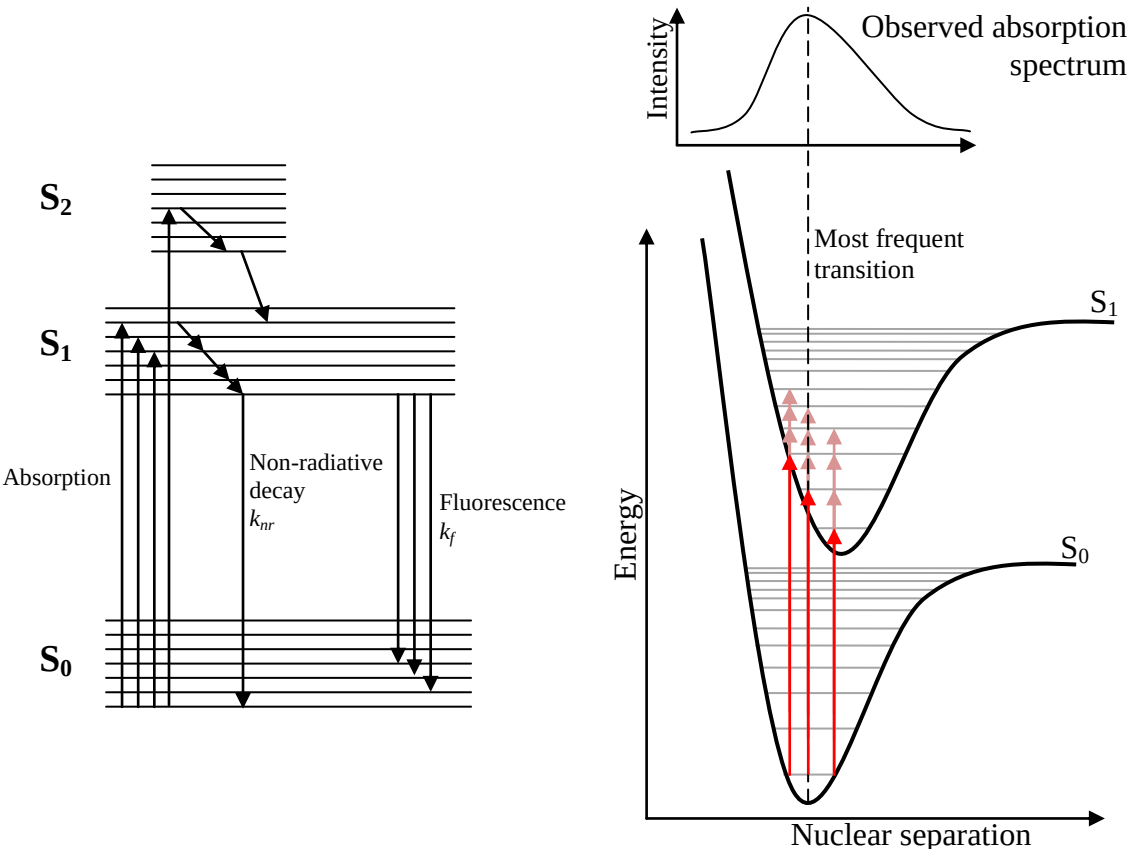


Figure 18 *– The processes of fluorescence:* ***(a)*** *Schematic representation of the available vibrational and electronic energy levels in a typical fluorophore and the key transitions between them;* ***(b)*** *Illustration of the Frank-Condon principle for a typical electronic absorption and the resultant absorption spectrum*

these transitions usually take place between a singlet ground state known as S_0 and one or more excited states labelled as S_1, S_2, S_3... in order of relative energy. The exact wavelengths at which absorption and fluorescence take place are determined by difference in energy of the molecular orbitals and as a result depend on the structure of the fluorescent molecule or 'fluorophore' (Figure 19). In addition, the absorption and fluorescence processes are selective as they correspond to specific electronic transitions and only a narrow range of wavelengths are absorbed.

The range of wavelengths corresponding to the absorption and emission processes is explained by the Franck-Condon principle[94] which assumes that the movement of electrons caused by the absorption is much faster than the movement of the nuclei, limiting the transitions to those where the positions of the atoms have a corresponding vibrational energy level above them (Figure 18b). As the equilibrium separation of the atoms involved is longer due to the higher energy of the excited state,

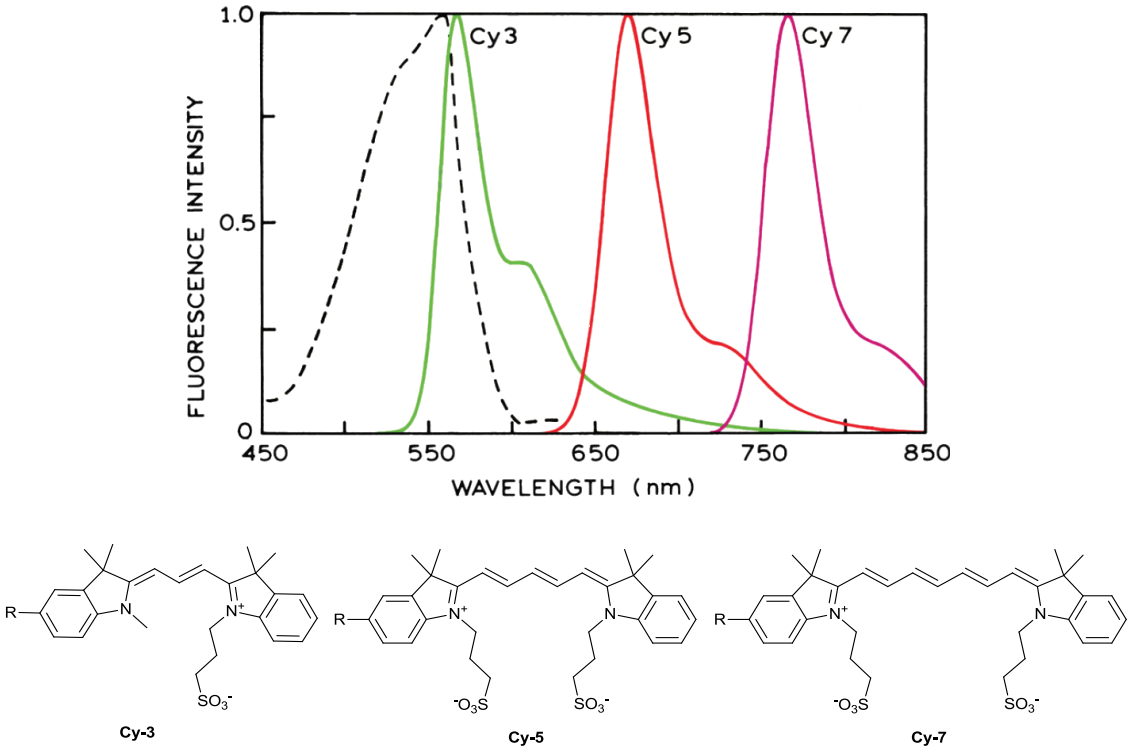


Figure 19 *- Variations in the emission spectra (coloured lines) of Cyanine family dyes with structure and the absorption spectrum of Cy3 (dotted line)*[93]

the transition usually also includes promotions into higher vibrational energy levels of S_1, creating a structured absorption band. In many cases, this structure is averaged out by differences in energy of both states due to solvent interactions and thermal distribution, yielding a broad, asymmetric peak. Similarly the process of fluorescence is also an electronic transition, the Frank-Condon principle still applies and structure can be observed representing the range of higher vibrational energy levels of S_0 populated from the vibrational ground state of electronic excited state S_1.

Despite the injection of energy into higher vibrational energy levels, fluorescence is normally only observed from the vibrational ground state of the excited state with the least energy (S_1) (Figure 18 a). These vibrational energy losses are responsible for the shift in wavelength observed between absorption and fluorescence, known as the Stokes shift. The close spacing of vibrational energy levels means that the excess vibrational energy of the fluorophore is rapidly transferred into the vibrational levels of the solvent, a process known as relaxation. In contrast once a vibrational ground state has been reached, the gap between available states becomes much larger and energy transfer much more difficult, and as a result the lifetime of the excited state S_1 is typically around 10^{-6} to 10^{-12} seconds[94] for organic fluorophores.

$$\phi = \frac{Number\ of\ events}{Number\ of\ photons\ absorbed} = \frac{Rate\ of\ process}{Rate\ of\ absorption} \qquad \sum_i \phi_i = 1$$

***Equation 1**- Definition of quantum yield*[94]

Although it is a relatively slow process, collisions with the solvent provide an alternative pathway to release the energy of the excited state, reducing the amount of fluorescence that can be observed, a process known as quenching. This energy loss is quantified by calculating the quantum yield (ϕ) using Equation 1 which defines the quantum yield for a process as the ratio of the amount of energy lost in that manner to

the total energy input e.g. if half the number of photons absorbed are reemitted through fluorescence, the fluorescence quantum yield is 0.5. The sum of all the energy loss pathways is always equal to 1, but many of them are hard to quantify, and the energy loss pathways are usually divided into two categories – the measurable radiative decay processes which involves the release of a photon, and non-radiative decay processes where the energy is lost to the vibrational and rotational energy levels of the surroundings. The rate of non-radiative decay (which strongly determines a fluorophores lifetime) and the regions of absorption and fluorescence are both determined by structure of the fluorophore which allows its properties to be tailored for a particular application in a similar way to other molecular components.

In addition to transferring energy into solvent molecules through collisions, energy can also be transferred directly into nearby molecular orbitals with the same energy (Figure 20). This pathway is mediated by dipole-dipole interactions and results in a return to the ground state for the donor fluorophore and the simultaneous promotion of an electron into an excited sate of the acceptor molecule, a process known as fluorescence resonance energy transfer or Förster resonance energy transfer (FRET).[95, 96] The interactions can operate over relatively long distances up to around 10 nm[93] independent of steric and electrostatic considerations, but a strong distance dependence allows it to be used as a 'molecular ruler' to determine the distances between two

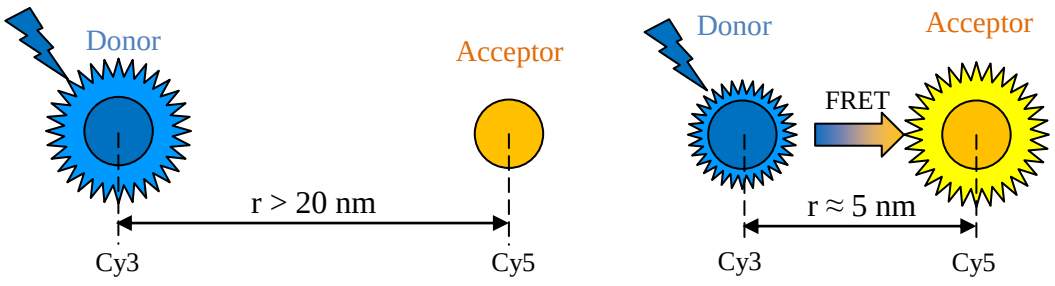


Figure 20 *- Distance dependence fluorescence resonance or Förster resonance energy transfer (FRET) between the donor(Cy3) and acceptor (Cy5) fluorophores where* R_0 *= 5.6 nm*

fluorophores. The efficiency of the energy transfer (E) for a given separation (r) is calculated using Equation 3, where R_0 is the characteristic Förster distance for a given pair of fluorophores defined as the separation at which 50% energy transfer takes place (Figure 21).

$$R_0{}^6 \propto \frac{\phi_f \kappa^2 J(\lambda)}{n^4}$$

***Equation 2**- Dependence of Förster distances (R_0)*

Although the Förster distance is a constant for a given fluorophore pair and solvent system, its calculation is non-trivial (Equation 2) as it is dependent on the fluorescent quantum yield of the donor (ϕ_f), an orientation factor (κ^2), the refractive index of the solvent (n) and the degree of overlap between the emission spectrum of donor and the absorption spectrum of the acceptor ($J(\lambda)$). The quantum yield of the donor and the spectral overlap are directly influenced by the choice of fluorophore with typical values of R_0 ranging from 20 to 60 Å. This 'adjustment' of R_0 is vital as the efficiency of the transfer rapidly changes from 1.5% at twice R_0 to 98.5% at half R_0

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

***Equation 3**- Distance dependence of FRET*

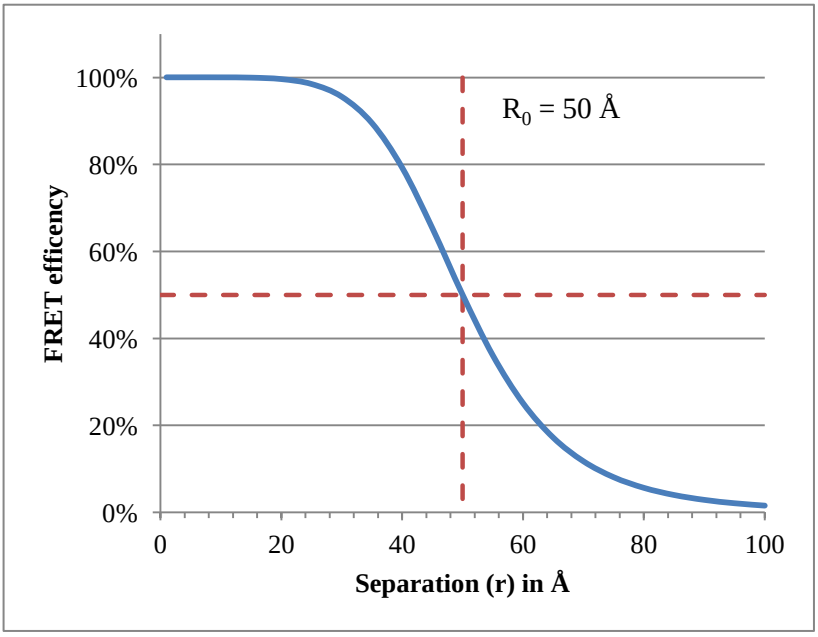


***Figure 21** - Variation of FRET efficiency with distance for a fluorophore pair for which R_0 = 50 Å*

(Figure 21), making FRET very sensitive to small changes in separation on the nanometre scale.

1.5.2 Three colour FRET and photonic wires

The ability of FRET to sense these small changes in distance has made it an invaluable tool for the study of the movements of protein domains and other biological motions.[97] The demands of having both good spectral overlap and separation of the excitation and fluorescence wavelengths led to the development of three colour FRET, where a second FRET transfer step takes place with the intermediate 'bridging' acceptor acting as donor (Figure 22). The overall efficiency of the complete transfer is normally lower due to the increased opportunities for non-radiative energy losses, but the extra shift in wavelength away from the excitation light eliminates direct excitation of acceptor fluorophore and decreases the background signal, allowing much greater sensitivity to be achieved.

The addition of a third fluorophore also increases the distances over which the energy can be transmitted and opens up new possibilities.[98, 99] Using a series of fluorophores as a 'photonic wire' capable of transporting energy along a specific pathway would fulfil one of the key requirements for creating useful devices with the bottom-up approach by providing a mechanism for selective interactions between

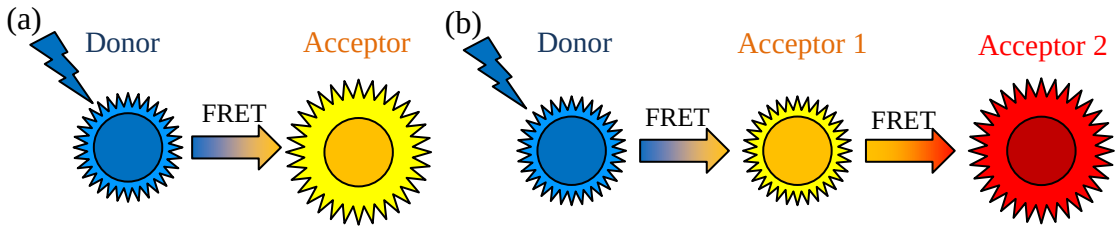


Figure 22 *- Observed fluorescence of two and three colour FRET* ***(a)*** *two colour FRET where weak fluorescence of the donor and strong fluorescence of the acceptor are observed;* ***(b)*** *three colour FRET where weak fluorescence from the donor and acceptor 1 with strong fluorescence observed from acceptor 2*

nearby components. In order to create a useful photonic wire, a rigid scaffold is required on which to position the fluorophores with very high precision that also allows careful control of its assembly. DNA fulfils all of these requirements, would allow networks to be formed through the use of junctions and has been extensively used to control the assembly of other building blocks, giving it a dual role in these bottom-up devices.

Several examples of photonic wires have been produced using DNA with two different templating mechanisms being used and two distinct approaches to the design of the wire. The simplest method to create a wire which was used during the initial[100] work, and for many subsequent studies, is to use end and internally modified single strands of DNA to position a series of different fluorophores where the irradiating energy 'cascades' along the wire and is re-emitted at a much longer wavelength (Figure 23).[101] The problem with this design is that as the energy moves along the wire a significant amount of energy is lost in the rapid vibrational relaxation of the excited state (Stokes shift) between each transfer step. Even an 'ideal' wire absorbing 500 nm light operating at 100% FRET efficiency and with complete re-emission from the final acceptor at 700 nm will have lost 29% of the total energy introduced into the wire. Not only is this inefficient, the range of absorbance and emission wavelengths that have been reported for fluorophores also imposes a maximum shift in wavelength onto the system, making the combined Stokes shift of all the fluorophores a fundamental limitation on the length of wire that could be formed using this approach. The relative

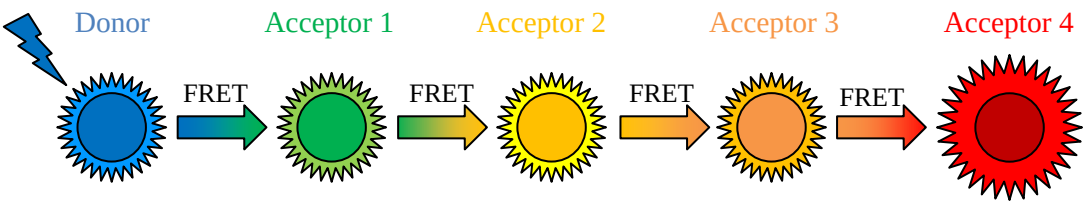


Figure 23 *– Design of a 5 colour cascade photonic wire*

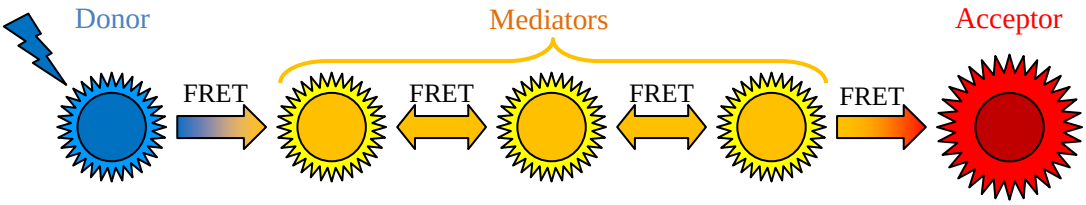


Figure 24 *– Design of a three colour diffusive photonic wire*

complexity of the cascade wire and the multiple energy transfer pathways that can take place have already been found to greatly complicate the analysis of the wire due to the deactivation of fluorophores and energy transfers 'skipping' some of the positions,[102, 103] making the use of low Stokes shift fluorophores a problematic solution to this limitation.

An alternative design to the 'cascade' approach exploits the ability of a fluorophore with significant overlap between its absorbance and emission spectra and a small Stokes shift to undergo FRET with another fluorophore of the same type, a process known as homo-FRET. As the energy levels of the donor and acceptor are the same, there is little loss of energy due to relaxation and the process can be repeated many times within the same wire (Figure 24). Unlike a 'cascade' molecular wire, the transfers are reversible making the energy transfer a very diffusive process where the energy is 'pushed' along the wire by the rapid 'filling' of nearby excited states by the donor. In common with the 'cascade' design there are multiple pathways for the energy transfer but as there are far fewer unique energy transfer steps that can take place, analysing the output of the wire is less problematic. The first designs[104, 105] of this type were constructed using the same modified DNA approach as previous 'cascade' wires but there are severe limitations associated with this method of construction.

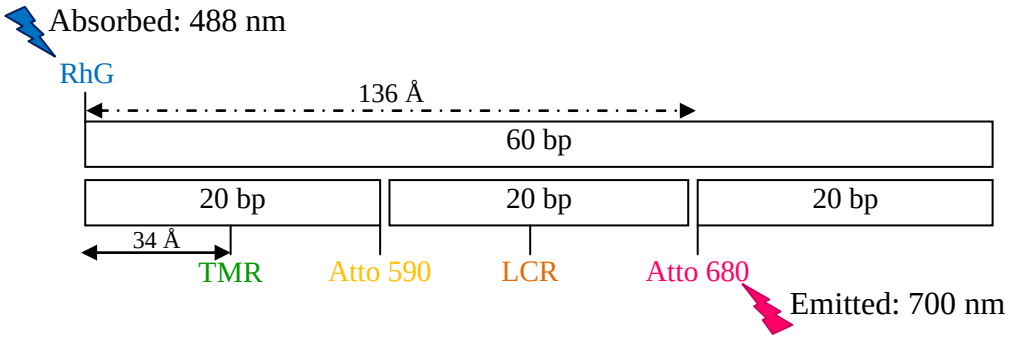


Figure 25 *– A five colour DNA template cascade photonic wire with 136 Å length*[101] *(RhG: Rhodamine Green; TMR: tetramethylrhodamine; LCR: LightCycler Red)*

Although a five colour 'cascade' photonic wire has been reported (Figure 25),[101] it has since been found that in solution 77% of wires do not behave as expected, with 40% showing emission exclusively from the donor and the remainder showing strong emission from intermediate fluorophores.[103] Part of this was attributed to deactivation of fluorophores, contacts with the surface on which the DNA was bound and the formation of secondary structures, but the authors acknowledge that incomplete assembly of the system plays a significant role. Achieving complete assembly of multiple strands remains a fundamental problem when using short strands of modified DNA but some progress has been made by incorporating additional internal modifications which allows fewer and longer strands to be used with higher affinity.[105, 106] The use of ligating enzymes to connect the shorter lengths of DNA into a single, more strongly binding strand after assembly has been suggested as an alternative method of addressing this problem,[103] but this would require enzymes capable of accepting the modifications to be identified which might prove problematic with the predominance of end-labelled DNA.

A rather different approach to constructing photonic wires has been reported which offers some solutions to these problems.[107] Rather than changing the DNA in order to position modifications, standard DNA can be used as a template to direct the assembly of the fluorophores. The first photonic wire to incorporate this approach still

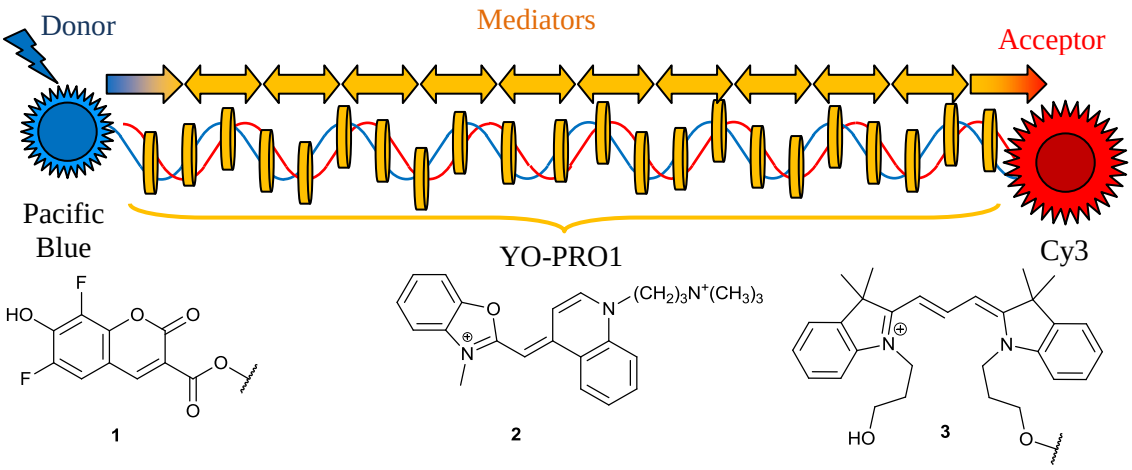


Figure 26 *– An intercalator based diffusive photonic wire*[107]

used two long end-labelled DNA strands to act as donor and acceptor, but used an intercalating dye (YO-PRO1, **2**) to act as mediator to create a diffusive design (Figure 26).[107] Energy transfer was observed over a distance of approximately 20 nm which represents a significant improvement over previously reported cascade and diffusive photonic wires (13.6 nm[101] and 13.2 nm[105] long respectively).

Despite this achievement the use of intercalators imposes significant limitations on the system, particularly during construction. The binding of intercalators shows little sequence specificity and as a result saturating conditions are required in order to create a predictable distribution of fluorophores. This is achieved through the use of a large excess of the intercalator, leaving a significant proportion of the molecules free in the solution which allows energy to 'leak' into unbound intercalators where it does not propagate along the wire. Such a high level of binding with approximately one molecule for every two base pairs also causes significant unwinding of the DNA double helix which reduces the confidence with which the dimensions of the strands can be predicted, undermining one of the key advantages of using DNA. The lack of sequence specificity also means that this approach does not allow selective connections to be formed – the excess of intercalator must be maintained to preserve binding and any DNA strand present will also become saturated. This behaviour can be exploited to

assemble large numbers of intercalators within DNA structures to 'harvest' light which is then re-emitted at a higher wavelength by a fixed fluorophore.[108, 109] The lack of directed assembly with intercalators and the incomplete assembly of multiple DNA 'sticky ends' illustrate how controlling the interactions between molecular building blocks remains a substantial barrier for the bottom-up approach.

1.6 Limitations of current approaches for the construction of DNA based photonic wires

The problems of forming molecular wires within a larger network using both the cascade and diffusive approaches are significant. Achieving complete assembly of a DNA end-labelled system has been shown to be problematic, whilst the lack of sequence specificity with intercalators based systems makes it impossible to selectively create a specific connection. The key limitation of both approaches as with many bottom-up techniques is controlling the assembly of the individual components within a larger framework. What is needed is a simple and general method of attaching these molecules to a specific site on a DNA structure with high precision and accuracy which uses the information contained within the sequence as an 'address' without significantly disrupting the framework. This need is not entirely novel, for millions of years since life began nature has achieved a similar goal of directing the biological machinery of cells to specific genes within the 3 billion base pairs of the human genome. Almost all of this is achieved through the use of proteins, but evolutionary pressures have produced some organisms which secrete small molecules which bind to DNA and interfere with this process. Modifying these natural products to create a programmable system of molecular recognition would yield the desired mechanism for addressing DNA which would have much wider uses in the area of nanotechnology and fundamental research.

Not only would it allow a high efficiency molecular wire to be created easily, it would also allow the creation of a 'toolbox' of molecular components to be assembled into the more complex networks and devices. The relative simplicity and direct 'readout' of each individual component that make up a photonic wire make it an ideal starting point for creating this toolbox to solve one of the greatest challenges for bottom-up nanotechnology.

1.7 Aims and Objectives

This project aims to provide proof of principle for this toolbox approach to DNA nanotechnology by establishing a general synthetic route to functionalised and programmable small molecules capable of directing a wide range of nanometre scale 'payloads' onto a specific site on DNA. In order to maximise the utility of the toolbox a wide variety of DNA sequences should be targetable with high affinity and precise control over the final position of the payload. To demonstrate these principles, a photonic wire will be constructed using this toolbox approach where it acts as a direct replacement for existing DNA end labelling methodology.

Chapter 2 DNA binding polyamides

2.1 Introduction, aims and objectives

Despite the much broader general meaning, polyamides is the adopted name of a family of DNA binding small molecules containing pyrrole and imidazole rings developed largely by Dervan *et al*..[110-113] These molecules are modular, readily programmed and have high binding affinity, making them an ideal mechanism for directing functionality onto a wide range of DNA sequences. As a result, it was decided to produce and establish a general route to a range of modified polyamides containing functional handles suitable for the envisaged 'mix and match' approach for attaching the 'payload'. The key objective within these aims is the synthesis of a modified polyamide suitable for the assembly of the photonic wire that will act as proof of principle for this 'toolbox' approach.

2.2 Molecular recognition of double stranded DNA: the major groove, minor groove and intercalation

2.2.1 Addressing DNA

The central role of DNA in biological systems is as a long term store of the genetic information that acts as the blueprints for the complex biological processes taking place within each cell vital for its survival. A wide range of proteins have evolved to read and process this stored information with several families of motifs which recognise binding sites ranging from a specific sequence of three bases in the case of zinc finger proteins[114] (Figure 27a) through to the non-specific binding of the DNA repair enzymes. Many protein recognition motifs such as leucine zippers[115] and

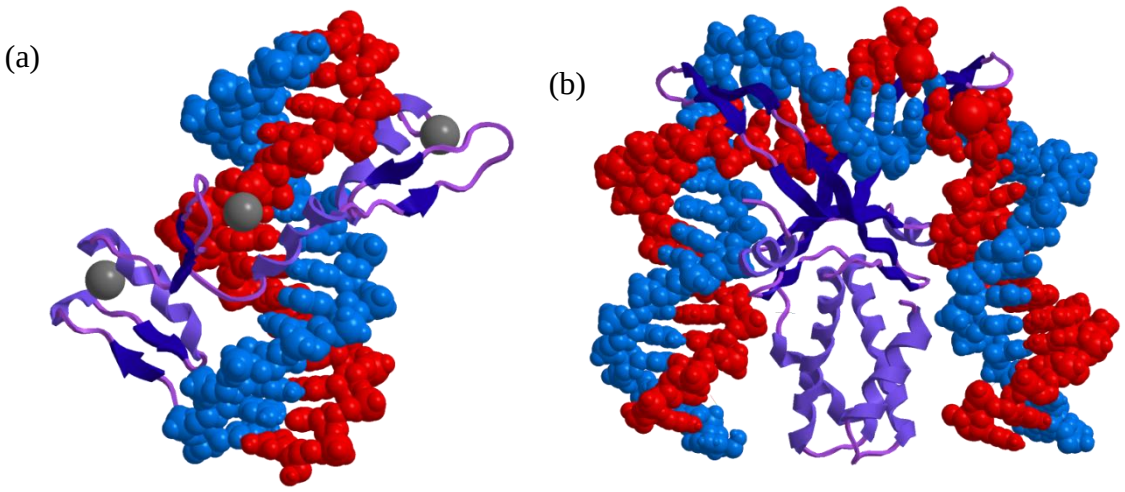


Figure 27*- Selected DNA binding proteins bound to matching strands;* ***(a)*** *zinc finger protein containing 3 recognition motifs;*[116] ***(b)*** *IHF protein inducing a 160° bend in DNA*[117]

helix-turn-helix domains[118, 119] achieve their specificity by positioning a rigid α-helix motif within the major groove allowing the residues facing the DNA to form hydrogen bonds with the exposed edges of the bases. Compared to the minor groove, major groove binding is much more common as it is broad enough to accommodate the proteins recognition structure and exhibits greater sequence specific variations in size, shape and hydrogen bonds to help discriminate potential binding sites. The high affinity of proteins comes from a large number of relatively weak independent interactions, making it very difficult to predictably design or modify the binding site of a particular protein. Some success has been achieved by recombining existing motifs from zinc finger proteins to target new sequences,[30] but in general protein design remains a matter of trial and error.

Filling the width of the major groove is the underlying reason for this myriad of small interactions and it might be thought that minor groove binding proteins might be simpler in structure and easier to design. Although less common, many of these motifs have been identified such as the TATA binding protein,[120, 121] the HMG box family[122, 123] and the IHF protein[117] which are all involved in controlling the replication of DNA. The minor groove is normally too small to accommodate these motifs and upon binding

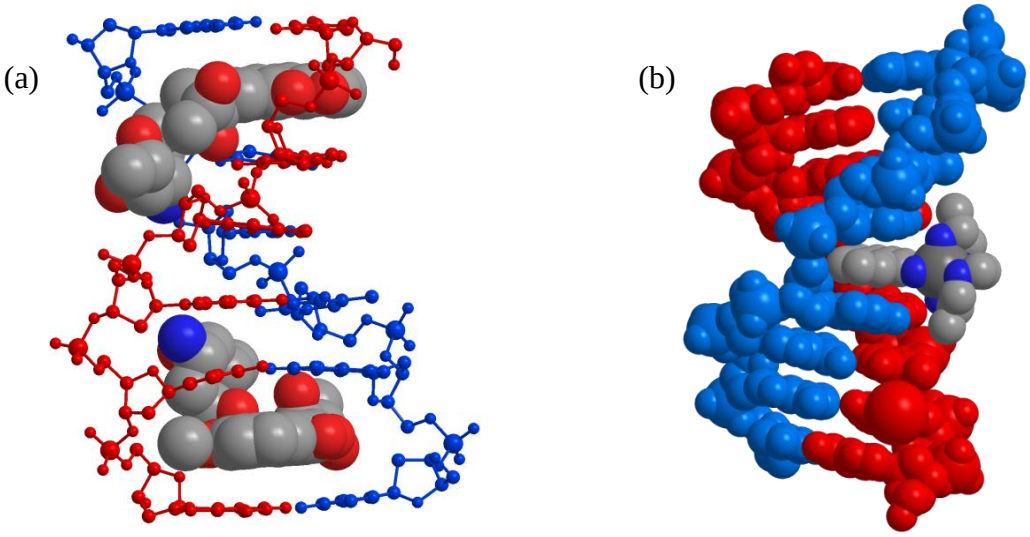
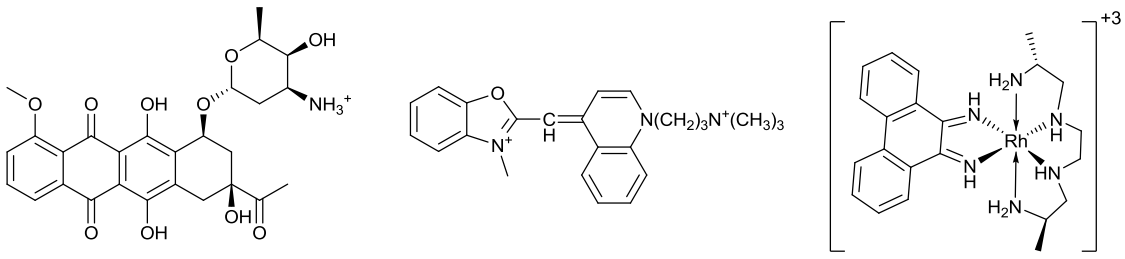


Figure 28 *- DNA binding modes of intercalators:* ***(a)*** *a simple intercalator Daunomycin* ***4***[124] *(PDB ID: 1D11);* ***(b)*** *a metallo-intercalator* ***5***[125] *(PDB ID: 454D)*

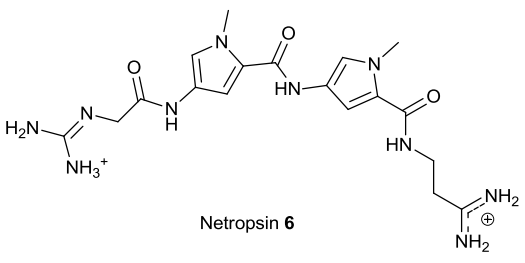
these proteins distort the DNA by bending or unwinding which allows other larger proteins to bind. An extreme example of this is the IHF protein (Figure 27b) which induces a 160° bend[117] but most proteins induce some change over an extended length of the helix upon binding which is highly undesirable if DNA is to be used as a defined scaffold. When the additional problems of synthesis and purification are also considered, proteins are ill-suited for bottom-up nanotechnology applications where ease of mass production is essential for moving the devices outside of the research lab.

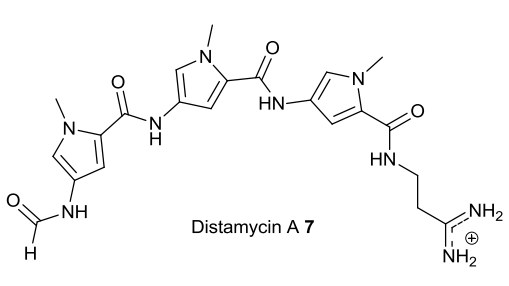
In addition to proteins, some species have evolved to produce DNA binding natural products which are secreted from cells to act as antimicrobial agents and reduce competition for resources. These compounds are frequently positively charged and invariable contain aromatic rings which interact favourably with the anionic phosphate backbone and hydrophobic bases of the DNA respectively, inspiring the synthesis of a



wide range of analogues. The major groove is too wide and requires too many interactions for these relatively small molecules to bind effectively, with minor groove binding and intercalation being the preferred binding modes. Intercalating molecules such as daunomycin **4** and the fluorophore YO-PRO **2** are highly aromatic and insert into the hydrophobic spaces between the stacked bases of the helix (Figure 28). A small degree of specificity arises from sequence specific variations in the dimensions of the helix but in general intercalators bind in any available site, limited only by the distortions in the double helix caused by the intercalation which prevent adjacent 'gaps' between base pairs from being occupied, commonly referred to as the neighbourhood exclusion principle. The lack of interaction with the exposed edges of the bases has been addressed using metallo-intercalators such as $[Rh(phi)(Me_2trien)]^{3+}$ **5** where coordination of the intercalating ring to a metal centre allows ligands to be positioned in the major groove (Figure 28b) but the limited sequence recognition that be achieved using intercalators means that they are fundamentally unsuited for targeting specific sites on DNA.

A more promising avenue of natural products are those that bind in the minor groove.[58, 126, 127] Some of the first compounds of this type that were isolated from Streptomyces bacteria were netropsin **6**[128] and the structurally related distamycin A **7**[129] which both consist of a series of pyrrole rings connected by amide linkages that allow the molecules to match curve of the double helix. Both molecules bind preferentially in the narrower minor groove associated with A:T rich sequences and at low





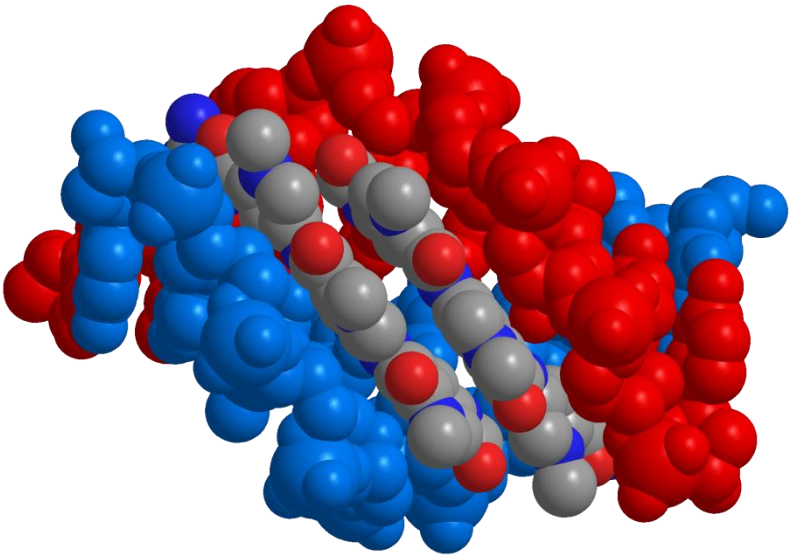
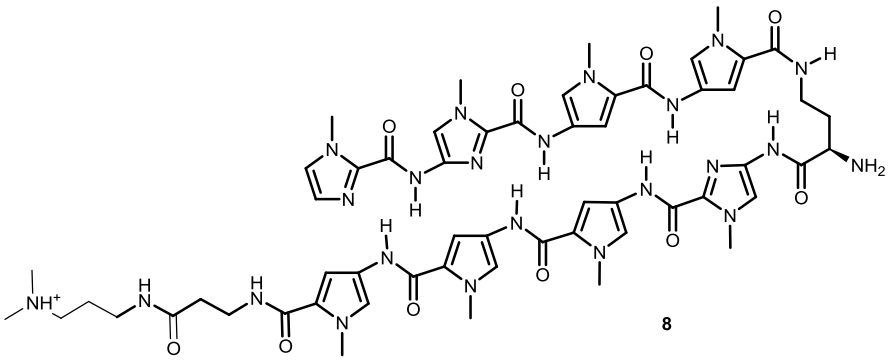


Figure 29 *- Crystal structure of the 2:1 binding mode of distamycin A* ***7***[130]

concentrations bind in a 1:1 ratio with the DNA. At increased concentrations distamycin A **7** was found to have an unexpected antiparallel 2:1 binding mode[130, 131] where two drug molecules stacked on top of each other (Figure 29) with the positive charges at opposite ends. The relative simplicity of these structures combined with their proven ability to interfere with DNA replication in their role as antimicrobial agents made them attractive targets for pharmaceutical development, which led to the optimisation of these polyamides into a complete system of DNA recognition by the Dervan group.[132]

2.3 DNA binding polyamides

2.3.1 The basic building blocks



8

The term ‘polyamide’ can be applied to a wide range of molecules, but has been adopted to describe the distamycin derived natural products such as **8** containing several aromatic rings connected by amide linkages and with a 2:1 preferred binding mode. The curvature of the molecule closely matches that inside the minor groove and positions each ring adjacent to a single base, creating a modular system where the DNA recognition properties at each position can be modified independently and allow any DNA sequence to be targeted. Numerous building blocks have been reported[113, 133-140] for incorporation into polyamides which afford varying DNA binding properties through both passive and active mechanisms but only the limited number shown in Table 1 are in common use. The modular nature, site specificity and ease of modification of polyamides have already been exploited to control gene expression in biological systems, but these properties also make them an ideal mechanism to extract the sequence ‘address’ from a DNA structure with nanometre scale accuracy and a high degree of control over the final position of the nanotechnology ‘payload’. Despite this, polyamides are not without limitations, most notably in their lack of ability to discriminate A and T bases and the detrimental effect of many modifications on DNA binding affinity.

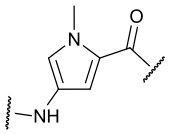
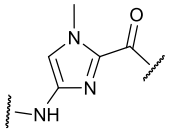
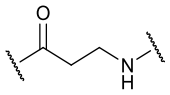
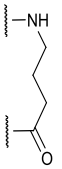
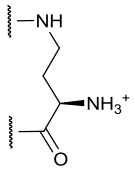
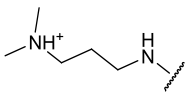
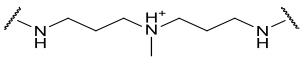
	Symbol	Structure	Sequence Affinity	Relative Binding affinity
Py			A/T	Medium
Im			G	Strong
β			A/T	Weak
Uncharged turn (**γ**)			A/T	Weak
Charged turn			A/T	Strong
Tail (**Dp**)			A/T	Strong
Linking tail			A/T	Strong

Table 1 *- The symbols, structures and binding affinity of the common polyamide building blocks*

When the DNA binding modes of netropsin **6** and distamycin **7** were characterised by x-ray crystallography, it was recognised that substituting a pyrrole ring with an imidazole might create interactions with the G:C base pair[141] and override the existing binding preference for long A:T/T:A regions, allowing a much wider range of sequences to be targeted. These modified natural products were soon shown to have this altered selectivity[142] and the preferred binding mode characterised by NMR as an anti-parallel 2:1 ratio where two molecules stack on top of each other,[110] in contrast to the natural products from which they were derived. The crystal structures of

polyamides including the methyl imidazole **Im** building block show the nitrogen of the imidazole forming a hydrogen bond to the N-2 of the guanine residue of the G:C base pair (Figure 30) in addition to the amide interactions observed for the pyrrole **Py** unit.[143] The ability of the **Im-Py** pair to select G:C base pairs over A:T base pairs comes from the need to form the additional hydrogen bond to overcome the added energetic penalty of desolvating the imidazole nitrogen.[144] In contrast to this active discrimination the A/T preference of paired pyrrole **Py-Py** building blocks as in the parent natural products is more passive and stems from the narrower width of the minor groove in A:T rich regions. As a result of this, the pair has minimal ability to discriminate between A:T and T:A base pairings.[145] In addition to the interactions of the polyamide rings, the amide links also form hydrogen bonds to any exposed lone pairs within the minor groove (Figure 30), but these impart no sequence specificity.

Early polyamides using the **Im** building block closely resembled the linear parent natural products and due to their reversible orientation were limited to

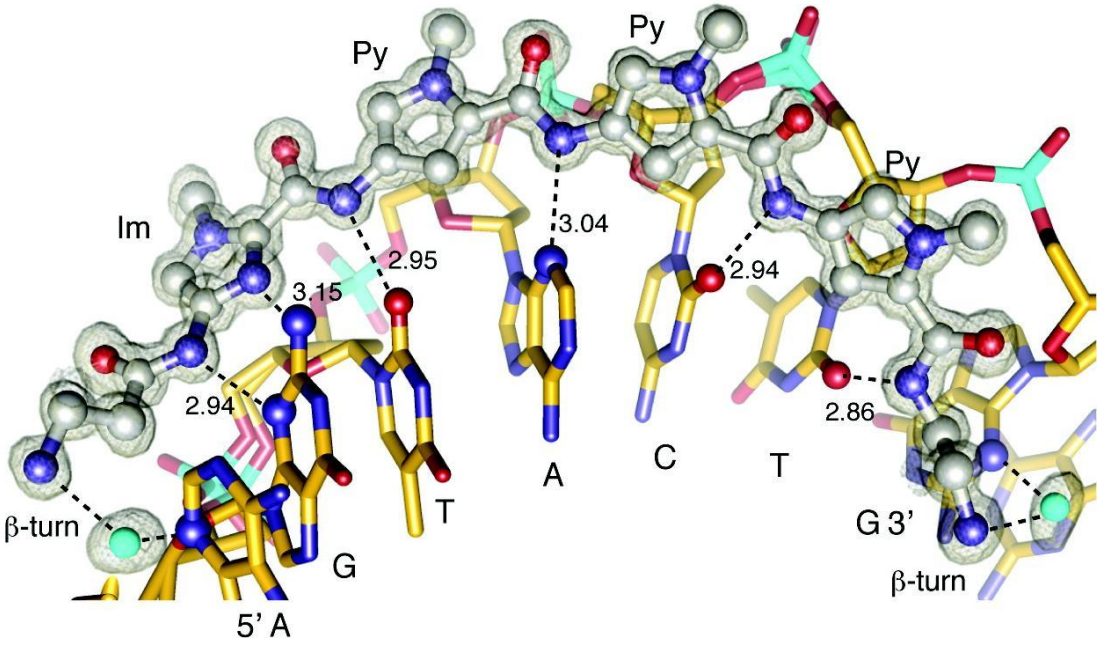
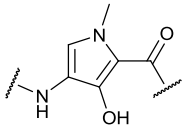


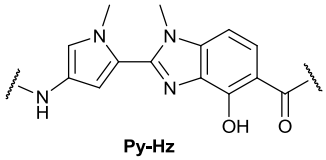
Figure 30 *– Hydrogen bonding of a DNA bound polyamide, including specific recognition of the G base by the* ***Im*** *building block[146] (PDB: 3OMJ)*

palindromic sequences with misalignment of molecules in the 2:1 binding mode also causing significant off-target binding.[112] Linking two polyamides together using a flexible alkyl chain to create a hairpin structure prevents this 'slippage' and fixes the binding orientation, allowing asymmetric binding sites to be targeted and increasing the overall binding affinity.[134] The hydrophobic nature of the 'turn' building block confers a weak A:T/T:A preference onto the base pair following the central binding sequence recognised by the **Im** and **Py** building blocks. This extends the length of sequence recognised but as no G or C targeting equivalent has been reported also imposes restrictions on the sequences which can be targeted. The addition of an amino group to the turn causes a dramatic increase in binding affinity as it becomes positively charged in solution and interacts strongly with the phosphate backbone of the DNA, but introduces a chiral building block to the synthesis. The α-position remains the most common site for the amino group as it corresponds to an amino acid,[147] but other substitutions have been reported with the amino group in the γ-position[140, 148] and attached to a short alkyl chain[139, 149] through the use of expensive chiral building blocks which has limited their uptake for general use.

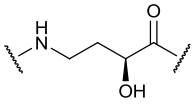
Selective recognition of the A:T base pair within the central binding sequence is possible using a hydroxyl group attached to one of the rings (the **Hp**[150] and **Hz**[151] building blocks) or an alkyl chain[138] **9** which projects into the asymmetric cleft of the A:T base pair, producing both shape and hydrogen bonding based discrimination. Unlike the **Im** building block, substitution of the **Py** building block results in a decrease in the binding affinity of the polyamide which was originally thought to be due to



Hp



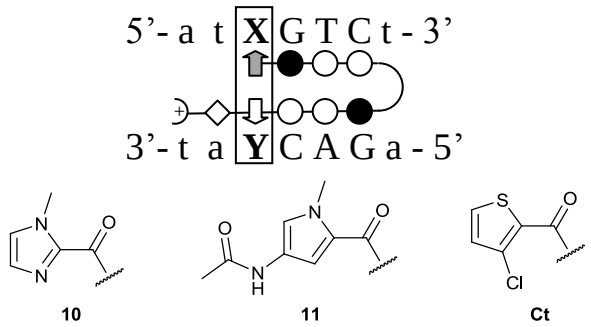
Py-Hz



9

distortions created in the DNA double helix[152] but this is now in doubt[153] and subsequent calculations have suggested that the energetic penalty of desolvating the hydroxyl group upon binding is not balanced by the formation of a new hydrogen bond to the O2 of thymine.[154] In addition to this loss in affinity, the presence of the hydroxyl group requires additional protecting groups during the synthesis of polyamides and decreases the products stability to both acids and free radicals.[151] Due to these problems and because **Im** based G:C recognition is normally sufficient to target a particular DNA binding site, the **Hp** and **Hz** building blocks are rarely used. As with biological systems, for nanotechnology applications the stability of both the polyamide and the DNA complex are both key requirements in addition to the specificity of binding which rules out the use of these building blocks.

In order to compensate for the lack of practical methods to achieve A:T discrimination, a longer sequence must be targeted to ensure a unique binding site. Similarly to the 'turn' segment, the hydrophobic 'tail' and **β** building blocks incorporated in most polyamides impart a minor A/T binding preference for the bases preceding the central binding sequence. This extends the recognition site for an eight ring polyamide to the seven base pair sequence 5'-wwXXXXw-3', where 'w' indicates a weak A/T preference and 'X' corresponds to the central binding sequence programmed using **Im** and **Py** building blocks. Attempts to expand the binding site by increasing the number of pairs of building blocks beyond four found no increase in binding affinity[155] which was attributed to the over tight curvature of the polyamide compared to the DNA.[143] Substituting the **β** building block for one of the **Py** building blocks on each half of the polyamide allows the curvature to be 'reset',[133] resulting in improved binding affinity for longer polyamides and enabling sequences up to 11 base pairs long to be targeted.[156] This substitution is not without cost as the **β** building block



⇧	10	11	Ct
⇩	Py	Im	Py
X : Y	G : C	C : G	T : A

Table 2 *- Sequence selectivity of polyamide 'stopper' units*

does not have the same favourable interactions with the minor groove as **Py**, resulting in a net loss in binding affinity for short polyamides where there is no over curvature to be corrected.

The **Py**, **Im**, **β** and various hairpin turn building blocks allow a wide range of sequences to be targeted but there are still some important restrictions on the sequence of the binding site. The preference for A and T bases before and after the central recognition sequence cannot be easily altered, though it can be reduced by changing the **β** and tail building blocks.[134] It is also desirable to have active recognition at the beginning of the central recognition sequence to prevent ‘slipping’ of the polyamide, which is most commonly achieved using the **Im** stopper **10** paired with **Py** to recognise G:C bases, with the reversed C:G recognition obtained using an acylated **Py** building block **11** paired with **Im** (Table 2). Targeting of T:A bases is also possible in this position using the **Ct** building block[135, 157-159] but no ‘stopper’ unit has been reported to allow the equivalent A:T base pair to be used as the start of the central recognition sequence. For biological uses these are requirements directly impact which sites can be targeted, but as in most nanotechnology purposes the sequence of the DNA binding site and positioning of the polyamide will be designed, these restrictions can be circumvented.

2.4 Site specific deposition of functionality along DNA duplexes using polyamides

2.4.1 Types and Binding Affinities of Existing Modified Polyamides

Polyamides were originally conceived of as a means of influencing gene expression in living cells and the majority of reports have focused on their applications for regulating gene expression.[160] Although polyamides are capable of blocking gene transcription when unmodified,[161] the medicinal potential of increasing gene expression by recruiting proteins and of covalently attaching polyamides at their binding site has led to the development of a variety of methods and sites to attach functionality (Figure 31). The most common and facile site for modification is through the use of the 'linking tail' building block where the functionality is introduced during or immediately after synthesis, allowing a single batch of polyamide to be divided to create a library of modifications. The amino group of the charged 'turn' building block is also a frequent site of attachment which, in common with tail modification, positions the attached 'payload' within the minor groove of the DNA. In order to ensure that the attached functionality projects out of the double helix and is capable of interacting with other molecules in solution, the **Py**[162, 163] and **Im**[164] building block can also be modified with

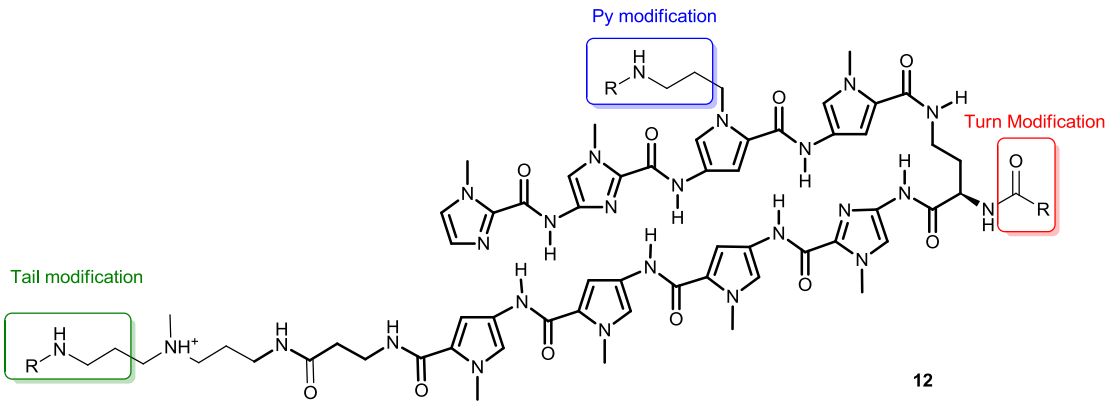
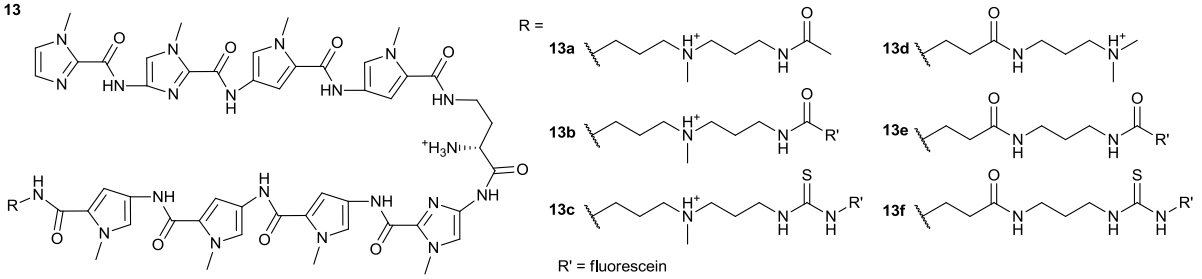


Figure 31 *- Potential sites for polyamide modification*

a flexible chain. A wide variety of fluorophore modified polyamides have already been synthesised to study the effect of the payload and linkages on uptake into a variety of cells[165, 166] but only a small minority have had DNA binding constants reported in the literature and appear to be strongly influenced by both the nature of the linkage and the 'payload' attached. The availability of DNA binding data for a variety of fluorescein-polyamide conjugates allows direct comparisons to be made on the role of the linker, overall charge and position of the modification and permits some general conclusions to be reached on the role of each.

Fluorescein is a negatively charged fluorophore and the reduction in binding affinity after tail modification (Table 3) can be attributed not only to the steric repulsion within the minor groove but also to the repulsion from the phosphate groups along the DNA backbone. The differences between tail modified polyamides with an amide linkage and thio-urea tail linked conjugates (Table 3 - **13b**, **13c**; **13e**, **13f**) suggest that the positioning of fluorophore by the linker may have help reduce these effects, but

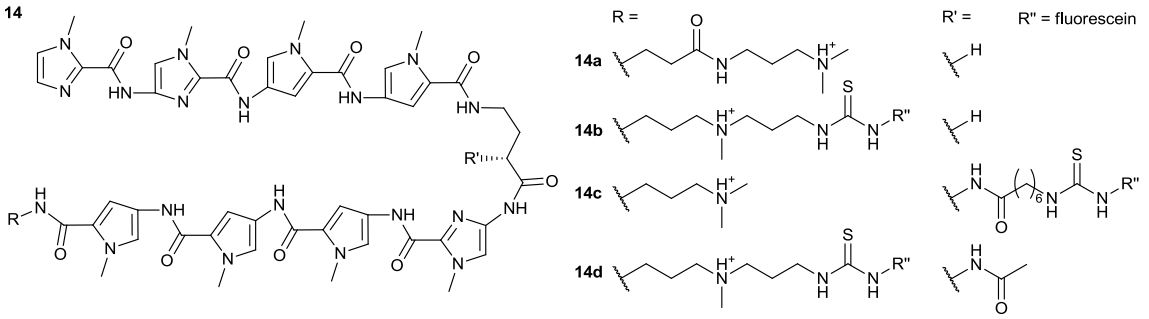


	Equilibrium binding constants at match site 5'-aat**GGTC**a-3'					
Linkage			K_D (nM)			K_D (nM)
-	**13a**	Ac-N	0.032	**13d**		0.027
Amide	**13b**	FITC-N	0.15	**13e**	FITC-N	0.31
Thio-urea	**13c**	FITC-S	0.059	**13f**	FITC-S	0.16

***Table 3**- Variations in binding affinity for tail modified polyamides due to payload and linker[167] with central recognition site in bold. FITC – Fluorescein isothiocyanate*

numerous favoured and disfavoured interactions between the linker and the DNA may also be present. The deletion of the cationic regions within the linker in contrast has a very clear effect of reducing the DNA binding affinity substantially (Table 3 - **13b**, **13e**; **13c**, **13f**), but the effect is not as severe as the difference observed between the charged and uncharged hairpin turn in unmodified polyamides (Table 3 - **13d** , Table 4 - **14a**). Tail modification of polyamides represents a quick and easy method to attach a wide variety of functionality, but the impact on the conjugates' affinity for DNA is hard to predict and can vary significantly between payloads.

The free amine on the hairpin turn of polyamides has also been used as an alternative modification site for polyamides using the same standard amide bond formation chemistry as for tail modification. As this amine is always protected during solid phase synthesis, the modification is relatively simple to introduce at several stages, but being a secondary amine the reaction proceeds slower than attachment at the tail.



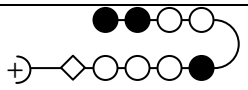
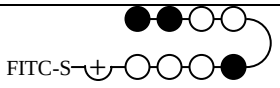
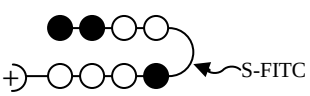
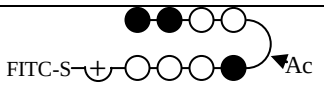
	Polyamide	Binding site	K_D (nM)
14a		5'-tat**GGTC**atg-3'	0.21
14b		5'-cat**GGTC**ata-3'	0.63
14c		5'-att**GGTC**aag-3'	53
14d		5'-att**GGTC**aag-3'	0.83

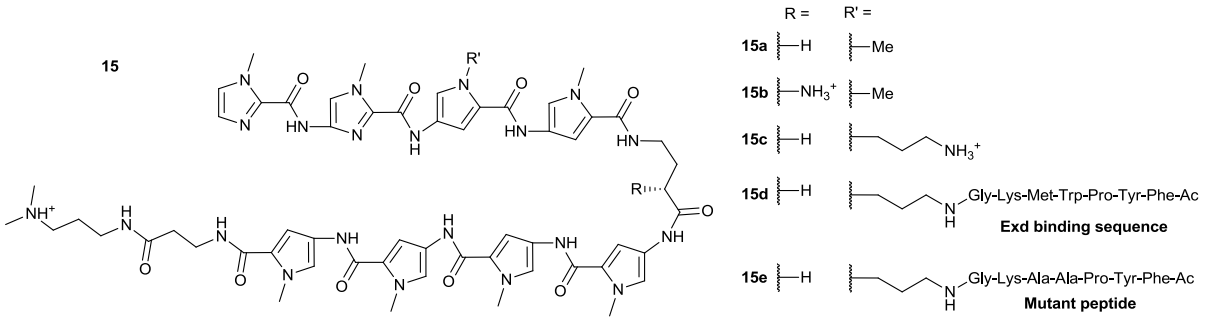
***Table 4** - The effect of turn and tail modification on binding affinity[164] with central recognition site in bold. FITC – Fluorescein isothiocyanate*

Despite the relative ease of this modification it is rarely used due to the dramatic 100-fold reduction in binding affinity that is observed compared to the parent polyamide (Table 4 - **14a**, **14c**) and to the equivalent tail modified polyamide (Table 4 - **14b**, **14c**). The addition of an acetate protecting group to the tail modified equivalent results in little loss in affinity (Table 4 - **14c**, **14d**), suggesting that the final positioning of the payload may be greatly disfavoured despite the longer linker attached to the turn. This effect has also been observed for the attachment of other fluorophores and is accompanied by a loss in binding specificity,[164] making it an unsuitable modification site for most applications.

One of the drawbacks for both turn and tail modification is the need for an extended flexible linker to allow the payload to be positioned outside of the minor groove as this prevents the exact positioning of the payload being known. Fluorophores are generally planar organic molecules and as such may preferentially remain within the hydrophobic minor groove rather than the hydrophilic environment around DNA which may cause increased quenching or a shifting in the fluorescence spectrum. Internal modification of polyamides positions the payload in the centre of the binding sequence and with the minor groove occupied by the main body of the polyamide there is limited scope for undesirable interactions. Modification from the pyrrole[162, 163] and imidazole stopper building blocks[164] have both been reported using flexible *N*-alkyl linkers with both alcohol and amine functionalised intermediates produced. A derivatised pyrrole building block would allow positioning anywhere within the sequence and would allow multiple attachment sites, whereas the modified stopper has a fixed usage and the synthesis involves unstable intermediates.

In principle, internal modification directs the payload straight out of the minor groove into the solvent and should have minimal effect on the binding affinity. The

introduction of the alkyl chain as the free amine appears to support this idea, showing an intermediate binding affinity between the singly and doubly charged parent polyamides (Table 5 - **15c**, **15a**, **15b**). The importance of the payload is shown by the addition of a functional peptide to bind the protein Exd to the end of this linker resulting in a dramatic 100-fold decrease in binding, but a non-functional peptide of the same length results in a only a 10-fold reduction (Table 5 – **15d**, **15a**, **15e**). This was originally attributed[162] to secondary structures of the active peptide which formed part of a protein binding domain, but the results can also be interpreted in terms of the



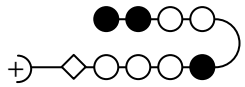
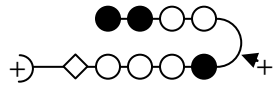
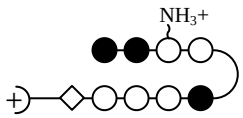
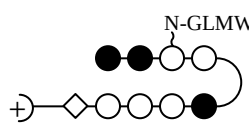
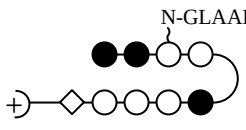
	Equilibrium Binding Constants - K_D (nM)		
	Match Site	Mismatch Sites	
	5'-tGG**T**Ca-3'	5'-tGG**C**Ca-3'	5'-tGG**G**Ca-3'
15a[150] 	0.077	-	-
15b[139] 	0.020	-	-
15c[162] 	0.048	0.97	0.76
15d[162] 	5.8	7.9	>100
15e[162] 	0.86	12	>100

Table 5*- Match and mismatch binding constants for internally modified polyamides with modification site underlined and mismatch sites in bold*

lipophilicity of the amino acids within the peptide domain. The short length of the peptide would make any secondary structure easy to disrupt and the active peptide contains several aromatic residues that would have favourable π-stacking interactions with the main body of the polyamide which would have to be disrupted before DNA binding could take place. The non-functional peptide replaced a methionine and a tryptophan residue with alanines which are much less hydrophobic and would not have the same energetic penalty to overcome, making DNA binding more favourable. If this explanation does play a part in these results, the severe loss in binding affinity would also affect highly hydrophobic fluorophores such as cyanine dyes and pyrene derivatives which would not have the favourable solvent interactions that can be expected for more polar fluorophores such as pacific blue **1** to offset the loss in π-stacking interactions.

Looking beyond the match site binding affinity, attaching a payload can also have other affects. Both the active and non-functional peptides show a greatly reduced affinity for G:C match sites over C:G and T:A base pairs (Table 5 column 3 vs. columns 1 and 2), granting sequence specificity much greater than shown by both the unfunctionalised chain and that shown by unmodified polyamides (Table 5 - **15c**, **15e**).[133] This sequence discrimination suggests that the addition of the payload causes shifts the positioning of the modified pyrrole within the minor groove and it has been suggested the specificity arises from steric clashes with the exocyclic amine of the G base.[162] The lack of available DNA binding data makes it difficult to determine the general effects of internally modifying polyamides, but the identity of the payload appears to play an important role in determining both the affinity and specificity of the conjugates.

The ability to use the modified **Py** building block as part of any pair along the length of the binding site allows for much finer control over the final positioning of the payload and minimises its interactions with the DNA, an important consideration for fluorophores which may be quenched in close proximity to the bases and for larger modifications such as nanoparticles which might otherwise clash with the DNA. The simplicity and ability to create a library of compounds from a single synthesis also makes 'tail' modification an attractive method for attaching functionality. In recent years, the potential role of polyamides in nanotechnology has been recognised[168] and several applications have been reported which show how they can be used to improve on existing methods.

The ability of polyamides to replace DNA end labelling has been demonstrated using a biotin labelled polyamide to selectively recruit the tightly binding streptavidin protein onto specific sites of a DNA structure[169, 170] and discrete pieces of DNA architecture have been combined using a pair of covalently linked polyamides 100 times smaller in size.[171] Also of particular interest are the numerous papers describing polyamide-fluorophore conjugates synthesised in order to study their uptake into live cells which cover a wide range of fluorophores and attachment chemistries.[158, 166, 172, 173] These reports illustrate the immense potential of polyamides not only as a building block to create photonic wires but also as a general method for localising nanometre scale objects onto a DNA scaffold.

2.5 Improving Photonic wire design

Polyamides have been proven capable of selectively targeting a specific DNA binding site over a wide range of posibilities[174] with high affinity and of being able to direct a 'payload' onto a specific site, features which allow the creation of a 'mix and

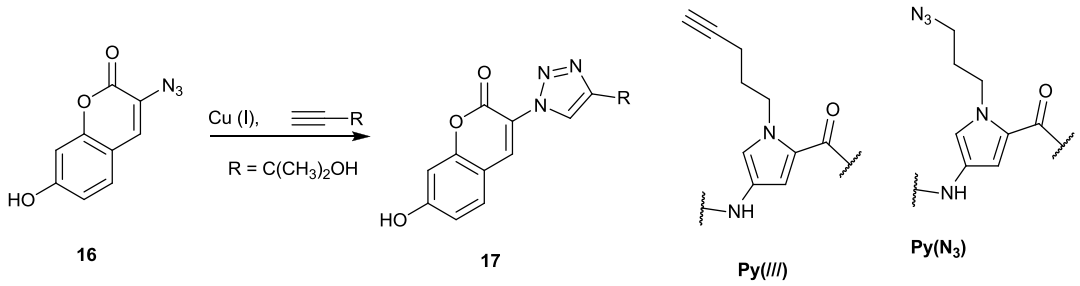
match toolbox' which allows any material or functionality to be positioned onto any structure with nanometre precision. The construction of simple photonic wire using polyamides allows all of these attributes to be demonstrated in a nanotechnology context in addition to its practical applications. As a direct replacement for the end-labelling of DNA strands, a design based on the previously published[107] example of a diffusive photonic wire was created using the intercalating YO dye where the use of a polyamide confers the added benefit of controlling the positioning and stoichiometry of this fluorophore. Working in parallel, Dr Wu Su[175] developed the necessary polyamide-intercalator hybrid compound utilising the same tail modification chemistry previously reported for similar fluorophore-polyamide conjugates,[166] and so work was initially focused on developing new, more flexible attachment chemistry using internally modified polyamides to acts as substitutes for the DNA end labelled donor and acceptor fluorophores within this design which illustrate the wider applicability of this approach.

2.5.1 Identifying Suitable Modifications

In order to create the desired nanotechnology 'toolbox' the chemistry used to attach the desired functionality must be versatile to maximise the range of conjugates that can be produced, selective in order to minimise chemical incompatibilities and ideally amenable to post-synthetic modification to allow the rapid creation of a library of functionality targeted using the same polyamide. The term 'click reaction' was first coined[176] as a general term to describe reactions of this type which were wide in scope, modular, high yielding, specific and had minimal by-products. In addition, the reactions must also use readily available starting materials and operate under simple conditions in benign solvents with easy purification. Several reactions were exemplified as having 'spring-loaded' components that remained inert until a specific

set of conditions were met, but the term click chemistry is now almost exclusively used to describe the copper catalysed Huisgen 1,3-dipolar cycloaddition between azides and alkynes to form 1,2,3-triazoles. This reaction normally proceeds very rapidly and with high yield due to the large quantity of energy stored within the triple bonds and the same groups also form the basis of several other click type reactions such as the azide based traceless Staudinger ligation[178, 179] and the formation of isoxazoles from the reaction of nitrile oxides with alkynes.[180]

Although other groups can be used for click type chemistry, such as epoxides[176] and α-keto acids,[181, 182] importantly azides and alkynes are inert towards a wide variety of reagents and conditions such as the strong acids and bases, coupling agents and extended periods of heating which are encountered during the synthesis of polyamides, making them ideal modifications to incorporate into the building blocks. The synthesis and incorporation of alkyne and azide modified **Py** building blocks would provide a significant step forward towards achieving the desired 'toolbox' approach to nanotechnology. In particular, the alkyne modified building block **Py(///)** would have



***Scheme 1**- Fluorogenic Click reaction of coumarin azide **16** with an alkyne*

	Pacific Blue (1)	**YO-PRO-1 (2)**	**Cy3 (3)**	'Clicked' coumarin azide **(17)**
Maximum excitation	403 nm	491 nm	552 nm	440 nm[177]
Maximum emission	455 nm	509 nm	570 nm	492 nm[177]

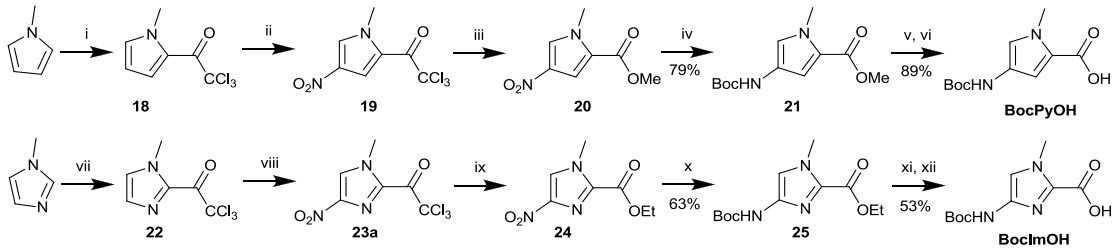
***Table 6** - Fluorescence properties of the previously reported molecular wire[107] and the Pacific Blue replacement **17** (R =-C(CH$_3$)$_2$OH)*

the additional advantage of reacting directly with the fluorogenic azido-coumarin **16**[177] in the presence of a copper catalyst to yield the fluorescent conjugate **17** (Scheme 1) which would act as a direct replacement for the pacific blue donor fluorophore[107] in the improved photonic wire design. In order to obtain the proper excitation wavelength and to maintain similarity with previously reported internally modified polyamides,[162] a short alkyl linker was selected for both the azide **Py(N_3)** and alkyne **Py(///)** building blocks. The immediate applications of the alkyne modified **Py(///)** building block make it an ideal starting point to show how a 'toolbox' of components can be applied to an existing nanotechnology problem.

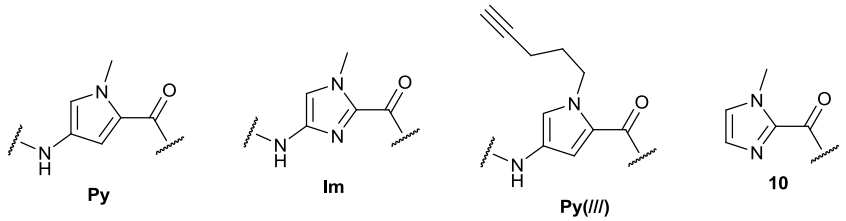
2.6 Synthesis of Polyamide Building Blocks

2.6.1 Synthesis of the BocPyOH and BocImOH building blocks

There are several methods for the synthesis of polyamides, but as all polyamides are derived from the same basic units (Table 1) and all of the methods involve the iterative coupling of the building blocks, the precursors required are the same. The alkyl chain **β** building block along with both the charged and uncharged turn building blocks are commercially available with a variety of protecting groups, as are the **Dp** and

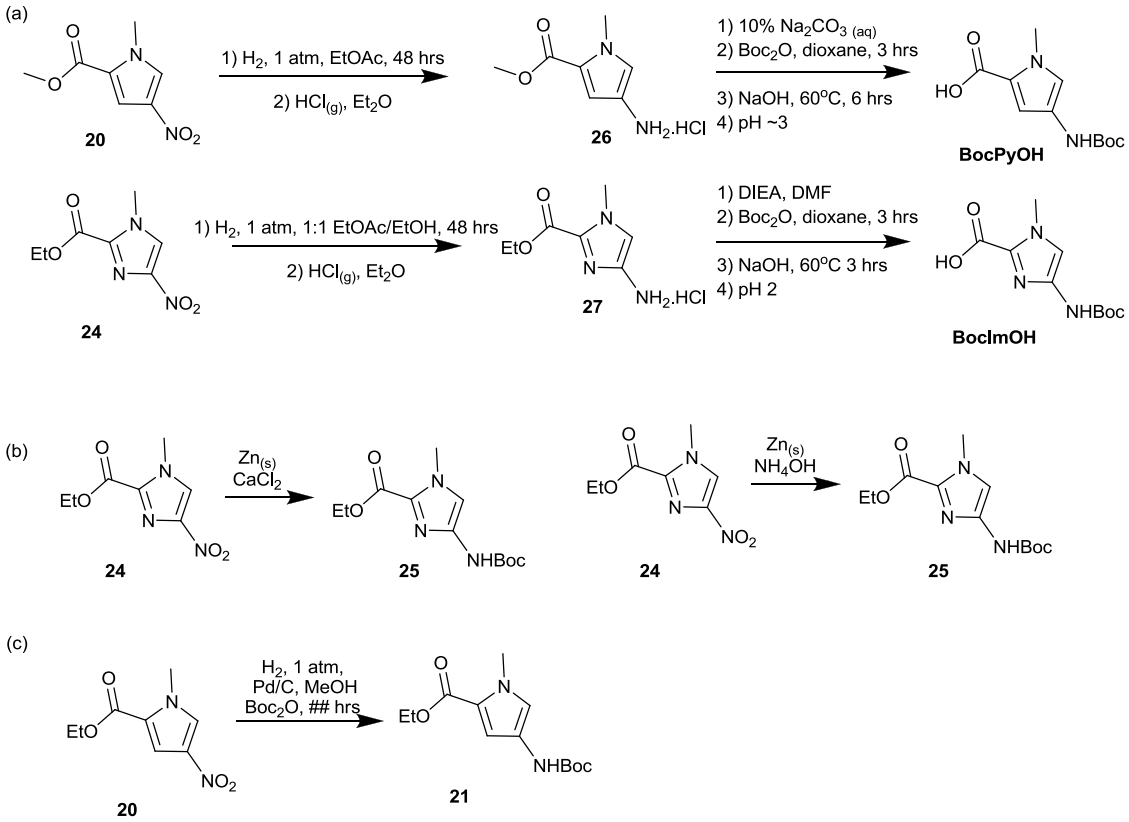


Scheme 2 *- Synthesis of the Boc protected **Py** and **Im** building block: (i) CCl_3COCl,DCM (ii) HNO_3, Ac_2O[183] (iii) NaOMe, MeOH (iv) $NaBH_4$, MeOH, $NiCl_2.6H_2O$, Boc_2O, -20°C (v) LiOH, H_2O, MeOH (vi) pH 2-3; (vii) CCl_3COCl, DCM (ii) HNO_3, Ac_2O (iii) NaOEt, EtOH (iv) H_2, Pd/C, Boc_2O, 48 hrs (v) LiOH, H_2O, EtOH (vi) pH 7;*

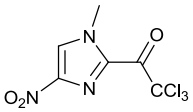


extended linking tail units. In contrast, the aromatic **Py**, modified **Py(///)**, **Im** and **Im** stopper **10** building blocks must be synthesised from simpler starting materials and the desired protecting groups introduced (Scheme 2). Although synthetic methods have been reported using both Fmoc[184] and Boc[183, 185-188] chemistry, Boc protecting chemistry was selected due to its faster coupling times and the continued interest in the optimising the coupling protocols.[185]

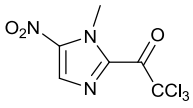
The literature synthesis[183, 184] of the **BocPyOH** building block begins with the acylation of 1-methyl-pyrrole (Scheme 2), but intermediate **20** is commercially available and was used to shorten the synthetic route. Similarly, **Im** precursor **24** is also



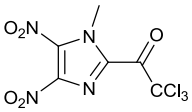
*Scheme 3 – Approaches to the reduction of nitro-precursors **20** and **24**: (a) Original literature procedure[183]; (b)Trial reactions using zinc metal; (c) Hydrogenation in methanol*



23a



23b



23c

commercially available and also allows the avoidance of the problematic nitration of **22** which was found to yield a near equimolar distribution of nitration isomers **23a** and **23b** and in some cases the dinitro derivative **23c**. The original literature procedure[183] for the reduction and protection of nitro precursors **20** and **24** (Scheme 3a) involves requires an extended hydrogenation (48 hours, 1.1 atmosphere) in ethyl acetate and 1:1 ethyl acetate : ethanol respectively with a subsequent precipitation of the reduced amine as its hydrochloride salt (**26**, **27**) before the installation of the Boc protecting group and saponification of the crude product in a second reaction (64% overall yield from **20** to **BocPyOH**, 42% overall yield from **24** to **BocImOH)**. In order to shorten the reaction time and improve the reaction yields alternative procedures were sought. Test reactions using zinc metal to reduce **20** and **24** under basic and neutral[189] buffered conditions (Scheme 3b) showed significant polymerisation occurred and were not pursued. A revised hydrogenation procedure for **20** in methanol using 10% palladium on carbon as catalyst with in-situ protection using di-tert-butyl dicarbonate (Boc_2O) was optimised by Dr Glenn Burley[190] and produced faster results under a hydrogen atmosphere (Scheme 3c) compared to the original literature method (4 hours, cf. 48 hours in EtOAc[183]), but further improvements in reaction times were sought to speed up the synthesis.

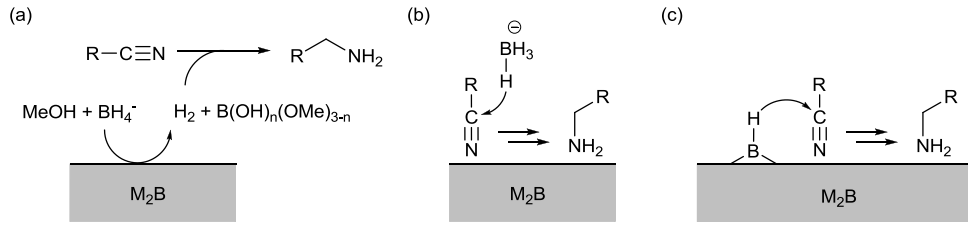
The nickel catalysed reduction of the nitro group using sodium borohydride reported by Kudo *et al.*[191] appeared promising and was found to give the required product in 76% yield (cf. original procedure to HCl salt **26** 82%[183]) when reacted at -20^{o}C with only 3 additional extraction steps for purification following an overnight reaction and with no additional protection step required. This reaction was an extension

Precursor salt (solvent)	% surface area Ni	Surface Ni : B ratio	% surface B B-Ni	% surface B $B-O_2^-$
$Ni(OAc)_2$ (water)	61	1.85	51	49
$NiBr_2$ (water)	77	2.63	56	44
$NiCl_2$ (water)	83	2.13	56	44
$NiCl_2$ (ethanol)	77	1.96	33	67

Table 7 *- Variations in the surface characteristics of nickel boride catalysts due to changes in the preparation method*[195]

of early work aimed at expanding the scope of sodium borohydride[192] and lithium aluminium hydride[193] as reducing agents which found that the addition of transition metal salts allowed the reduction of alkynes, alkyl halides, nitriles and amides which would otherwise be untouched by these reducing agents.[194] In all of these examples, the reaction is catalysed by the in-situ formation of the metal boride or aluminide precipitate which can be isolated and partially characterised, with the most studied system being the reaction of cobalt (II) chloride with sodium borohydride, but even in this system little is known about the mechanism of the reaction or the exact nature of the catalyst.

One of the complications in studying this family of reactions is that the catalyst activity is strongly dependent on the metal salt and solvent used for its preparation. After XPS analysis of a variety of precursor salts and preparation solvents, these differences have been ascribed[196] to the level of oxidation on the surface (Table 7) and contamination of the surface with spectator ions such as chloride. The hydrogenation activity of the nickel catalysts towards alkenes found to be linearly dependent on the level of surface boron bound to nickel.[195] These types of study are of limited use in understanding this reaction as the catalyst is amorphous and the surface area varies significantly depending on the solvent used[197] making generalisations difficult. The



Scheme 4 *- Potential mechanisms for the catalysed reduction of nitriles by sodium borohydride and metal borides: (a) homogeneous reduction by released hydrogen; (b) coordination and attack by solution phase borohydride; (c) coordination and attack by surface bound species*

mechanistic studies that have been done on the cobalt chloride – sodium borohydride system[198, 199] and other papers reporting unexpected reactions would seem to indicate these catalysts are capable of three different mechanisms of action which may be competitive or side reactions for different substrates.

The simplest mode of action for these catalysts is as a catalyst for the generation of hydrogen from sodium borohydride (Scheme 4a), a mechanism generally invoked for the reduction of alkenes and alkynes[194] as the reactions are also catalysed by these compounds under a hydrogen atmosphere and fail when conditions are used where hydrogen generation is suppressed.[199] Studies on the reduction of alkyl halides have also suggested a free radical mechanism with an initial oxidative addition step on the catalyst surface followed by release of the alkyl fragment as a radical which rapidly abstracts a proton from the reaction solvent (Scheme 4b).[199] In this case, the role of the borohydride or aluminium hydride is to regenerate the active catalyst.

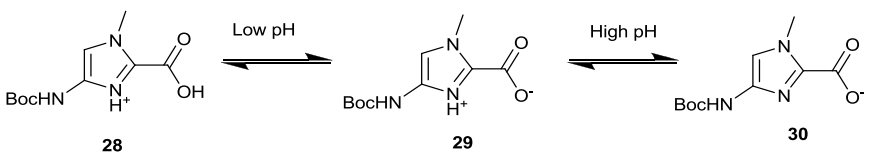
The most general mechanism is that for the reduction of nitriles which involves rapid coordination to the catalyst surface followed by a rate determining attack by the borohydride anion (Scheme 4b, c).[198, 199] The kinetic isotope effect observed when using deuterated borohydride with preformed cobalt boride to reduce benzonitrile shows that a B-H bond is broken in the limiting step but cannot identify whether the reacting species is in solution or surface bound.[199] The study also showed a first order rate dependence on the concentration of borohydride over a wide range of concentrations

which is most consistent the borohydride remaining in solution and no dependence on the concentration of benzonitrile at low catalyst loadings (5-10 mol%). This may not be the only mechanism at work as amino-borane complexes and diborane have both been reported as active reduction catalysts for nitriles[199, 200] and nitro groups[201] respectively which are unreactive in the absence of the catalyst, suggesting the presence of a surface bound reducing species. Most of the mechanistic work has used ethanol as the reaction catalyst due to the slower rate for the decomposition of sodium borohydride,[202] but methanol is a much more common solvent for synthetic reactions and reacts rapidly to produce large quantities of hydrogen that could also be an active reducing agent. Given the heterogeneous nature of the system, the variety of competing mechanisms and reducing agents available and the sensitivity of the catalyst to slight changes in its preparation, only a limited understanding of the mechanism(s) predominant in the nickel chloride - sodium borohydride system can be reached, but it is likely to be a combination of all three mechanisms operating at different rates as the reaction progresses.

The interplay of these possible mechanisms and the initial catalyst formation steps are of particular concern due to the apparent temperature sensitivity of the reaction. The reaction was found to be very exothermic during the addition of the sodium borohydride, making temperature control problematic which may cause limited polymerisation or side reactions which would lower the overall yield. Attempts to apply this reaction to the synthesis of the **BocImOH** precursor **25** using the same conditions resulted in polymerisation but continued investigations by other researchers[203] using similar conditions at -70^{o}C have shown that careful control of temperature allows the product **25** to be recovered in 84% yield (cf. 78% to HCl salt **27** in literature proccedure[183]) and have also noted improved yields and less gas expansion

of the solvent during the synthesis of the **BocPyOH** precursor **21**. As a result, the reduction of **24** was achieved using a revised procedure of bubbling hydrogen for 48 hours through a dry methanol solution of **24** in the presence of Boc_2O and 10% Pd/C catalyst which was found to require addition in two portions for the catalysis to take place and give **25** in 63% yield.

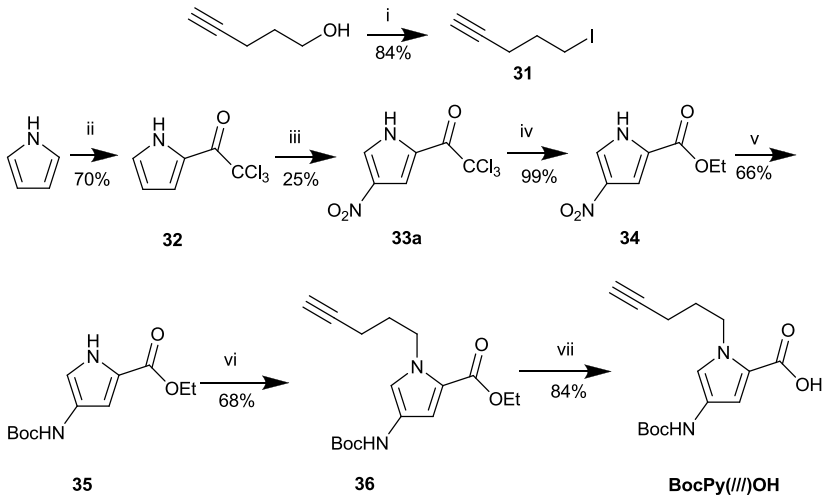
The esters **21** and **25** formed by the reduction and protection of **20** and **24** were initially saponified using a dilute sodium hydroxide solution as per literature protocols,[183] but the use of lithium hydroxide instead was found to improve the efficiency of the initial back-extraction steps during the purification. In order to isolate **BocPyOH** free from any salts, the organic solvent was removed from the crude saponification mixture and any unreacted starting material in the remaining lithium hydroxide solution extracted using ethyl acetate. Careful acidification of the aqueous layer to pH 2-3 produced the desired product which was then extracted into ethyl acetate and the solvent removed under reduced pressure. In order to minimise the presence of water during the coupling reactions and produce a finely divided crystalline product, the crude **BocPyOH** was dissolved, precipitated from petrol and finally freeze dried. Isolating the **BocImOH** product was far more problematic due to its poor phase distribution during the initial extraction and the formation of a glue-like precipitate upon acidification beyond pH 7, which may be due to the formation of insoluble zwitterions (Scheme 5). Isolation of this precipitate using centrifugation followed by freeze drying allowed the product to be recovered, but the difficulty of handling this precipitate resulted in very high mechanical losses.



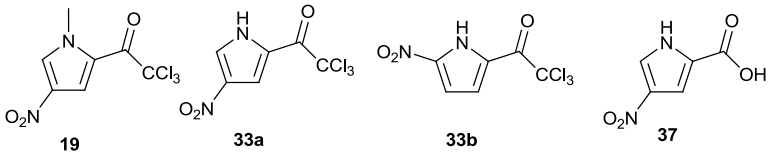
Scheme 5 *– Proposed Zwitterions formed by the* ***BocImOH*** *building block*

2.6.2 Synthesis of Modified Polyamide Building Blocks

In addition to the standard **BocPyOH** and **BocImOH** building blocks, the modified building block **BocPy(///)OH** was also required. Modified **Py** building blocks have been reported previously with amino groups[162, 163] and provided the initial steps in Scheme 6. The synthesis closely resembles that of both **BocPyOH** and **BocImOH** with the synthesis beginning with the formation of the trichloroacyl derivative through the slow addition of pyrrole to trichloroacetyl chloride, which after treatment with activated charcoal to remove polymerised pyrrole and recrystalisation from hexane gave intermediate **32** in 70% yield. In common with in the original synthesis of the analogous *N*-methyl pyrrole intermediate **19,**[183] the trichloroacetyl group has a strong directing effect in the subsequent nitration with nitric acid at -30°C, but the ratio of the favoured **33a** to impurity **33b** in the crude product was found to improve from to 3:1 to 32:1 with careful control of the temperature of the reaction. The two product were separated by recrystallisation to give the desired isomer **33a** in 25% overall yield. The low yield is in part due to the formation of **33b** but as the crude yield from the reaction

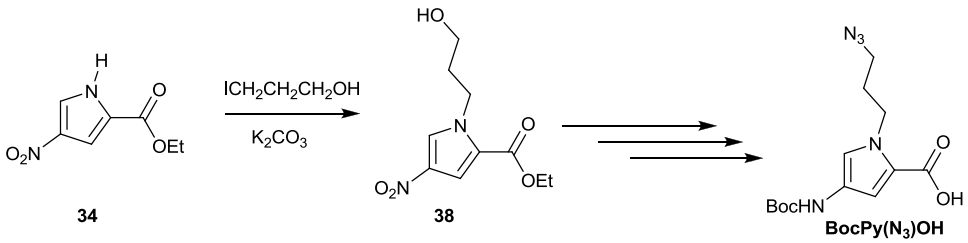


Scheme 6 *- Synthesis of the Boc protected* ***Py(///)*** *building block: (i) PPh_3, I_2, imidazole,; (ii) CCl_3COCl, Et_2O (iii) HNO_3, Ac_2O (iv) NaOEt, EtOH (v) H_2, 10% Pd/C, MeOH Boc_2O, (vi)* ***31****, Cs_2CO_3 (vii) LiOH, H_2O, EtOH*



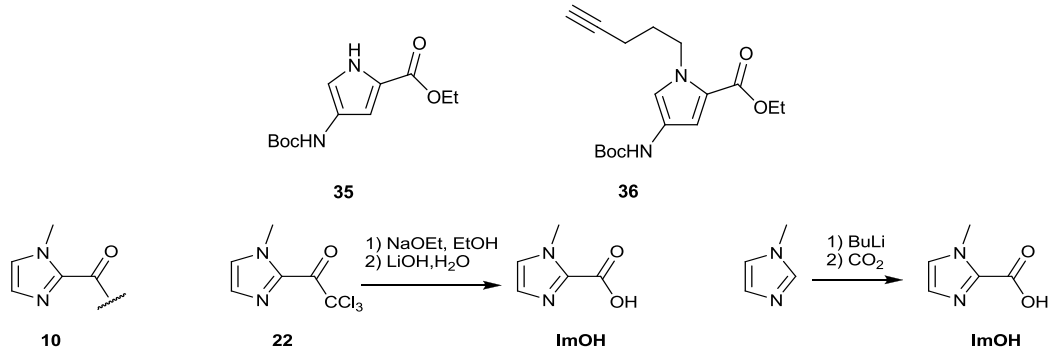
is just 50% the precipitation of the product from the reaction mixture using water may be limiting, despite the low solubility expected of both isomers in water under acidic conditions and even the hydrolysed product **37**, which may be formed as the water is added, would also be expected to be highly insoluble at low pH. This explanation is given credence by the much higher yields (70% overall) that have been reported where the acetic anhydride and nitric acid were removed under reduced pressure to yield the crude product and the product purified by column chromatography.[204] Although a higher yield is desirable, the considerable length of time that would be required to remove the acetic anhydride and the corrosive nitric acid from the reaction mixture makes the precipitation procedure more attractive, especially on a large scale.

After the isolation of **33a**, treatment with sodium ethoxide rapidly yielded the ethyl ester **34** in near quantitative yield (99%) which readily deprotonates under basic conditions to produce a bright yellow solution. The ease with which this can be achieved means that the alkylation provides an easy pathway to a range of potential derivatives and an azide modified **Py** building block has been synthesised via the hydroxyl intermediate **38** using this route.[205] However the selective reduction of the nitro group and subsequent protection step are not necessarily compatible with all the functionality that might be introduced, thus in order to develop a general route to install any type of modification the Boc protected alkylation precursor **35** was a desirable



intermediate. The nickel catalysed reduction of the nitro group of **34** using the same protocol as for the **Py** precursor **20** failed, but reduction under a hydrogen atmosphere using 10% palladium on carbon catalyst over 4 hours yielded the desired intermediate **35** in a satisfactory 66% yield without the need for chromatography.

The lack of electron withdrawing effects caused by the nitro group compared to **34** made the alkylation of **35** much more problematic. Initial reactions using potassium carbonate required extended period of heating (130°C, 72 hours, 46% yield) and scaling up of the test reactions was detrimental to the purity and yield of the reaction which is not uncommon for heterogeneous reactions of this type. Several different base (K_2CO_3, Cs_2CO_3, KOH, NaOEt) and solvent (THF, DMF, EtOH, MeCN) combinations were trialled before cesium carbonate and dry DMF were selected. The application of microwave heating and the use of polyethylene glycol as phase transfer catalyst significantly reduced the reaction temperature and time to 85°C for 90 minutes in addition to improving the yield to 68%. Although this yield remains low, the cleanliness of the reaction allows unreacted starting to material to be recovered and the method allows the use of several small scale reactions to be used to scale up. No attempt was made to preserve dry conditions when adding the reagents before the reactions as the presence of small amounts of water has been shown to be of significant benefit for these types of solid-liquid heterogeneous reactions.[206]



Scheme 7** - Synthesis of the **Im** stopper precursor **ImOH

The final step to produce **BocPy(///)OH** after the alkylation is saponification of **36** using lithium hydroxide to give the free acid in 84% yield. As with the unmodified **BocPyOH** and **BocImOH** building blocks, the product was freeze dried to remove any traces of water that might interfere with the coupling during the solid phase synthesis. In addition to these building blocks, the **Im** stopper precursor **ImOH** was also required. Initial work followed the route reported by Dervan *et al.* which involves the esterification of **BocImOH** intermediate **22** which was then saponified to give the desired compound.[111] As with the **BocImOH** building block, this saponification and subsequent acidification were problematic[190] and in addition signs of decomposition by loss of CO_2 were observed in the NMR of the product. As a result, a more direct route to the product was used based on the addition of CO_2 gas after the formation of the imidazole anion using butyl lithium which gave the product in good yield (89%) and removed the need to heat the product under basic conditions.

These building blocks with the addition of the commercially available **Boc-γ-OH**, **Boc-β-OH** and **39** building blocks and the cleavage amines **Dp** and **40** (Figure 32) allow synthesis of a diverse range of polyamides with capability to produce tail, turn or internal modifications. Two different approaches have been reported for polyamide synthesis – solution phase synthesis which is ideal for gram scale production[188] and solid phase synthesis[183, 185-187] which is better suited to the creation of a library of compounds on a small scale.

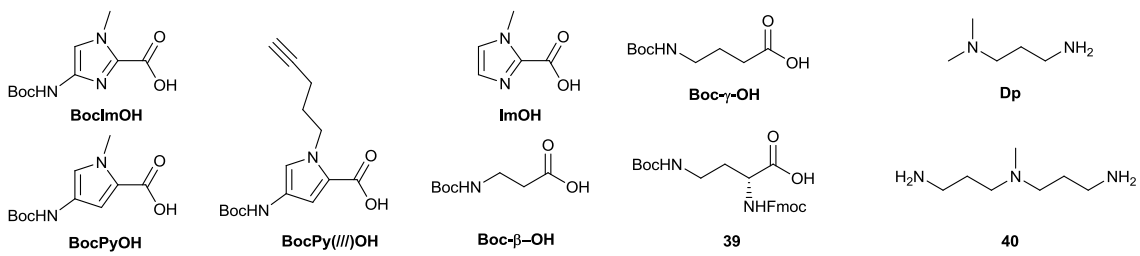
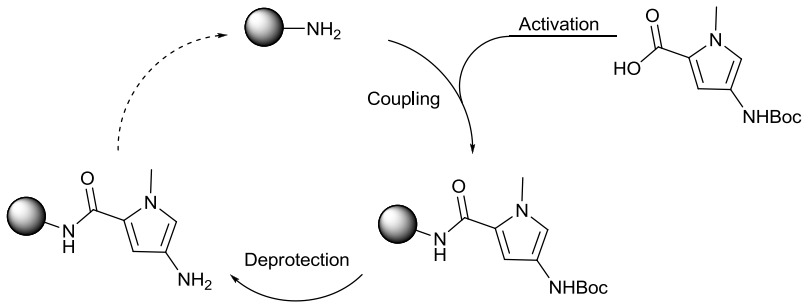


Figure 32 - Assorted Polyamide Building Blocks for solution and solid phase synthesis

2.7 Solid phase Synthesis Protocols of Polyamides

Early polyamides were synthesised in a similar fashion to their natural product precursors using laborious solution phase chemistry[112] with numerous purification steps which severely limited the speed at which new molecules could be produced. A key advance in the use of polyamides was the development of automated and manual solid phase synthesis protocols for their synthesis using both Boc[183, 185-187] and Fmoc[184] protection chemistry as it allows research quantities of material to be produced very quickly. The techniques of solid phase synthesis have been developing since the principle was first introduced over 40 years ago[207] and it has been very successful in the synthesis of peptide sequences up to 30 amino acids on both a small[208] and industrial scale.[209] Solid phase synthesis addressed the problems of purifying the products and of driving the reaction to completion using excess reagents, whilst also providing a means to create a library of compounds during cleavage from the resin. However, several issues remain with this approach to synthesis.

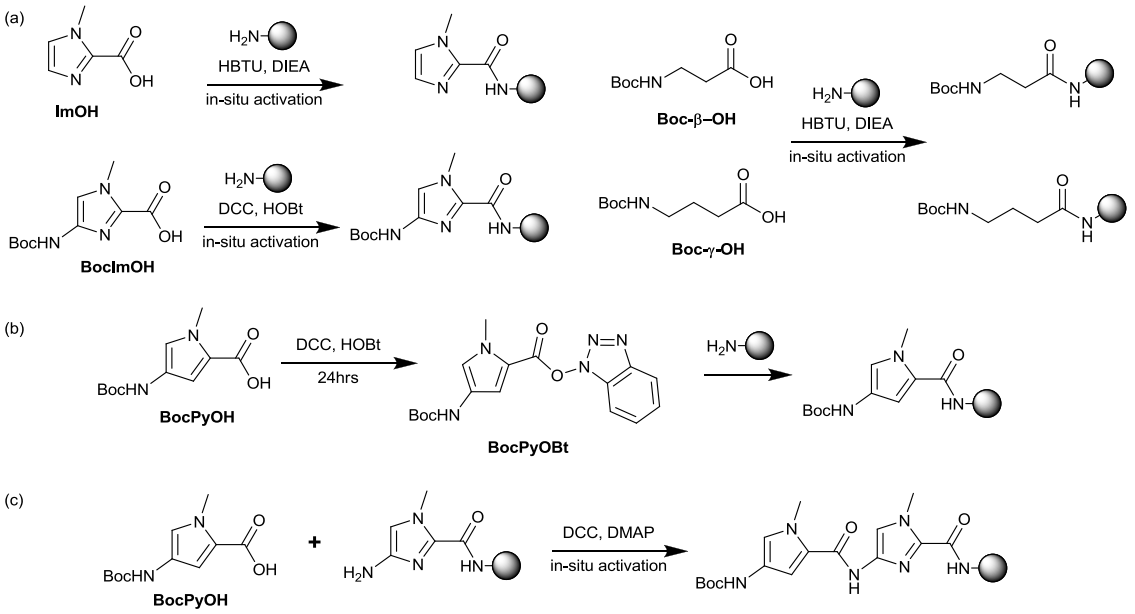
As the product remains attached to the resin throughout the synthesis, this greatly simplifies purification after each step but severely limits the methods available to monitor the progress of the reaction. Tests on resin samples using reagents such as ninhydrin[210] or chloranil[211] can provide some insight into the progress of the reaction, but for a quantitative measure sample cleavage followed by HPLC analysis is often required. This process is time consuming and can obscure small drops in yield which are particularly important due to the sensitivity of the overall yield to each stepwise reaction. The linear nature of solid phase synthesis means that in the 18 step synthesis of a typical polyamide targeting 4 base pairs on PAM resin using standard protocols[185] a drop in the stepwise yield from 99.9% to 99.4% would result in a reduction of 8.5% in the overall yield, whilst a 1% drop in the stepwise yield from 99.9% to 98.9% reduces



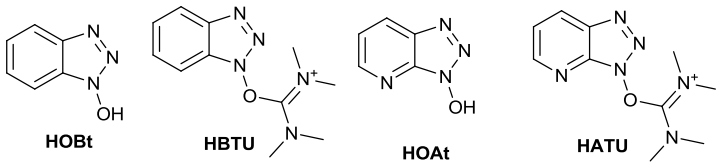
Scheme 8 *- Solid phase synthesis cycle for polyamides shown using the* ***BocPyOH*** *building block*

the yield by 16%. Despite these problems, solid phase synthesis remains the method of choice for producing small quantities of polyamides and protocols using different resins[186] have been developed to allow a range of groups to be introduced in the tail position and enable a wider range of attachment chemistry to be used to attach a 'payload'.

The basic cycle of solid phase synthesis is simpler than that of DNA (Figure 12), containing only activation and deprotection stages (Scheme 8) and does not routine include a capping step (cf. DNA). The original Boc chemistry solid phase protocol developed by Dervan *et al.*[183] used HOBt esters to activate the building blocks before

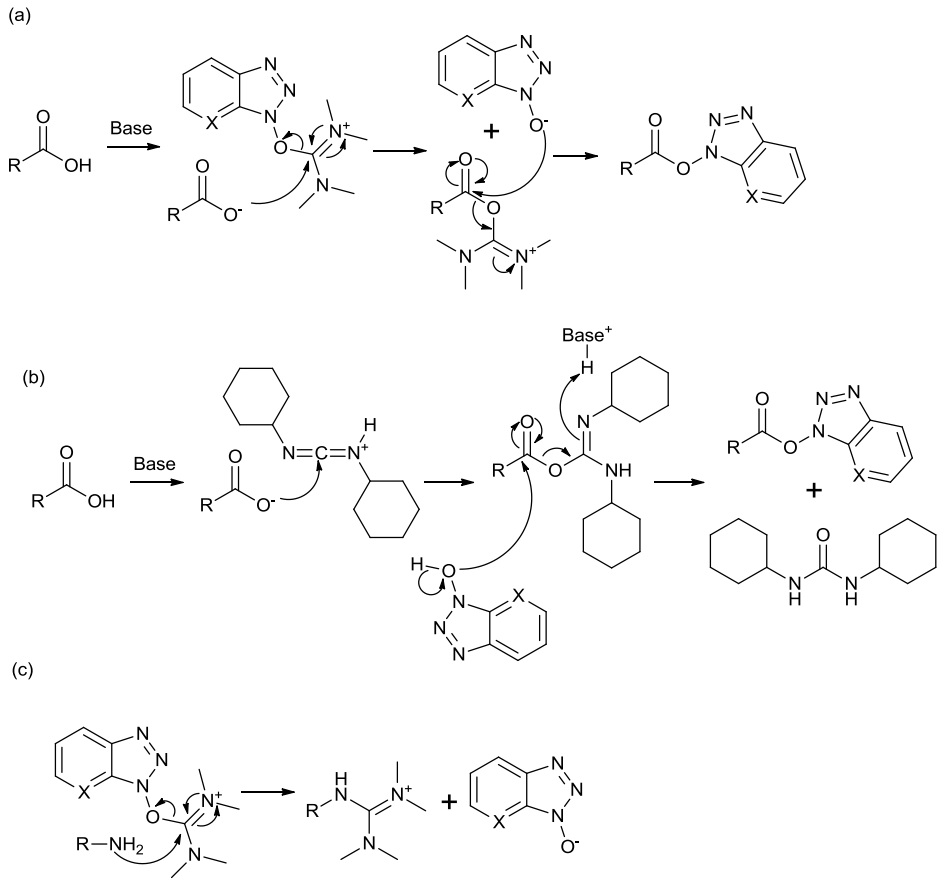


Scheme 9 *- Activation and solid phase synthesis of the* ***ImOH***, ***BocImOH***, ***Boc-β-OH***, ***Boc-γ-OH*** *and* ***BocPyOH*** *building blocks*



their introduction to the solid phase resin and Boc protection chemistry. Under this protocol, the stopper **ImOH** and alkyl building blocks **Boc-β-OH** and **Boc-γ-OH** are activated in-situ using HBTU (Scheme 9a) for several minutes and the mixture added directly to the resin. The initial activation of the **BocImOH** building block using both **HOBt** and **HBTU** was reported to be very sluggish which can be attributed to electron withdrawing effect of the imidazole ring as both activations require nucleophilic attack as the initial step (Scheme 10a,b). In order to prevent the reacting amine forming guanidinium adducts (Scheme 10c), all of the HBTU must be consumed before the activated acid is added to the resin[212, 213] which is not easy to achieve given the reduced reactivity of the imidazole acid. As a result the **BocImOH** building block was activated using DCC and HOBt instead.

The HOBt activated ester **BocPyOBt** of the **BocPyOH** building block was used in this report in preference to in-situ activation, presumably for convenience, and was formed using DCC (dicyclohexylcarbodiimide) and HOBt over 24 hours and isolated (Scheme 9b) before addition to the resin. However, this approach was found to be unsatisfactory for couplings to **Im** bases due to the reduced reactivity of the deprotected amine and although a revised activation procedure to form an anhydride was proposed, the formation of **BocPyImOH** dimers using solution phase chemistry was preferred to avoid this problem. In contrast, once the **BocImOH** activated ester is formed, it is more reactive due to the electron withdrawing effect of the ring which counteracts the low reactivity of the imidazole amine, and **Im** to **Im** couplings were reported to proceed smoothly without changes in the activation procedure.



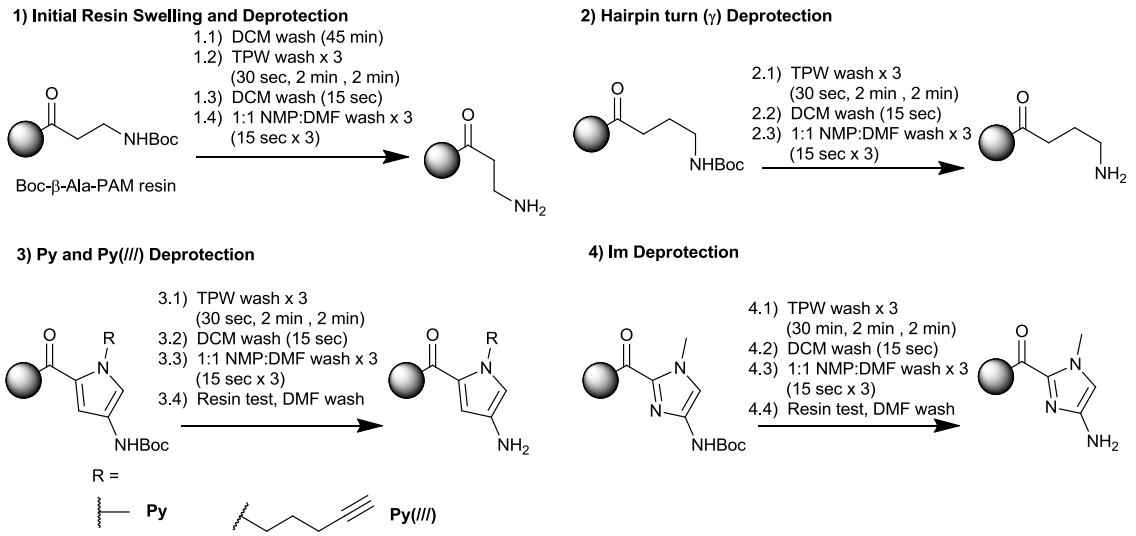
Scheme 10 *- Mechanisms of carboxylic acid activation with a selection of coupling agents: (a) Formation of activated esters with HATU and HBTU; (b) formation of activated esters using DCC and HOAt or HOBt; (c) formation of guanidinium side products due to unreacted coupling reagent*

A subsequent paper by Chamberlin *et al.*[185] introduced several modifications to the procedure to reduce coupling times and remove the need to pre-form activated esters. The use of HBTU and HOBt was replaced by in-situ activation with HATU for simple **Py** to **Py** couplings and an extended pre-activation with DCC and HOAt for the alkyl and imidazole building blocks. HATU and HOAt have been successfully applied to difficult peptide couplings[214-216] and it was hoped to improve the efficiency of the problem imidazole couplings and reduce coupling times for the other building blocks. The reduced level of racemisation compared to HOBt and other coupling agents[214] is also relevant when using chiral building blocks for the 'turn' segment of hairpin polyamides where the two enantiomers have been shown to have dramatically different binding properties.[149]

Another modification introduced by Chamberlin *et al.*[185] was to the Boc deprotection procedure which originally used a 0.5 M solution of thiophenol diluted with 80% TFA in DCM.[183] The thiophenol acts as a scavenger for the t-butyl cations formed during the deprotection that would otherwise react with the growing polyamide chain, but the pungent odour of this mixture makes it particularly undesirable for regular use and several alternatives have been described.[217, 218] The improved deprotection mixture was reported as a solution of phenol (5% w/v) in water (2.5% v/v) and trifluoroacetic acid (92.5% v/v) and is known as TPW. Much faster deprotection was observed using this reagent, significantly reducing the time cycle required for the polyamide synthesis. Extended coupling times are still required for the **Im** building block, but these are much reduced from the original procedure.

2.7.1 Improved Solid phase Synthesis of Polyamides

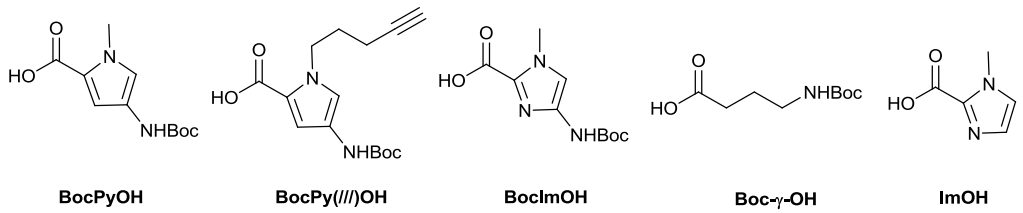
The improved procedures developed by Chamberlin *et al.*[185] were used as the basis for all solid phase synthesis on the Boc-β-Ala-PAM resin with a number of minor modifications based on practical experience. In place of 1:1 NMP:DMF as solvent for the coupling reactions, DMF was used in line with the original solid phase protocols[183] and used as wash solution instead. The duration of the first deprotection wash using TPW (Scheme 11, steps 1.2, 2.1 and 3.1) was reduced slightly for the **Py**, **Py(///)** and hairpin turn **γ** after 30 sec was found to give satisfactory results (cf. 1 min[185]). In contrast, the **Im** deprotection (Scheme 11, step 4.1) was extended (30 min, 2 min, 2 min; cf. 1 min, 2 min, 2min, 10 min[185]) to ensure complete removal of the Boc protecting group. In addition, after an exothermic reaction between residual trifluroacetic acid and trace amine impurities in the DMF:NMP wash solution was observed through the vessel feeling hot upon the addition of the wash solution

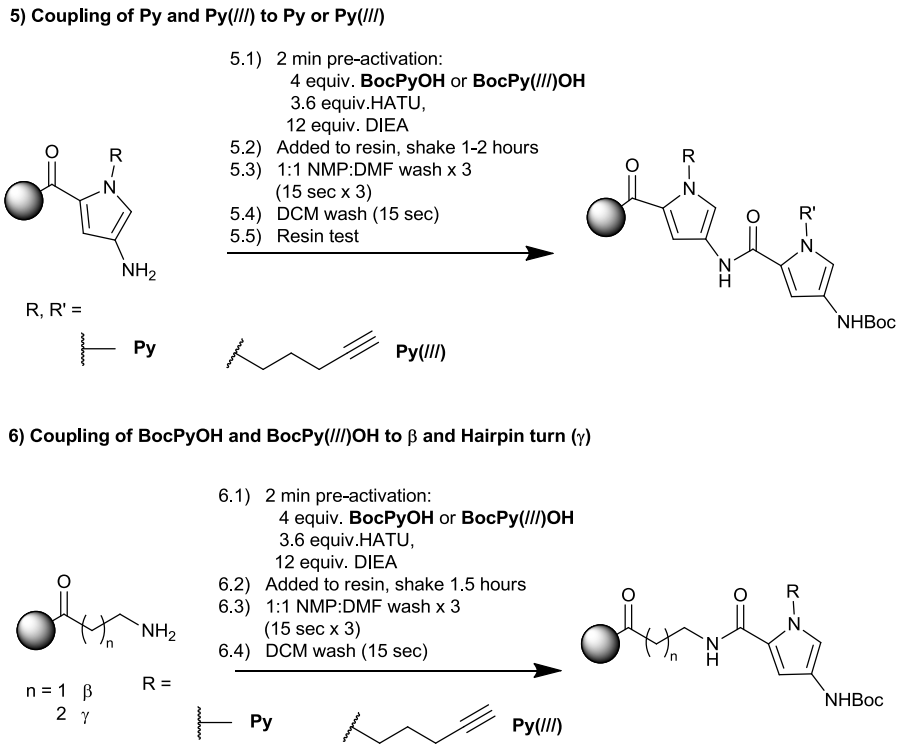


Scheme 11 – *Solid phase protocols for the deprotection of resin bound polyamide building blocks*

immediately after the TPW deprotection, an intermediate DCM wash was added to all the deprotection steps (Scheme 11, steps 1.3, 2.2, 3.2 and 4.2) which appeared to be successful in removing the residual acid. This wash has the added benefit of swelling the resin (cf. Scheme 11, step 1.1), aiding the removal of the TPW and improving the flow of reagents to reactive sites within the pores of the resin. Monitoring of the deprotection was achieved by testing resin samples with chloranil[211] which was found to be a reliable indicator of the free amine of the aromatic building blocks (**Py, Py(///), Im**) but very poor for the alkyl amines exposed after the deprotection of the Boc-β-Ala-PAM resin and the hairpin turn **γ** (Scheme 11, protocols 1 and 2). As a result, resin tests were only used for the **Py**, **Py(///)** and **Im** deprotections (Scheme 11, steps 3.4 and 4.4) with any resin disturbed by the sampling washed to the base of the vessel using DMF.

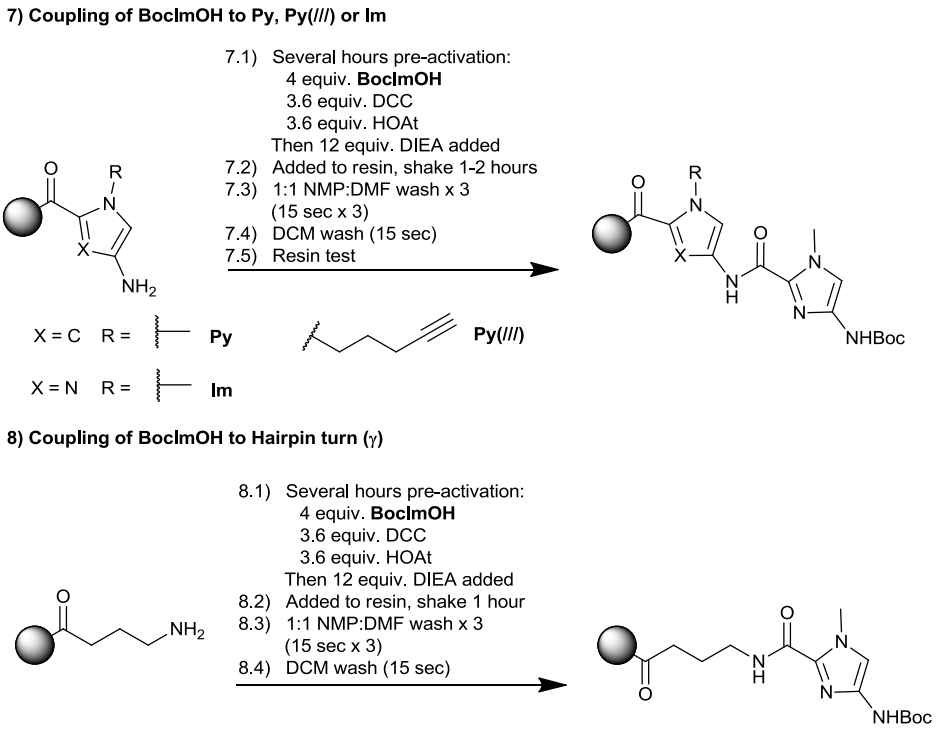
A small number of modifications were also found to be necessary for the coupling reactions reported by Chamberlin *et al.*[185] The activation steps for the





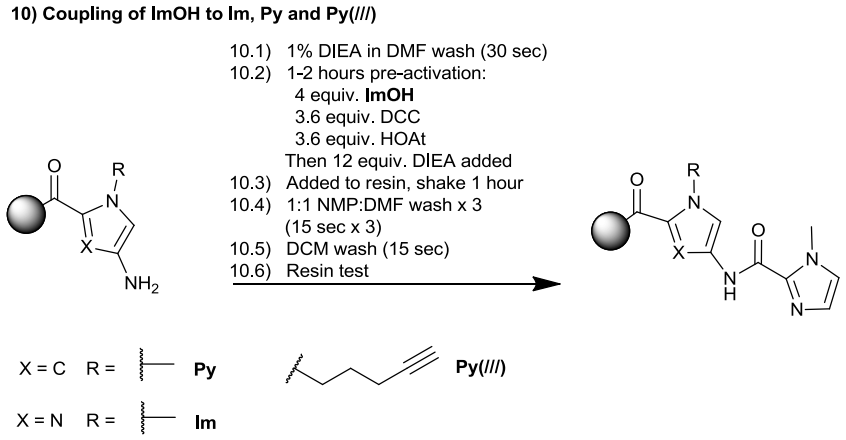
Scheme 12 *- Solid phase protocols for the coupling of the* ***BocPyOH*** *and* ***BocPy(///)OH*** *building blocks*

BocPyOH, **BocPy(///)OH** and **BocImOH** couplings (Scheme 12 and Scheme 13, protocols 5-8) were unaltered, but extended coupling times were found to be required for consistent results. During the first 5 coupling cycles (up to the addition of the hairpin turn **γ** for 8 ring polyamides) 1 hour was found to be sufficient, but for subsequent couplings 2 hours was required for completion (cf. 15-20 min[185]). The time required for the activation of the **BocImOH** building block (Scheme 13, steps 7.1 and 8.1) was found to be variable and was judged from the formation of the DCU precipitate which was separated by centrifugation before the activated solution was added to the resin. As with the deprotection protocols, a DCM wash was also added (Scheme 12 and Scheme 13, steps 5.4, 6.4, 7.4, 8.4 and 9.4) as the final step to remove any traces of residual amines from the 1:1 DMF:NMP washes prior to the subsequent TPW deprotections. Similarly, resin tests also were omitted after couplings to the **β** and **γ** alkyl building blocks (Scheme 11 and Scheme 12, protocols 6 and 8) due to chloranils poor performance indicating the presence of the uncoupled free amine.



Scheme 13 *- Solid phase synthesis protocols for the coupling of the* ***BocImOH*** *building block*

In contrast to the slow activation of the **BocImOH** building block, the **ImOH** stopper building block activates readily in 1-2 hours (Scheme 14, step 10.1), but the coupling step was found to be more problematic. Initially, it was found that frequently in place of the **Im** stopper **10**, trifluoroacetic had been coupled (observed by MALDI and ESI mass spectrometry) despite the presence of excess base, effectively capping the polyamide. This recurring problem was partly solved by the direct synthesis of high purity **ImOH** from imidazole using butyl lithium and carbon dioxide (Scheme 6) which



Scheme 14 *- Solid phase synthesis protocols for the coupling of the* ***ImOH*** *stopper* ***10***

meant that all of the coupling agent was consumed, but the formation of this by-product at the final step was normally fatal for the entire synthesis as its similarity to the final product prevented clean separation from the product peak using HPLC. As a result, an additional washing step with a solution of DIEA was introduced before this coupling to minimise this side reaction

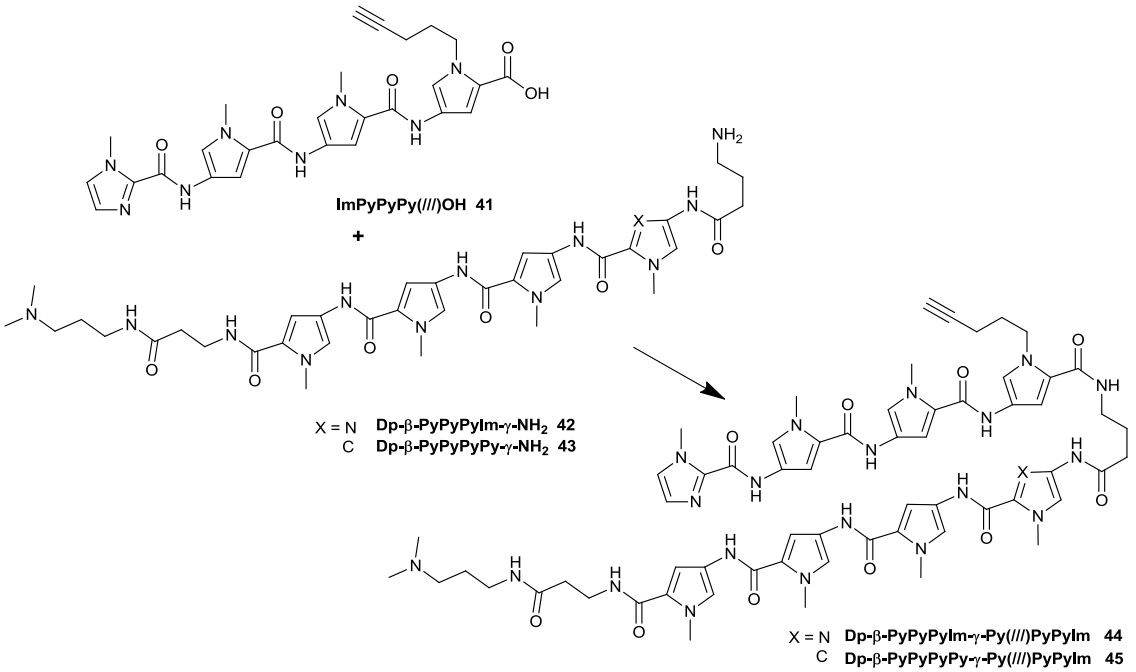
These modified protocols were used to attempt synthesis of full length polyamides incorporating the **Py(///)** modification and unmodified polyamide containing terminal amines suitable for off-resin tail modification. The combination of reaction monitoring using chloranil and selective HPLC analysis allowed effective monitoring of the reaction during the synthesis cycle and as a result, truncated sequences were generally not observed by HPLC or mass spectrometry except for trifluoroacetic acid capped products.

The HPLC purification of the polyamides after cleavage from the resin using the desired cleavage amine was optimised for more difficult separations but routinely used an 18 minute ramp from 10 to 90% acetonitrile in water with 0.1% TFA as buffer. To allow discrimination of polyamides from other UV active impurities, detection at both 254 nm and 310 nm was used, allowing the distinctive polyamide absorption around 310 nm to be observed. After a suitable method was identified using analytical HPLC, the remainder of the resin was cleaved using a large excess of primary amine and the product precipitated using 10 equivalent volumes of diethyl ether to remove traces of the cleavage amine which have been reported to contaminate the product after purification.[185, 187] This protocol yielded the crude polyamide as a fluffy off-white precipitate which was dissolved in the starting eluent, with up to 20% v/v DMSO being added to aid solubility, and purified by semi-preperative HPLC. After collection of the

fractions, the solvent was removed by a combination of reduced pressure and freeze-drying to yield the final product as the TFA salt.

Analysis of the final yields of the overall synthesis is complicated by these trifluoroacetic acid salts in the final product, but apparent yields of 27-46% were achieved which comparable favourably to polyamides synthesised using similar protocols.[187, 219] This corresponds to a stepwise yield of around 90% for a full length polyamide, significantly below that which might be suggested by studies of the performance of coupling agents over shorter reaction times (95% yield within 20 minutes).[187] This anomaly can be explained by losses during the polyamide precipitation procedure and HPLC purification, but the presence of the trifluoroacetic acid salts in the final product makes thorough analysis of the yields difficult. There have been reports where the distinctive absorption around 310 nm has been used to determine polyamide concentration,[140] but it was observed that the strength of this absorption was very solvent dependent and overlapped with the expected absorption region of the coumarin dye and so was not used at this stage of the project. The relatively small amounts of the product being produced in solid phase synthesis also mean that relatively minor changes can have a major impact and yields as low as 0.1% have been reported.[220]

Solid phase synthesis is very good for producing small quantities of polyamides, but has significant limitations. Its linear nature does not lend itself to the production of multiple compounds and any drop in yield is exaggerated by the repeated cycles. In addition to full polyamides, a number of 'half' polyamides such as **42** and **43** consisting of four aromatic building blocks coupled with the uncharged hairpin loop **γ** were also produced to be reacted with an acid functionalised 'half' such as **41** to form a complete polyamide (Scheme 15). This would allow a convergent synthesis using a 'mix and



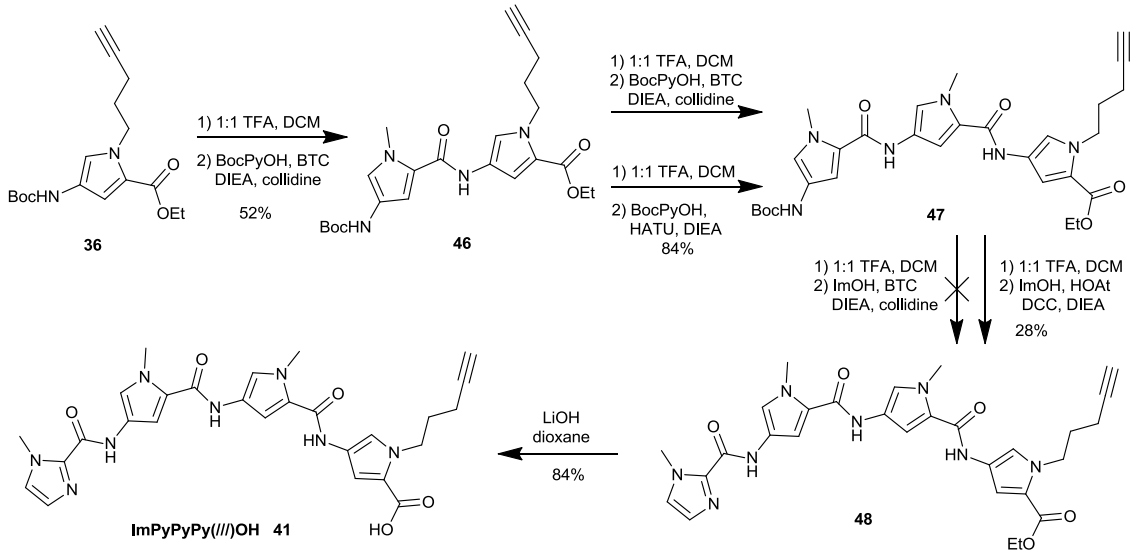
Scheme 15 - *Convergent synthesis using polyamide halves*

match' system of polyamide 'halves' which would be combined to produce the desired polyamide and also avoids the more difficult second half of the polyamide synthesis. The solid phase synthesis of the acid terminated 'half' polyamides (e.g. **41**) was attempted on the oxime resin which has been reported previously[186] as a route to several different 'head' modifications using the improved activation protocols. A modified deprotection step was required to account for the reduced acid stability of the resin and the TPW deprotection mixture was diluted to 25% v/v in DCM to prevent premature cleavage from the resin, but HPLC and mass spectrometry analysis of the couplings identified several truncated species consistent with incomplete removal of the Boc protecting group. As a result, no further synthesis was attempted using the oxime resin and instead the corresponding polyamide acid halves were synthesised using solution phase synthesis.

2.8 Solution phase synthesis

The original solution phase synthesis of polyamides[112] was primarily limited by the time consuming purification steps, making this method impractical for full length polyamides. The need for a method to produce gram scale quantities of polyamides for medicinal applications quickly and cost effectively drove the development of a precipitation based protocol which removed the need for additional purification steps.[188]

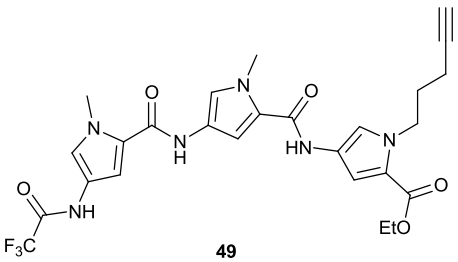
The greatly improved monitoring capabilities using methods such as NMR and TLC during solution phase synthesis compared to resin based reactions allows easier optimisation of the reactions, but the stepwise yield is lower without the large excess of reagents to drive the reaction to completion. The polyamide 'half' **ImPyPyPy(///)OH 41** was initially synthesised using solution phase protocols with BTC based activation and purification using extractions (Scheme 16). BTC, also known as triphosgene,[221, 222] has been shown to be a highly efficient coupling agent for polyamide synthesis[187] but was thought to be too activating to use during solid phase synthesis except for the problematic **BocPyOH** coupling to imidazole residues (Scheme 9c) due to the formation of the acid chloride. It was selected in order to maximise the yield of the



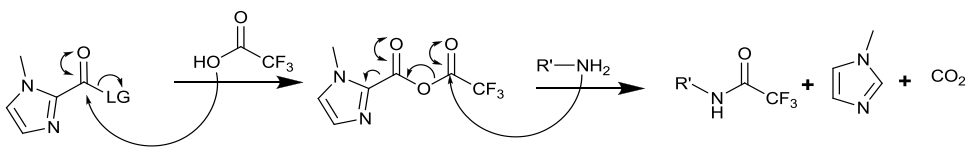
Scheme 16 *– Solution phase synthesis of alkyne polyamide 'half'* ***41***

reaction as it was thought that the use of a single equivalent of the activated acid (cf. four equivalents in solid phase synthesis) and easier monitoring of the reaction would allow this problem to be overcome which was found to be the case in the majority of coupling reactions. Unfortunately, the coupled product was invariably contaminated with collidine and was used crude for subsequent reactions to avoid the need for column chromatography after each step.

Attempts to deprotect **36** using hydrochloric acid in diethyl ether as described in the reported solution phase synthesis[188] failed with no free amine visible by mass spectrometry or NMR, and as a result a 1:1 mixture of TFA in laboratory grade DCM was used as an alternative. This grade of DCM was selected due to the presence of trace amounts of water which would help 'mop-up' the *tert*-butyl cations produced in the Boc deprotection (cf. 92.5% TFA, 2.5% water and 5% phenol in the TPW deprotection mixture used during solid phase synthesis). The synthesis continued using the crude reaction products up to the trimer **47** with no significant by-products being formed. When the addition of the **ImOH** stopper was attempted using BTC based protocols, only the TFA coupled product **49** was produced from the trace amounts of acid remaining after the deprotection, despite the presence of large amounts of excess base (3 equiv. collidine, 6 equiv. DIEA) and the separate pre-activation of **ImOH**



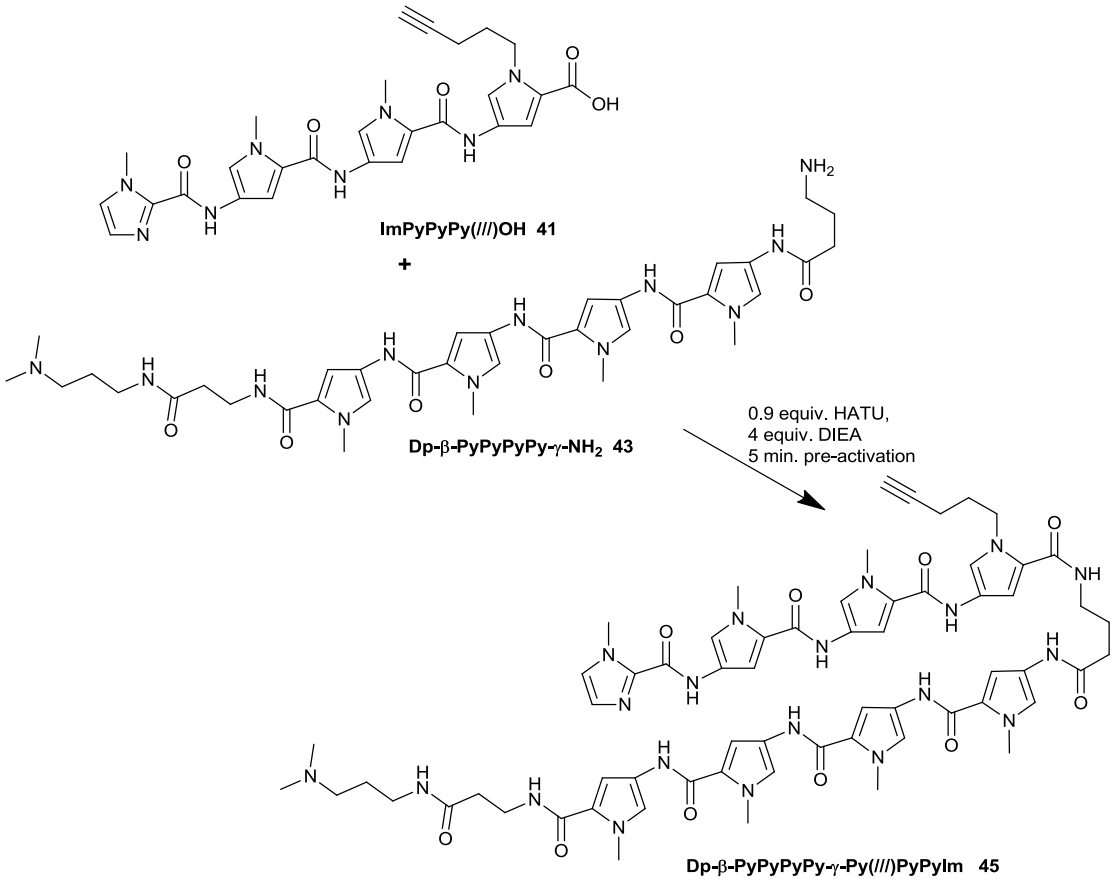
49



Scheme 17 *- Proposed mechanism for the formation of a mixed anhydride from activated* ***ImOH*** *(LG: Leaving Group – Cl or OAt) and subsequent decomposition of* ***ImOH***

building block. Why this should have occurred only for this reaction rather than for any of the previous solution phase couplings using the same protocol is not clear. Some possibilities are the presence of unreacted BTC activating the TFA or the formation of a mixed anhydride which went on to react with the deprotected amine of **47** (Scheme 17), but both of these proposals rely on the presence of free TFA which is unrealistic in the presence of the large excess of base.

The solution phase synthesis was repeated using HATU activation for the second **BocPyOH** coupling reaction to produce **41** (Scheme 16) whilst avoiding contamination with collidine. When HOAt and DCC were used to activate the **ImOH** stopper a mixture of the desired tetramer **48** and impurity **49** were produced and recovered in 28% and 9% yield respectively after purification using normal phase chromatography (EtOAc gradient in DCM, 5% v/v NEt_3). The rather low yield for this reaction can only



Scheme 18 *- 'Half and half' polyamide coupling using HATU activation*

be attributed to losses during the purification as after initial trials the crude reaction mixture was purified directly without extractions or other purification steps and none of the deprotected intermediate was observed in the crude mixture by mass spectrometry. Finally, the ethyl ester **48** was saponified using lithium hydroxide in dioxane to give the polyamide acid 'half' **41** suitable for coupling to the corresponding amine 'halves' such as **43** in 84% yield. Trial reactions monitored using HPLC and mass spectrometry between the **BocPyOH** building block and polyamide 'half' **43** (produced using solid phase synthesis) were used to identify suitable coupling conditions (Scheme 18). This 'half and half' approach was envisioned to be an ideal route to producing a library of polyamides by combining different fragments to vary the binding site in addition to the type and positioning of the functionality (Figure 33), allowing a diverse range of

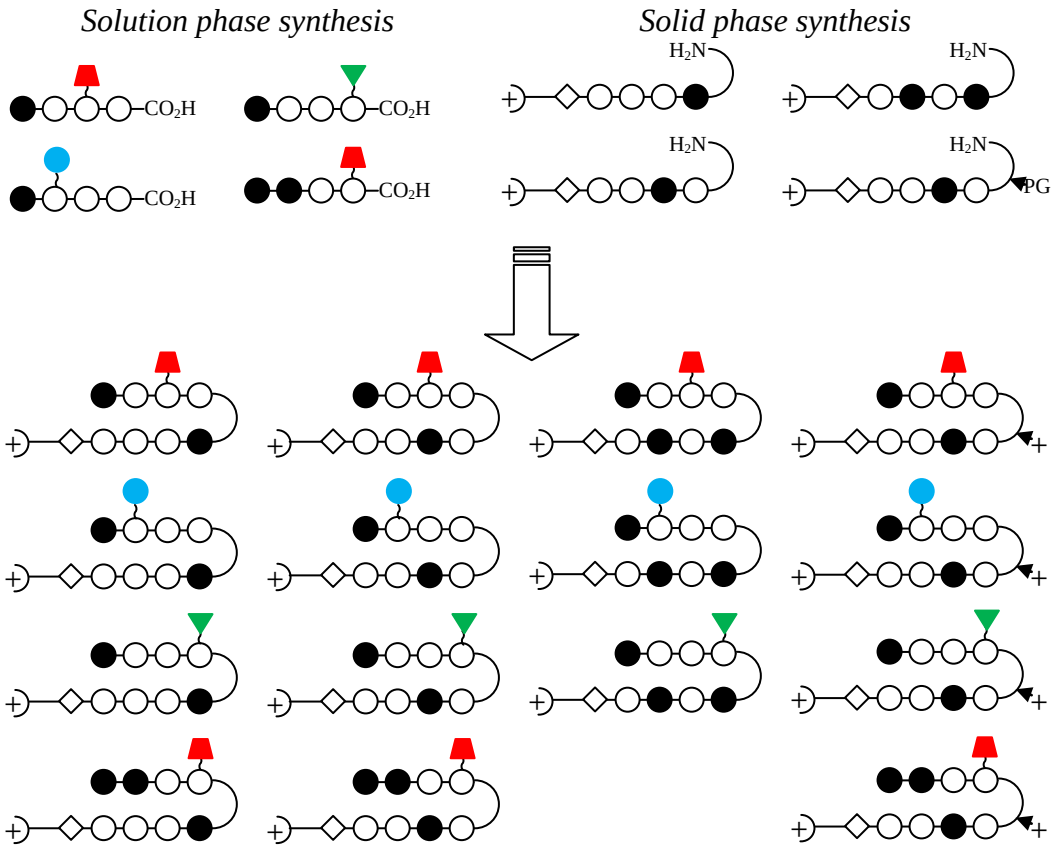


Figure 33 *- Combination of polyamide 'halves' to produce a library of functionality, positioning and binding site; PG – protecting group*

polyamide to be produced whilst avoiding the more difficult later stages of the solid phase synthesis.

2.8.1 Combination Synthesis of Polyamides

The 'half and half' approach demonstrates how solution and solid phase can be used to complement each other. The improved monitoring, greater flexibility and closer control of the reaction conditions associated with solution phase synthesis make it ideal for producing polyamide 'halves' containing a wide variety of functionality and allow reaction conditions to be tailored to protect more sensitive functionality. Solid phase synthesis excels at producing a wide range of the corresponding 'halves' with high yield and purity which can be used to alter the polyamide binding site, but can be limited by its linear nature and relatively small scale. Solution phase synthesis has been shown to be a viable method for the synthesis of gram scale quantities of full length polyamides[188] but is ill-suited for producing the several milligram scale quantities often required for research. Between these two extremes, a combination of the two methods allows the best of both methods to be used and the worst of the drawbacks avoided.

2.9 Conclusions

Polyamides are a versatile, readily modified, readily programmed and relatively accessible method of targeting a specific site on DNA, but majority of reports so far have concentrated on their well established biological activity.[157, 161, 163, 173, 204] With the development of improved synthetic methods[185, 187, 188] to aid development projects and the early indications of their broader applications,[169, 171, 219, 223] it is clear that polyamides have much greater potential beyond the cell that is yet to be explored. The development of a general synthetic route to produce building blocks for creating

internally modified polyamides fulfils the original aim of allowing a range of modified polyamides to be produced and the additional flexibility in the positioning of the functionality represents a genuine advantage over the more popular tail modification of polyamides. Improvements to the synthesis of the polyamide building blocks and solid phase protocols have also made access to this family of molecules simpler and easier. With the synthesis of alkyne functionalised polyamides achieved, methods for the attachment of the 'payload' using this functionality can now be investigated.

2.10 Experimental

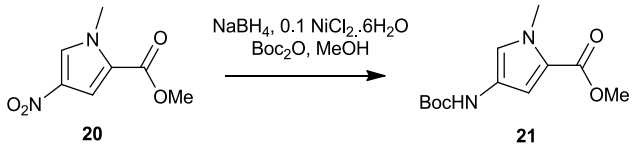
Some experimental procedures and additional characterisation data were added between the first submission in July 2011 and final submission. These additions are collected in Appendix 1.

All chemicals and solvents were purchased from Acros, Alfa Aesar, Fisher, Nova Biochem or Sigma Aldrich unless otherwise stated and were used as received. Boc-β-Ala-PAM resin was purchased from Peptides International. Analytical TLC was carried out using Polygram 0.20mm silica gel plates with detection by UV at 254 nm and visualisation using PMA or ninhydrin. Preparative column chromatography was performed using Merk Kieselgel $60_{(230\text{-}400\ mesh)}$ silica gel. ^{1}H, ^{19}F and ^{13}C spectra were recorded on a *Bruker DPX 300, Bruker DRX 400* or *Bruker AV 500* spectrometer as stated. Chemical shifts reported are in parts per million (ppm, δ). *J*-coupling constants are given in Hertz (Hz) and signal multiplicities are abbreviated as follows: s (singlet), d (doublet), t (triplet), q (quartet), qu (quintet), m (multiplet) and br for broad signals. ESI mass spectra were recorded on a micromass Quattra LC spectrometer. MALDI-TOF accurate mass spectra were recorded on a Voyger-DE STR using α-cyano-4-hydroxycinnamic acid as the matrix. LC-MS spectra and accurate ESI mass spectra were recorded using a Xevo QTof mass spectrometer (Waters) coupled to an Acquity LC system (Waters) using an Acquity UPLC BEH C18 column (2.1 x 50 mm, Waters). The flow rate was 0.6 ml min^{-1} and the gradient was as follows: 95% Solvent A (0.1% formic acid in water) with 5% solvent B (0.1% formic acid in acetonitrile) was held constant for 0.5 min, then there was a linear gradient to 100% B over the next 2.1 min. After 1 min at 100% solvent B, the gradient was returned to 95% solvent A and 5% solvent B over 0.2 min. Mass accuracy was achieved using a reference lock mass scan, once per second. The ES cone voltage was 30 V and collision energy was 4 eV. The MS

acquisition rate was 10 spectra per second and m/z data ranging from 50 to 2000 Da was collected. IR spectra were recorded on a neat sample using a Perkin Elmer FT-IR with ATR attachment. Microwave reactions were conducted on a CEM Discover system fitted with an IR temperature sensor. Dry methanol was obtained using CaH_2, then distilled and stored over molecular sieves until use.

2.10.1 Synthesis of the BocPyOH Building Block

2.10.1.1 Synthesis of methyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxylate (21)[224]

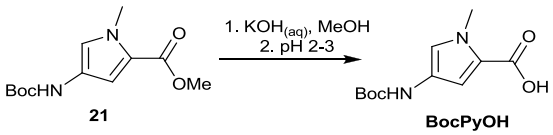


A suspension of **20** (10.0052g, 54.34 mmol), di-tert-butyl dicarbonate (23.745 g, 108.8 mmol) and nickel(II) chloride hexahydrate (1.2946 g, 5.45 mmol) in HPLC grade methanol (200 mL) was cooled to -20^{o}C in a 1 L 3 neck flask and sodium borohydride (14.3741 g, 380.0 mmol) added slowly in small portions to avoid excessive foaming. After the addition, the black solution was allowed to warm to room temperature under a nitrogen atmosphere and stirred for 16 hours. The crude reaction mixture was diluted with ethyl acetate (200 mL) and water (30 mL) to which celite was added and the resulting slurry filtered to remove a fine black precipitate. The organic solvent was removed from the filtrate under reduced pressure and the remaining aqueous suspension diluted with ethyl acetate (200 mL) which was washed with $NaHCO_3$, brine, a 1% aqueous solution of syn-diethylenetriamine (x 3) to remove any soluble nickel salts and brine. The solvent was removed from the organic layer under reduced pressure to yield the crude product which was redissolved in DCM (19 mL) and precipitated using

hexane (250 mL) to yield the title compound (10.569 g, 76%) as a white solid consistent with previously reported characterisation.[224] Mp. 114-115°C (lit. 115-116°C[224])

NMR: δ_{1H}(300 MHz, $CDCl_3$): 1.43 (9H, s, Boc), 3.69 (3H, s, C**H**$_3$), 3.77 (3H, s, C**H**$_3$), 6.60 (1H, s, Ar-**H**), 7.09 (1H, s, Ar-**H**), 9.09 (1H, s, N**H**Boc); δ_{13H}(100 MHz, DMSO): 28.1, 36.0, 50.8, 78.5, 107.3, 118.6, 119.2, 123.1, 152.7, 160.7; IR (neat) υ_{max}: 3349 (w), 1713 (m), 1683 (s), 1587 (m), 1555 (m), 1457 (m), 1388 (m), 1243 (s), 1150 (s), 779 (m). ESI: $M+Na^+$ 277.1172 (calculated 277.1164);

2.10.1.2 Synthesis of 4-((tert-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxylic acid (BocPyOH)[183]

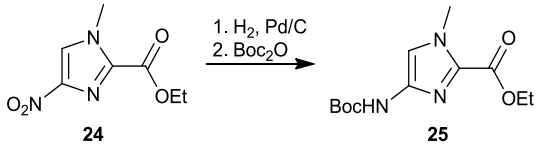


A solution of **21** (9.00 g, 35.4 mmol) in 1:1 methanol : $LiOH_{(aq)}$ (1 M, 70.8 mmol) was heated at 55°C for 2 hours and the organic solvent removed under reduced pressure. The remaining aqueous solution was then washed with EtOAc and acidified to pH 2-3 using 1 M $HCl_{(aq)}$ then extracted 3 times into EtOAc. The combined organic layers were dried with $MgSO_4$, the solvent removed under reduced pressure and the crude product redissolved in DCM before being precipitated with petrol 40-60°C and the resulting solid suspended in water and freeze dried to yield the title compound (7.5865 g, 89%) as a finely divided white solid consistent with previously reported characterisation[183, 224] with the exception of the unexpectedly low melting point. Mp. 140-142°C (lit. 160-161°C[224])

NMR: δ_{1H}(400 MHz, DMSO): 1.44 (9H, bs, Boc), 3.77 (3H, s, N-**CH**$_3$), 6.58 (1H, bs, Ar-H), 7.04 (1H, s, Ar-H), 9.04 (s, 1H, N**H**Boc), 12.08 (s, 1H, COO**H**); δ_{13C}(100 MHz, DMSO): 28.1, 36.0, 107.4, 78.4, 118.7, 119.6, 122.8, 152.7, 161.8; IR (neat) υ_{max}: 3348

(w), 2978 (br, w), 1694 (s), 1667 (s), 1449 (s), 1390 (s), 1284 (s), 1157 (s), 1146 (s), 1062 (m); ESI: M+Na^+ 263.1015 (calculated 263.1008);

2.10.1.3 Synthesis of ethyl 4-(tert-butoxycarbonylamino)-1-methyl-1*H*-imidazole-2-carboxylate (25)



24 **25**

The reduction and Boc protection of **24** was problematic and many different procedures were attempted in order to try and reduce the time required for the reaction.

Method 1 – H_2, Pd/C, MeOH, 1 atm, 48 hours

A 500 mL 3-neck flask containing a solution of **24** (5.172 g, 26.0 mmol) and di-tert-butyl dicarbonate (6.8233 g, 31.3 mmol) in dry methanol (250 mL) was fitted with a glass frit gas bubbler and hydrogen bubbled through for 48 hours. The crude reaction mixture was filtered through celite, and the solvent removed under reduced pressure. The resulting solid was redissolved in the minimum amount of DCM and a large volume of petrol 40-60^{o}C added to precipitate the title compound (1.2179 g, 17%). The DCM was removed under reduced pressure to give a further precipitate of the title compound (3.2041 g, 46%) and the products combined into a single fraction of white solid (4.422g, 63%).

NMR: δ_{1H}(400 MHz, DMSO): 1.28 (3H, t, ^3J = 7.1 Hz, $OCH_2C\mathbf{H_3}$), 1.44 (9H, s, Boc), 3.88 (3H, s, N-$C\mathbf{H_3}$), 4.25 (2H, q, ^3J = 7.1 Hz, $OC\mathbf{H_2}CH_3$), 7.30 (1H, bs, Ar-H), 9.67 (1H, bs, ArN**H**Boc).

Method 2 – H_2, Pd/C, THF, 50 psi, 18 hours

Dry THF (50 mL) was degassed by bubbling hydrogen through for 10 minutes then transferred into a metal hydrogenation flask containing **24** (0.1046 g, 0.53 mmol) and

palladium on carbon (0.0021 g, 10% by weight) and the reaction vessel quickly sealed. The vessel was purged with hydrogen, pressurised with hydrogen at 50 psi and stirred for 18 hours. The reaction mixture was filtered through celite and triethylamine (0.22 ml, 1.58 mmol) and di-tert-butyl dicarbonate (0.1151 g, 0.53 mmol) added. After stirring for 2 hours, the reaction mixture was added to water and neutralised before extraction into EtOAc. The combined organic layers were washed with brine, dried with $MgSO_4$ and the solvent removed by under reduced pressure to yield the starting material.

Method 3 – H_2, Pd/C, DMF, 50 psi, 22 hours

Dry DMF (50 mL) was degassed by bubbling hydrogen through for 10 minutes then transferred into a metal hydrogenation flask containing **24** (2.0078 g, 10.0 mmol) and palladium on carbon (0.0435 g, 10% by weight) and the reaction vessel quickly sealed. The vessel was purged with hydrogen, pressurised with hydrogen at 50 psi and stirred for 22 hours. The reaction mixture was filtered through celite and triethylamine (4.2 ml, 30.1 mmol) and di-tert-butyl dicarbonate (2.2271 g, 10.0 mmol) in dry DMF (5 mL) added. After stirring for 2 hours, the reaction mixture was added to water and neutralised before extraction into EtOAc. The combined organic layers were washed with brine, dried with $MgSO_4$ and the solvent removed by under reduced pressure to yield the starting material.

Method 4 – H_2, Pd/C, THF, 150 psi, 66 hours

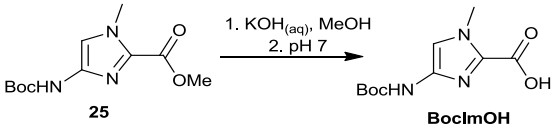
Dry THF (50 mL) was degassed by bubbling hydrogen through for 15 minutes then transferred into a metal hydrogenation flask containing **24** (0.0536 g, 0.25 mmol) and palladium on carbon (0.0043 g, 10% by weight) and the reaction vessel quickly sealed. The vessel was purged with hydrogen, pressurised with hydrogen at 150 psi and stirred for 66 hours. The reaction mixture was filtered through celite and triethylamine (0.0761

g, 0.75 mmol) and di-tert-butyl dicarbonate (0.0548 g, 0.25 mmol) added. After stirring for 2 hours, the reaction mixture was added to water and neutralised before extraction into EtOAc. The combined organic layers were washed with brine, dried with $MgSO_4$ and the solvent removed by under reduced pressure to yield the starting material.

Method 5 – H_2, EtOAc, 150 psi, 22 hours

EtOAc (50 mL) was dried using MgSO4 and degassed by bubbling hydrogen through for 15 minutes then transferred into a metal hydrogenation flask containing **24** (0.0542 g, 0.25 mmol) and palladium on carbon (0.0030 g, 10% w/w) and the reaction vessel quickly sealed. The vessel was purged with hydrogen, pressurised with hydrogen at 150 psi and stirred for 22 hours. The reaction mixture was filtered through celite and triethylamine (0.1 ml, 0.75 mmol) and di-tert-butyl dicarbonate (0.0554 g, 0.25 mmol) added. After stirring for 2 hours, the reaction mixture was added to water and neutralised before extraction into EtOAc. The combined organic layers were washed with brine, dried with $MgSO_4$ and the solvent removed by under reduced pressure to yield the starting material.

2.10.1.4 Synthesis of 4-((tert-butoxycarbonyl)amino)-1-methyl-1*H*-imidazole-2-carboxylic acid (BocImOH)[183]

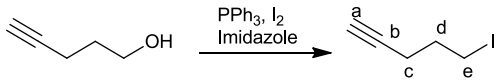


A solution of **25** (1.8071 g, 6.71 mmol) in 30 mL 1:1 ethanol:$LiOH_{(aq)}$ (1 M, 15 mmol) was heated at 60°C for 2 hours and the organic solvent removed under reduced pressure. The remaining aqueous solution was acidified to pH 7 using 1 M $HCl_{(aq)}$ and the resulting glue-like precipitate recovered and washed thoroughly with distilled water,

then resuspended in water and freeze dried to yield the title compound (0.8625 g, 53%) as a finely divided white solid consistent with previously reported characterisation.[183]

NMR: δ_{1H}(400 MHz, DMSO): 1.43 (9H, s, Boc), 3.86 (3H, s, N**C**$\mathbf{H_3}$), 7.22 (1H, s, Ar-H), 9.56 (1H, s, N**H**Boc).

2.10.1.5 Synthesis of 5-iodopent-1-yne (31)[225]

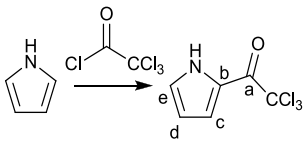


26

To a stirred solution of triphenyl phosphine (11.842 g, 45.1 mmol) in DCM (200 mL) were added in quick succession imidazole (3.0726 g, 45.1 mmol) and iodine (11.4598 g, 45.2 mmol). The solution was stirred for 30 minutes and a solution of pent-4-yn-1-ol (3 mL, 32.2 mmol) in DCM (50 mL) was added slowly. The solution was stirred for 48 hours, then washed with dilute sodium sulphite solution and brine. The organic layer was separated and the solvent removed carefully under reduced pressure to yield the crude product was diluted with petrol 40-60^{o}C and the resultant precipitate of triphenylphosphine oxide removed by filtration. The organic solvent was removed carefully under reduced pressure and the product purified by vacuum distillation (40-51 mbar, 60-75^{o}C) to yield the title compound (5.2243 g, 84%) as a clear liquid consistent with previously reported characterisation.[225]

NMR: δ_{1H}(400MHz, $CDCl_3$): 1.99 (1H, t, ^4J = 2.6 Hz, H_a), 2.01 (2H, quintet, ^3J = 6.7 Hz, H_d), 2.28 (2H, td, ^3J = 6.7 Hz, ^4J = 2.6 Hz, H_c), 3.25 (2H, t, ^3J = 6.7 Hz, H_e); δ_{13C}(75 MHz, $CDCl_3$): 5.0 (C_e), 19.4 (C_c), 31.8 (C_d), 69.4 (C_a), 82.3 (C_b). IR (neat) υ_{max}: 3294 (m), 1427 (w), 1221 (m), 634 (s)

2.10.1.6 Synthesis of 2,2,2-trichloro-1-(1*H*-pyrrol-2-yl)ethanone (32)[226]

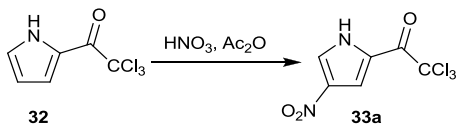


32

Pyrrole (10 mL, 144 mmol) was dissolved in dry diethyl ether (80 mL) in a dropping funnel and added slowly over 2 hours to a solution of trichloroacetyl chloride (16.5 mL, 148 mmol) in dry diethyl ether (25 mL) in a 250 mL 3-neck flask fitted with a reflux condenser and a nitrogen bubbler. The reaction mixture was stirred for an additional 60 minutes before a solution of potassium carbonate (12.6 g in 40 mL water) was slowly added through the dropping funnel. The organic layer was separated off, dried with $MgSO_4$ and 0.75 g activated charcoal was added. The solution was filtered through celite, the solvent removed under reduced pressure and the crude product recrystallised from hot hexane to yield the title compound (21.43g, 70 %) as a cream solid consistent with previously reported characterisation. [227]

NMR: δ_{1H}(400MHz, $CDCl_3$): 6.39 (1H, dt, H^d, ^3J = 4.1 Hz, ^4J = 2.7 Hz), 7.17 (1H, td, H^c, ^3J = 2.7 Hz, ^4J = 1.3 Hz), 7.39 (1H, ddd, H^e, J = 4.1 Hz, 2.7 Hz, 1.3 Hz), 9.49 (1H, bs, NH); δ_{13C}(75 MHz, DMSO): 95.2 (CCl_3), 111.3 (C-H), 121.3 (C-H), 121.7, 129.4 (C-H), 171.4 (C=O).

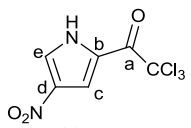
2.10.1.7 Synthesis of 2,2,2-trichloro-1-(4-nitro-1*H*-pyrrol-2-yl)ethanone (33a)[172]



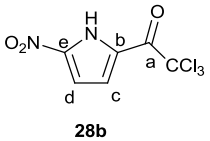
32 **33a**

A solution of **32** (20.0832 g, 94.5 mmol) in acetic anhydride (100 mL) was cooled to -20^{o}C and fuming nitric acid (6 mL) added slowly whilst this temperature was

maintained. The solution was then allowed to warm to room temperature and stirred for 4 hours before being cooled again to -30^{o}C and water added. After stirring for 30 minutes and being allowed to warm to room temperature, the mixture was allowed to stand for a further 30 minutes and the precipitate collected by filtration. The crude precipitate was dried by azeotropic distillation with toluene then dissolved in ethyl acetate and treated with $MgSO_4$ to remove the remaining traces of water. Removal of the organic solvent under reduced pressure yielded the product as a variable mix of isomers (**33a**, **33b,** 1:3 to 32:1 ratio) (11.91 g, 49%).



28a



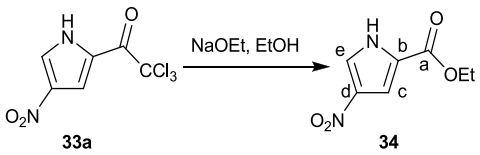
28b

The desired isomer **33a** is the major product of this reaction and was isolated by recrystallisation (ethyl acetate, hexane) to yield the title compound as a cream solid with only a single isomer present (5.996 g, 25% overall yield) consistent with previously reported characterisation.[172]

33a: NMR: δ_{1H}(400MHz, DMSO): 7.71 (1H, d, H^e, J = 1.4 Hz), 8.38 (1H, d, H^c, J = 1.4 Hz), 13.64 (1H, bs, NH); δ_{13C}(75 MHz, DMSO): 95.2 (CCl_3), 111.3 (C-H), 121.3 (C-H), 121.7, 129.4 (C-H), 171.4 (C=O)

33b: NMR: δ_{1H}(300MHz, DMSO): 7.24 (1H, d, Ar-H, ^{3}J = 4.5 Hz), 7.35 (1H, d, Ar-H, ^{3}J = 4.5 Hz).

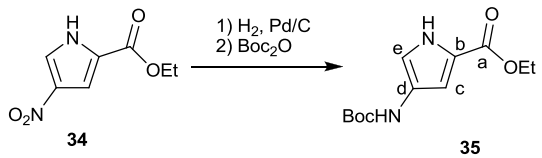
2.10.1.8 Synthesis of ethyl 4-nitro-1*H*-pyrrole-2-carboxylate (34)[228]



A solution of sodium ethoxide was produced by dissolving a 60% suspension of sodium hydride (0.3070 g) in absolute ethanol (5 mL) and added to a solution of **33a** (3.0337 g, 11.8 mmol) in absolute ethanol (50 mL). The solution was stirred at room temperature for 1 hour, then the solvent removed under reduced pressure. The crude product was redissolved in ethyl acetate and washed with an acidified brine solution. The aqueous layer was reextracted twice with ethyl acetate and the combined organic fractions dried using MgSO4, filtered and the solvent removed under reduced pressure to yield the title compound (2.167 g, 99%) as a white powder constistent with previously reported characterisation.[228]

NMR: $\delta_{\mathbf{1H}}$(400 MHz, DMSO): 1.30 (3H, t, ^{3}J = 7.1 Hz, $OCH_2C\mathbf{H}_3$), 4.29 (2H, q, ^{3}J = 7.1 Hz, $OC\mathbf{H}_2CH_3$), 7.27 (1H, d, ^{4}J = 1.7 Hz, CH^{c}), 8.07 (1H, d, ^{4}J = 1.7 Hz, CH^{e}), 13.13 (1H, bs, NH); $\delta_{\mathbf{13H}}$(75 MHz, DMSO): 14.1 ($OCH_2\mathbf{C}H_3$), 60.7 ($O\mathbf{C}H_2CH_3$), 109.4 (C^{c}), 123.0 (C^{b}), 124.3 (C^{e}), 136.6 (C^{d}), 159.4 (C^{a}).

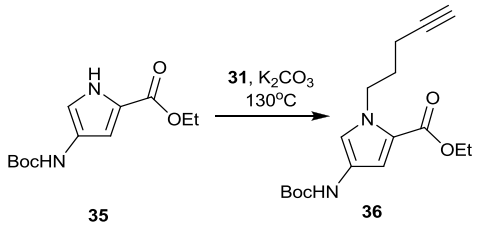
2.10.1.9 Synthesis of ethyl 4-(tert-butoxycarbonylamino)-1*H*-pyrrole-2-carboxylate (35) [228]



A mixture of **34** (1.80 g, 9.77 mmol) and 10% Pd/C (0.2018g) in a dry 500 mL 3-neck flask was placed under a nitrogen atmosphere and dry methanol (250 mL) added using a

cannular. Hydrogen was then bubbled through the solution using a frit for 2 hours, then di-tert-butyl dicarbonate (2.5607 g, 11.73 mmol) added and the solution stirred for a further 2 hours under a static hydrogen atmosphere. The reaction mixture was diluted with methanol, filtered through celite and the solvent removed under reduced pressure. The crude product was dissolved in DCM and washed with saturated $NaHCO_3$ solution to remove any unreacted **34** and brine, then dried using $MgSO_4$, filtered and the solvent removed under reduced pressure until the minimum amount of DCM remained. The addition of petrol 40-60^oC yielded the title compound **35** as a fine white solid (1.6458 g, 66%) consistent with all previously reported characterisation.[228] Mp. 189-190°C NMR: δ_{1H}(400 MHz, DMSO): 1.26 (3H, t, ^3J = 7.1 Hz, $OCH_2C\mathbf{H}_3$), 1.44 (9H, s, Boc), 4.19 (2H, q, ^3J = 7.1 Hz, $OC\mathbf{H}_2CH_3$), 6.60 (1H, bs, CH^c), 6.94 (1H, bs, CH^e), 9.07 (1H, bs, NH), 11.49 (1H, bs, NH); δ_{13C}(75MHz, DMSO): 14.3, 18.1, 45.6, 59.4, 105.3, 112.5, 119.2, 124.9, 152.7, 160.3. IR (neat) υ_{max}: 3295 (m), 1677 (s), 1562 (m), 1391 (m), 1249 (m), 1217 (m), 1158 (m), 1118 (m), 771 (m); ESI: MH^+ 255.1348 (calculated 255.1345);

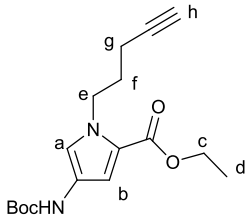
2.10.1.10 Synthesis of ethyl 4-((tert-butoxycarbonyl)amino)-1-(pent-4-yn-1-yl)-1*H*-pyrrole-2-carboxylate (36)



This reaction was carried out sequentially five times and the products combined during the workup due to significant drop in yield after scale-up during optimisation.

To a solution of **35** (0.48 g, 1.9 mmol) and **31** (0.7525 g, 3.8 mmol) in dry dioxane (5 mL) was added Cs_2CO_3 (2.4603 g, 7.6 mmol) and PEG-400 (1.0 g). The reaction

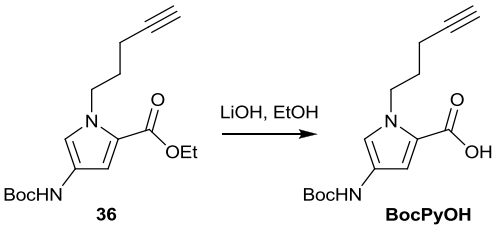
flask was sealed and microwave heated at 85°C for 90 minutes and combined with the other reaction products. The combined mixture was washed with saturated $NaHCO_3$ solution and brine, dried using $MgSO_4$, filtered and the solvent removed under reduced pressure. The crude reaction mixture was purified by column chromatography (EtOAc in DCM, 0%-5%) to yield the title compound **36** as a white solid (2.07 g, 68%). Mp. 92-94°C



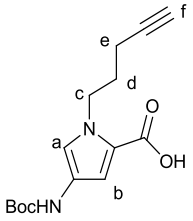
36

NMR: δ_{1H}(400 MHz, DMSO): 1.26 (3H, t, ^3J = 7.1 Hz, H^d), 1.45 (9H, s, Boc), 1.82 (2H, qu, ^3J = 7.1 Hz, H^f), 2.09 (2H, td, ^3J = 7.1 Hz, ^4J = 2.6 Hz, H^g), 2.84 (1H, t, ^4J = 2.6 Hz, H^h), 4.19 (2H, q, ^3J = 7.1 Hz, H^c), 4.26 (2H, t, ^3J = 7.1 Hz, H^e), 6.67 (1H, bs, $CH^{a/b}$), 7.11 (1H, bs, $CH^{a/b}$), 9.12 (1H, bs, N**H**Boc); δ_{13C}(100 MHz, CD_3OD): 14.7, 16.2, 28.8, 31.4, 48.7, 61.0, 70.4, 80.5, 83.9, 110.2, 120.3, 120.4, 124.4 155.6, 162.4; IR (neat) υ_{max}: 3289 (m), 2979 (w), 2946 (w), 1685 (s), 1593 (s), 1526 (m), 1453 (m), 1396 (m), 1368 (m), 1274 (m), 1252 (s), 1219 (m), 1191 (m), 1092 (s), 1075 (m), 1027 (w), 988 (m), 886 (w), 829 (w), 780 (m), 759 (w), 692 (m); ESI: MH^+ 321.1822 (calculated 321.1814);

2.10.1.11 Synthesis of 4-(tert-butoxycarbonylamino)-1-(pent-4-ynyl)-1*H*-pyrrole-2-carboxylic acid (BocPy(///)OH)



A suspension of **36** (0.39 g, 1.21 mmol) in 5 mL 1:1 absolute ethanol: $LiOH_{(aq)}$ (1 M, 2.5 mmol), was heated at 50°C for 2 hours until the solution became clear. The organic solvent was removed under reduced pressure and the remaining aqueous solution diluted, then extracted with EtOAc. The aqueous phase was then acidified to pH 3 using $HCl_{(aq)}$ and extracted three times with EtOAc. These combined organic fractions were dried using $MgSO_4$, filtered and the solvent removed under reduced pressure. The crude product was suspended in water and freeze dried to yield the title compound **BocPy(///)OH** as cream powder (0.30 g, 84%). Mp. 121-134°C

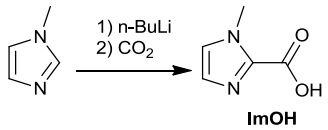


BocPy(///)OH

NMR: δ_{1H}(400 MHz, DMSO): 1.43 (9H, s, Boc), 1.81 (2H, qu, ^3J = 7.0 Hz, H^d), 2.02 (2H, td, ^3J = 7.0 Hz, ^4J = 2.6 Hz, H^e), 2.76 (1H, t, ^4J = 2.6 Hz, H^f), 4.31 (2H, t, ^3J = 7.0 Hz, H^c), 6.19 (1H, bs, $CH^{a/b}$), 6.67 (1H, bs, $CH^{a/b}$), 8.74 (1H, bs, N**H**Boc); δ_{1H}(400 MHz, CD_3OD): 1.39 (9H, s, Boc), 1.83 (2H, qu, ^3J = 6.9 Hz, H^d), 2.03 (2H, td, ^3J = 6.9 Hz, ^4J = 2.6 Hz, H^e), 2.17 (1H, t, ^4J = 2.6 Hz, H^f), 4.26 (2H, t, ^3J = 6.9 Hz, H^c), 6.64 (1H, bs, $CH^{a/b}$), 6.97 (1H, bs, $CH^{a/b}$); δ_{13C}(100 MHz, CD_3OD): 16.2 28.7, 31.5, 70.3, 80.5, 83.9, 110.5, 120.3, 120.7, 124.3, 155.6, 164.0; IR (neat) υ_{max}: 3369 (w), 3294 (w),

2969 (w), 2936 (w), 2616 (w), 1693 (m), 1671 (s), 1586 (m), 1549 (s), 1462 (m), 1390 (s), 1366 (m), 1293 (m), 1283 (m), 1442 (s), 1161 (s), 1093 (m), 1064 (m), 995 (m), 939 (w), 886 (m), 795 (m), 766 (m), 718 (m); ESI: MH^+ 293.1514 (calculated 293.1501).

2.10.1.12 Synthesis of 1-methyl-1*H*-imidazole-2-carboxylic acid (ImOH)[111]



ImOH

A solution of N-methyl imidazole (0.97mL, 12.2 mmol) in dry THF (20 mL) was cooled to -78^{o}C and a solution of n-butyl lithium (3.3 mL, 2.5 M in hexanes, 13.0 mmol) added then stirred at -78^{o}C for 90 minutes. Carbon dioxide gas was generated using dry ice in a side arm flask and dried by passing the gas through calcium chloride pellets before being bubbled through the reaction solution for 30 minutes after which the solution was allowed to warm to room temperature and quenched using 1 M $HCl_{(aq)}$ (3 mL, 3 mmol). The yellow oil which separated underneath the water was separated and freeze dried to yield the title compound as a white powder (1.368 g, 89%) consistent with previously reported characterisation data.[111]

NMR: δ_{1H}(400 MHz, DMSO): 3.94 (3H, s, N-C$\mathbf{H_3}$), 6.82 (1H, s, Ar-H), 7.13 (1H, s, Ar-H).

2.10.2 Solid Phase Synthesis Protocols

2.10.2.1 General procedure for Boc-β-Ala-PAM resin

Polyamides were synthesised using modified literature protocols.[185] For full length polyamides a low loading 0.26 mequiv/g resin was used and the higher loading 0.58 mequiv/g resin for shorter fragments. TPW (5% phenol w/v, 2.5% H_2O v/v, 95% trifluoroacetic acid v/v) was used for all Boc deprotections and 1:1 NMP:DMF used to

wash away excess reagents. Except during the addition of reagents, the solid phase synthesis flask was mechanically shaken with sufficient liquid present to allow the resin to flow freely. Removal of the reagents between washes was achieved using a suction pump.

Resin preparation

Before use the resin was swelled using DCM for 1 hour, then deprotected by washing with TPW (30s, 2 min, 2 min). The resin was then washed with DCM (15s) and 1:1 NMP:DMF (15s, 15s, 15s).

Coupling Protocol – BocPyOH, BocPy(///)OH, BocPy(N_3)OH

The appropriate building block (4 equiv.) and HATU (3.6 equiv) were dissolved in dry peptide grade DMF and DIEA added (12 equiv.). After activation in a sealed vial for 2 minutes, the mixture was added to the flask and the shaken for 1 hour. The resin was then washed with 1:1 NMP:DMF (15s, 15s, 15s) and a few grains of resin taken for analysis.

Coupling Protocol – BocImOH

The **BocImOH** building block (4 equiv.), HOAt (4 equiv.) and DCC (3.6 equiv.) were dissolved in dry peptide grade DMF in a sealed vial and allowed to stand for at least 3 hours. After the precipitate had been formed, the mixture was centrifuged and the supernatant liquid added to the flask with DIEA (12 equiv.) and shaken for 1-2 hours. The resin was then washed with 1:1 NMP:DMF (15s, 15s, 15s) and DCM (15s) then a few grains of resin taken for analysis.

Coupling Protocol – Boc-γ-OH

The **Boc-γ-OH** building block (4 equiv.) and HATU (3.6 equiv) were dissolved in dry peptide grade DMF and DIEA added (12 equiv.). After activation in a sealed vial for 2 minutes, the mixture was added to the flask and the shaken for 1.5 hours. The resin was

then washed with 1:1 NMP:DMF (15s, 15s, 15s) and DCM then a few grains of resin taken for analysis.

Coupling Protocol – ImOH Stopper

The **ImOH** building block (4 equiv.), HOAt (4 equiv.) and DCC (3.6 equiv.) were dissolved in dry peptide grade DMF in a sealed vial and allowed to stand for at least 1 hour. After the precipitate had been formed, the mixture was centrifuged and the supernatant liquid added to the flask with DIEA (12 equiv.) and shaken for 1 hour. The resin was then washed with 1:1 NMP:DMF (15s, 15s, 15s) and DCM (15s) then a few grains of resin taken for analysis.

Deprotection Protocol

Deprotection of the resin bound polyamide was achieved by washing with TPW (30 s, 2 min, 2 min) except for the **Im** building block which required extended washing (20 min, 2 min, 2 min). The resin was then washed with DCM (15s) and 1:1 NMP:DMF (15s, 15s, 15s) to remove the remaining TFA. A few grains of the resin were then removed for analysis. It was found that after the addition of around 6 building blocks, extended reaction times were required for complete deprotection to take place.

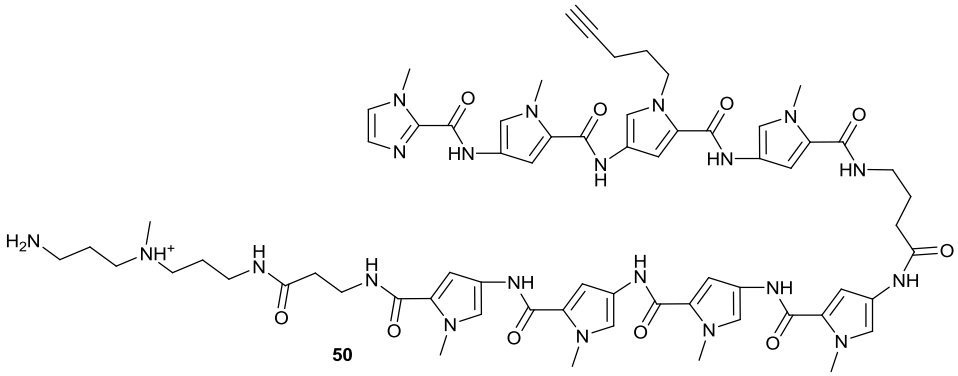
Resin cleavage

After washing with 1:1 NMP:DMF, the resin was then shrunk using methanol (4 x 10 minutes). The recovered resin was suspended in 3-(dimethylamino)-1-propylamine **40** and heated at 50^{o}C for 16 hours then centrifuged to separate off the supernatent. The crude supernatant was then diluted (10:1) with diethyl ether and the resulting precipitate isolated by centrifugation. The supernatant was removed and the precipitation procedure repeated to yield the crude product as a coloured precipitate ready for HPLC purification.

Reaction Monitoring

A few drops of chloranil and acetaldehyde were added to each resin sample and the mixture heated in an oil bath at 90°C for 30 seconds. In the presence of a free secondary amine, the resin changed from clear to dark red after heating. Similarly, resin samples for HPLC were suspended in 3-(dimethylamino)-1-propylamine and heated at 90°C for 20 minutes in an oil bath. After centrifugation, the supernatant was diluted with diethyl ether and the resulting precipitate washed again with diethyl ether. The remaining precipitate was then dissolved in 100 μL of HPLC grade methanol and analysed by HPLC at 254 nm and 310 nm.

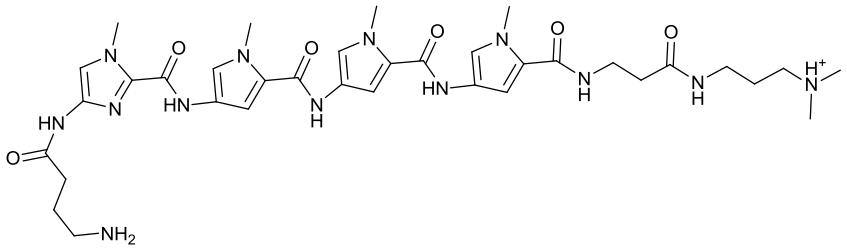
2.10.2.2 Synthesis of ImPyPy(///)Py-γ-PyPyPyPyβDp (50)



50

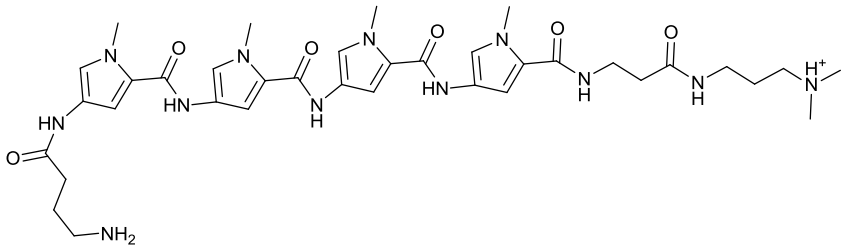
The product was synthesised using standard protocols on low loading Boc-β-Ala-PAM resin (0.050g, 0.26 mequiv/g), cleaved using 3-(dimethylamino)-1-propylamine **40** and purified by HPLC. No matching peaks were found by MALDI or electrospray mass spectrometry and a sample prepared by Dr Wu Su was used for further studies.

2.10.2.3 Synthesis of H_2N-γ-ImPyPyPy-β-Dp (42)



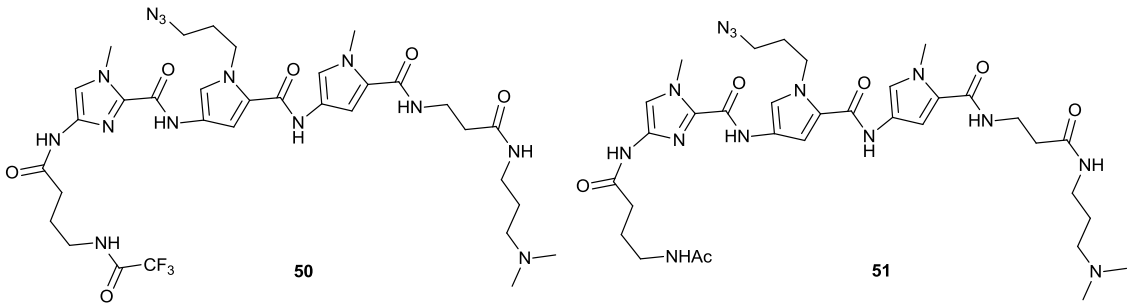
The product was synthesised using standard protocols on high loading Boc-β-Ala-PAM resin (0.199 g, 0.58 mequiv/g), cleaved using 3-(dimethylamino)-1-propylamine **40** and purified by HPLC (10-90% 18 min, MeCN in H_2O, 0.1%TFA) to yield the title compound as a white solid (23.3 mg, 27%). MALDI: MH^+ 748.885 (calculated 748.400).

2.10.2.4 Synthesis of H_2N-γ-PyPyPyPy-β-Dp (43)



The product was synthesised using standard protocols on high loading Boc-β-Ala-PAM resin (0.1070 g, 0.58 mequiv/g), cleaved using 3-(dimethylamino)-1-propylamine (**Dp**) and purified by HPLC (10-20% 2 min, 20-50% 20 min, MeCN in H_2O, 0.1%TFA) to yield the title compound as a white solid (13.6 mg, 29%). MALDI: MH^+ 747.6106 (calculated 747.4049).

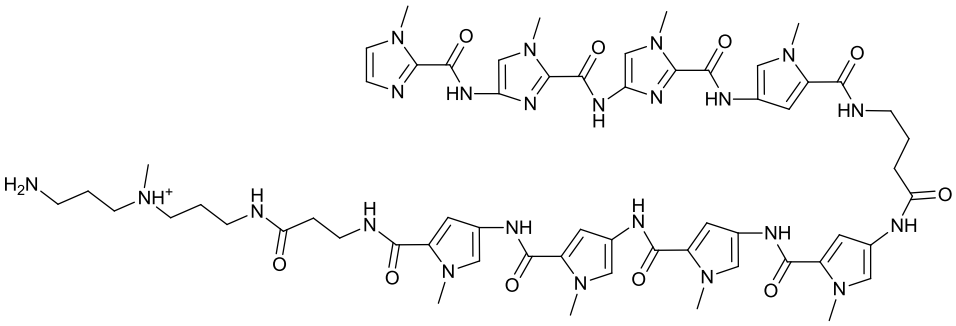
2.10.2.5 Synthesis of CF_3-γ-ImPy(N_3)Py-β-Dp (50)



The synthesis of polyamide **51** was attempted using standard protocols on high loading Boc-β-Ala-PAM resin (0.050g, 0.58 mequiv/g) with the terminal acetyl group installed using acetic anhydride (100μL in 1.5 mL DMF, 30 minutes) followed by standard post

coupling washes. The **BocPy(N_3)OH** building block was supplied by Rugerro Dondi. The crude polyamide was cleaved from the resin using 3-(dimethylamino)-1-propylamine (**Dp)** and purified by HPLC (10-90% MeCN in H_2O, 18 minutes, 0.1% TFA) to yield a mixture of acetyl capped polyamide **51** (2.5 mg, 12%) and TFA capped polyamide **50** (1.3 mg, 6%). **50**: MALDI: MH^+ 791.6286 (calculated 791.3671), **51**: MALDI: MH^+ 737.2277 (calculated 737.3955)

2.10.2.6 Synthesis of ImImImPy-γ-PyPyPyPy-β-NH_2 (52)

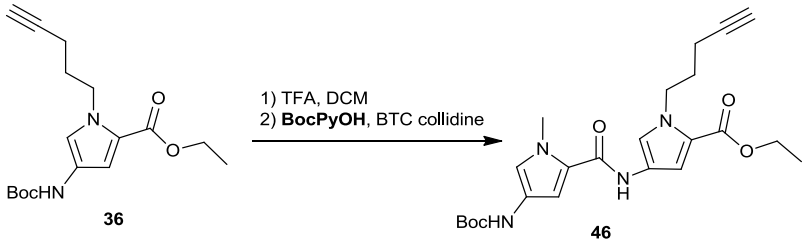


The product was synthesised using standard protocols on low loading Boc-β-Ala-PAM resin (0.1028 g, 0.26 mequiv/g), cleaved using 3,3'-diamino-*N*-methyldipropylamine **40** and purified by HPLC (MeCN in H_2O, 0.1%TFA, 10-20% 5 min, 20-45% 25 min, 45-90% 2 min) to yield the title compound **52** as a white solid (15.6 mg, 46%). ESI: MH^+ 1266.6064 (calculated 1266.6164); MALDI: MH^+ 1266.7088 (calculated 1266.6164). NMR: δ_{1H}(400 MHz, CD_3OD): 1.90 (2H, qu, J = 6.3 Hz), 2.01 (2H, qu, J = 7.1 Hz), 2.32 (2H, t, J = 6.8 Hz), 2.43 (2H, t, J = 6.2 Hz), 2.76 (3H, s), 2.96 (2H, t, J = 7.4 Hz), 3.32 (2H, t, J = 6.3 Hz), 3.76 (3H, s, Py-Me), 3.77 (3H, s, Py-Me), 3.77 (3H, s, Py-Me), 3.79 (3H, s, Py-Me), 3.92 (3H, s, Im-Me), 3.96 (6H, s, Im-Me), 6.61-6.63 (1H, m), 6.72- 6.79 (4H, m), 6.85-6.87 (1H, m), 6.98-7.05 (4H, m), 7.11-7.14 (1H, m), 7.19 (1H, s), 7.32 (1H, s), 7.40 (1H, s). Additional multiplets at 3.40-3.55 (2H?, appears as

numerous sharp peaks), 3.00-3.15 (2H?, broad signal with sharp quintet on top) and 1.79-1.82 (broad signal merging with quintet at δ 1.92).

2.10.3 Solution Phase Synthesis

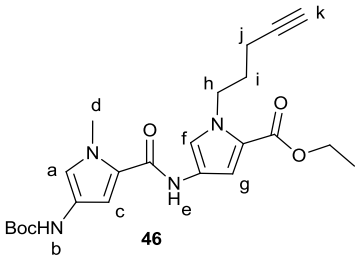
2.10.3.1 Synthesis of BocPyPy(///)OEt (46) (BTC Methodology)



A sample of **36** (0.1283 g, 0.4 mmol) was dissolved in dry diethyl ether (25 mL) and HCl gas bubbled through for 5 minutes during which two distinct precipitates were observed. The HCl gas was again bubbled through for 5 minutes and the solution allowed to stir for 30 minutes. After removal of the reaction solvent under reduced pressure, NMR and mass spectrometry showed no deprotection had taken place. The crude reaction mixture was then stirred in 1:1 mixture of triflouroacetic acid and wet DCM (50 mL) for 30 minutes and the solvent removed under reduced pressure to yield the Boc deprotected intermediate.

A mixture of **BocPyOH** (0.0968 g, 0.4 mmol) and BTC (0.0396 g, 0.13 mmol) was dissolved in dry THF (1 mL) and collidine (160 µL) added dropwise. After mixing for 1 minute, the solution was added to the deprotected intermediate dissolved in dry THF (1 mL) and DIEA (140 µL, 0.8 mmol) added. After stirring for 2 hours, the reaction was diluted with saturated $NaHCO_3$ solution and ethyl acetate, the organic layer separated, then washed with hydrochloric acid (pH 2.15) and brine. Drying with $MgSO_4$ and removal of the solvent under reduced pressure gave an oil due to the presence of collidine. Precipitation was attempted using toluene, but both the solid and

the supernatant were found to contain the product. The solvent in the supernatant was removed under reduced pressure and the product (0.0408 g) used crude for the next reaction.

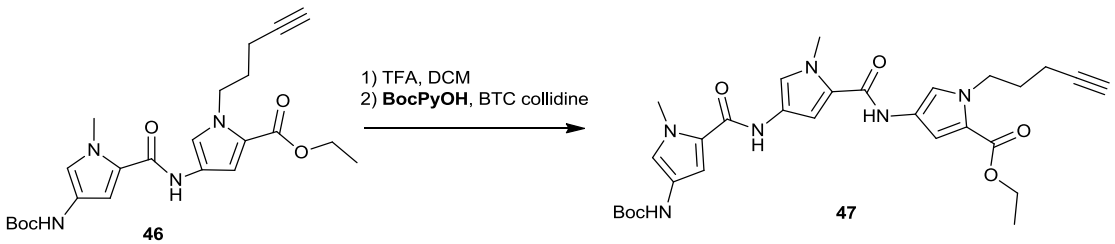


46

NMR: δ_{1H}(400 MHz, DMSO): 1.18 (2H, t, OC**H**$_2$CH$_3$, J = 7.0 Hz), 1.46 (9H, s, Boc), 1.85 (2H, qu, H^i, J = 7.0 Hz), 2.12 (2H, td, H^j, J = 7.0 Hz, J = 2.6 Hz), 2.85 (1H, t, H^k, J = 2.6 Hz), 3.81 (3H, s, H^d), 4.21 (2H, q, OCH$_2$C**H**$_3$, J = 7.0 Hz), 4.31 (2H, t, H^h, J = 7.0 Hz), 6.84 (1H, bs, H$^{a/c}$), 6.89 (1H, bs, H$^{a/c}$), 6.96 (1H, d, H$^{f/g}$, J = 2.0 Hz), 7.47 (1H, d, H$^{f/g}$, J = 2.0 Hz), 9.10 (1H, bs, H^b), 9.85 (1H, s, H^e); NMR: δ_{13C}(100 MHz, DMSO): 14.2, 14.9, 17.3, 18.4, 21.7, 28.2, 29.9, 36.0, 47.1, 59.4, 71.7, 83.3, 103.8, 109.0, 119.7, 122.4, 137.4, 146.1, 152.8, 158.4, 160.1; ESI: MH$^+$ 443.2308 (calculated 443.2294)

2.10.3.2 Synthesis of BocPyPyPy(///)OEt (47) (BTC and HATU methodology)

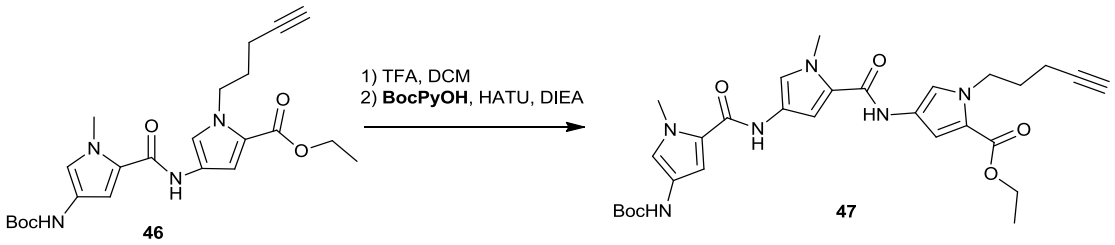
BTC Methodology



Crude **46** (0.0408 g) was dissolved in a 1:1 mixture of trifluoroacetic acid and wet DCM (25 mL), stirred for 30 minutes, then the solvent removed under reduced pressure to

yield the deprotected intermediate. A mixture of **BocPyOH** (0.0987 g, 0.4 mmol) and BTC (0.0400 g, 0.13 mmol) was dissolved in dry THF (1 mL) and collidine (160 µL) added dropwise. After mixing for 1 minute, the solution was added to the deprotected intermediate dissolved in dry THF (1 mL) and DIEA (140 µL, 0.8 mmol) added. After stirring for 2 hours, the reaction was diluted with a $NaHCO_3$ solution and ethyl acetate, the organic layer separated then washed twice with acidifed brine (pH 2.15) and brine. Drying with $MgSO_4$ and removal of the solvent under reduced pressure gave a crude oil (0.2664 g) which was used for the next step without purification.

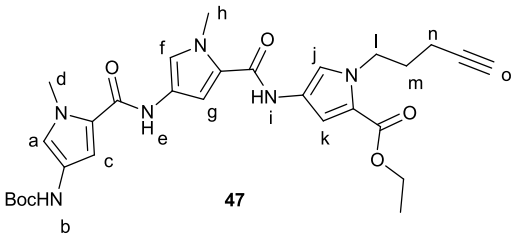
HATU Methodology



Dimer **46** (0.1361g, 0.31 mmol) was dissolved in a 1:1 mixture of trifluoroacetic acid and laboratory grade DCM (10 mL), stirred for 30 minutes and the solvent removed under reduced pressure. The crude mixture was diluted in toluene which was then removed under reduced pressure to remove additional residual trifluoroacteic acid and give the deprotected intermediate as an orange oil.

Separately, the **BocPyOH** building block (0.0741g, 0.31 mmol) and HATU (0.1055g, 0.28 mmol) were dissolved in dry DMF (2 mL) and DIEA (200 µL, 1.15 mmol) added. After stirring for 5 minutes, this solution was added to a mixture of the deprotected intermediate and DIEA (230 µL, 1.32 mmol) dissolved in dry DMF (3 mL) and stirred for 2 hours. The reaction mixture was diluted with EtOAc and washed with a saturated $NaHCO_3$ solution, dilute HCl (pH ≈ 2.5) and brine. The organic layer was dried using MgSO4 and the solvent removed under reduced pressure. The crude product was

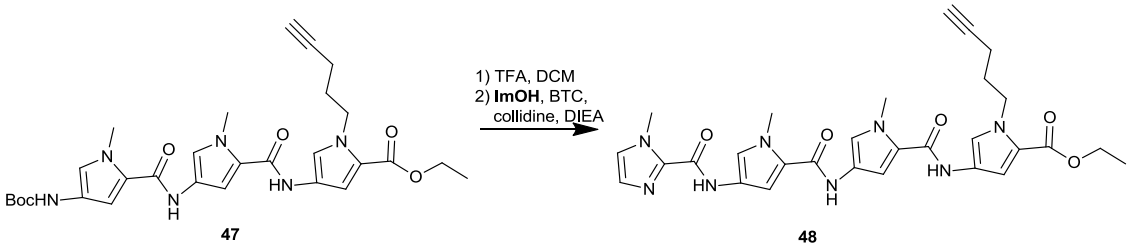
diluted in a small amount of DCM (2 mL) and precipitated with petrol 40-60°C to give the product as a finely divided brown solid (0.1458 g) which NMR analysis showed remained contaminated with DMF and was used crude for further reactions.



NMR: δ_{1H}(400 MHz, DMSO): 1.18 (2H, t, $OCH_2C\mathbf{H_3}$, J = 7.0 Hz), 1.45 (9H, s, Boc), 1.86 (2H, qu, H^m, J = 7.0 Hz), 2.12 (2H, td, H^j, J = 7.0 Hz, J = 2.5 Hz), 2.86 (1H, t, H^o, J = 2.5 Hz), 3.81 (3H, s, H^d), 3.85 (3H, s, H^h), 4.22 (2H, q, $OC\mathbf{H_2}CH_3$, J = 7.0 Hz), 4.32 (2H, t, H^l, J = 7.0 Hz), 6.84 (1H, bs, $H^{a/c}$), 6.90 (1H, bs, $H^{a/c}$), 6.97 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 7.07 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 7.22 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 7.49 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 9.08 (1H, bs, H^b), 9.85 (1H, s, $H^{e/i}$), 9.92 (1H, s, $H^{e/i}$) ; ESI: MH^+ 565.2798 (calculated 565.2775).

2.10.3.3 Synthesis of ImPyPyPy(///)OEt (48)

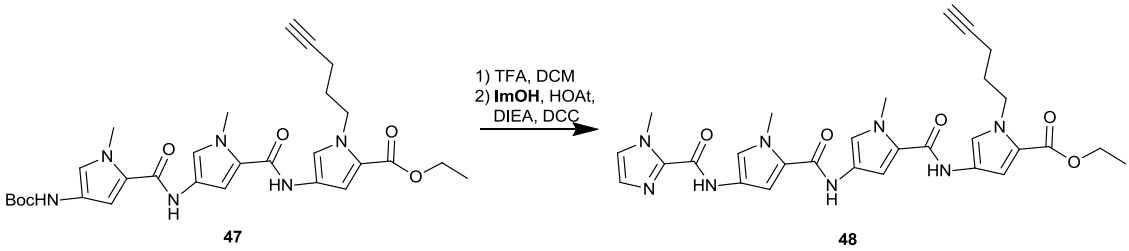
BTC Methodology



Crude **47** (0.2664 g) was dissolved in a 1:1 mixture of trifluoroacetic acid and wet DCM (25 mL), stirred for 30 minutes, then the solvent removed under reduced pressure to yield the deprotected intermediate. A mixture of **ImOH** (1-methyl-1*H*-imidazole-2-carboxylic acid) (0.0507 g, 0.4 mmol) and BTC (0.0401 g, 0.13 mmol) was dissolved in

dry THF (1 mL) and collidine (160 µL) added dropwise. After mixing for 1 minute, the solution was added to the deprotected intermediate dissolved in dry THF (1 mL) and DIEA (420 µL, 2.40 mmol) added. After stirring for 2 hours, the reaction was diluted with a $NaHCO_3$ solution and ethyl acetate, the organic layer separated then washed twice with acidifed brine (pH 2.15) and brine. Drying with $MgSO_4$ and removal of the solvent under reduced pressure gave a crude oil which was purified by column chromatography which yielded two fractions with a combined weight of 40.4 mg. The product of the reaction was identified by mass spectrometry and NMR as the trifluroacetic acid amide **49**.

HATU Methodology

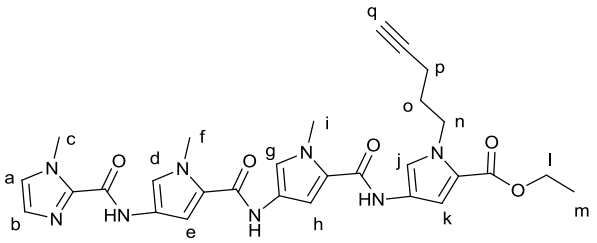


47 48

Crude **47** (0.1265 g, ~0.22mmol) was dissolved in a 1:1 mixture of trifluoroacetic acid and wet DCM (25 mL), stirred for 30 minutes, then the solvent removed under reduced pressure. The crude mixture was diluted in toluene which was then removed under reduced pressure to remove additional residual trifluoroacteic acid and give the deprotected intermediate as an orange oil. Separately, a mixture of DCC (0.0421 g, 0.2 mmol), **ImOH** (1-methyl-1*H*-imidazole-2-carboxylic acid, 0.0284 g, 0.22 mmol) and HOAt (0.0303 g, 0.22 mmol) were dissolved in dry DMF (2 mL) and stirred for 45 minutes until the DCU precipitate was observed. This solution was then added to a mixture of the deprotected intermediate and DIEA (300 µL, 1.79 mmol) dissolved in dry DMF (3 mL) and stirred for 18 hours. The reaction mixture was filtered and the solvent removed under reduced pressure to yield the crude product (0.3158 g). Initial attempts to purify the product using manual reverse phase chromatography (MeCN,

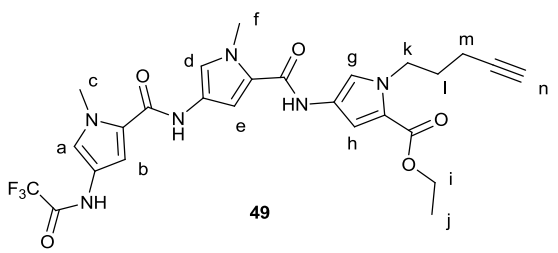
0.1% TFA) and HPLC failed to separate the product **48** from the trifluoroacetic acid conjugate **49** and purification was achieved using normal phase chromatography (1:1 EtOAc:DCM, 5% NEt_3 to 10% MeOH, 85% EtOAc, 5% NEt_3) which gave the product **48** as a orange oil (0.0358 g, 28%) and the trifluoroacetic acid conjugate **49** as a brown oil (0.0112 g, 9%). The NMR of both products showed some contamination by residual NEt_3.

Product 48



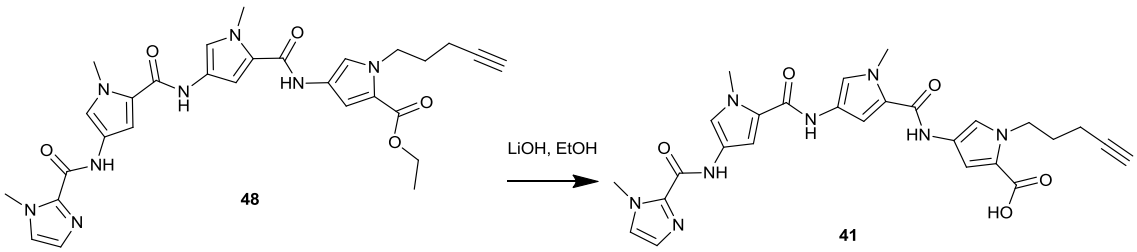
NMR: δ_{1H}(400 MHz, CD_3OD): 1.35 (3H, t, H^m, J = 7.1 Hz), 1.96 (2H, qu, H^o, J = 6.9 Hz), 2.17 (2H, td, H^p, ^3J = 6.9 Hz, ^4J = 2.5 Hz), 2.30 (1H, t, H^q, ^4J = 2.5 Hz), 3.92 (3H, s, $H^{f/i}$), 3.93 (3H, s, $H^{f/i}$), 4.06 (3H, s, H^c), 4.27 (2H, q, H^l, J = 7.1 Hz), 4.41 (2H, t, H^n, J = 6.9 Hz), 6.94 (2H, m, $H^{a/b/d/e/g/h/j/k}$), 6.99 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.9 Hz), 7.05 (1H, s, $H^{a/b/d/e/g/h/j/k}$), 7.22 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.3 Hz), 7.25 (1H, s, $H^{a/b/d/e/g/h/j/k}$), 7.31 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.5 Hz), 7.43 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.8 Hz); δ_{13C}(100 MHz, CD_3OD): 13.3, 14.8, 30.0, 34.4, 35.4, 59.6, 69.0, 104.5, 105.1, 109.6, 119.2, 119.5, 120.3, 121.9, 125.9, 127.0; ESI: MH^+ 573.2579 (calculated 573.2574)

Impurity 49

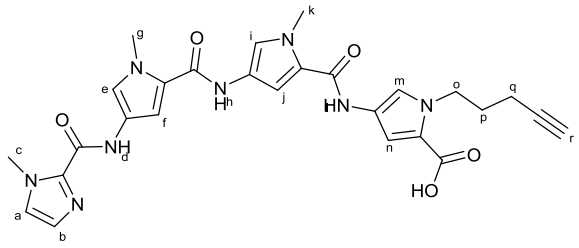


NMR: δ_{1H}(400 MHz, CD_3OD): 1.86 (2H, qu, H^l, J = 6.8 Hz), 2.07 (2H, td, H^m, ^3J = 6.8 Hz, ^4J = 2.6 Hz), 2.22 (1H, t, H^n, ^4J = 2.6 Hz), 3.82 (3H, s, $H^{c/f}$), 3.83 (3H, s, $H^{c/f}$), 4.17 (2H, q, H^i, J = 7.2 Hz), 4.31 (2H, t, H^k, J = 6.8 Hz), 6.92 (1H, br s, $H^{a/b/d/e/g/h}$), 6.93 (1H, d, $H^{a/b/d/e/g/h}$, J = 2.0 Hz), 6.98 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.4 Hz), 7.17 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.4 Hz), 7.19 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.5 Hz), 7.19 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.8 Hz), H^j obscured by residual NEt_3 from the column chromatography at 1.32 ppm; δ_{13C}(100 MHz, CD_3OD): 14.8, 16.3, 31.5, 33.7, 36.9, 37.0, 54.9, 61.1, 63.7, 70.6, 106.6, 111.1, 121.0, 121.2, 121.7; δ_{19F}(376.5 MHz, CD_3CN): -75.0; ESI: MH^+ 561.2058 (calculated 561.2073)

2.10.3.4 Synthesis of ImPyPyPy(///)OH (41)



To a solution of **48** (0.0253g, 4.4 μmol) in 1,4-dioxane (5 mL) was added aqueous lithium hydroxide (1 M, 2 mL) and the reaction heated at 40°C for 24 hours. The reation mixture was acidified using aqueous hydrochloric acid (1 M, 4 mL) and evapourated to dryness under reduced pressure. The crude product was diluted with hydrochloric acid (1M) and extracted three times with with *n*-butanol, then the organic fraction combined and the solvent removed under reduced pressure. HPLC purification (30-50% MeCN in H_2O over 20 minutes, 0.1% TFA) yielded the product which was then freeze dried to give a pale yellow solid (0.0201, 84%).



NMR: δ_{1H}(400 MHz, DMSO): 1.87 (2H, qu, H^p, J = 6.9 Hz), 2.11 (2H, td, H^q, ^{3}J = 6.9 Hz, ^{4}J = 2.5 Hz), 2.85 (1H, t, H^r, ^{4}J = 2.5 Hz), 3.85 (3H, s, $H^{g/k}$), 3.86 (3H, s, $H^{g/k}$), 4.00 (3H, s, H^c), 4.33 (2H, t, H^o, J = 6.9 Hz), 6.87 (1H, s, $H^{a/b/e/f/i/j/m/n}$), 7.05 (1H,s, $H^{a/b/e/f/i/j/m/n}$), 7.08 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.9 Hz), 7.17 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.9 Hz), 7.25 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.6 Hz), 7.31 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.6 Hz), 7.40 (1H, s, $H^{a/b/e/f/i/j/m/n}$), 7.43 (1H, s, $H^{a/b/e/f/i/j/m/n}$), 9.92 (1H, s, $H^{h/l}$), 9.98 (1H, s, $H^{h/l}$), 10.44 (1H, s, H^d), δ_{13C}(100 MHz, DMSO): 7.1, 30.1, 35.1, 36.0, 48.6, 62.5, 71.7, 83.5, 104.8, 104.9, 108.7, 116.1, 116.2, 118.5, 118.6, 119.0, 121.4, 122.2, 122.6, 123.0, 126.3, 127.0, 138.8, 156.1, 157.7, 158.0, 158.5; IR (neat) υ_{max}: 3435 (w), 3301 (w), 1662 (s), 1563 (m), 1433 (m), 1196 (s), 1130 (s), 841 (m), 801 (m), 724 (s); ESI: MH^+ 545.2252 (calculated 545.2261).

Chapter 3 Synthesis of polyamide fluorophore conjugates

3.1 Introduction, aims and objectives

With the establishment of a synthetic route to alkyne modified polyamides, a general method for the attachment of a 'payload' using this functional handle is a key part of achieving the proposed toolbox approach. Click chemistry is ideal for this purpose as this family of reactions is defined by their high yield, specificity and wide scope and would provide a general method of attaching a wide variety of functionality. Several different click type reactions using alkynes and azides were investigated, with particular focus on the copper catalysed click reaction due to the availability of a fluorogenic coumarin azide **16** which once reacted would allow direct substitution of a DNA labelled fluorophore in an existing photonic wire design (Table 6),[107, 219] providing a simple comparison between the two approaches, making this fluorophore polyamide conjugate a key objective.

3.2 Attachment Chemistry

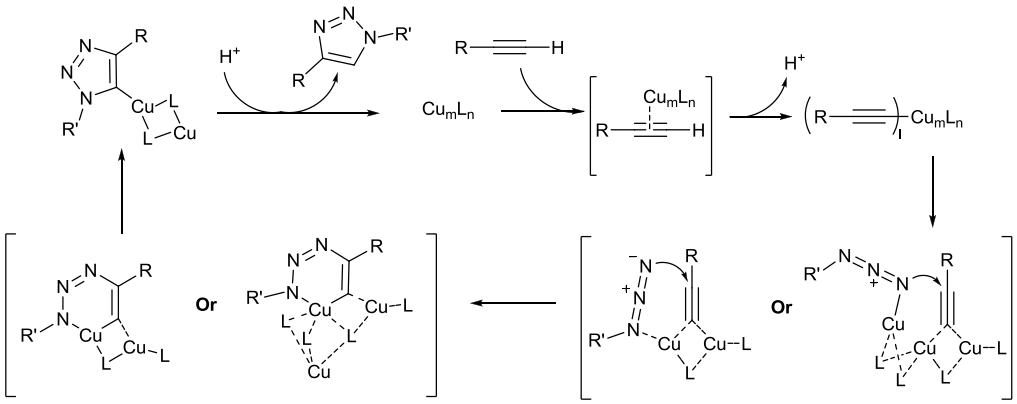
The selection of azide and alkyne functionality as chemical 'handles' allows the use of a range of different attachment chemistry in order to attach fluorophores and other functionality of interest simply and efficiently. The azide triple bond has the potential to release large amounts of energy but as it is kinetically slow to react with most chemical reagents organic azides are ideal starting materials for 'click' type reactions.[229] The term 'click reaction' was first coined[176] as a general term to describe reactions that were wide in scope, modular, high yielding, specific and had minimal by-

products. In addition, these reactions must also use readily available starting materials and operate under simple conditions in benign solvents with easy purification. These are all highly desirable properties for the 'toolbox' approach which would provide a much simpler method for the attachment of a payload to polyamides compared to existing work which would also have utility in wider applications. Several reactions were exemplified as having 'spring-loaded' components that remained inert until a specific set of conditions were met, but the term click chemistry is now almost exclusively used to describe the copper catalysed Huisgen 1,3-dipolar cycloaddition of azides and alkynes to afford 1,2,3-triazoles.

3.3 Applications of Copper Catalysed Click chemistry

The simplicity and versatility of the copper (I) catalysed Huisgen 1,3-dipolar cycloaddition of azides and alkynes means that it has now found a role not only in drug discovery[230, 231] as was originally envisioned[176] but also in a wide range of fields[232] such as polymer chemistry, dendrimer formation and post synthetic modification of DNA.[233-235] The copper catalyst allows organic azides to rapidly react with terminal alkynes to produce the five membered triazole ring regiospecifically, cleanly and in high yield. In addition the click reaction is tolerant of a wide range of functional groups, making it ideal chemistry for attaching a 'toolbox' of payloads. The previous work on the post synthetic functionalisation of DNA[233-235] also suggests the possibility of introducing functionality after the polyamide has bound to the DNA scaffold.

The detailed mechanism of the click reaction remains unclear and is still the subject of discussion[232, 236-238] primarily because of uncertainty over the nature of the copper catalyst which is suspected to contain multiple metal ions in solution, the nature of the interaction between the copper and the alkyne and conflicting kinetic data

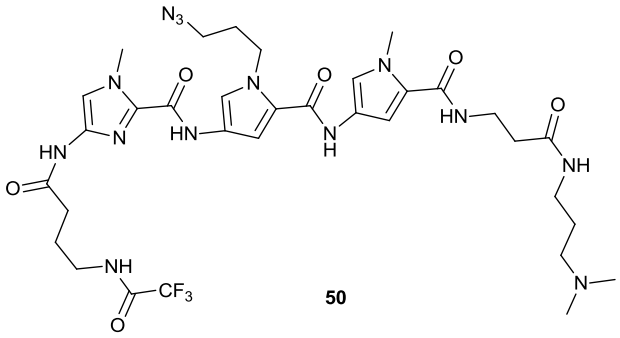


Scheme 19 *- Proposed mechanism of the copper catalysed click reaction*[232]

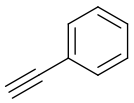
suggesting variable or second order dependence with respect to copper.[236] The first step of the generally accepted mechanism (Scheme 19) is the binding of the alkyne to the catalyst and adoption of an 'end-on' σ binding mode. Copper coordinated azide subsequently attacks at the C-2 position to form a 6 membered ring intermediate which contracts to form the triazole. Addition of a proton allows release of the products and regeneration of the catalyst. The ligand 'L' is normally the counter ion of the copper salt introduced into the reaction but several alternative polydentate ligands have been developed with the aim of accelerating the reaction, improving water solubility and preventing side reactions with sensitive substrates. The polydentate nature of these ligands might be thought to help simplify the species involved in the reaction mechanism, but similarly variable kinetics and complex inhibitory effects have also been observed.[239, 240]

3.3.1 Initial Click Chemistry Trials

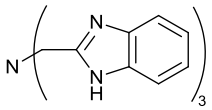
In order to test the compatibility of the click reaction with the functional groups present in a typical polyamide the 'half' polyamide **50** was synthesised by solid phase synthesis and reacted with phenyl acetylene **54**. It was feared that the presence of the **Im** building block which is likely to be protonated at neutral pH, might prevent the azide from approaching the cationic copper complex or inhibit the catalysis by binding more strongly than the azide or alkyne functionalities. A range of different conditions were tested in order to optimise the reaction with this in mind. The use of buffered conditions during the reaction was tested due to concerns over the pH of the resulting solution after the addition of the reducing agent. This was also tested through the use of both sodium ascorbate and the alternative reducing agent TCEP.HCl which has also been used during click reactions which give basic and acidic solutions respectively.



50



54



BIM

Reducing agent	Ligand	20 mM HEPES buffer				H_2O (unbuffered)			
		Click product	**50**	**54**	Other†	Click product	**50**	**54**	Other†
Sodium Ascorbate	BIM	**92**	**5**	**0**	**3**	**76**	**14**	**3**	**7**
	-	**90**	**0**	**4**	**6**	**86**	**0**	**6**	**7**
TCEP.HCl	BIM	**0**	**0**	**73**	**27**	**0**	**66**	**0**	**34**
	-	**0**	**90**	**5**	**5**	**0**	**97**	**3**	**0**

Table 8 *– Products of the click reaction between **50** and **54** (1 eq. $CuSO_4$ (1.48 mM), 1 eq. **50**, 1 eq. **54**, 2 eq. reducing agent, 1 eq. ligand, 70% EtOH, 30% H_2O, 40°C, 1 hour) as a proportion of the total integrated area of HPLC peaks detected at 254 nm eluting after 5 minutes. †- additional unidentified peaks eluting after 5 minutes*

Finally, the ligand BIM[239] was tested as a means to control the complexation and reactivity of the copper catalyst during the reaction.

Encouragingly, the click reaction proceeded cleanly and efficiently as indicated by HPLC under all of the conditions tested using sodium ascorbate as the reducing agent (Table 8), but no reaction at all was observed using TCEP.HCl. The relatively large quantities of 'other' peaks in the reaction products suggested that some degradation of the starting materials was taking place, particularly in the presence of the BIM ligand and for all further click chemistry tests, sodium ascorbate was used exclusively. The presence of the HEPES buffer was found to have a beneficial effect on both conversion and the purity of the products, indicating that pH does play an important role in this reaction. Interestingly, in contrast to the reactions using the hydrochloric acid salt of tris(2-carboxyethyl)phosphine (TCEP.HCl) the presence of the BIM ligand during the sodium ascorbate tests consistently increased the amount of azide starting material remaining after 1 hour and under unbuffered conditions appeared to slow down the reaction with a small quantity of both starting materials remaining. This is consistent with the ligand complexing with the copper which reduces its activity, but also minimising the known side reaction in which the azide starting material is reduced to the amine which would contribute to the impurity peaks observed by HPLC.

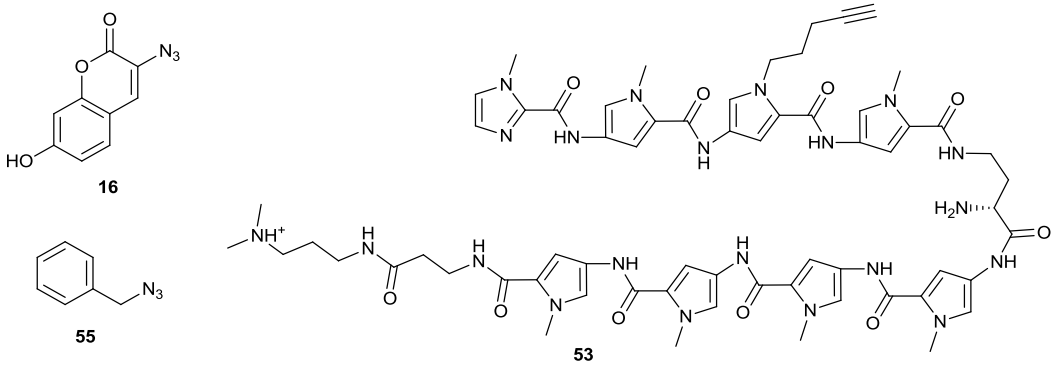
3.3.2 Click Chemistry on Full Length Polyamides

After the success of these trials, a number of trials were undertaken with the full length alkyne polyamide **53**[†] and coumarin azide **16** using these optimised conditions as a starting point (Table 9). Polyamide **53** contains the **Im** stopper **10** and a charge on the hairpin turn, common features of polyamides which were not present in the initial test

[†] Supplied by Dr Wu Su, University of Leicester

Cu source	Equiv. **16**	Equiv. Cu	Equiv. NaAsc	Organic co-solvent	% Organic co-solvent	[HEPES] (mM)
$CuSO_4$	1	0.1	0.2	EtOH	30	27
		1	2	EtOH	30	20
		16	32	EtOH	30	10
		16†	32	EtOH	30	28
	4	1	5	ButOH	50	57
	10	0.5	0	EtOH	33	62
CuOTf	4	1	5	ButOH	49	57
CuBr	10	0.5	0	EtOH	30	52
		1	0	EtOH	30	52
		2	0	EtOH	30	52
	5	5†	5	EtOH	62	0
		5†	5	EtOH	86	0

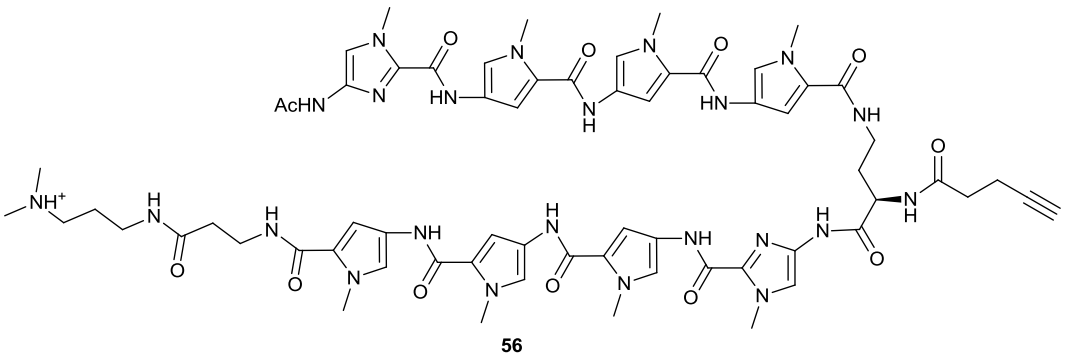
*Table 9 - Summary of conditions attempted for the click reaction between **16** and **53** at 40°C using the BIM ligand; † - no BIM ligand was used for these reactions; NaAsc- sodium ascorbate*



reactions, and the fluorogenic coumarin azide **16** was selected to allow fluorescence detection of the product. In all of the test reactions, analysis by HPLC showed no formation of any distinct product peaks, normally observed with complete consumption of the alkyne and the formation of a large number of inseparable smaller peaks containing the distinctive polyamide absorption at 310 nm. Variations in the

solvent, buffer, copper source, and the proportions of the reaction failed to improve this, suggesting that in contrast to the 'half' polyamide **50**, polyamide alkyne **53** is inherently unstable to click chemistry conditions and degrades into a variety of closely related compounds. Test reactions between phenyl acetylene and coumarin azide **16** using the previously optimised conditions yielded the correct click product by mass spectrometry with complete consumption of the alkyne by HPLC and comparable tests using pentyn-1-ol showed similar results, but the reaction of **53** with benzyl azide **55** yielded only the starting materials. The reason for this behaviour is unclear, but may be related to complexation of the alkyne during the click reaction (Scheme 19). Unlike the azide component, the alkyne becomes covalently bound to the copper catalyst and remains so for much of catalytic cycle, presenting much more opportunity for alternative ligand interactions to take place.

It is conceivable that a large number of weak interactions might take place with the amide bonds of the polyamides and the **Im** building block leading to a stable complex of several copper ions due to the chelate effect, but such a complex would be directly competing with the BIM ligand which is also polydentate and contains similar benzimidazole functionality. Similarly, if the formation of a complex is the cause of the degradation, this should also be observed to the same extent during the reaction with benzyl azide **55**. Since these trials were completed, one example of copper catalysed click chemistry on a turn modified alkyne polyamide **56** has been reported (template

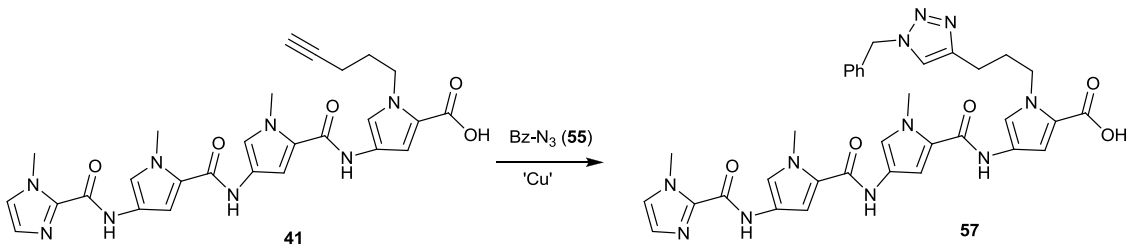


56

driven examples have previously been reported[241]) under similar conditions (1.5 equiv. azide, 1.1 equiv. $CuSO_4$, 2.1 equiv. sodium ascorbate, H_2O, [**56**] 159 mM) with a reported yield of 60%.[242] The close resemblance of these conditions to many of the trials undertaken suggests that the position of the modification may play an unexpectedly large role in determining the outcome of the reaction, allowing or preventing the formation of a complex with the copper catalyst. Further investigation is needed to establish the exact nature of these results, but the pressing need for the fluorophore-polyamide conjugate and the success of the click reaction on the 'half' polyamide meant that trials on the full length polyamide were dropped in favour of functionalising an alkyne polyamide 'half' which would then be coupled with a corresponding amine to produce the final product.

3.3.3 Click Chemistry on 'Half' Polyamides

After the unexpected failure of click chemistry on the full length polyamide **53**, the conditions were more thoroughly optimised for the alkyne 'half' polyamide **41** using benzyl azide **55** with variations in reaction temperature, catalyst, ligand and the reaction solvent (Table 10). In all of the trials, no benzyl azide **55** was observed after 1 hour when 1 equivalent of ligand was used for each equivalent of copper catalyst except where piperidine was used as ligand. The trials show the solvent plays a complex role in this reaction, with some conversions improving with increased organic solvent and in other cases decreasing after the same substitution. Inspection of the conversion data



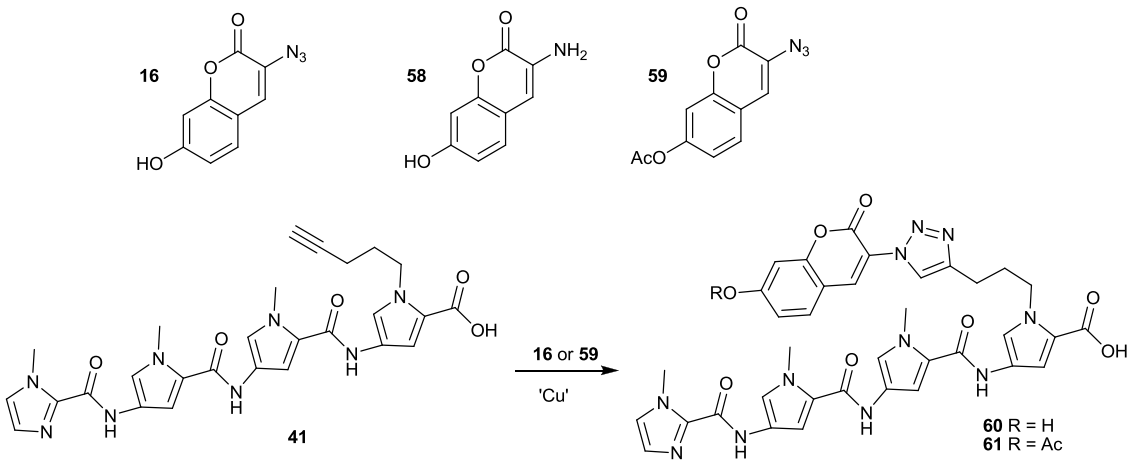
reveals several key factors, most notably that higher conversions were obtained in the absence of the BIM ligand (Table 10, entries 5 vs. 7; 6 vs. 8), which is consistent with the slower reaction observed during the initial click chemistry trials (**Table 8**), but surprising given the lack of benzyl azide starting material, the reduced form of which would be expected to contribute towards the impurity 'other' peaks. A similar contradiction with the same trials is also observed for buffered reactions, where the buffer causes a decrease in conversion and an increase in impurities rather than an

Entry No.	Stoichiometry					Solvent			Products		
	'Cu'	55	Equiv. Cu	Equiv. NaAsc	Ligand	EtOH	DMSO	H_2O	57	41	other†
1	CuBr	1	1	0	BIM	30	0	70‡	**0**	**79**	**21**
2		1	1	0	BIM	100	0	0	**0**	**75**	**25**
3	$CuSO_4$	1	0.2	1	pip††	94	0	6	**6**	**85**	**9**
4		1	1	2	BIM	30	0	70‡	**19**	**69**	**12**
5		1	1	2	BIM	30	0	70	**28**	**72**	**0**
6		1	1	2	BIM	5	25	70	**28**	**51**	**21**
7		1	1	2	-	30	0	70	**60**	**40**	**0**
8		1	1	2	-	0	30	70	**37**	**63**	**0**
9		1	1	5	BIM	30	0	70	**42**	**58**	**0**
10		1	1	5	BIM	5	25	70	**52**	**48**	**0**
11	$CuSO_4$	2	1	2	-	30	0	70	**75**	**25**	**0**
12		2	1	2	-	30	0	70	**0**‡‡	**100**	**0**
13		2	1	4	-	30	0	70	**92**	**8**	**0**
14		2	1	4	-	30	0	70	**0**‡‡	**100**	**0**
15		2	2	5	BIM	0	54	46	**4**	**48**	**48**

Table 10 *- Products of the click reaction between **55** and **41** (40°C, 1 hour, [**41**] 1.27 mM) as a proportion of the total integrated area of HPLC peaks detected at 254 nm eluting after 5 minutes. †- additional unidentified peaks eluting after 5 minutes; ‡ - Reactions carried out with 50 mM HEPES buffer;†† - pip – 5% by volume piperidine, [**41**] 16.2 mM; ‡‡ - Reactions carried out at 60°C*

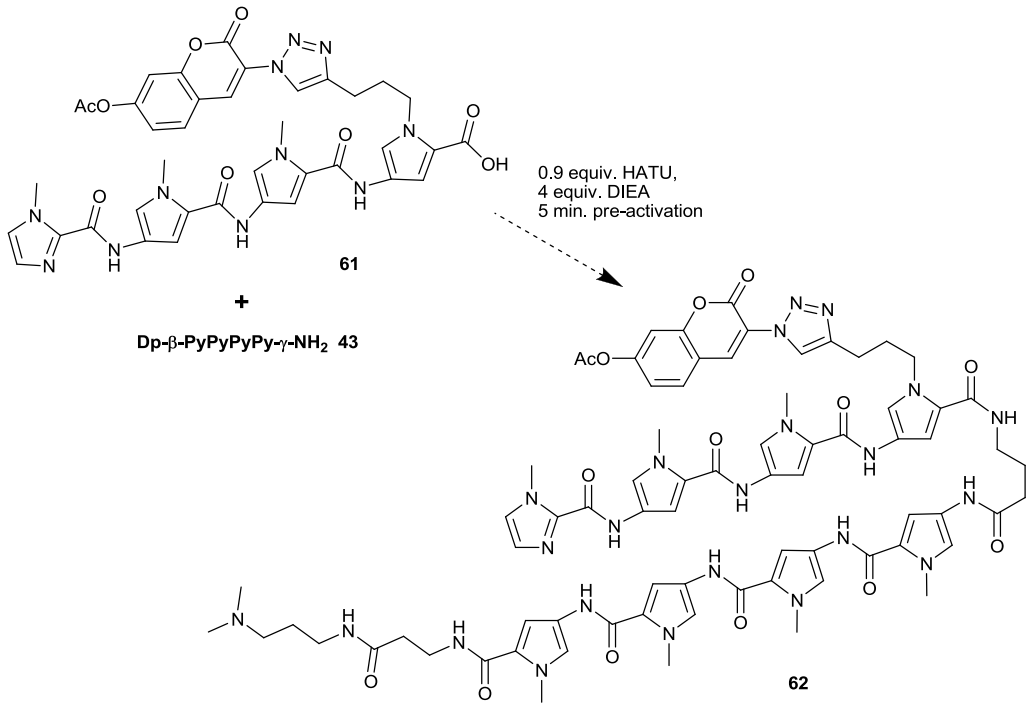
improvement (Table 10 entry 4 vs. 5; cf. Table 8 row 1 vs. 2).

The identity and stoichiometry of the copper catalyst also appears to be a key determinant of the reaction, with no conversion observed for copper (I) bromide (Table 10 entries 1, 2) even in completely organic solvent and greatly reduced conversion upon the addition of a second equivalent of copper (II) sulphate with a significant increase in impurities (Table 10 entry 15 vs. entries 9-14). Interestingly, the addition of a second equivalent of the azide does improve conversion of the starting material to the product, suggesting that the benzyl azide **55** is being consumed in a side reaction which is not being observed by HPLC after the sampling procedure. The presence of side reactions is also suggested by the improvement in conversion after the addition of extra sodium ascorbate which is used to regenerate the oxidised copper catalyst. The complete failure of the reaction upon additional heating without any conversion to impurities suggests that these side reactions increase dramatically at higher temperatures but are not responsible for the formation of the by-products observed at lower temperatures. Using these optimised conditions, the click reaction was trialled using the coumarin azide **16** (Scheme 20) which yielded a major product (77% conversion by HPLC) with the expected UV absorption at 340 nm along with the alkyne starting material **41** (13%)



Scheme 20 *- Trial Click reactions between alkyne 'half' polyamide* ***51*** *and coumarin azides* ***16*** *and* ***54***

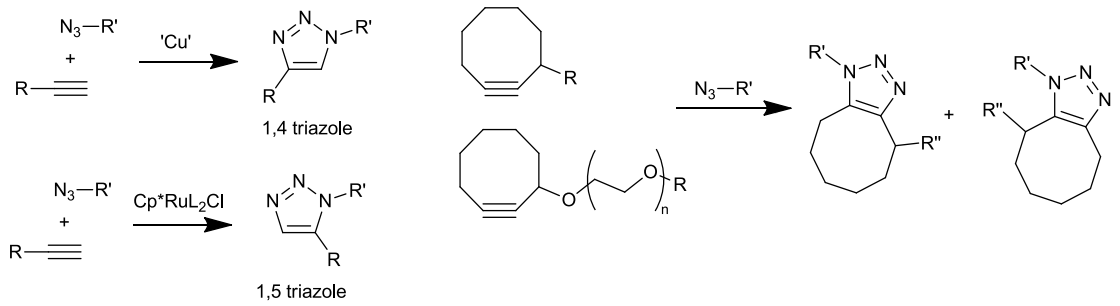
and the reduced coumarin **58** (10%). The presence of the hydroxyl group on the coumarin fluorophore is incompatible with the HATU coupling of the two polyamide fragments (Scheme 18) and further trials were undertaken using the *O*-acylated coumarin azide **59**. As expected, under the same conditions a major peak was detected by HPLC with an altered UV profile (λ_{max} = 327 nm) and the reaction scaled up. Attempts to characterise the product by electrospray and MALDI mass spectrometry failed to record the correct molecular ion even in the negative spectrum and it was decided to continue the synthesis using product **61** which was thought to be not susceptible enough to ionisation to be detected and that characterisation would be possible after the introduction of the cationic polyamide tail. Using previously optimised conditions (Scheme 18) **61** was reacted with polyamide 'half' **43** and analysed by mass spectrometry for the expected product **62** (Scheme 21), but this could not be observed by any technique and it was concluded that the click reaction between



Scheme 21 *– Attempted 'Half and half' polyamide coupling between the expected Coumarin modified acid* ***61*** *and amine* ***43*** *to produce full length polyamide* ***62***

coumarin azides and alkyne polyamides had in fact failed and that this route was not viable. The copper catalysed click reaction has been exemplified as an 'ideal' reaction in which the reagents react quickly and cleanly to give the desired product,[176, 243] but these studies show that this not always the case. Like many catalytic reactions, a wide variety of factors can affect the outcome, not all of them predictable or readily rationalised. Although not as widely applied, other click type reactions have been developed which could also be applied to functionalise polyamides.

3.3.4 Alternatives to Copper Based Click Reactions



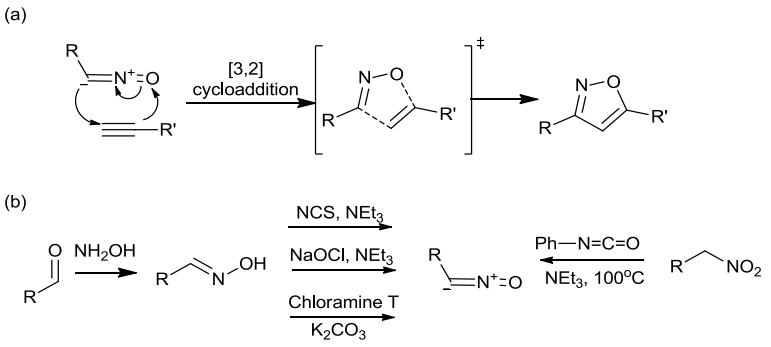
Scheme 22 *- Alternative click type reactions*

A number of alternatives to the copper catalysed click reaction have been reported with distinctive properties which make them preferable in some applications (Scheme 22). It has been shown that the 1,5-triazole derivative can be produced using ruthenium catalysis[244, 245] under harsher conditions but as with the copper catalysed reaction, the heavy metal catalysts used are too toxic for *in vivo* applications and strain promoted variations have been developed for biological applications with living cells. Introduction of a triple bond into a cyclooctane ring causes ring strain which increases the energy of the molecule by 42 kJmol^{-1} compared to the linear octyne.[246] This strain is not present in the transition state, effectively reducing the energy barrier for the reaction (estimated as 109 kJmol^{-1} for the uncatalysed reaction[247]) and allowing the

normally very slow thermal cycloaddition to take place over the course of 12 hours.[248] Several variations on this method were introduced to further improve the kinetics such as the introduction of electron withdrawing groups[249, 250] and additional ring strain,[251] but the formation of the strained ring is synthetically very challenging and is characterised by low yields. The use of a polyethylene glycol (PEG) linker greatly simplified the synthesis with some loss in kinetics,[252] but as with the other metal free click reactions, the products show the very limited regioselectivity observed for the uncatalysed reaction and a mixture of the 1,5 derivative along with the 1,4 isomer is formed. In biological tagging applications, this mixture of products is not detrimental to the final outcome, but where short linkers are being used to control the position of the payload the 1,5 regioisomer would direct the fluorophore back towards the DNA, causing steric clashes which could reduce the binding affinity and it is highly likely that quenching of an attached fluorophore would be increased. These problems mean that none of these methods is suitable for the modifying polyamides for a photonic wire, but other uncatalysed click type reactions have been reported using alkyne substrates.

3.3.5 Isoxazole Click chemistry

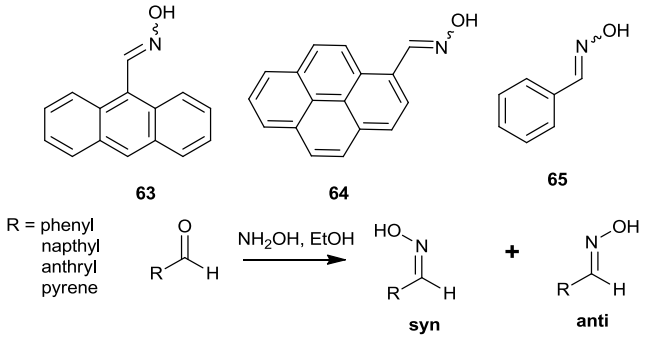
One example of these uncatalysed click reactions is the [3,2] cycloaddition between a terminal alkyne and a nitrile oxide to produce isoxazoles (Scheme 23a). Copper[247] and ruthenium[253] catalysed reactions have been reported that selectively produce the 1,4 and 1,5 isomers respectively but in contrast to the click reaction, the uncatalysed reaction occurs at an appreciable rate and has a high degree of regioselectivity for the 1,4 isomer.[254-256] The initial formation of the nitrile oxide is the key stage of this reaction as it is highly reactive and is usually formed in-situ to minimise side reactions such as the formation of dimers. Synthesis from the nitroalkane



Scheme 23 *- Synthesis of isoxazoles by a click type reaction: (a) uncatalysed reaction mechanism; (b) generation of the nitrile oxide precursers*

has been reported[180, 256] but chlorination of the oxime followed by elimination of HCl is more widely used[253-257] because of the milder conditions that can be employed (Scheme 23b). The oxidising agent chloramine-T is mild enough to be compatible with a wide range of functional groups and can be used in a water-ethanol mixture ideally suited for solubilising both hydrophilic polyamides and the more hydrophobic payloads.[254]

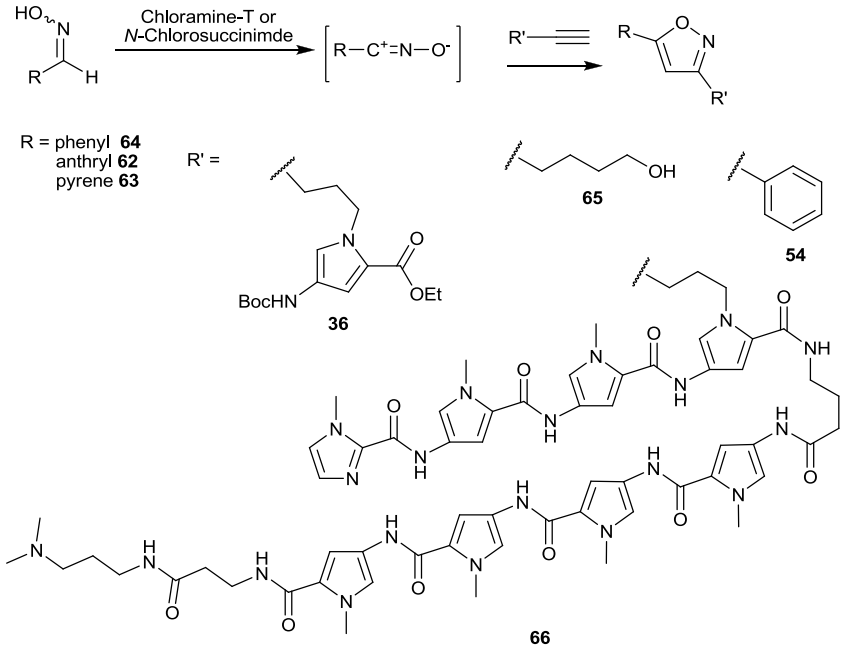
A fluorescent naphthyl derivative has been produced using chloroamine-T methodology[254, 258] and as this type of functionality is unreactive and simple to monitor during the reaction, the appropriate anthryl **63** and pyrene oxime **64** precursors were synthesised using NH_2OH in ethanol (Scheme 24).[259] This reaction can give rise to a mixture of *syn* and *anti* oxime isomers, which were observable for both **63** (1.5:1, 49% yield) and **64** (11:1, 17% yield). Recrystallisation of **63** from ethanol yielded a single compound corresponding to the majority of the crude product and was tentatively assigned as the *syn* isomer after NMR experiments showed no NOE cross peaks were



Scheme 24 *- Synthesis of syn and anti oximes **63** and **64** using NH_2OH*

observed between the hydroxyl and oxime C-H protons. Similarly, NMR experiments on **64** showed no NOE cross peaks between the hydroxyl and oxime C-H protons of the major isomer, also allowing it to be tentatively assigned as the *syn* isomer. Previous literature reports on this reaction achieved similar results and as the mixture of isomers has been shown to react satisfactorily to form isoxazoles,[255, 258, 260] the crude reaction product was used in all cases without further purification.

Test reactions using the commercially available benzyl oxime **65**, naphyl oxime **63** and pyrene oxime **64** with the alkynes hexyn-1-ol **66**, phenyl acetylene **5**, alkyne pyrrole intermediate **36** and polyamide **66**‡ (Scheme 25) based on the chloroamine-T literature protocol[258] all yielded a large number of unidentifiable peaks despite variations in temperature, solvent, the identity and equivalents of oxime, the presence of additional base and altering the oxidising agent (Table 11). As a result, this chemistry was not pursued any further and after the failure of multiple click type reactions, previously reported methods for tail modification were investigated as a means to



Scheme 25 *- Trial isoxazole click reactions between oximes and alkynes*

‡ Synthesised by Dr Wu Su

produce the desired fluorophore-polyamide conjugates.[164-167, 219]

Alkyne		Oxime		Oxidising agent		Solvent			Notes
	[alkyne] mM		Equiv.		Equiv.	EtOH	H_2O	Sat $NaHCO_3$	
66	54	**65**	1	Chl-T	1	80	-	20[†]	
66	54	**65**	2.5	Chl-T	6	80	-	20[†]	
66	0.81	**65**	2.5	Chl-T	6	80	-	20	
66	0.81	**65**	2.5	NCS	6	80	-	20	
66	8.4	**65**	2.5	Chl-T	6	80	-	20	
66	0.84	**65**	2.5	Chl-T	6	80	-	20	
66	0.57	**65**	5	Chl-T	7.5	72	28	-	
66	0.68	**64**	2.5	Chl-T	6	80	-	20	
66	0.68	**64**	2.5	Chl-T	6	80	-	20	‡
66	0.68	**64**	2.5	Chl-T	6	80	-	20	††
66	54	**63**	1	Chl-T	1	80	-	20	
66	54	**63**	2.5	Chl-T	6	80	-	20	
67	2.57	**64**	2.5	Chl-T	6	82	-	18	
67	3.9	**64**	5	Chl-T	7.5	85	15	-	
36	0.57	**65**	5	Chl-T	7.5	72	28	-	
66	0.57	**65**	5	Chl-T	7.5	72	28	-	40°C
66	0.57	**64**	5	Chl-T	7.5	72	28	-	40°C
66	0.84	**65**	2.5	Chl-T	6	80	-	20	40°C
66	8.4	**65**	2.5	Chl-T	6	80	-	20	40°C
54	8.4	**65**	2.5	Chl-T	6	80	-	20	40°C
54	8.4	**65**	2.5	Chl-T	6	50	50	-	40°C
54	8.4	**65**	2.5	Chl-T	6	50	50	-	40°C

Table 11 *– Summary of conditions and substrates attempted for the isoxazole forming click chemistry reaction (Scheme 25). Chl-T – Chloramine T; NCS – N-chlorosuccinimide*

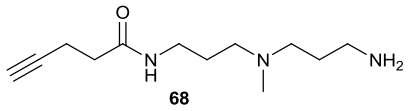
[†] *- Reactions using 4% w/v $NaHCO_{3\,(aq)}$ (cf. ~9.6% w/v saturated $NaHCO_{3\,(aq)}$);*

[‡] *- Reaction carried out with the addition of 1 equiv. N-methyl imidazole;*

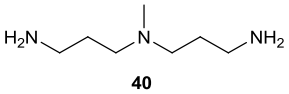
[††] *- Reaction carried out with the addition of 1 equiv. 3,3'-Diamino-N-methyldipropylamine* ***40***

3.4 Synthesis of polyamide fluorophore conjugates

3.4.1 Polyamide modification using copper catalysed Click chemistry

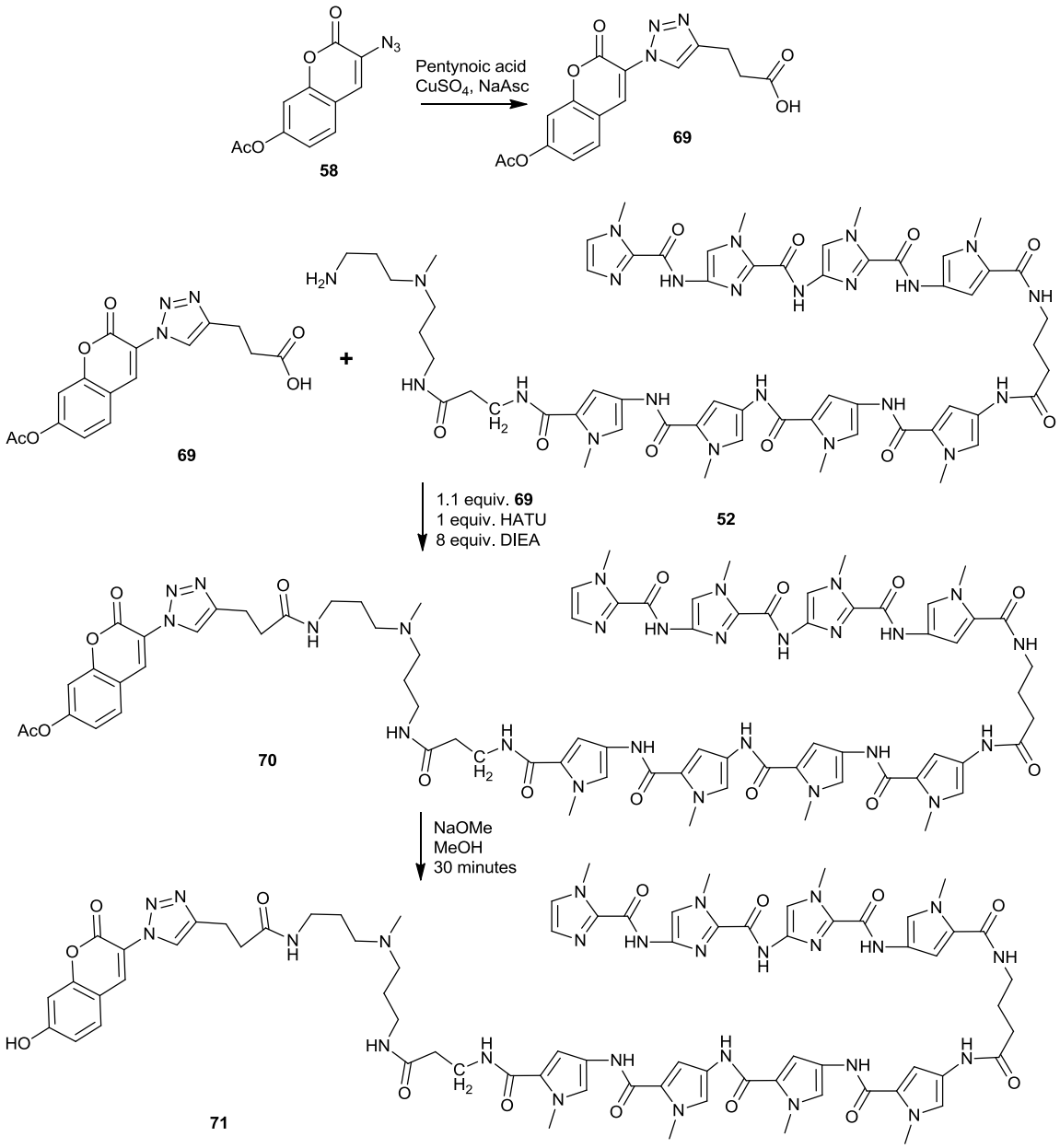


68



40

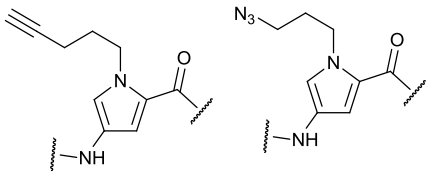
Tail modification is a well established method of introducing functionality into polyamides and numerous fluorophore-polyamide conjugates of this type have been reported.[165, 166] Although it is possible to cleave the completed polyamide from solid phase synthesis resin using a modified amine such as **68**, a large excess is required and the modification is subjected to an extended period of heating under basic conditions making it unsuitable for expensive or sensitive functionality. An alternative method is to cleave from the resin the polyamide using the diamine **40**, purifying the polyamide and then attaching the desired modification using standard amide coupling chemistry (Scheme 26). This approach to polyamide tail functionalisation has been widely reported and literature protocols[219] were used to couple the amine functionalised polyamide **52** with the protected coumarin acid **69** formed from the click reaction of the protected coumarin azide **59** with pentynoic acid. Unlike the problematic click reactions on full length polyamides (Table 9), this click reaction proceeded using only 0.2 equivalents of copper and a total of 1 equivalent of sodium ascorbate, which after acidic extractions to remove the copper catalyst and HPLC purification, gave **69** in 28% yield. The coupling of acid **69** and polyamide amine **52** (Scheme 26) was monitored by MALDI mass spectrometry and HPLC, and once complete immediately followed by removal of the coumarin protecting group using sodium methoxide in methanol before HPLC purification to give the fluorescently active polyamide-coumarin conjugate **71** in 17% overall yield (**52** to **71**).



Scheme 26 *- Synthesis of Coumarin modified polyamides using tail functionalisation*

3.5 Overview of the production of polyamide conjugates

Initially it was hoped that functionalised building blocks such as **Py(///)** and **Py(N_3)** would open up a new method for creating internally modified polyamides with greater control over the positioning of the 'payload' and easier access to a wide range of modifications, but the failure of click chemistry on full length polyamides represents a severe set-back to this approach. Efficient attachment chemistry is essential in library synthesis and copper catalysed click chemistry was selected due to its wide utility, high



Py(///) Py(N_3)

yield and tolerance of functional groups, qualities which few other reactions share.[176] The generic route developed to the modified building blocks (Scheme 6) should however allow rapid access to a variety of internal modifications once more suitable coupling chemistry can be identified. Despite these setbacks, a coumarin functionalised polyamide **71** has been produced that is suitable for inclusion into an improved photonic wire design and can now be tested to determine its suitability as a replacement for DNA end labelling.

3.6 Experimental

Some experimental procedures and additional characterisation data were added between the first submission in July 2011 and final submission. These additions are collected in Appendix 1.

All chemicals and solvents were purchased from Acros, Alfa Aesar, Fisher, or Sigma Aldrich unless otherwise stated and were used as received. Oxygen and moisture sensitive substances and reactions were carried out under a nitrogen or argon atmosphere as stated in the experimental in pre-dried glassware using Schlenk techniques. Reagents and solvents for solid phase synthesis were biotech or HPLC grade. Analytical TLC was carried out using Polygram 0.20mm silica gel plates with detection by UV at 254 nm and visualisation using PMA or ninhydrin. Preparative column chromatography was performed using Merk Kieselgel $60_{(230\text{-}400\ mesh)}$ silica gel. ^{1}H, ^{31}P and ^{13}C spectra were recorded on a *Bruker DPX300, Bruker DRX 400* or *Bruker AV 500* spectrometer as stated. Chemical shifts reported are in parts per million (ppm, δ). *J*-coupling constants are given in Hertz (Hz) and signal multiplicities are abbreviated as follows: s (singlet), d (doublet), t (triplet), q (quartet), qu (quintet), m (multiplet) and br for broad signals. IR spectra were recorded on a neat sample using a Perkin Elmer FT-IR with ATR attachment. ESI mass spectra were recorded on a micromass Quattra LC spectrometer. MALDI-TOF accurate mass spectra were recorded on a Voyger-DE STR using α-cyano-4-hydroxycinnamic acid as the matrix. LC-MS spectra and accurate ESI mass spectra were recorded using a Xevo QTof mass spectrometer (Waters) coupled to an Acquity LC system (Waters) using an Acquity UPLC BEH C18 column (2.1 x 50 mm, Waters). The flow rate was 0.6 ml min^{-1} and the gradient was as follows: 95% Solvent A (0.1% formic acid in water) with 5% solvent B (0.1% formic acid in acetonitrile) was held constant for 0.5 min, then there was a linear gradient to 100% B

over the next 2.1 min. After 1 min at 100% solvent B, the gradient was returned to 95% solvent A and 5% solvent B over 0.2 min. Mass accuracy was achieved using a reference lock mass scan, once per second. The ES cone voltage was 30 V and collision energy was 4 eV. The MS acquisition rate was 10 spectra per second and m/z data ranging from 50 to 2000 Da was collected. Analytical and semipreparative RP-HPLC was performed at room temperature on the ULTIMAT 3000 Instrument (DIONEX). UV absorbance was measured using a photodiode array detector at 260 and 310nm. An ACE C18 column (4.6 X 250 mm, 5 μm, 300 Å) was used for analytical HPLC and an ACE C18 column (10 X 250 mm, 5 μm, 300 Å) for semipreparetive HPLC.

3.6.1 Click chemistry trials (Table 8, Table 9, Table 10)

3.6.1.1 General Procedure for Click Chemistry Trials Using $CuSO_4$

To a solution of the required azide was added a solution of the appropriate alkyne and $CuSO_{4(aq)}$. Immediately afterwards, the ligand solution was added and the mixture diluted to the desired concentration using solvent or a HEPES buffer solution. Where any component was not used, the appropriate volume of additional solvent was also added. Sodium ascorbate or TCEP.HCl was quickly added and the reaction stirred at 40^{o}C for 1 hour before sampling.

3.6.1.2 General Procedure for Click Chemistry Trials Using Cu(I) Catalysts

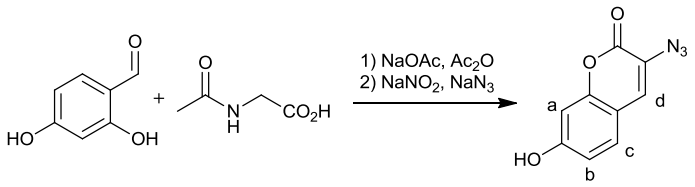
To a solution of the required azide was added a solution of the appropriate alkyne and the mixture diluted to the desired concentration using solvent or a HEPES buffer solution. Where any component was not used, the appropriate volume of additional solvent was also added. In a separate flask, a solution of the appropriate ligand was added to the solid copper catalyst and the solution sonicated and warmed until the

catalyst had completely dissolved. The appropriate volumes of the two solutions were then mixed and the reaction stirred at 40^{o}C for 1 hour before sampling.

3.6.1.3 Click Chemistry Trials Sampling Procedure

The crude reaction mixture was sonicated to suspend any precipitate and a 15 μL sample removed. This was then diluted with 135 μL of HPLC grade methanol and after mixing centrifuged for 30s at 5400 rpm to allow direct analysis of the supernatant by HPLC. HPLC analysis took place using a 15 minute 20-50% gradient of acetonitrile in water with 0.1% TFA and peak analysis was performed at 254 nm using the built in integration function with peaks identified before the start of the gradient at 5 minutes manually removed to eliminate injection and DMSO solvent peaks.

3.6.1.4 Synthesis of Coumarin azide (3-azido-7-hydroxy-2*H*-chromen-2-one) 16[177]



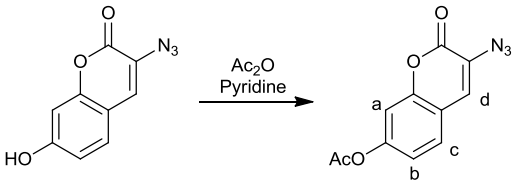
A solution of 2,4-dihydroxybenzaldehyde (2.7626 g, 20 mmol), N-acetyl glycine (2.3421 g, 20 mmol) and anhydrous sodium acetate (4.943 g, 60 mmol) in acetic anhydride (100 mL) was refluxed for 18 hours, then allowed to cool. The reaction mixture was poured onto ice and the precipitate collected, then refluxed in a 2:1 mixture of concentrated hydrochloric acid (10.2 M) and ethanol (30 mL total) for 1 hour. Ice water (40 mL) was added and the mixture cooled in an ice bath before sodium nitrIte (2.7789 g, 40 mmol) was added and the reaction stirred for 10 minutes. Sodium azide (1.3431 g, 20 mmol) was added to the solution, then stirred for 15 minutes. The precipitate was collected and washed with water, then freeze dried to recover the title

compound as a brown solid (0.3644 g, 9%) broadly consistent with previously reported characterisation.[177, 261, 262] Mp. >320°C (lit. 118-120°C[262])

NMR: δ_{1H}(400 MHz, DMSO): 6.77 (1H, d, H^d, J = 2.3 Hz), 6.81 (1H, dd, H^c, J = 8.5 Hz, J = 2.3 Hz), 6.49 (1H, d, H^b, J = 8.5 Hz), 7.61 (1H, s, H^a), 10.52 (1H, s, OH); δ_{13C}(100 MHz, DMSO): 102.0, 111.2, 113.8, 121.0, 127.8, 129.0, 152.7, 157.3, 160.4;

IR (neat) υ_{max}: 3048 (br, w), 2115 (s), 1706 (s), 1679 (m), 1614 (s), 1318 (m), 1259 (m), 1220 (s), 1025 (m), 996 (m), 835 (s).

3.6.1.5 Synthesis of Acylated Coumarin azide 59[177]

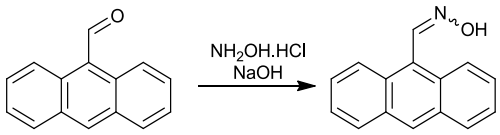


To a solution of coumarin azide **16** (0.0198 g, 0.15 mmol) in acetic anhydride (0.5 mL, 5.3 mmol) was added pyridine (0.4 mL, 4.9 mmol) (0.1227 g, 0.89 mmol) and the reaction stirred for 16 hours. The reaction mixture was diluted with EtOAc and the organic fraction washed with saturated $NaHCO_3$ solution, $HCl_{(aq)}$ (0.1 M) and brine. The organic solvent was dried with $MgSO_4$ and the solvent removed under reduced pressure to yield the title compound as a dark brown solid (0.003 g, 126%) consistent with previously reported characterisation.[177] Mp. 120-122°C

NMR: δ_{1H}(400 MHz, DMSO): 2.30 (3H, s, O**Ac**), 7.17 (1H, dd, H^c, J = 8.5 Hz, J = 2.2 Hz), 7.33 (1H, d, H^d, J = 2.2 Hz), 7.69 (1H, s), 7.71 (1H, s); δ_{13C}(100 MHz, DMSO): 20.8, 109.9, 117.1, 119.2, 125.0, 126.1, 128.5, 151.2, 151.6, 156.8, 168.8;

IR (neat) υ_{max}: 2125 (m), 1752 (m), 1722 (s), 1619 (m), 1430 (m), 1367 (m), 1328 (m), 1213 (s), 1142 (m), 1120 (m), 918 (s), 877 (m), 758 (s).

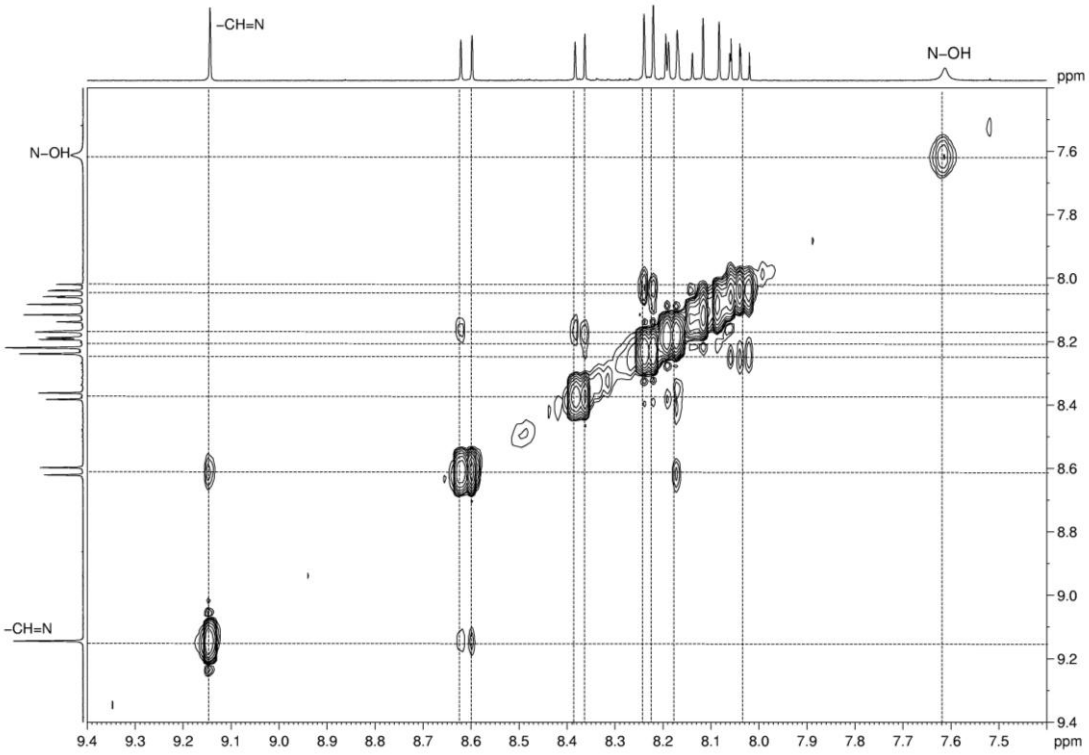
3.6.1.6 Synthesis of 1-anthraldehyde oxime (63)[263]



To mixture of 1- anthraldehyde (0.5008g, 2.42 mmol) and $NH_2OH.HCl$ (0.5627 g, 8.08 mmol) in ethanol (15 mL) was slowly added powdered NaOH (0.8746 g, 21.8 mmol) and the reaction stirred for 30 minutes at room temperature before refluxing for a further 30 minutes. After cooling, the crude reaction mixture was added to a 4:1 mixture (10 mL) of H_2O and concentrated $HCl_{(aq)}$ and the ethanol removed under reduced pressure. The remaining aqueous solution was extracted with DCM and the organic phase washed with brine before drying with $MgSO_4$ and removal of the solvent under reduced pressure. The crude product recovered as bright yellow powder consisting of a mixture of syn and anti isomers (0.2522 g, 47%, 1:1.5) which on recrystallisation from ethanol was recovered as pale tan needles containing a single compound (major isomer) tentatively assigned as the *syn* isomer due to the absence of NOE cross peaks between the hydroxyl group and the oxime C-H. This compound was found to be consistent with previously reported characterisation.[264]

***Syn* isomer:** NMR: δ_{1H}(400 MHz, DMSO): 7.53-7.62 (4H, m), 8.14 (2 H, d, J = 7.6 Hz), 8.45 (2 H, d, J = 8.8 Hz), 8.68 (1H, s), 9.21 (1H, s), 11.71 (1H, s). δ_{1H}(400 MHz, $CDCl_3$): 7.61 (1H, br s, N-O**H**), 8.01-8.25 (7H, m), 8.37 (1H, d, J = 8.1 Hz), 8.61 (1H, d, J = 9.3 Hz, R-C**H**=NOH); δ_{13C}(100 MHz, $CDCl_3$): 122.8, 125.0, 125.2, 125.7, 125.9, 126.2, 127.4, 128.4, 128.7, 149.7; IR (neat) υ_{max}: 3205 (br m), 1920 (br w), 1598 (m), 1584 (w), 1460 (m), 1343 (m) 1299 (m), 1237 (m), 1186 (s), 1056 (m), 963 9s), 928 (m), 899 (m), 834 (s) 818 (s), 790 (m), 749 (m), 711 (s), 674 (m); ESI: MH^+ 222.0921 (calculated 222.0919);

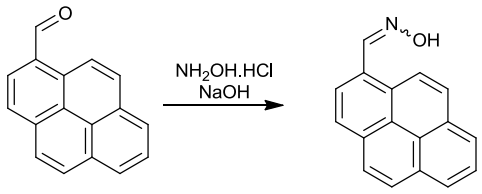
***Syn* NOE Enhancements:**



***Anti* isomer:**

NMR: δ_{1H}(400 MHz, DMSO): 7.71 (2H, t, J = 7.7 Hz), 7.84-7.90 (2H, m), 8.28-8.33 (4H, m), 9.08 (1H, s).

3.6.1.7 Synthesis of pyrene-1-carbaldehyde oxime (64)[259]

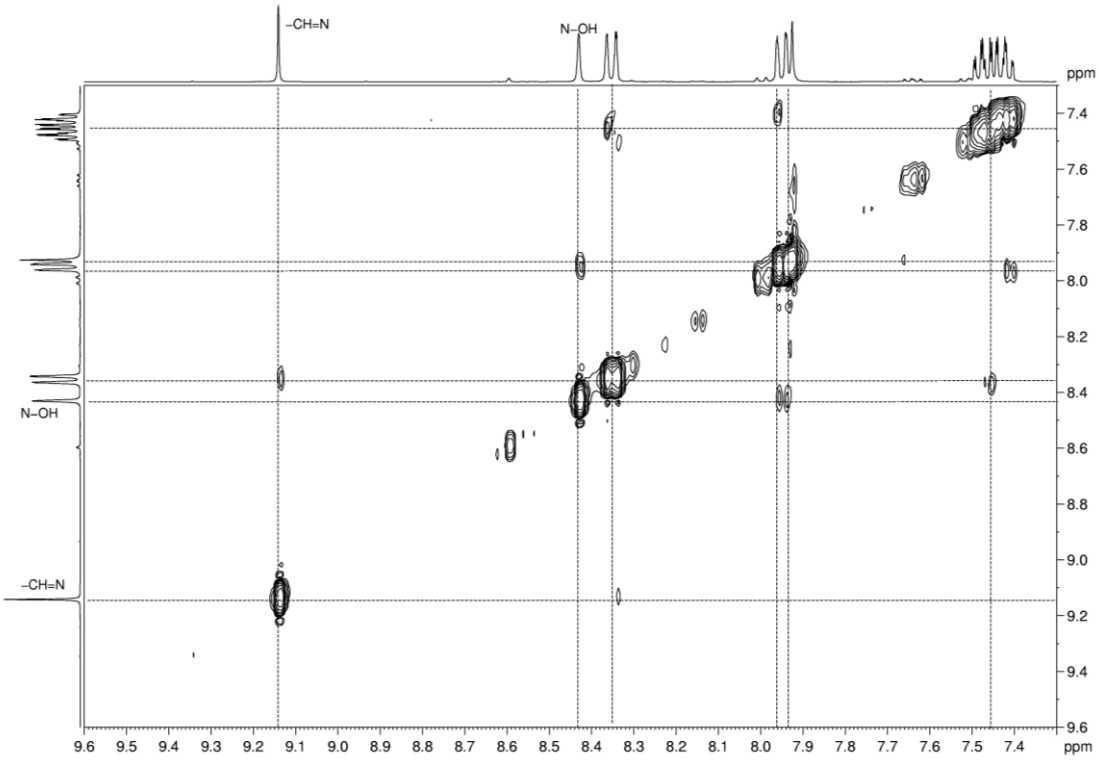


To mixture of 1-pyrenealdehyde (0.2502 g, 1.09 mmol) and $NH_2OH.HCl$ (0.2563 g, 3.62 mmol) in ethanol (8 mL) was slowly added powdered NaOH (0.3915 g, 9.77 mmol) and the reaction stirred for 30 minutes at room temperature before refluxing for a further 30 minutes. After cooling, the crude reaction mixture was added to a 4:1 mixture (5 mL) of H_2O and concentrated $HCl_{(aq)}$ and the ethanol removed under reduced

pressure. The remaining aqueous solution was extracted with DCM and the organic phase washed with brine before drying with $MgSO_4$ and removal of the solvent under reduced pressure to yield the title compound as bright yellow powder consisting of an 11:1 mixture of isomers (0.046 g, 17%). The major isomer was tentatively assigned as *syn* due to the absence of NOE interactions between the hydroxyl group and the oxime C-H (NOESY NMR experiment shown below) and found to be consistent with previously reported characterisation.[260]

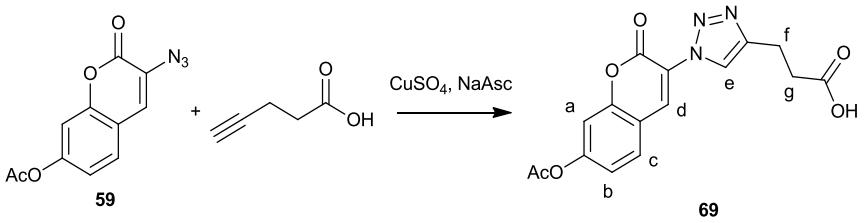
***Syn* isomer:** NMR: δ_{1H}(400 MHz, $CDCl_3$): 7.40-7.50 (4H, m), 7.91-7.97 (3H, m), 8.33-8.37 (2H, m), 8.43 (1H, s), 9.14 (1H, s); δ_{13C}(100 MHz, $CDCl_3$): 123.7, 125.0, 125.4, 126.4, 126.8, 128.8, 128.9, 129.0, 129.4, 130.3, 131.3, 149.0; IR (neat) υ_{max}: 3274 (br m), 1625 (w), 1522 (w), 1442 (m), 1299 (w), 1257 (w), 1179 (w), 1159 (w), 1071 (w), 975 (s), 939 (m), 914 (m), 885 (m), 876 (s), 842 (m), 783 (m), 728 (s), 659 (m);

***Syn* NOE Enhancements:**



3.6.2 Tail modified Polyamides

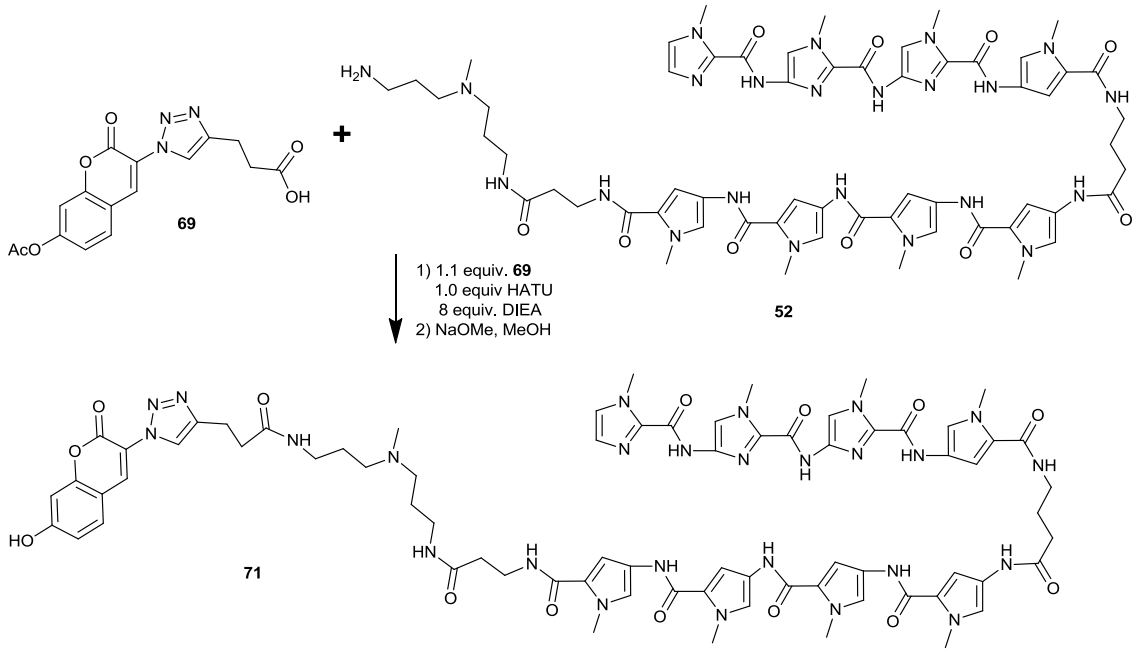
3.6.2.1 Synthesis of 3-(1-(7-acetoxy-2-oxo-*2H*-chromen-3-yl)-*1H*-1,2,3-triazol-4-yl)propanoic acid (69)



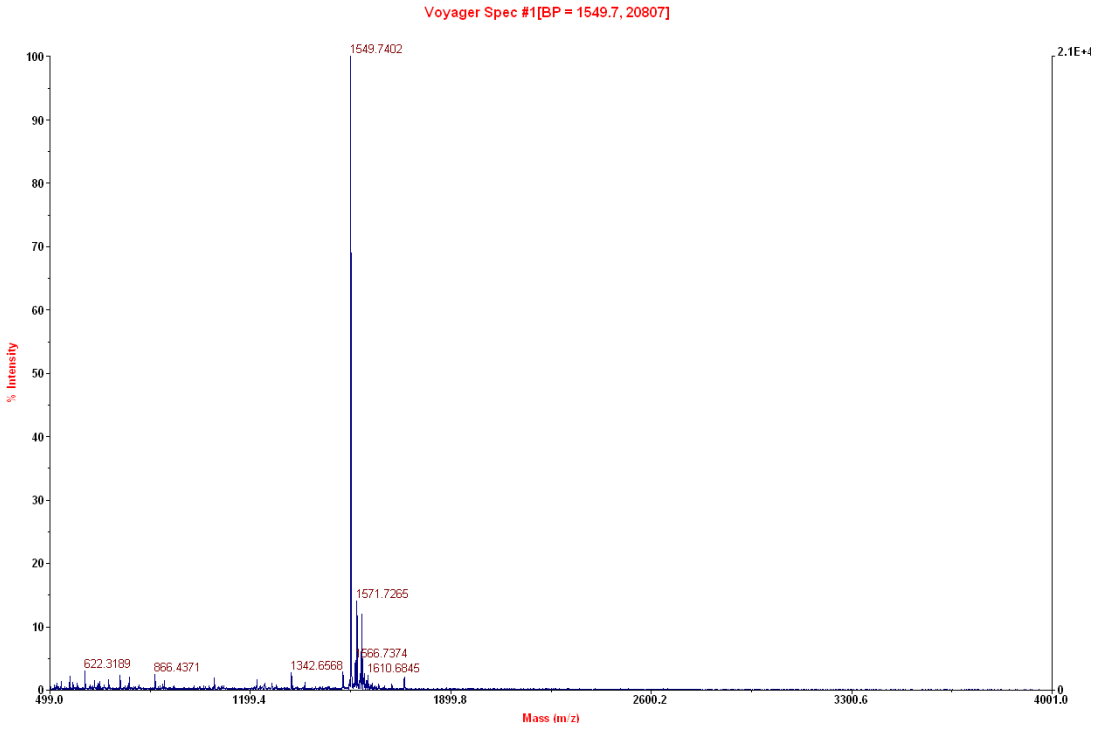
To a solution of **59** (0.0300 g, 0.12 mmol) in ethanol (4 mL), solutions of pent-4-ynoic acid (0.0121g, 0.12 mmol) in ethanol (4 mL) and $CuSO_4$ (0.0064 g, 0.024 mmol) in water (4 mL) were added. Sodium ascorbate (0.0121 g, 0.061 mmol) was quickly added and the reaction stirred at room temperature for 18 hours. Monitoring by TLC showed that the reaction was incomplete and a second addition of sodium ascorbate was made (0.0126g, 0.063 mmol) and the reaction allowed to stir until all of the azide starting material was consumed. The reaction mixture was diluted with EtOAc and washed with HCl (0.1 M) and brine, then the organic layer dried using $MgSO_4$ and the solvent removed under reduced pressure. The crude product was purified by HPLC to yield the title compound as an off white solid (0.0117 g, 28%).

NMR: δ_{1H}(500 MHz, DMSO): 2.34 (3H, s, O**Ac**), 2.67 (2H, d, CH_2^g, J = 7.3 Hz), 2.98 (2H, d, CH_2^f, J = 7.3 Hz), 7.29 (1H, dd, H^c, J = 8.5 Hz, J = 2.2 Hz), 7.45 (1H, d, H^d, J = 2.2 Hz), 7.98 (1H, d, J= 8.5 Hz, H^b), 8.41 (1H, s), 8.71 (1H, s) 12.2 (1H, s, COO**H**); δ_{13C} (126 MHz, $CDCl_3$): 20.5 (CH_2), 20.9 (CH_2), 33.0 (OAc), 110.1, 116.1, 119.0, 119.5, 122.9, 130.3, 133.9, 146.2, 152.9, 153.4, 155.7, 168.7, 173.5; ESI: MH^+ 344.0893 (calculated 344.0883);

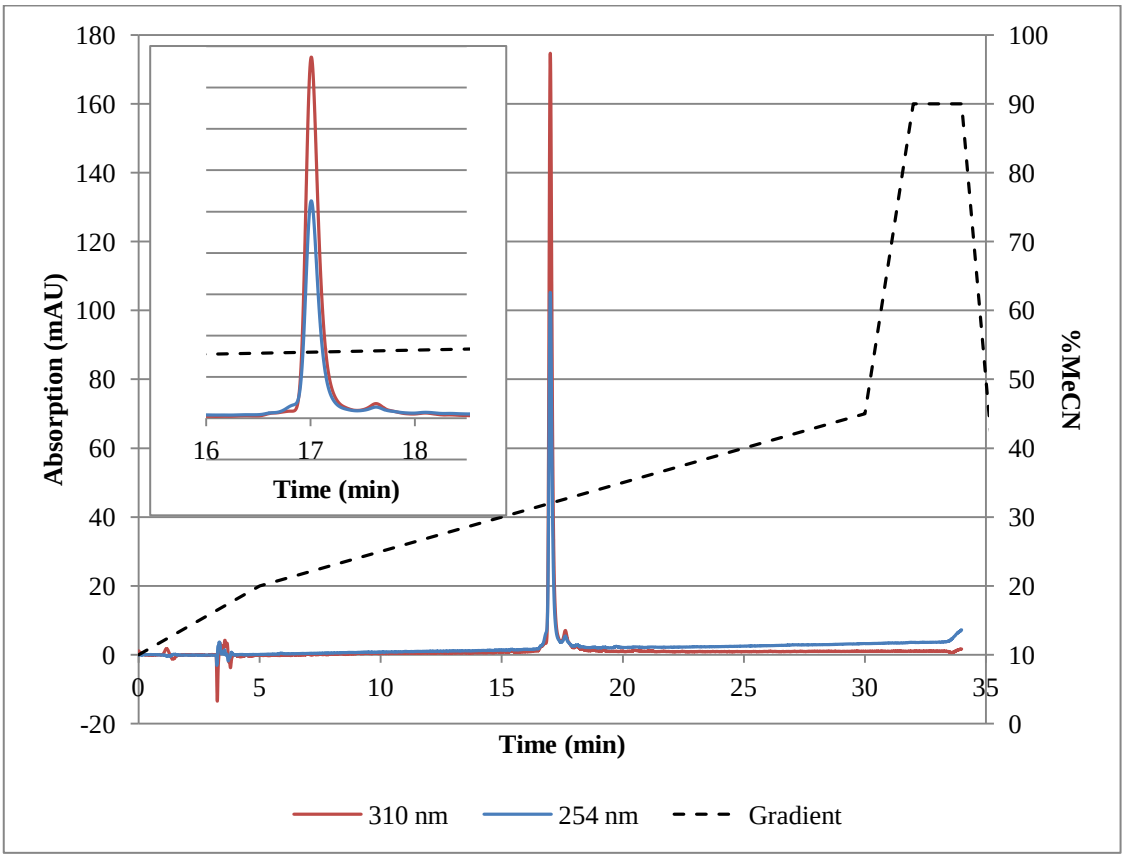
3.6.2.2 Synthesis of the Coumarin-polyamide Cou-Dpβ-PyPyPyPy-γ-PyImImIm (71)



To a solution of **69** (0.0038 g, 11.1 umol) in dry DMSO (2.525 mL) was added solutions of DIEA (14.07 uL, 80.8 umol) in dry DMF (2.511 mL) and HATU (0.038 g, 10.1 umol) in dry DMF (2.525 mL) and the reaction stirred for 5 minutes. This was then added to a solution of **ImImImPy-γ-PyPyPyPy-β-NH_2 (52)** (0.0140 g, 10.1 umol) in DMSO (2.525 mL) and the mixture allowed to stir at room temperature for 72 hours. After confirmation of product formation by HPLC, the reaction mixture was added directly to a solution of sodium methoxide (0.0543 g, 1.0 mmol) in dry methanol (10 mL) and stirred for 30 minutes. The reaction mixture was then neutralised using dilute $HCl_{(aq)}$ and the methanol removed under reduced pressure. HPLC purification (0-20% 5 min, 20-45% 30 min, 45-90% 2 min, MeCN in H_2O, 0.1% TFA) yielded the title compound **71** (0.0027 g, 17%) as a white solid. MALDI: 1549.8048, 1571.7843 (Calculated $M+H^+$ 1549.6739, $M+Na^+$ 1571.6558). ESI: MH^+ 1549.6791(calc. 1549.6739);

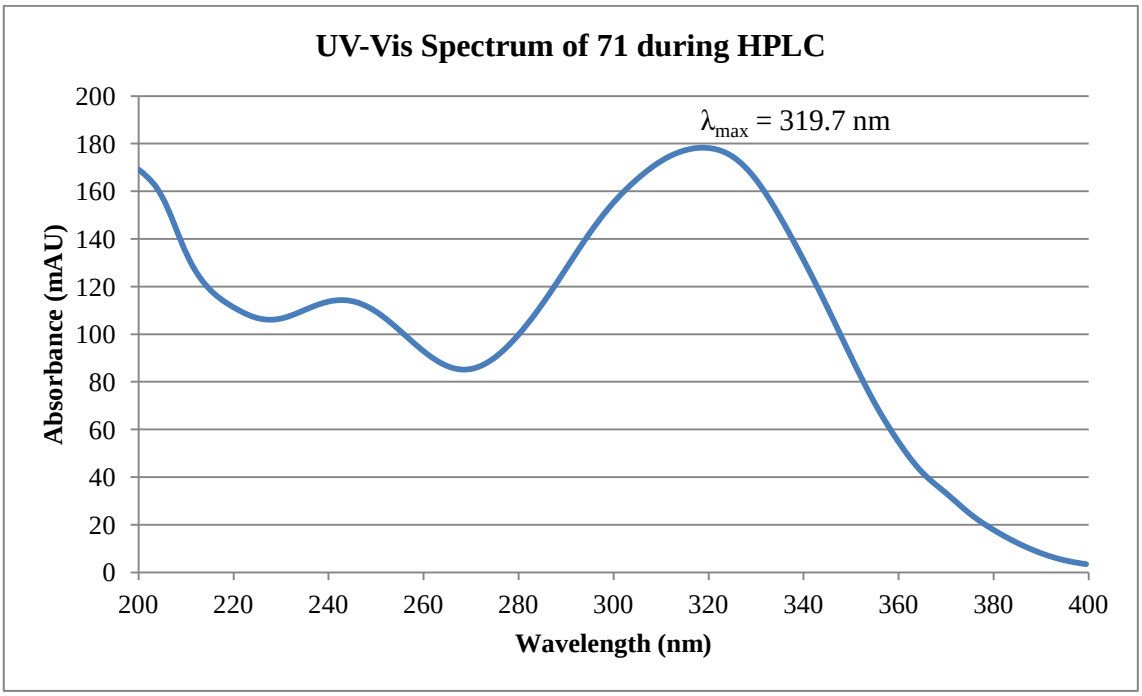


Graph 1 *- MALDI spectrum of polyamide-coumarin conjugate* ***64***



Graph 2 *- HPLC of polyamide-coumarin conjugate* ***71*** *after purification at 310 nm and 254 nm (Left axis), HPLC gradient shown on right axis; Insert: expansion of the 16.5-18 minute region*

NMR: δ_{1H}(500 MHz, DMSO): 1.72-1.83 (6H, m), 2.30 (2H, t, J = 7.4 Hz), 2.36 (2H, t, J = 7.4 Hz), 2.69-2.76 (4H, m), 2.97 (4H, t, J = 7.4 Hz), 3.06-3.16 (8H, m), 5.23 (2H, t, J = 6.2 Hz), 3.81 (3H, s, Py-Me), 3.82 (3H, s, Py-Me), 3.84 (3H, s, Py-Me), 3.85 (3H, s, Py-Me), 3.86 (3H, s, Py-Me), 4.01 (3H, s, Im-Me), 4.02 (3H, s, Im-Me), 4.04 (3H, s, Im-Me), 6.85 (1H, d, J = 2.1 Hz), 6.99 (1H, s), 7.02 (1H, d, J = 1.8 Hz), 7.08 (1H, s), 7.09-7.11 (2H, m), 7.20 (1H, s), 7.46 (1H, s), 7.59 (1H, s), 7.66 (1H, s), 7.75 (1H, d, J = 8.7 Hz), 8.30 (1H, s), 8.56 (1H, s), 9.15 (1H, br s, Cou-OH), 9.64 (1H, s), 9.85 (1H, s), 10.09 (1H, s), 10.33 (1H, s), 10.91 (1H, s).



Graph 3 *- UV-Vis spectrum of purified polyamide-coumarin conjugate* ***71*** *obtained during HPLC purification (Graph 2)*

Chapter 4 DNA based photonic wire using a polyamide programming approach

4.1 Introduction, aims and objectives

With a suitable method established for the production of functionalised polyamides, their application as a replacement for DNA end labelling can now be tested as part of an improved photonic wire (Figure 34). This design is based on a previously reported wire containing a YO functionalised polyamide **72**[219] and directly substitutes coumarin polyamide **71** for an end labelled Pacific Blue dye (Table 6), allowing easy comparison of their relative performance. Initial tests focussed on establishing whether **71** has a strong DNA binding affinity, an essential requirement to validate the toolbox approach as incomplete assembly has been shown to be a key weakness of existing photonic wires[102, 103] that should be readily improved upon.

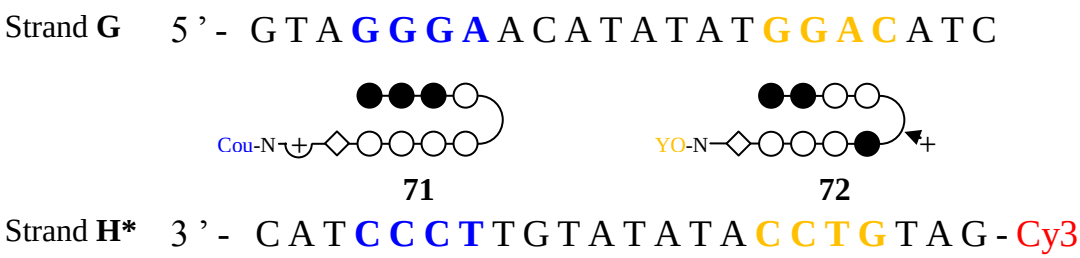


Figure 34 *-An improved photonic wire design using polyamides **71** and **72**, where **71** directly replaces Pacific blue end labelled DNA in previously reported designs.* [219]

4.2 DNA Binding Properties of Polyamide-Fluorophores

4.2.1 Characterisation of DNA binding Using DNA Melting Temperatures

DNA binding affinity is a key characteristic of a polyamide but there are limited methods available to determine this. Historically, the affinity of a DNA binding molecule is determined using a footprinting technique[265] in which an end-labelled duplex of DNA containing several suspected binding sites is exposed to an indiscriminate double stranded cleavage agent such as a chemical or enzyme at a concentration such that on average each DNA strand is only cleaved once. Where a DNA binding molecule or protein is located, the strands are protected from cleavage and result in 'voids' when the mix of DNA fragments is analysed by length afterwards using electrophoresis (Figure 35). There are two commonly used protocols: hydroxyl radicals generated using an iron complex[266] and the enzyme DNase,[267] but both use similar analyses which require that the minimum concentration of ligand tested to be much higher than that of the DNA, effectively limiting the maximum binding affinity that can be determined for polyamides to 2×10^{10} M^{-1}.[268] As the affinity of polyamides being produced approaches this upper limit, the use of DNA melting temperatures has been proposed as a semi-quantitative alternative which would allow mismatch sites and

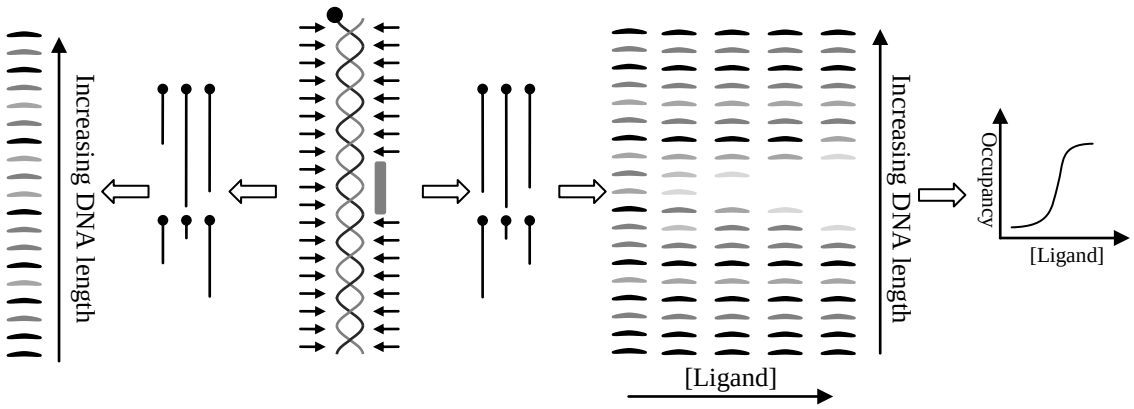
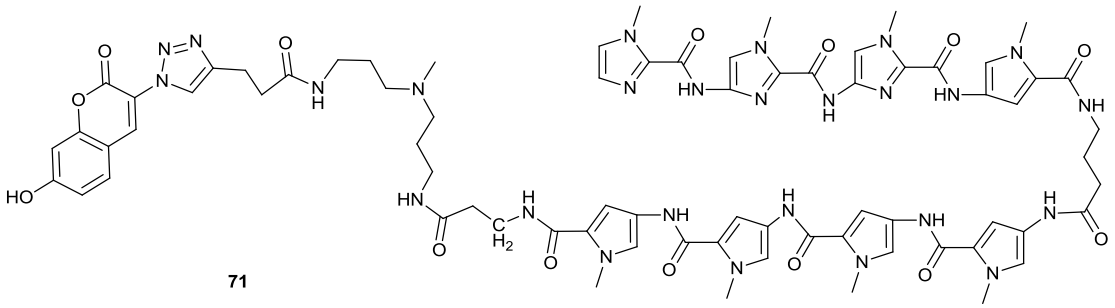


Figure 35 *- Schematic representation of a DNA footprinting assay to determine the DNA binding sites and binding affinity for a ligand*

the relative binding affinity of related polyamides to be assessed.[269-271]

The temperature at which a given piece of double stranded DNA dissociates, its melting temperature, is dependent on degree of stabilisation provided by the duplex compared to the individual strands in solution. Most of this stabilisation comes from the DNA base pairing, resulting in a strong sequence dependency due to the increased hydrogen bonding between G:C base pairs, temperature reductions for mismatched DNA strands and a general increase in melting temperature as the strands are lengthened. When small molecules such as polyamides bind to the DNA duplex the additional interactions introduced must also be overcome, increasing the melting temperature and allowing thermodynamic information about the binding process to be deduced from the degree of stabilisation observed.[269] Typically this information is not extracted as several assumptions and additional pieces of thermodynamic data are required and the degree of duplex stabilisation is used as a semi-quantitative measure of binding affinity. It has been shown that the expected trends in binding affinity are also reflected in the change in DNA melting temperature,[140, 161] allowing a much more rapid assessment of the DNA binding properties of a polyamide to be made compared to conventional footprinting techniques.

4.2.2 DNA Binding Properties of the Polyamide-Coumarin Conjugate



71

Verifying that the modified polyamide **71** does bind to DNA as expected is key not only for its role as a part of a photonic wire, but also adds to the limited information available on the polyamide modifications which are tolerated that is vital for understanding the scope of the 'toolbox' approach. Using the same experimental design previously reported when testing polyamides,[219] complementary DNA strands 14 bases long were synthesised containing both match (strands **A** and **B**) and mismatch binding sites (strands **C** and **D**, **E** and **F**) (Table 12). The mismatch binding sites were positioned within the central binding sequence and chosen to test the G vs. C and G vs. T selectivity of the polyamide respectively. As in previous studies,[219] the concentrations of both the DNA and polyamide were determined by UV-Vis spectroscopy and sodium cacodylate buffer used due to the superior stability of its

Strand	Sequence	Binding site	
A **B**	5'-GCA TAG GGA ATC GC-3' 3'-CGT ATC CCT TAG CG-5'	5'- taGGGAat-3'	Match
C **D**	5'-GCA TAG G**C**A ATC GC-3' 3'-CGT ATC C**G**T TAG CG-5'	5'- taGG**C**Aat-3'	G→C mismatch
E **F**	5'-GCA TAG G**T**A ATC GC-3' 3'-CGT ATC C**A**T TAG CG-5'	5'- taGG**T**Aat-3'	G→T mismatch

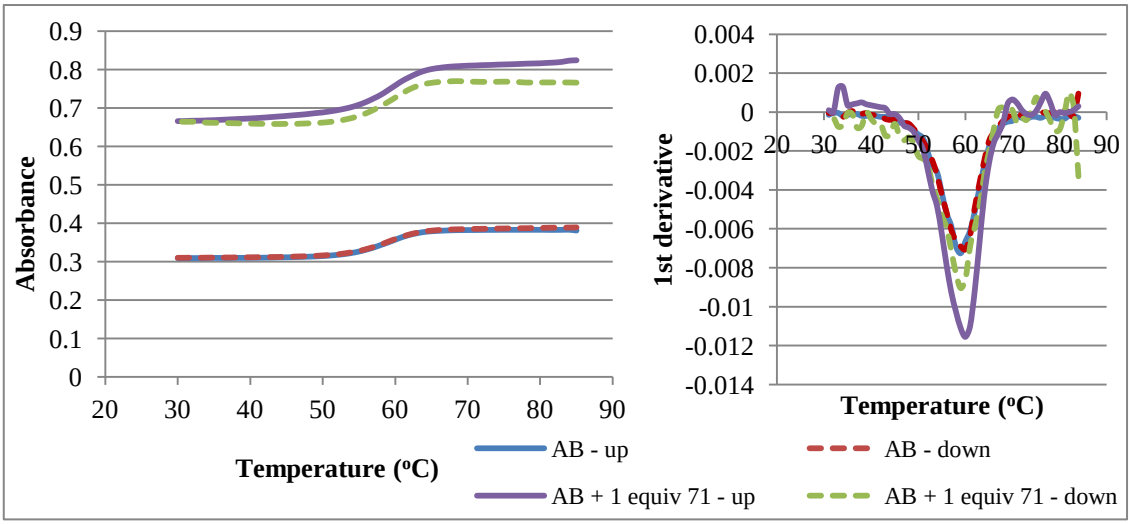
Table 12 *- Sequence and polyamide **71** binding sites of synthesised DNA strands*

	CD	AB + 1 equiv. **71**	AB + 1.5 equiv. **71**
T_m (^oC)	60	60	65
ΔT_m (^oC)	-	0	5

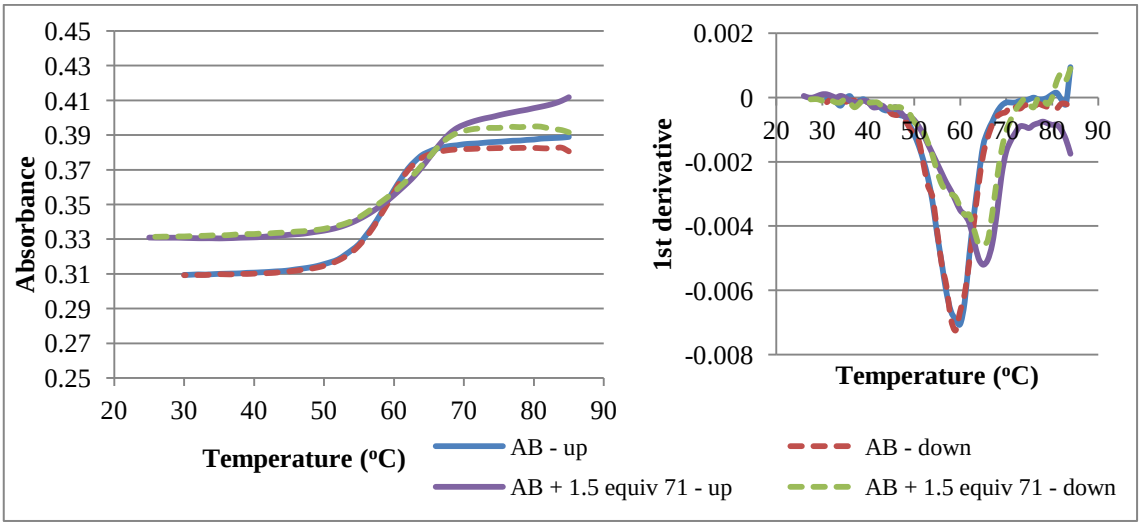
Table 13 *- DNA melting temperatures for match strand* ***AB*** *in the presence of polyamide fluorophore* ***71***

background signal during the heating and cooling cycle.

Initial tests using one equivalent of polyamide **71** with match DNA duplex **AB** caused great concern as no discernable change in melting temperature could be observed (Graph 4). Reassuringly, when the same experiment was carried out using 1.5 equivalents of polyamide **71**, the match site duplex was stabilised by 5^{o}C (Graph 5), which whilst relatively low confirms that DNA binding does take place (Table 13). Experiments using 16 equivalents of **71** failed to give any further increase the melting temperature and given the low level of duplex stabilisation, no mismatch binding data was collected as an even smaller change in melting temperature would be expected, rendering it undetectable. It is unclear why adding just 0.5 equivalents more of the polyamide would cause such a larger increase in melting temperature compared to 1:1 stoichiometry, but it is unlikely that this question can be answered satisfactorily without



Graph 4 *- DNA melting curve of duplex* ***AB*** *in presence and absence of 1 equivalent of polyamide* ***71****; Absorbance measured at 260 nm; Up - Upward temperature ramp, Down - downward temperature ramp*



***Graph 5**- DNA melting curve of duplex AB in presence and absence of 1.5 equivalents of polyamide **71**; Absorbance measured at 260 nm; Up - Upward temperature ramp, Down - downward temperature ramp*

the use of footing printing assays which are much more difficult and time consuming to conduct. The use of 14 base pair lengths of DNA represents a small difference from the 12 bases used in previous reports[219, 269-271] but this would not have affected the results significantly.

Using previously reported ΔT_m values[140] for polyamides which have also been characterised by DNA footprinting analysis,[220, 272] an approximate range for the DNA binding affinity of **71** can be established. Comparison to polyamides **73** and **74** suggests a K_D between 5 and 35 nM (Table 14), consistent with the 10 fold drop in DNA binding affinity expected for tail modifications from the unmodified poylamide (K_D = 2.5 nM[273]) (Table 3). In addition to the destabilisation of the bound polyamide

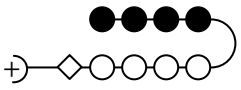
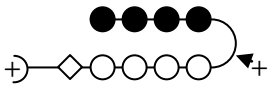
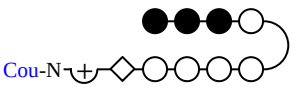
	 **73**	 **74**	 **71**
ΔT_m (°C)	3.4[140]	6.7[140]	5
K_D (nM)	5[220]	35[272]	-

***Table 14** - Comparison of previously reported ΔT_m values with DNA binding affinities (K_D) determined by DNA footprinting experiments*

due to steric clashes between the modification and DNA, the flexibility of the tail linker may also allow relatively hydrophobic 'payloads' to fold back and interact with polyamide rings to form a stable complex which must be broken up before DNA binding can take place. These effects lower the net gain in energy upon binding and shift the DNA binding equilibrium back towards the polyamide in free solution, increasing the value of K_D. The formation of such complexes has been suspected in previous studies on fluorophore modified polyamides due to the unexpected quenching of fluorescence[274] and there is some evidence for this in the shift in absorbance maxima of **71** from the 310 nm typical in unmodified polyamides to 320 nm (Graph 3).

4.3 Fluorescence properties of polyamide-coumarin conjugate

4.3.1 Determination of Quantum yield

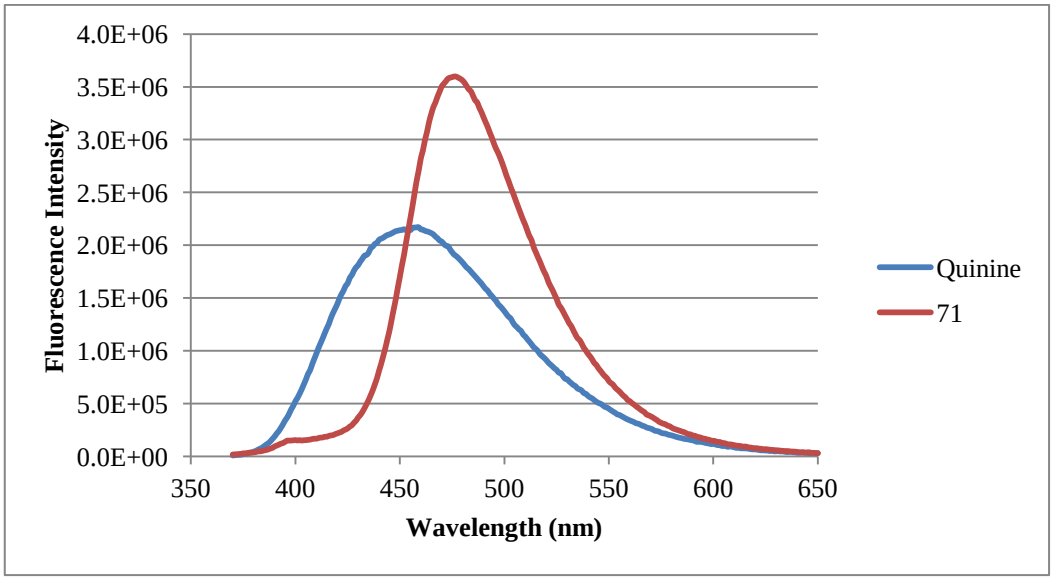
The original report describing the fluorogenic click reaction of 3-azido coumarins with alkynes quoted a quantum yield of 0.6-0.7 for conjugates similar to **71**,[177] but the attachment of the coumarin fluorophore to a polyamide and localisation onto DNA dramatically increases the opportunities for non-radiative decay which directly influence this figure. In addition, the necessity of using buffer solutions when constructing the photonic wire also introduces additional species into the solution which also increase the amount of solvent quenching that that takes place. As a result, the quantum yield of the coumarin-polyamide **71** was determined independently under the experimental conditions by calibration against the fluorescence standard quinine (Graph 6). The full calculation (Equation 4) to calculate the samples fluorescent quantum yield

$$Q = Q_R \frac{F}{F_R} \frac{OD_R}{OD} \frac{n^2}{n_R^2}$$

Equation 4 *- Full equation for the standardisation of quantum yield*

$$Q = Q_R \frac{F}{F_R}$$

Equation 5 *- Simplified calculation of the standardisation of quantum yield*



Graph 6 *- Background corrected fluorescence spectra of 200 nM quinine in 0.1 M H_2SO_4 and 200 nM Coumarin-polyamide* ***71*** *in PBS buffer with excitation at 350 nm and 370 nm respectively*

(Q) requires the quantum yield of the reference (Q_R), the integrated fluorescence spectra of the sample (F_{Cou}) and reference (F_R), the optical density (absorption) of the sample (OD) and reference (OD_R) and the refractive index of the sample (n) and reference (n_R) solutions.[93] By using a low concentration in the solutions such that the absorption (OD and OD_R) is less than 0.05 absorption units to minimise depletion of the irradiating light as it passes through the solutions and assuming there is no change in the refractive index, the fluorescence quantum yield can be calculated using the simpler Equation 5. The steady state fluorescence spectra of **71** in PBS buffer and quinine in 0.1 M H_2SO_4 were recorded with excitation at 200 nM (Graph 6), numerically integrated over the complete range of wavelengths and the quantum yield of **71** was calculated using Equation 5 and $Q_R = 0.58$[93] to give a value of 0.70 which was used in all further calculations for the photonic wire.

4.4 Analysis of the Improved Photonic Wire

4.4.1 Theoretical Treatment of the A-B-C Model Photonic Wire

Calculating the end to end efficiency of a molecular wire, that is the proportion of the absorbed energy at the donor is emitted from the acceptor, is non-trivial and requires several additional pieces of information to enable the analysis.[107] Consider a molecular wire constructed using 3 fluorophores A, B and C respectively such that A absorbs light as the donor, B acts as the 'bridge' and C behaves as the acceptor (Figure 36). The emission profiles of each fluorophore may have significant overlap but the individual contribution of each fluorophore can be extracted by measuring the standard fluorescence spectra for each component ($F_A(\lambda)$, $F_B(\lambda)$ and $F_C(\lambda)$) and fitting the

$$F_T(\lambda) = aF_A(\lambda) + bF_B(\lambda) + cF_C(\lambda)$$

Equation 6 *- General fitting model for the analysis of the general photonic wire A-B-C*

$$F_C = \int_0^\infty F_C(\lambda)\ d\lambda$$

Equation 7 *- Quantification of the total fluorescence of fluorophore C (F_C) by integration where $F_C(\lambda)$ is the deconvoluted fluorescence spectrum of C at wavelength λ*

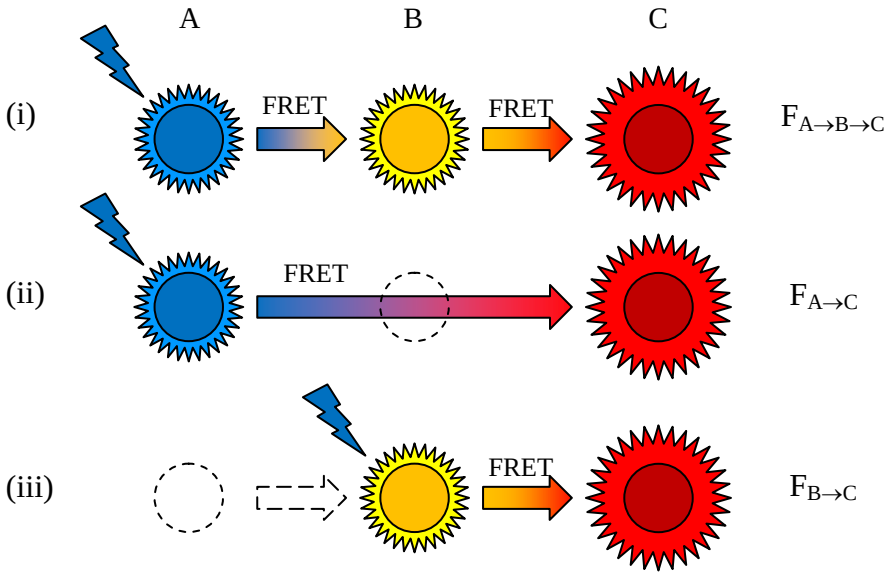


Figure 36 *- Design of the 3 component molecular wire using fluorophores A, B and C and the FRET processes contributing to emission (F_C) at fluorophore C*

observed spectrum ($F_T(\lambda)$) as a linear combination of the individual components (Equation 6) using least squares regression, a process known as deconvolution. Once the spectrum for the acceptor fluorophore C ($F_C(\lambda)$) has been isolated it is numerically integrated over the entire range of wavelengths to yield F_C, the total area enclosed by the fluorescence peak which is a measure of the total fluorescence from C over all wavelengths (Equation 7).

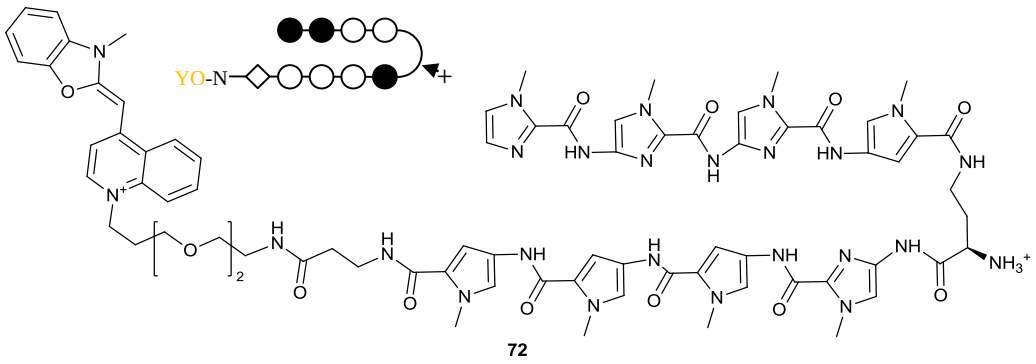
The total fluorescence emission observed from the acceptor fluorophore (F_C) is dependent on the excitation of fluorophore C which can occur through several pathways in addition to the desired FRET 'chain' (Figure 36i) from A to B to C ($F_{A\rightarrow B\rightarrow C}$). The simplest pathway, direct excitation by the irradiating light is normally negligible as the acceptor is chosen to have very weak absorption in that region. Similarly, the fluorescence due to direct FRET transfer from A to C ($F_{A\rightarrow C}$) is also negligible due to the weak spectral overlap and large distance between the two fluorophores (Figure 36 ii). However, the need for good spectral overlap between A and B in order for the wire to operate efficiency means that the absorption region of the bridging fluorophore B often overlaps with the range of wavelengths used to excite the donor, resulting in direct excitation of B followed by FRET transfer to the acceptor ($F_{B\rightarrow C}$) (Figure 36 iii). This contribution to the overall fluorescence at C (F_C) can be measured by performing the experiment under the same conditions using only fluorophores B and C and subtracting the observed fluorescence signal at C in this simplified system from that observed in the full wire, leaving only the fluorescence due to the full FRET transfer ($F_{A\rightarrow B\rightarrow C}$) and also subtracting any fluorescence caused by direct excitation of the acceptor fluorophore C. As with all FRET processes only a proportion of all the energy arriving at a acceptor will be converted into fluorescence emission, so the observed fluorescence must be divided by the fluorophores quantum yield (Q_C) to give an accurate figure for the actual

$$Efficency = \frac{N_{C^*}}{N_{A^*}} = \frac{F_{A\rightarrow B\rightarrow C}/Q_C}{F_A/Q_A} = \frac{(F_C - F_{B\rightarrow C})/Q_C}{F_A/Q_A}$$

Equation 8 *- Calculation of the efficiency of photonic wire A-B-C*

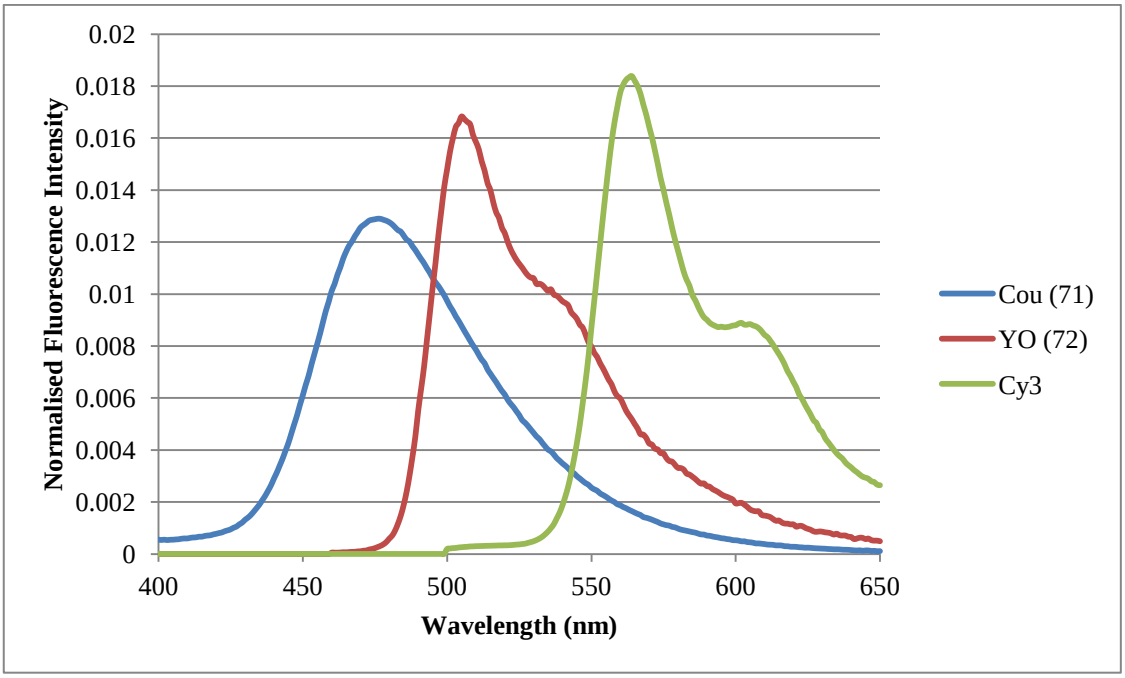
number of excited molecules of C present in the system. The efficiency of a photonic wire is taken as the proportion of the number of excited acceptor molecules (N_C*) in the wire compared to the number of excited donor molecules (N_A*) measured for fluorophore A in the absence of the wire, i.e. without any FRET transfer processes taking place. This quantity is calculated in the same way by integrating the emission spectrum of fluorophore A alone (F_A) and dividing this by its quantum yield (Q_A), as summarised in Equation 8.

4.4.2 Construction and Analysis of the Improved Photonic Wire



72

A photonic wire was assembled in accordance with the planned design (Figure 34) using coumarin labelled polyamide **71** (Cou) as donor (A), the YO labelled polyamide **72** (YO) synthesised by Dr Wu Su[219] as bridging fluorophore (B) and Cy3 5' labelled DNA (**H***) (Cy3) purchased from Invitrogen as the acceptor (C). In light of the lower estimated binding affinity of **64**, the concentration of the DNA and the polyamides was increased to 500 nM (cf. 50 nM in previous studies[219]) in order to promote binding. To provide reference spectra for the deconvolution of the photonic wire, individual fluorescence spectra were recorded for each of the fluorophores in the



Graph 7**- Normalised reference spectra of fluorophores **71, 72** and the Cy3 labelled DNA duplex (**GH)*

presence of matching DNA under experimental concentrations at their respective maximum excitation wavelengths then normalised (Graph 7) to enclose a total area of 1 by numerical integration ($F_{Cou} = F_{YO} = F_{Cy3} = 1$). In accordance with previous work,[107, 219] steady state fluorescence spectra were recorded under the experimental conditions with excitation at 380 nm for the complete wire (Graph 12), the wire in the absence of the coumarin donor **71** (Graph 11) and the fluorescence of the coumarin donor **71** alone (Graph 9) using unlabeled DNA. In addition, the spectra of the wire in the absence of the bridging YO fluorophore **72** (Graph 10) and of the Cy3 labelled DNA alone (Graph 8) were recorded to determine the contribution from direct donor to acceptor FRET and confirm the presence of Cy3 respectively.

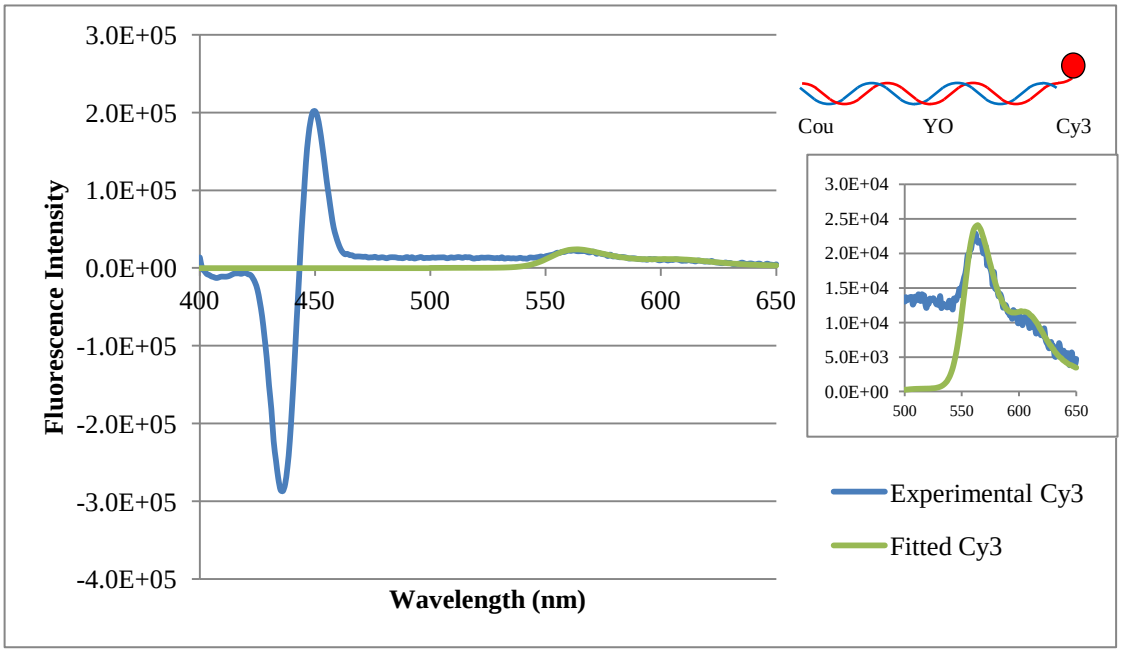
After collection, the raw fluorescence spectra were deconvoluted using the individual normalised spectra and fitted using the least squares method to Equation 6 with the coefficients for absent fluorophores fixed at zero. Analysis of the wire uses the derived integrated fluorescence area of each component (F_A, F_B, F_C) rather than the unprocessed spectral data, so the difference in area (Δ%) between fitted and

$$\Delta\% = \frac{\left|F_{Experimental} - F_{Fitted}\right|}{F_{Experimental}}$$

Equation 9 *- Calculation of Δ% - the difference in area between the experimental and fitted integrated fluorescent spectra*

experimental spectra (F_{Fitted} and $F_{Experimental}$) before deconvolution was used as an additional indirect measure of the goodness of fit (Equation 9). In contrast to the least squares fitting procedure which matches the data to a model without bias, Δ% is heavily weighted towards the key emission region used to analyse the wire. In all cases where this measure was calculated over the fitting region, it corresponded to less than 3% of the enclosed area and was found to be just 1.12% for the full wire, confirming that the least squares fitting procedure does allow accurate modelling of the fluorescence data.

4.4.2.1 In-situ Reference Fluorescence Spectra of the Cy3 and Coumarin Fluorophores

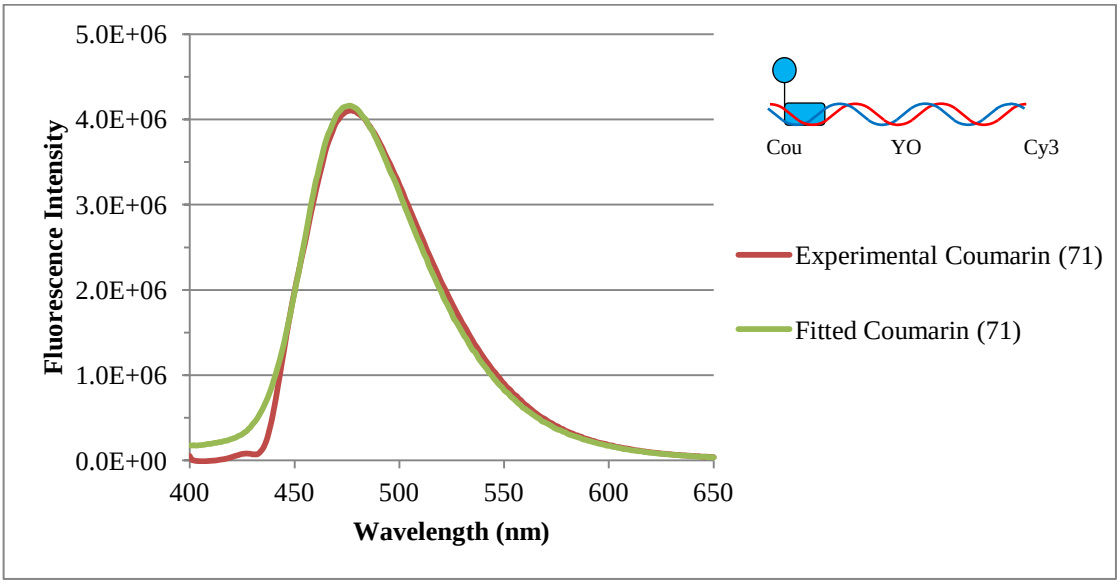


Graph 8 *- Experimental and fitted fluorescence spectra of 500 nM Cy3-DNA (**GH***) with irradiation at 475 nm . Insert – expansion of the 500-650 nm region*

A key requirement of for a light based nanotechnology systems such as a photonic wire is the selective excitation of individual fluorophores. In order to quantify

any direct excitation of Cy3 acceptor fluorophore which would give potentially misleading results and confirm the presence of the Cy3 fluorophore, the fluorescence spectrum of the Cy3 labelled DNA was recorded in the absence of the other fluorophores (Graph 8) with excitation at 475 nm. The level of fluorescence observed was of the same order of magnitude as the background signal (Graph 8 insert), proving that this design allows strong discrimination between the donor and acceptor fluorophores even after excitation with wavelengths much closer to the absorption region of the acceptor dye.

The major peak observed around 450 nm in Graph 8 is not caused by the DNA or any component of the photonic wire but is a result of background scattering by water. As a blank is recorded for the buffer solution and subtracted during data collection in principle no signal should be observed, but due to instrument drift and small shifts in the scattering wavelength after the addition of the wire components this subtraction process is not always completely effective. Fortunately as this is ordinarily a relatively weak signal and the concentration and volume of buffer are constant, the size of this

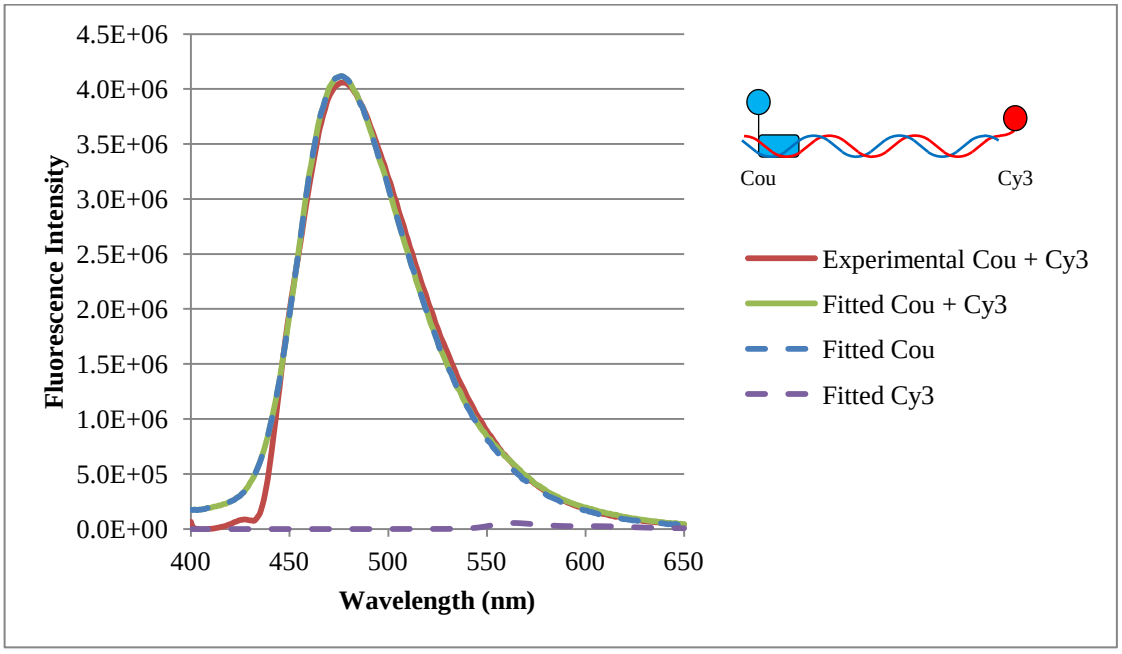


Graph 9 *- Experimental and fitted fluorescence spectra of coumarin-polyamide* ***71*** *(1.0 equivalents) with 500 nM unlabelled DNA (****GH****) and irradiation at 380 nm*

peak is normally much smaller than the fluorescence from the wire components. However, the same factors in addition to incomplete blanking of the sample can also affect the shape of the reference peak, particularly in the region of the scattering peak below 450 nm. This is particularly apparent in the fluorescence spectrum (Graph 9) of the coumarin-polyamide partial wire (used to determine F_A for Equation 8) as the reference and experimental systems have identical components and were recorded under the same conditions.

4.4.2.2 Analysis of Cou – Cy3 partial wire

Recording of the fluorescence spectrum of a partial wire consisting of only the Coumarin-polyamide **71** and Cy3 labelled DNA shows the important role of the bridging YO dye within the wire. The deconvoluted spectrum (Graph 10) is completely dominated by the coumarin fluorophore (λ_{max} 477 nm, >99%, Table 15) with little fluorescence observed from Cy3 which shows that the positioning of the bridging



Graph 10 *- Experimental and fitted fluorescence spectra of coumarin-polyamide (1.0 equivalents) with 500 nM Cy3-DNA (**GH***) and irradiation at 380 nm*

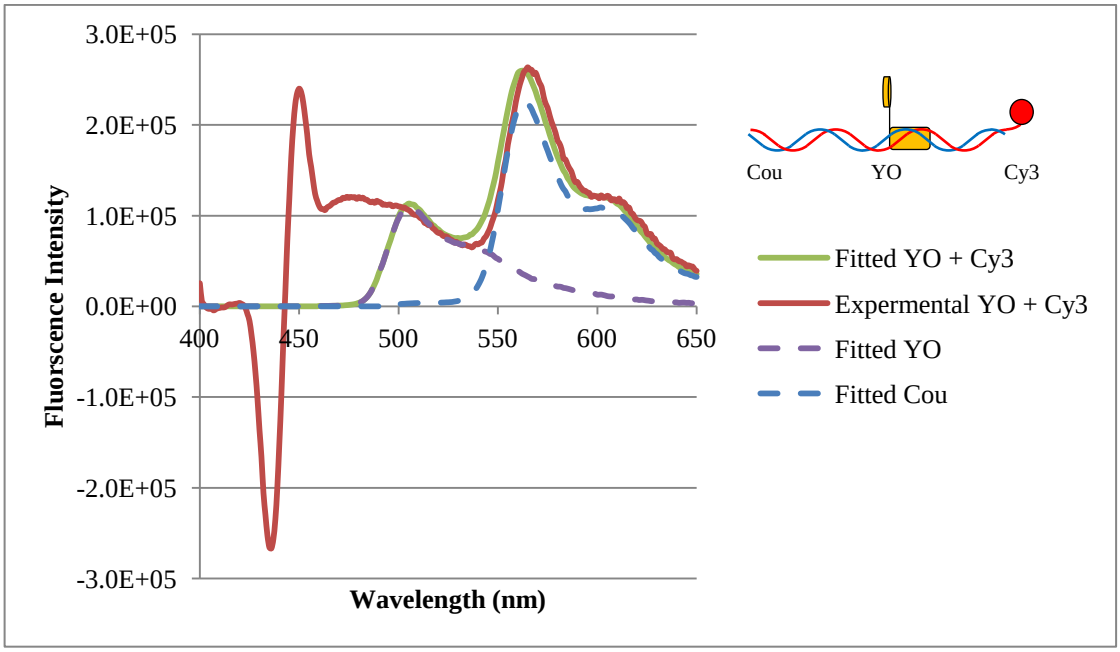
	Cou λ_{max} 477 nm	YO	Cy3 λ_{max} 564 nm	Total fitted area	Experimental area	Δ%
Fitted Area	3.19×10^8	0	2.96×10^6	3.13×10^8	3.21×10^8	2.56%
% Area	99.1%	0	0.92%	-	-	-

***Table 15**- Analysis of the Cou-Cy3 partial photonic wire*

fluorophores allows control over the energy pathways even where the donor and acceptor are in relatively close proximity. In common with the coumarin only partial wire (Graph 9), the fitted curve deviates from the experimental data most below 450 nm, with Δ% halving to 1.13% when this region is excluded from the calculation. Due to the complete domination of the spectrum by the coumarin fluorophore, a near identical value for Δ% as the coumarin only system would be expected (Δ% = 1.85%), but over the entire spectral range it is significantly higher (Table 15). The sizes of these errors in the total area are several times larger the area attributed to the Cy3 dye (λ_{max} 564 nm, 0.92%) which casts severe doubts on the accuracy of the deconvolution process in this system as the Cy3 component may have been erroneously increased to compensate for the previously observed differences between the experimental and reference coumarin

spectra during the fitting process. As a result, only the qualitative conclusion that the bridging YO dye is required for efficient FRET transfer between the Cy3 and coumarin fluorophores at this separation can be extracted from this experiment.

4.4.2.3 Analysis of YO – Cy3 partial wire



Graph 11 *- Experimental, fitted and deconvoluted fluorescence spectra of YO-Polyamide* ***72*** *(1.0 equivalents) with 500 nM Cy3-DNA (****GH****) and irradiation at 380 nm*

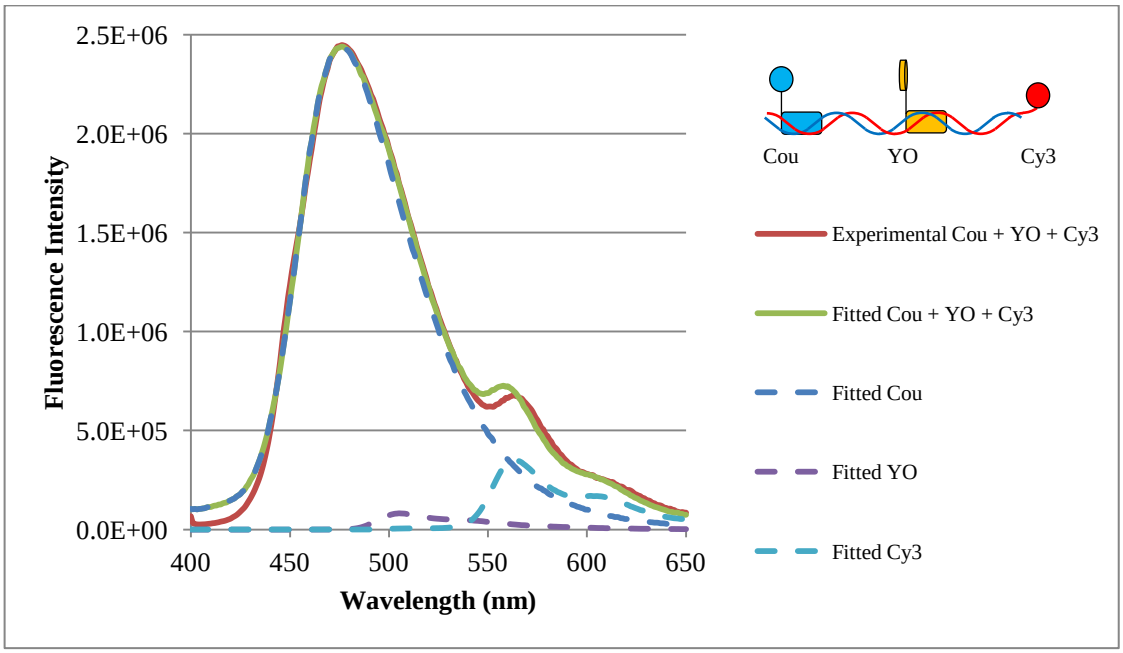
In order to assess the efficiency of the full wire, it is necessary to quantify any Cy3 fluorescence (F_{Cy3}) due to direct excitation of the bridging YO fluorophore **72** ($F_{YO \rightarrow Cy3}$) by 380 nm light used during the experiment (Equation 8). The small amount of fluorescence due to direct excitation of the Cy3 will also be contribute in this partial wire but this has already been observed to be negligible (Graph 8) even with excitation using much higher wavelengths. Deconvolution of the recorded spectrum (Graph 11) is complicated by the modified emission profile of the YO dye due to excitation at 380 nm compared to the reference spectrum where excitation took place at 490 nm which is responsible for the additional signal in the 460-500nm region. Although in principle additional fitting using an extra reference spectrum could be used to correct this, deviations from reference spectra in this region have already been observed which would limit the reliability of the deconvolution and as a result, the experimental data was fitted using only the 500-650 nm region (Table 16). The poor fitting of the

	Cou	YO λ_{max} 505 nm	Cy3 λ_{max} 564 nm	Total fitted area (500-650 nm)	Experimental area (500-650 nm)	Δ%
Fitted Area	0	5.68 x10^6	1.26 x10^7	1.78 x10^7	1.76 x10^7	1.36%
% Area	0	31.1%	68.9%	-	-	-

***Table 16**- Analysis of the YO-Cy3 partial photonic wire using the 500-650 nm fitting region*

experimental data in the 400-500 nm region is undesirable but as only the deconvoluted Cy3 fluorescence is required to calculate the efficiency of the photonic wire, it is the 500-650 nm Cy3 fluorescence region which is key during the fitting process. Close agreement between the fitted and experimental data is observed within this region (Δ% = 1.36%) and as the YO dye shows relatively weak fluorescence between these wavelengths, any inaccuracies caused by modified emission profile will be minimised. Finally, the total fluorescence signal observed is an order of magnitude lower than that recorded for the coumarin donor alone (Graph 9) which suggests that although the design of the wire allows for some non-specific excitation of the wire, absorption by the coumarin fluorophore should remain the dominant mechanism through which energy is introduced.

4.4.2.4 Analysis of the Full Wire



Graph 12 *- Experimental, fitted and deconvoluted fluorescence spectra of Coumarin-polyamide **71** (1.0 equivalents) and YO-Polyamide **72** (1.0 equivalents) with 500 nM Cy3-DNA (**GH***) and irradiation at 380 nm*

As has been observed in previous studies,[219] the fluorescence spectrum of the full wire is dominated by the emission of the donor with a lesser contribution from the acceptor (Table 17). The Coumarin donor **71** (λ_{max} 477 nm) represents 89% of the total fluorescence with little contribution from the bridging YO fluorophore **72** (λ_{max} 505 nm, 2.3%) and just 9% of the total area attributed to the Cy3 acceptor (λ_{max} 564 nm), suggesting only a limited amount of energy is propagating along the full wire. Calculation using Equation 8[107] yields an overall wire efficiency of 8.9% which is

	Cou λ_{max} 477 nm	YO λ_{max} 505 nm	Cy3 λ_{max} 564 nm	Total fitted area	Experimental area	Δ%
Fitted Area	1.89×10^8	4.88×10^6	1.92×10^7	2.10×10^8	2.13×10^8	1.12%
% Area	88.7%	2.3%	9.0%	-	-	-

Table 17 *- Analysis of the Cou-YO-Cy3 full photonic wire*

significantly below the 49% observed for the Pacific Blue-YO-Cy3 wire which formed the basis for the design.[219] Efficient transfer of energy from the YO dye to Cy3 was demonstrated in this previous work and has also been demonstrated in the YO-Cy3 partial wire (Graph 11, Table 16) indicating that the initial FRET step between the coumarin donor **71** and the YO fluorophore **72** limits the overall efficiency of the wire. The coumarin donor shows strong fluorescence at the expected wavelengths (Graph 9) and was specifically chosen to have a similar excitation wavelength and overlap integral with the YO dye to minimise the alterations from the original design (Table 6).

In light of the results of the DNA melting temperature studies which indicated much weaker binding than expected, this strongly suggests that the majority of the **71** remains unbound in solution which would result in high coumarin fluorescence due to the absence of a suitable acceptor close enough for FRET during the lifetime of the excited state. Low affinity binding of **71** is particularly problematic due to the low concentrations of the wire components required for the use with the highly sensitive fluorescence detector, and despite the use of relatively high concentrations (500 nM) to promote binding to DNA it does not appear to have been sufficient during these experiments. The very high association rate typically observed for polyamides[275] would seem to rule out any alternative kinetic explanations, and the high level of fluorescence observed with the coumarin partial wire (Graph 9) limits the scope for additional non-FRET quenching pathways to be introduced. It is conceivable that the unwinding effect on the DNA following binding of any of the polyamides might cause this effect and the overall level of fluorescence is significantly reduced (44%) in the full wire compared to that of the coumarin alone, but this would not affect the relative proportions of the observable processes of energy propagation along the wire and fluorescence. The lack of direct information on the processes within the wire makes it impossible to draw firm

conclusions and the measurement of the fluorescence lifetimes might allow some separation of solution and DNA bound fluorophores, but this is beyond the scope of this investigation.

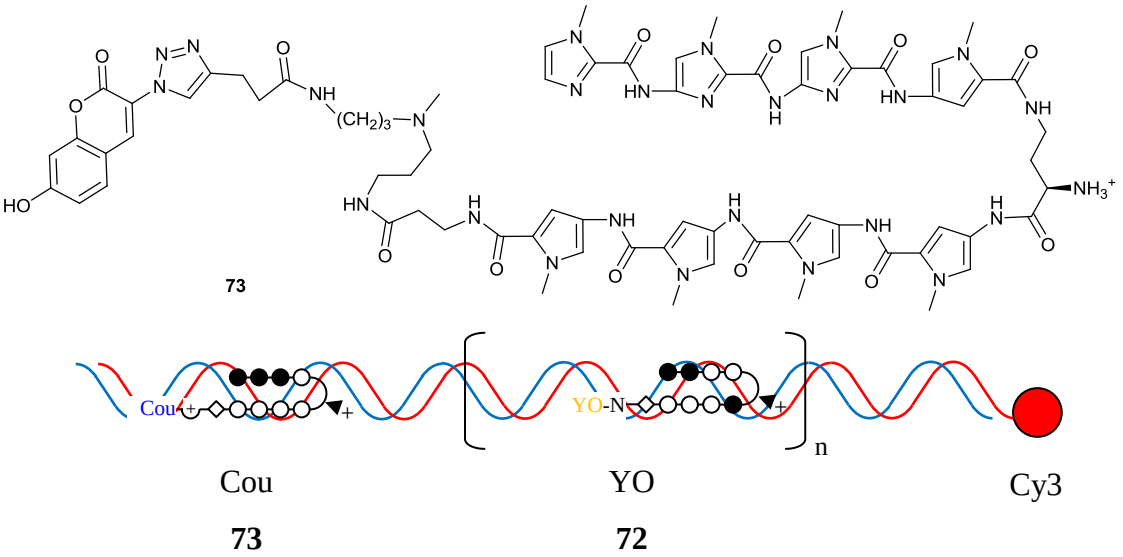
4.5 Conclusions and Future Work

The poor efficiency of the final photonic wire is a disappointing final result, but a significant amount of useful information has still been extracted during its production. An improved nickel catalysed reduction method has been established for the **Py** building block which replaces an extended hydrogenation and does not require chromatographic purification, improving a key step in a synthesis that has seen few innovations since it was first reported 15 years ago.[183] The same method has since been extended to the synthesis of the more problematic **Im** building block where this overnight reaction replaces 48 hours of continuous hydrogenation.[203] A novel alkyne modified building block has also been introduced and incorporated into polyamides via both solid and solution phase chemistry, adding a new route for the internal functionalisation of polyamides. The synthetic route taken in its production is versatile and should allow a much wider range of functionality to be introduced as internal modifications as the penultimate step in the building block synthesis where chemical incompatibilities should be minimised.

The viability of post-synthetic modification of polyamides using click chemistry has also been investigated thoroughly and found to be unexpectedly problematic for alkyne modified polyamides both as large molecules and much smaller fragments. Despite these failures, a tail modified fluorophore-polyamide was created using click chemistry and shown to bind to DNA. During the characterisation of this conjugate, the general applicability of the DNA melting temperature approach to the determination of

binding affinity has been called into question, a technique which as the reported binding constants of polyamides increase is likely to become more wide spread. Finally, despite the poor binding of the polyamide fluorophore conjugate **71**, the photonic wire that was constructed still showed energy transfer over a distance of approximately 6 nm showing that this approach to the fundamental design of photonic wires is a viable alternative to existing DNA end labelling methods. The original concept of a nanotechnology 'toolbox' has not been achieved during the course of this work, but some unexpected limitations have been found that will help inform future designs of this type.

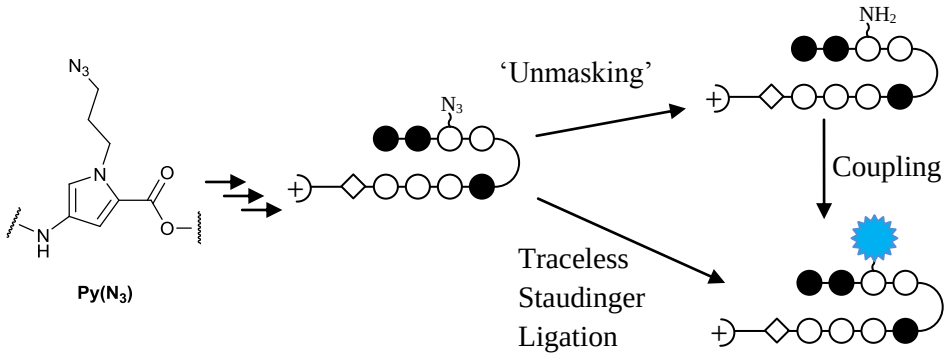
There are many unanswered questions around many of these results, particularly surrounding the underlying causes of the failure of the click reaction which may have important implications for future designs of this type. Further investigations are needed in this area, particularly in establishing whether the problem is more general or simply limited to internally modified alkyne polyamides. It has been assumed that the tail modified polyamide **71** has poor DNA binding properties, but without DNA footprinting analysis this is mostly conjecture. The addition of a charged amine on the



73 **72**

Figure 37 *- Revised photonic wire using polyamide **73** with YO functionalised polyamide **72** and Cy3 end labelled DNA, where n ≥1*

hairpin loop has been shown to improve the binding affinity by an order of magnitude[147] which may be enough to promote binding of **73**. It would also enable the construction of extended photonic wires (Figure 37) similar to those reported using DNA end-labelling, allowing further investigation into whether the coumarin fluorophore is tolerated and allow the true effectiveness of this design of photonic wire to be assessed when the donor is properly positioned on the DNA. Alternatively, the **Py(N_3)** building block could be used to produce internally modified azide polyamides containing a 'masked' amine suitable for coupling to fluorophores through the traceless Staudinger ligation or as a heavily protected amine suitable for coupling using standard peptide coupling chemistry (Scheme 27).[162] It is too early to tell if polyamides truly are a versatile tool for nanotechnology, but their versatility and flexibility offer a novel way out of the 'dead-end' that is the top-down approach.



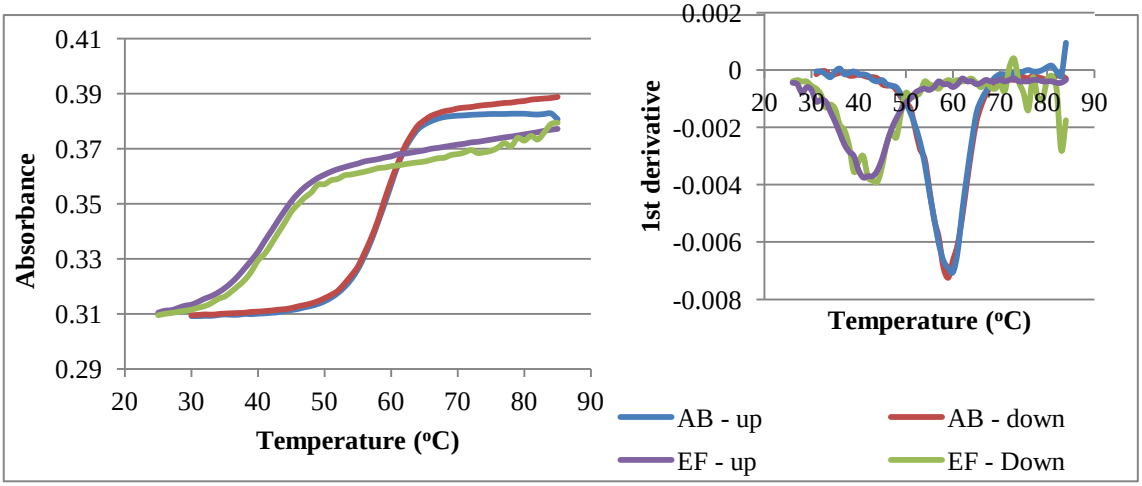
Scheme 27 *- Derivatisation of azide functionalised polyamides*

4.6 Experimental

Analytical and semipreparative RP-HPLC was performed at room temperature on the ULTIMAT 3000 Instrument (DIONEX). UV absorbance was measured using a photodiode array detector at 260 and 310nm. An ACE C18 column (4.6 X 250 mm, 5 µm, 300 Å) was used for analytical HPLC and an ACE C18 column (10 X 250 mm, 5 µm, 300 Å) for semipreparetive HPLC. MALDI-TOF accurate mass spectra were recorded on a Voyger-DE STR using α-cyano-4-hydroxycinnamic acid as the matrix.

4.6.1 DNA Melting Temperature Analysis

DNA strands with the required binding sites (Table 12) were synthesised using standard protocols and after purification by HPLC were characterised by MALDI mass spectroscopy. The concentration of the DNA strands was determined in water by UV-Visible spectroscopy and the concentration of **71** estimated using its absorption at 310 nm.[175] Melting temperature analysis was performed on a PerkinElmer Lambda 35 UV/Vis spectrophotometer equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. A degassed aqueous solution of 10 mM sodium cacodylate,



Graph 13 *- DNA melting curve of duplexes AB (T_m = 60°C) and EF (T_m = 42°C); Up - Upward temperature ramp, Down - downward temperature ramp*

10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 was used as the analysis buffer. The corresponding DNA duplex and compound **71** were mixed with 1:1 stoichiometry to a final concentration of 1.4 μM for each experiment. Prior to analysis, samples were heated to 90 °C, and cooled to a starting temperature of 25 °C with a heating rate of 5°C/min for each ramp. Denaturation profiles were recorded at λ = 260 nm from 40 to 85°C. The reported melting temperatures were defined as the minimum of the first derivative of the denature profile

4.6.2 Fluorescence Spectroscopy

Steady state fluorescence measurements were performed using a Horiba Fluorolog 3 fluorimeter. In the dark, 500 μL solutions were prepared in PBS buffer (50 mM Na^+, pH 7.5), of DNA duplex **GH** or **GH*** and compounds **71** and **72** up to a final concentration of 500 nM. The solutions were gently shaken and then allowed to sit at room temperature for 5 h. The measured sample was placed in a 500 μL quartz cell with 5 mm path length and kept at 20 °C during the measurement. Corrected emission spectra were collected from 400 nm to 650 nm using an excitation wavelength of 380 nm.

DNA Strands G, H, H*

Sequence:

G: 5'-GTA **GGG A**AC ATA TAT *GGA C*AT C-3'

H: 5'-GAT *GTC C*AT ATA TGT **TCC C**TA C-3'

H*: 5'-Cy3-GAT *GTC C*AT ATA TGT **TCC C**TA C-3'

Bold – Polyamide **71** binding site (Cou)

Itallics – Polyamide **72** binding site (YO)

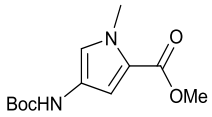
Appendix 1

Introduction

Between the initial submission in July 2011 and final submission, some compounds were re-synthesised to enable more thorough characterisation and in some cases, revised synthetic routes used. These additions have been inserted into appropriate experimental sections and have also been reproduced below.

Additions to Chapter 2

2.10.1.1 Synthesis of methyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxylate (21)[224]

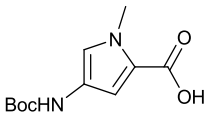


21

Recovered as a white solid consistent with previously reported characterisation.[224] Mp. 114-115°C (lit. 115-116°C[224])

NMR: δ_{1H}(300 MHz, $CDCl_3$): 1.43 (9H, s, Boc), 3.69 (3H, s, C**H**$_3$), 3.77 (3H, s, C**H**$_3$), 6.60 (1H, s, Ar-**H**), 7.09 (1H, s, Ar-**H**), 9.09 (1H, s, N**H**Boc); δ_{13H}(100 MHz, DMSO): 28.1, 36.0, 50.8, 78.5, 107.3, 118.6, 119.2, 123.1, 152.7, 160.7; IR (neat) υ_{max}: 3349 (w), 1713 (m), 1683 (s), 1587 (m), 1555 (m), 1457 (m), 1388 (m), 1243 (s), 1150 (s), 779 (m). ESI: M+Na^+ 277.1172 (calculated 277.1164);

2.10.1.2 Synthesis of 4-((tert-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxylic acid (BocPyOH)[183]

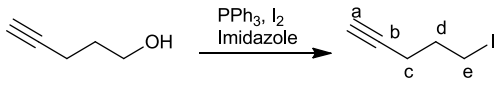


BocPyOH

Recovered as a finely divided white solid consistent with previously reported characterisation[183, 224] with the exception of the unexpectedly low melting point. Mp.140-142°C (lit. 160-161°C[224])

NMR: δ_{1H}(400 MHz, DMSO): 1.44 (9H, bs, Boc), 3.77 (3H, s, N-**C**$\mathbf{H_3}$), 6.58 (1H, bs, Ar-H), 7.04 (1H, s, Ar-H), 9.04 (s, 1H, N**H**Boc), 12.08 (s, 1H, COO**H**); δ_{13C}(100 MHz, DMSO): 28.1, 36.0, 107.4, 78.4, 118.7, 119.6, 122.8, 152.7, 161.8; IR (neat) υ_{max}: 3348 (w), 2978 (br, w), 1694 (s), 1667 (s), 1449 (s), 1390 (s), 1284 (s), 1157 (s), 1146 (s), 1062 (m); ESI: $M+Na^+$ 263.1015 (calculated 263.1008);

2.10.1.5 Synthesis of 5-iodopent-1-yne (31)[225]

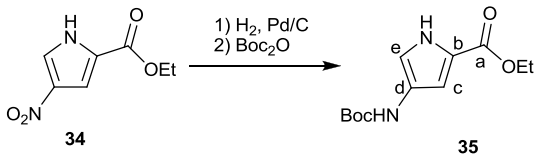


26

Recovered as a clear liquid consistent with previously reported characterisation.[225]

NMR: δ_{1H}(400MHz, $CDCl_3$): 1.99 (1H, t, ^4J = 2.6 Hz, H_a), 2.01 (2H, quintet, ^3J = 6.7 Hz, H_d), 2.28 (2H, td, ^3J = 6.7 Hz, ^4J = 2.6 Hz, H_c), 3.25 (2H, t, ^3J = 6.7 Hz, H_e); δ_{13C}(75 MHz, $CDCl_3$): 5.0 (C_e), 19.4 (C_c), 31.8 (C_d), 69.4 (C_a), 82.3 (C_b). IR (neat) υ_{max}: 3294 (m), 1427 (w), 1221 (m), 634 (s)

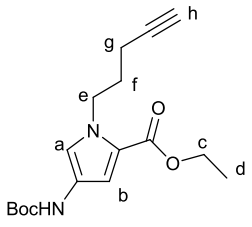
2.10.1.9 Synthesis of ethyl 4-(tert-butoxycarbonylamino)-1*H*-pyrrole-2-carboxylate (35)[228]



Recovered as a fine white solid consistent with all previously reported characterisation.[228] Mp. 189-190°C

NMR: δ_{1H}(400 MHz, DMSO): 1.26 (3H, t, ^3J = 7.1 Hz, $OCH_2C\mathbf{H}_3$), 1.44 (9H, s, Boc), 4.19 (2H, q, ^3J = 7.1 Hz, $OC\mathbf{H}_2CH_3$), 6.60 (1H, bs, CH^c), 6.94 (1H, bs, CH^e), 9.07 (1H, bs, NH), 11.49 (1H, bs, NH); δ_{13C}(75MHz, DMSO): 14.3, 18.1, 45.6, 59.4, 105.3, 112.5, 119.2, 124.9, 152.7, 160.3. IR (neat) υ_{max}: 3295 (m), 1677 (s), 1562 (m), 1391 (m), 1249 (m), 1217 (m), 1158 (m), 1118 (m), 771 (m); ESI: MH^+ 255.1348 (calculated 255.1345);

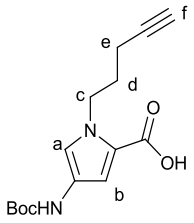
2.10.1.10 Synthesis of ethyl 4-((tert-butoxycarbonyl)amino)-1-(pent-4-yn-1-yl)-1*H*-pyrrole-2-carboxylate (36)



Mp. 92-94°C; NMR: δ_{1H}(400 MHz, DMSO): 1.26 (3H, t, ^3J = 7.1 Hz, H^d), 1.45 (9H, s, Boc), 1.82 (2H, qu, ^3J = 7.1 Hz, H^f), 2.09 (2H, td, ^3J = 7.1 Hz, ^4J = 2.6 Hz, H^g), 2.84 (1H, t, ^4J = 2.6 Hz, H^h), 4.19 (2H, q, ^3J = 7.1 Hz, H^c), 4.26 (2H, t, ^3J = 7.1 Hz, H^e), 6.67 (1H, bs, $CH^{a/b}$), 7.11 (1H, bs, $CH^{a/b}$), 9.12 (1H, bs, N**H**Boc); δ_{13C}(100 MHz, CD_3OD):

14.7, 16.2, 28.8, 31.4, 48.7, 61.0, 70.4, 80.5, 83.9, 110.2, 120.3, 120.4, 124.4 155.6, 162.4; IR (neat) υ_{max}: 3289 (m), 2979 (w), 2946 (w), 1685 (s), 1593 (s), 1526 (m), 1453 (m), 1396 (m), 1368 (m), 1274 (m), 1252 (s), 1219 (m), 1191 (m), 1092 (s), 1075 (m), 1027 (w), 988 (m), 886 (w), 829 (w), 780 (m), 759 (w), 692 (m); ESI: MH^+ 321.1822 (calculated 321.1814).

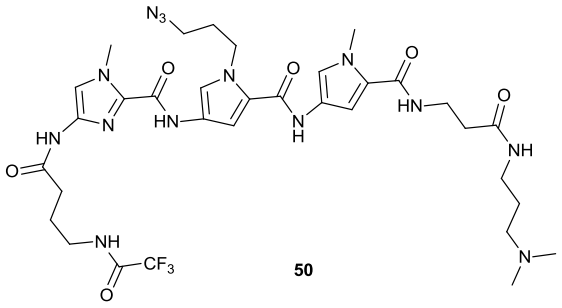
2.10.1.11 Synthesis of 4-(tert-butoxycarbonylamino)-1-(pent-4-ynyl)-1*H*-pyrrole-2-carboxylic acid (BocPy(///)OH)



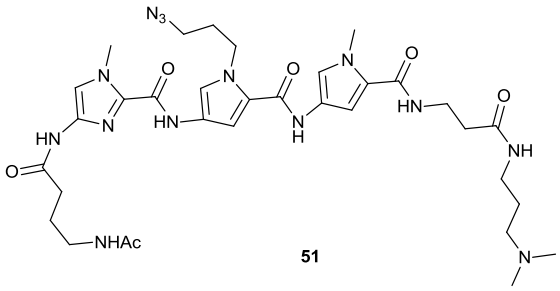
BocPy(///)OH

NMR: δ_{1H}(400 MHz, DMSO): 1.43 (9H, s, Boc), 1.81 (2H, qu, ^3J = 7.0 Hz, H^d), 2.02 (2H, td, ^3J = 7.0 Hz, ^4J = 2.6 Hz, H^e), 2.76 (1H, t, ^4J = 2.6 Hz, H^f), 4.31 (2H, t, ^3J = 7.0 Hz, H^c), 6.19 (1H, bs, $CH^{a/b}$), 6.67 (1H, bs, $CH^{a/b}$), 8.74 (1H, bs, N**H**Boc); δ_{1H}(400 MHz, CD_3OD): 1.39 (9H, s, Boc), 1.83 (2H, qu, ^3J = 6.9 Hz, H^d), 2.03 (2H, td, ^3J = 6.9 Hz, ^4J = 2.6 Hz, H^e), 2.17 (1H, t, ^4J = 2.6 Hz, H^f), 4.26 (2H, t, ^3J = 6.9 Hz, H^c), 6.64 (1H, bs, $CH^{a/b}$), 6.97 (1H, bs, $CH^{a/b}$); δ_{13C}(100 MHz, CD_3OD): 16.2 28.7, 31.5, 70.3, 80.5, 83.9, 110.5, 120.3, 120.7, 124.3, 155.6, 164.0; IR (neat) υ_{max}: 3369 (w), 3294 (w), 2969 (w), 2936 (w), 2616 (w), 1693 (m), 1671 (s), 1586 (m), 1549 (s), 1462 (m), 1390 (s), 1366 (m), 1293 (m), 1283 (m), 1442 (s), 1161 (s), 1093 (m), 1064 (m), 995 (m), 939 (w), 886 (m), 795 (m), 766 (m), 718 (m); ESI: MH^+ 293.1514 (calculated 293.1501).

2.10.2.5 Synthesis of CF_3-γ-ImPy(N_3)Py-β-Dp (50)



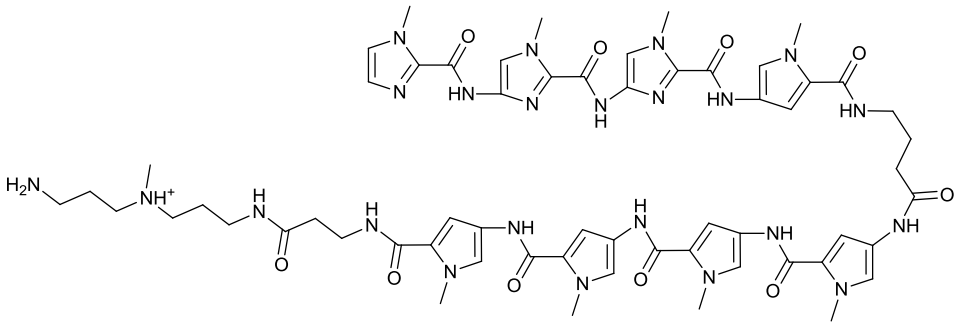
50



51

The synthesis of polyamide **51** was attempted using standard protocols on high loading Boc-β-Ala-PAM resin (0.050g, 0.58 mequiv/g) with the terminal acetyl group installed using acetic anhydride (100μL in 1.5 mL DMF, 30 minutes) followed by standard post coupling washes. The **BocPy(N_3)OH** building block was supplied by Rugerro Dondi. The crude polyamide was cleaved from the resin using 3-(dimethylamino)-1-propylamine (**Dp)** and purified by HPLC (10-90% MeCN in H_2O, 18 minutes, 0.1% TFA) to yield a mixture of acetyl capped polyamide **51** (2.5 mg, 12%) and TFA capped polyamide **50** (1.3 mg, 6%). **50**: MALDI: MH^+ 791.6286 (calculated 791.3671), **51**: MALDI: MH^+ 737.2277 (calculated 737.3955)

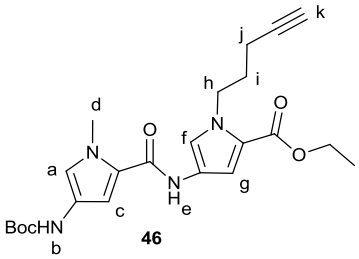
2.10.2.6 Synthesis of ImImImPy-γ-PyPyPyPy-β-NH_2 (52)



The product was synthesised using standard protocols on low loading Boc-β-Ala-PAM resin (0.1028 g, 0.26 mequiv/g), cleaved using 3,3'-Diamino-*N*-methyldipropylamine and purified by HPLC (MeCN in H_2O, 0.1%TFA, 10-20% 5 min, 20-45% 25 min, 45-

90% 2 min) to yield the title compound **52** as a white solid (15.6 mg, 46%). ESI: MH^+ 1266.6064 (calculated 1266.6164); MALDI: MH^+ 1266.7088 (calculated 1266.6164). NMR: δ_{1H}(400 MHz, CD_3OD): 1.90 (2H, qu, J = 6.3 Hz), 2.01 (2H, qu, J = 7.1 Hz), 2.32 (2H, t, J = 6.8 Hz), 2.43 (2H, t, J = 6.2 Hz), 2.76 (3H, s), 2.96 (2H, t, J = 7.4 Hz), 3.32 (2H, t, J = 6.3 Hz), 3.76 (3H, s, Py-Me), 3.77 (3H, s, Py-Me), 3.77 (3H, s, Py-Me), 3.79 (3H, s, Py-Me), 3.92 (3H, s, Im-Me), 3.96 (6H, s, Im-Me), 6.61-6.63 (1H, m), 6.72- 6.79 (4H, m), 6.85-6.87 (1H, m), 6.98-7.05 (4H, m), 7.11-7.14 (1H, m), 7.19 (1H, s), 7.32 (1H, s), 7.40 (1H, s). Additional multiplets at 3.40-3.55 (2H?, appears as numerous sharp peaks), 3.00-3.15 (2H?, broad signal with sharp quintet on top) and 1.79-1.82 (broad signal merging with quintet at δ 1.92).

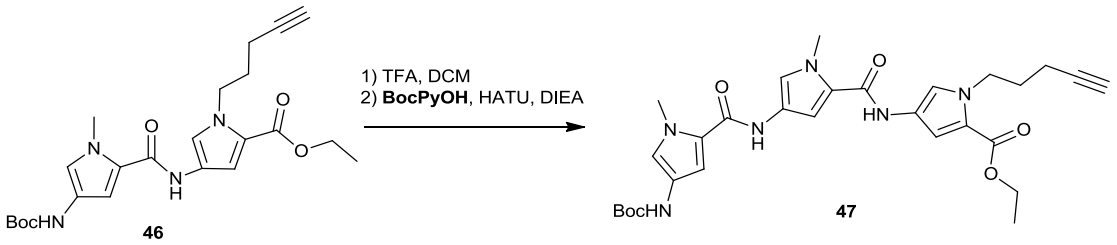
2.10.3.1 Synthesis of BocPyPy(///)OEt (46) (BTC Methodology)



46

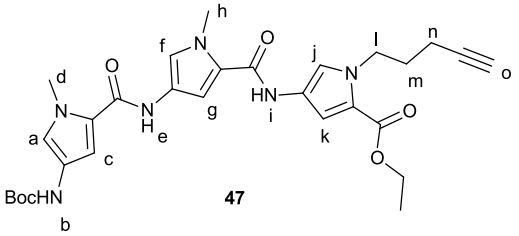
NMR: δ_{1H}(400 MHz, DMSO): 1.18 (2H, t, OC$\mathbf{H_2}$$CH_3$, J = 7.0 Hz), 1.46 (9H, s, Boc), 1.85 (2H, qu, H^i, J = 7.0 Hz), 2.12 (2H, td, H^j, J = 7.0 Hz, J = 2.6 Hz), 2.85 (1H, t, H^k, J = 2.6 Hz), 3.81 (3H, s, H^d), 4.21 (2H, q, $OCH_2C\mathbf{H_3}$, J = 7.0 Hz), 4.31 (2H, t, H^h, J = 7.0 Hz), 6.84 (1H, bs, $H^{a/c}$), 6.89 (1H, bs, $H^{a/c}$), 6.96 (1H, d, $H^{f/g}$, J = 2.0 Hz), 7.47 (1H, d, $H^{f/g}$, J = 2.0 Hz), 9.10 (1H, bs, H^b), 9.85 (1H, s, H^e); NMR: δ_{13C}(100 MHz, DMSO): 14.2, 14.9, 17.3, 18.4, 21.7, 28.2, 29.9, 36.0, 47.1, 59.4, 71.7, 83.3, 103.8, 109.0, 119.7, 122.4, 137.4, 146.1, 152.8, 158.4, 160.1

2.10.3.2 Synthesis of BocPyPyPy(///)OEt (47) - HATU methodology



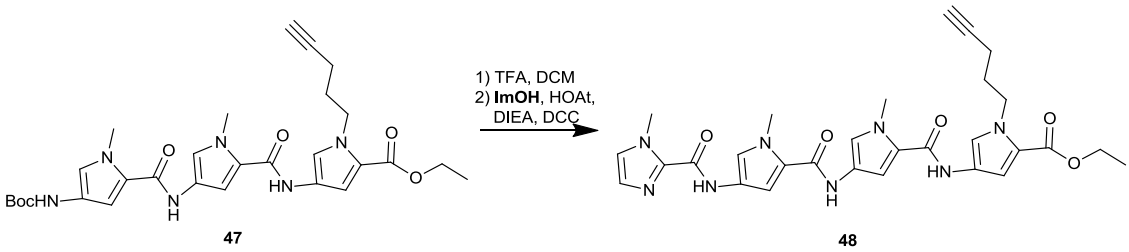
Dimer **46** (0.1361g, 0.31 mmol) was dissolved in a 1:1 mixture of trifluoroacetic acid and laboratory grade DCM (10 mL), stirred for 30 minutes and the solvent removed under reduced pressure. The crude mixture was diluted in toluene which was then removed under reduced pressure to remove additional residual trifluoroacteic acid and give the deprotected intermediate as an orange oil.

Separately, the **BocPyOH** building block (0.0741g, 0.31 mmol) and HATU (0.1055g, 0.28 mmol) were dissolved in dry DMF (2 mL) and DIEA (200 μL, 1.15 mmol) added. After stirring for 5 minutes, this solution was added to a mixture of the deprotected intermediate and DIEA (230 μL, 1.32 mmol) dissolved in dry DMF (3 mL) and stirred for 2 hours. The reaction mixture was diluted with EtOAc and washed with a saturated $NaHCO_3$ solution, dilute HCl (pH ≈ 2.5) and brine. The organic layer was dried using MgSO4 and the solvent removed under reduced pressure. The crude product was diluted in a small amount of DCM (2 mL) and precipitated with petrol 40-60°C to give the product as a finely divided brown solid (0.1458 g) which NMR analysis showed remained contaminated with DMF and was used crude for further reactions.



NMR: δ_{1H}(400 MHz, DMSO): 1.18 (2H, t, $OCH_2C\mathbf{H_3}$, J = 7.0 Hz), 1.45 (9H, s, Boc), 1.86 (2H, qu, H^m, J = 7.0 Hz), 2.12 (2H, td, H^j, J = 7.0 Hz, J = 2.5 Hz), 2.86 (1H, t, H^o, J = 2.5 Hz), 3.81 (3H, s, H^d), 3.85 (3H, s, H^h), 4.22 (2H, q, $OC\mathbf{H_2}CH_3$, J = 7.0 Hz), 4.32 (2H, t, H^l, J = 7.0 Hz), 6.84 (1H, bs, $H^{a/c}$), 6.90 (1H, bs, $H^{a/c}$), 6.97 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 7.07 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 7.22 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 7.49 (1H, d, $H^{f/g/j/k}$, J = 2.0 Hz), 9.08 (1H, bs, H^b), 9.85 (1H, s, $H^{e/i}$), 9.92 (1H, s, $H^{e/i}$).

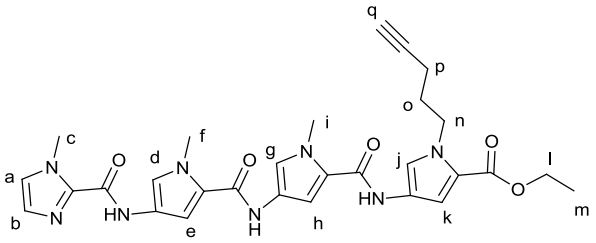
2.10.3.3 Synthesis of ImPyPyPy(///)OEt (48) – HATU Methodology



Crude **47** was dissolved in a 1:1 mixture of trifluoroacetic acid and wet DCM (25 mL), stirred for 30 minutes, then the solvent removed under reduced pressure. The crude mixture was diluted in toluene which was then removed under reduced pressure to remove additional residual trifluoroacteic acid and give the deprotected intermediate as an orange oil. Separately, a mixture of DCC (0.0421 g, 0.2 mmol), **ImOH** (1-methyl-1*H*-imidazole-2-carboxylic acid, 0.0284 g, 0.22 mmol) and HOAt (0.0303 g, 0.22 mmol) were dissolved in dry DMF (2 mL) and stirred for 45 minutes until the DCU precipitate was observed. This solution was then added to a mixture of the deprotected intermediate and DIEA (300 μL, 1.79 mmol) dissolved in dry DMF (3 mL) and stirred for 18 hours. The reaction mixture was filtered and the solvent removed under reduced pressure to yield the crude product (0.3158 g). Initial attempts to purify the product using manual reverse phase chromatography (MeCN, 0.1% TFA) and HPLC failed to separate the product **48** from the trifluoroacetic acid conjugate **49** and purification was achieved using normal phase chromatography (1:1 EtOAc:DCM, 5% NEt_3 to 10%

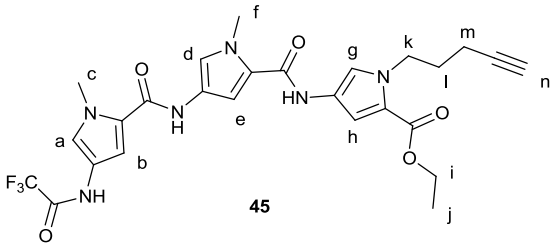
MeOH, 85% EtOAc, 5% NEt_3) which gave the product **48** as a orange oil (0.0358 g, 28%) and the trifluoroacetic acid conjugate **49** as a brown oil (0.0112 g, 9%). The NMR of both products showed some contamination by residual NEt_3.

Product 48



NMR: δ_{1H}(400 MHz, CD_3OD): 1.35 (3H, t, H^m, J = 7.1 Hz), 1.96 (2H, qu, H^o, J = 6.9 Hz), 2.17 (2H, td, H^p, ^3J = 6.9 Hz, ^4J = 2.5 Hz), 2.30 (1H, t, H^q, ^4J = 2.5 Hz), 3.92 (3H, s, $H^{f/i}$), 3.93 (3H, s, $H^{f/i}$), 4.06 (3H, s, H^c), 4.27 (2H, q, H^l, J = 7.1 Hz), 4.41 (2H, t, H^n, J = 6.9 Hz), 6.94 (2H, m, $H^{a/b/d/e/g/h/j/k}$), 6.99 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.9 Hz), 7.05 (1H, s, $H^{a/b/d/e/g/h/j/k}$), 7.22 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.3 Hz), 7.25 (1H, s, $H^{a/b/d/e/g/h/j/k}$), 7.31 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.5 Hz), 7.43 (1H, d, $H^{a/b/d/e/g/h/j/k}$, J = 1.8 Hz); δ_{13C}(100 MHz, CD_3OD): 13.3, 14.8, 30.0, 34.4, 35.4, 59.6, 69.0, 104.5, 105.1, 109.6, 119.2, 119.5, 120.3, 121.9, 125.9, 127.0; ESI: MH^+ 573.2579 (calculated 573.2574)

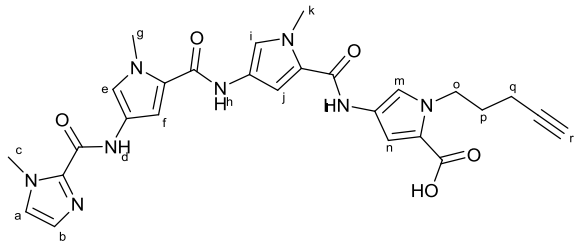
Impurity 49



NMR: δ_{1H}(400 MHz, CD_3OD): 1.86 (2H, qu, H^l, J = 6.8 Hz), 2.07 (2H, td, H^m, ^3J = 6.8 Hz, ^4J = 2.6 Hz), 2.22 (1H, t, H^n, ^4J = 2.6 Hz), 3.82 (3H, s, $H^{c/f}$), 3.83 (3H, s, $H^{c/f}$), 4.17 (2H, q, H^i, J = 7.2 Hz), 4.31 (2H, t, H^k, J = 6.8 Hz), 6.92 (1H, br s, $H^{a/b/d/e/g/h}$), 6.93 (1H, d, $H^{a/b/d/e/g/h}$, J = 2.0 Hz), 6.98 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.4 Hz), 7.17 (1H, d, $H^{a/b/d/e/g/h}$, J =

1.4 Hz), 7.19 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.5 Hz), 7.19 (1H, d, $H^{a/b/d/e/g/h}$, J = 1.8 Hz), H^{j} obscured by residual NEt_3 from the column chromatography at 1.32 ppm; δ_{13C}(100 MHz, CD_3OD): 14.8, 16.3, 31.5, 33.7, 36.9, 37.0, 54.9, 61.1, 63.7, 70.6, 106.6, 111.1, 121.0, 121.2, 121.7; δ_{19F}(376.5 MHz, CD_3CN): -75.0; ESI: MH^+ 561.2058 (calculated 561.2073)

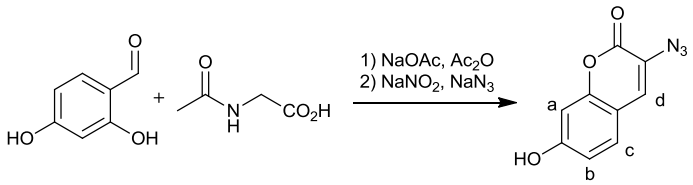
2.8.3.4 Synthesis of ImPyPyPy(///)OH (41)



NMR: δ_{1H}(400 MHz, DMSO): 1.87 (2H, qu, H^p, J = 6.9 Hz), 2.11 (2H, td, H^q, ^3J = 6.9 Hz, ^4J = 2.5 Hz), 2.85 (1H, t, H^r, ^4J = 2.5 Hz), 3.85 (3H, s, $H^{g/k}$), 3.86 (3H, s, $H^{g/k}$), 4.00 (3H, s, H^c), 4.33 (2H, t, H^o, J = 6.9 Hz), 6.87 (1H, s, $H^{a/b/e/f/i/j/m/n}$), 7.05 (1H,s, $H^{a/b/e/f/i/j/m/n}$), 7.08 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.9 Hz), 7.17 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.9 Hz), 7.25 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.6 Hz), 7.31 (1H, d, $H^{a/b/e/f/i/j/m/n}$, J = 1.6 Hz), 7.40 (1H, s, $H^{a/b/e/f/i/j/m/n}$), 7.43 (1H, s, $H^{a/b/e/f/i/j/m/n}$), 9.92 (1H, s, $H^{h/l}$), 9.98 (1H, s, $H^{h/l}$), 10.44 (1H, s, H^d), δ_{13C}(100 MHz, DMSO): 7.1, 30.1, 35.1, 36.0, 48.6, 62.5, 71.7, 83.5, 104.8, 104.9, 108.7, 116.1, 116.2, 118.5, 118.6, 119.0, 121.4, 122.2, 122.6, 123.0, 126.3, 127.0, 138.8, 156.1, 157.7, 158.0, 158.5; IR (neat) υ_{max}: 3435 (w), 3301 (w), 1662 (s), 1563 (m), 1433 (m), 1196 (s), 1130 (s), 841 (m), 801 (m), 724 (s).

Additions to Chapter 3

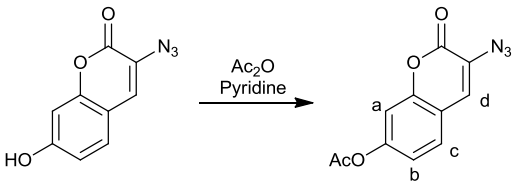
3.6.1.4 Synthesis of Coumarin azide (3-azido-7-hydroxy-2*H*-chromen-2-one) 16[177]



Recovered as a brown solid broadly consistent with previously reported characterisation.[177, 261, 262] Mp. >320°C (lit. 118-120°C[262])

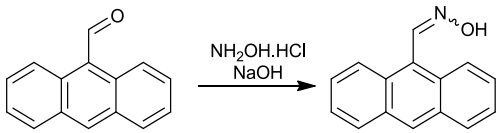
NMR: δ_{1H}(400 MHz, DMSO): 6.77 (1H, d, H^d, J = 2.3 Hz), 6.81 (1H, dd, H^c, J = 8.5 Hz, J = 2.3 Hz), 6.49 (1H, d, H^b, J = 8.5 Hz), 7.61 (1H, s, H^a), 10.52 (1H, s, OH); δ_{13C}(100 MHz, DMSO): 102.0, 111.2, 113.8, 121.0, 127.8, 129.0, 152.7, 157.3, 160.4; IR (neat) υ_{max}: 3048 (br, w), 2115 (s), 1706 (s), 1679 (m), 1614 (s), 1318 (m), 1259 (m), 1220 (s), 1025 (m), 996 (m), 835 (s).

3.6.1.5 Synthesis of Acylated Coumarin azide 59[177]



Recovered as a dark brown solid consistent with previously reported characterisation.[177] Mp. 120-122°C; NMR: δ_{1H}(400 MHz, DMSO): 2.30 (3H, s, O**Ac**), 7.17 (1H, dd, H^c, J = 8.5 Hz, J = 2.2 Hz), 7.33 (1H, d, H^d, J = 2.2 Hz), 7.69 (1H, s), 7.71 (1H, s); δ_{13C}(100 MHz, DMSO): 20.8, 109.9, 117.1, 119.2, 125.0, 126.1, 128.5, 151.2, 151.6, 156.8, 168.8; IR (neat) υ_{max}: 2125 (m), 1752 (m), 1722 (s), 1619 (m), 1430 (m), 1367 (m), 1328 (m), 1213 (s), 1142 (m), 1120 (m), 918 (s), 877 (m), 758 (s).

3.6.1.6 Synthesis of 1-anthraldehyde oxime (63) [263]

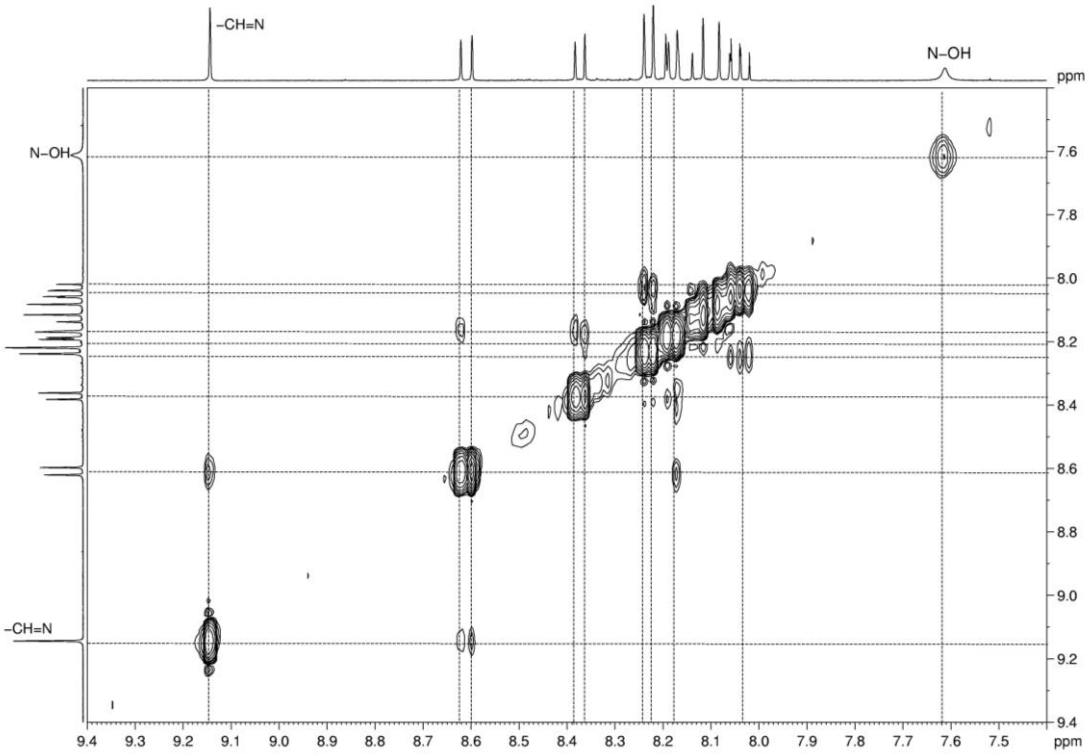


The crude product recovered as a mixture of syn and anti isomers (1.5:1) which on recrystallisation from ethanol was recovered as pale tan needles containing a single compound (major isomer) tentatively assigned as the *syn* isomer due to the absence of NOE cross peaks between the hydroxyl group and the oxime C-H. This compound was found to be consistent with previously reported characterisation.[264]

***Syn* isomer:**

NMR: δ_{1H}(400 MHz, DMSO): 7.53-7.62 (4H, m), 8.14 (2 H, d, J = 7.6 Hz), 8.45 (2 H, d, J = 8.8 Hz), 8.68 (1H, s), 9.21 (1H, s), 11.71 (1H, s). δ_{1H}(400 MHz, $CDCl_3$): 7.61 (1H, br s, N-O**H**), 8.01-8.25 (7H, m), 8.37 (1H, d, J = 8.1 Hz), 8.61 (1H, d, J = 9.3 Hz, R-C**H**=NOH); δ_{13C}(100 MHz, $CDCl_3$): 122.8, 125.0, 125.2, 125.7, 125.9, 126.2, 127.4, 128.4, 128.7, 149.7; IR (neat) υ_{max}: 3205 (br m), 1920 (br w), 1598 (m), 1584 (w), 1460 (m), 1343 (m) 1299 (m), 1237 (m), 1186 (s), 1056 (m), 963 9s), 928 (m), 899 (m), 834 (s) 818 (s), 790 (m), 749 (m), 711 (s), 674 (m); ESI: MH^+ 222.0921 (calculated 222.0919);

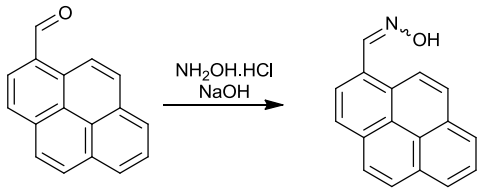
***Syn* NOE Enhancements:**



***Anti* isomer:**

NMR: δ_{1H}(400 MHz, DMSO): 7.71 (2H, t, J = 7.7 Hz), 7.84-7.90 (2H, m), 8.28-8.33 (4H, m), 9.08 (1H, s).

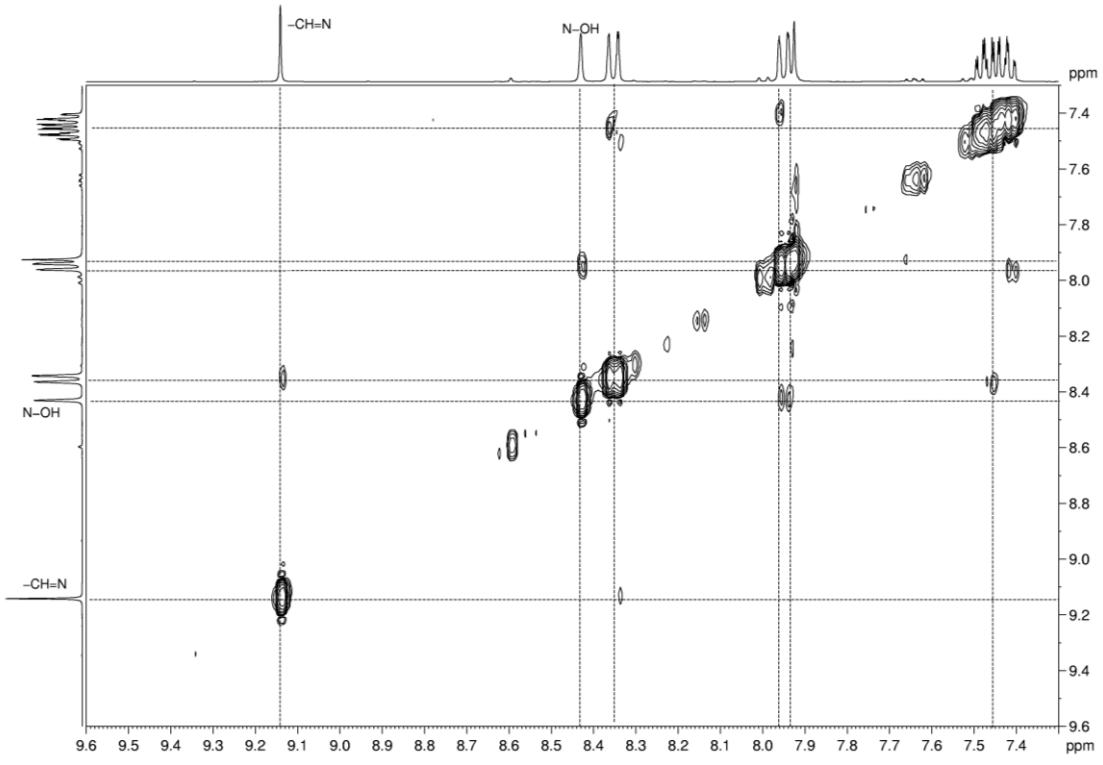
3.6.1.7 Synthesis of pyrene-1-carbaldehyde oxime (64)[259]



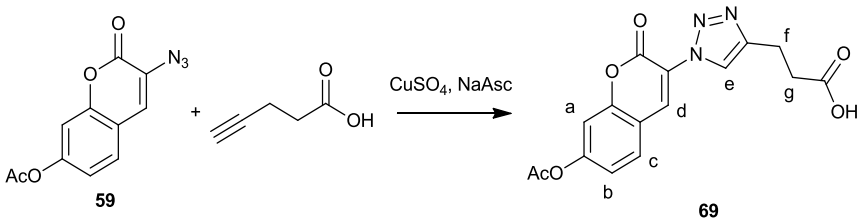
Recovered as bright yellow powder consisting of an 11:1 mixture of isomers. The major isomer was tentatively assigned as *syn* due to the absence of NOE interactions between the hydroxyl group and the oxime C-H (NOESY NMR experiment shown below) and found to be consistent with previously reported characterisation.[260]

***Syn* isomer:** NMR: δ_{1H}(400 MHz, $CDCl_3$): 7.40-7.50 (4H, m), 7.91-7.97 (3H, m), 8.33-8.37 (2H, m), 8.43 (1H, s), 9.14 (1H, s); δ_{13C}(100 MHz, $CDCl_3$): 123.7, 125.0, 125.4, 126.4, 126.8, 128.8, 128.9, 129.0, 129.4, 130.3, 131.3, 149.0; IR (neat) υ_{max}: 3274 (br m), 1625 (w), 1522 (w), 1442 (m), 1299 (w), 1257 (w), 1179 (w), 1159 (w), 1071 (w), 975 (s), 939 (m), 914 (m), 885 (m), 876 (s), 842 (m), 783 (m), 728 (s), 659 (m);

***Syn* NOE Enhancements:**



3.6.2.1 Synthesis of 3-(1-(7-acetoxy-2-oxo-*2H*-chromen-3-yl)-*1H*-1,2,3-triazol-4-yl)propanoic acid (69)



NMR: δ_{1H}(500 MHz, DMSO): 2.34 (3H, s, O**Ac**), 2.67 (2H, d, CH_2^g, J = 7.3 Hz) 2.98 (2H, d, CH_2^f, J = 7.3 Hz), 7.29 (1H, dd, H^c, J = 8.5 Hz, J = 2.2 Hz), 7.45 (1H, d, H^d, J = 2.2 Hz), 7.98 (1H, d, J= 8.5 Hz, H^b), 8.41 (1H, s), 8.71 (1H, s) 12.2 (1H, s, COO**H**); δ_{13C} (126 MHz, $CDCl_3$): 20.5 (CH_2), 20.9 (CH_2), 33.0 (OAc), 110.1, 116.1, 119.0, 119.5, 122.9, 130.3, 133.9, 146.2, 152.9, 153.4, 155.7, 168.7, 173.5; ESI: MH^+ 344.0893 (calculated 344.0883);

3.6.2.2 Synthesis of the Coumarin-polyamide Cou-Dpβ-PyPyPyPy-γ-PyImImIm (71)

NMR: δ_{1H}(500 MHz, DMSO): 1.72-1.83 (6H, m), 2.30 (2H, t, J = 7.4 Hz), 2.36 (2H, t, J = 7.4 Hz), 2.69-2.76 (4H, m), 2.97 (4H, t, J = 7.4 Hz), 3.06-3.16 (8H, m), 5.23 (2H, t, J = 6.2 Hz), 3.81 (3H, s, Py-Me), 3.82 (3H, s, Py-Me), 3.84 (3H, s, Py-Me), 3.85 (3H, s, Py-Me), 3.86 (3H, s, Py-Me), 4.01 (3H, s, Im-Me), 4.02 (3H, s, Im-Me), 4.04 (3H, s, Im-Me), 6.85 (1H, d, J = 2.1 Hz), 6.99 (1H, s), 7.02 (1H, d, J = 1.8 Hz), 7.08 (1H, s), 7.09-7.11 (2H, m), 7.20 (1H, s), 7.46 (1H, s), 7.59 (1H, s), 7.66 (1H, s), 7.75 (1H, d, J = 8.7 Hz), 8.30 (1H, s), 8.56 (1H, s), 9.15 (1H, br s, Cou-OH), 9.64 (1H, s), 9.85 (1H, s), 10.09 (1H, s), 10.33 (1H, s), 10.91 (1H, s).

References

1. International Roadmap Commitee in International Technology Roadmap for Semiconductors (2010 update)
2. E. Roduner, *Chem. Soc. Rev.*, 2006, **35**, 583-592.
3. V. Balzani, A. Credi and M. Venturi, *Chem.--Eur. J.*, 2002, **8**, 5524-5532.
4. P. W. K. Rothemund, *Nature*, 2006, **440**, 297-302.
5. N. C. Seeman, *Nature*, 2003, **421**, 427-431.
6. E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf and J. Kjems, *Nature*, 2009, **459**, 73-U75.
7. R. Chhabra, J. Sharma, Y. G. Ke, Y. Liu, S. Rinker, S. Lindsay and H. Yan, *J. Am. Chem. Soc.*, 2007, **129**, 10304-10305.
8. A. Kuzuya, M. Kimura, K. Numajiri, N. Koshi, T. Ohnishi, F. Okada and M. Komiyama, *Chembiochem*, 2009, **10**, 1811-1815.
9. B. Sacca, R. Meyer, M. Erkelenz, K. Kiko, A. Arndt, H. Schroeder, K. S. Rabe and C. M. Niemeyer, *Angew. Chem. Int. Ed.*, 2010, **49**, 9378-9383.
10. B. A. R. Williams, K. Lund, Y. Liu, H. Yan and J. C. Chaput, *Angew. Chem. Int. Ed.*, 2007, **46**, 3051-3054.
11. H. Yan, S. H. Park, G. Finkelstein, J. H. Reif and T. H. LaBean, *Science*, 2003, **301**, 1882-1884.
12. F. A. Aldaye and H. F. Sleiman, *J. Am. Chem. Soc.*, 2007, **129**, 4130-+.
13. Y. Q. Wen, C. K. McLaughlin, P. K. Lo, H. Yang and H. F. Sleiman, *Bioconjugate Chem.*, 2010, **21**, 1413-1416.
14. H. M. J. Carstairs, K. Lymperopoulos, A. N. Kapanidis, J. Bath and A. J. Turberfield, *Chembiochem*, 2009, **10**, 1781-1783.
15. *US Pat.*, US3261081, 1966.
16. T. Ito and S. Okazaki, *Nature*, 2000, **406**, 1027-1031.
17. M. Lundstrom, *Science*, 2003, **299**, 210-211.
18. J. V. Barth, G. Costantini and K. Kern, *Nature*, 2005, **437**, 671-679.
19. G. M. Whitesides, J. P. Mathias and C. T. Seto, *Science*, 1991, **254**, 1312-1319.
20. W. Lu and C. M. Lieber, *Nat Mater*, 2007, **6**, 841-850.
21. R. Kramer, J. M. Lehn and A. Marquis-Rigault, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5394-5398.
22. J. M. Lehn, *Angew. Chem., Int. Ed.*, 1988, **27**, 89-112.
23. R. Kramer, J. M. Lehn and A. Marquisrigault, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, **90**, 5394-5398.
24. D. P. Funeriu, J. M. Lehn, K. M. Fromm and D. Fenske, *Chem.--Eur. J.*, 2000, **6**, 2103-2111.
25. D. N. Reinhoudt and M. Crego-Calama, *Science*, 2002, **295**, 2403-2407.
26. J. H. Lee, N. Y. Wong, L. H. Tan, Z. Wang and Y. Lu, *J. Am. Chem. Soc.*, 2010.
27. S. Lyonnais, L. Goux-Capes, C. Escude, D. Cote, A. Filoramo and J. P. Bourgoin, *Small*, 2008, **4**, 442-446.
28. J. W. Liu and Y. Lu, *J. Am. Chem. Soc.*, 2007, **129**, 8634-8643.
29. W. Yoshida and Y. Yokobayashi, *Chem. Commun.*, 2007, 195-197.
30. M. Dhanasekaran, S. Negi and Y. Sugiura, *Acc. Chem. Res.*, 2006, **39**, 45-52.
31. P. G. Holder and M. B. Francis, *Angew. Chem. Int. Ed.*, 2007, **46**, 4370-4373.
32. R. S. Kane and A. D. Stroock, *Biotechnol. Progr.*, 2007, **23**, 316-319.

33. K. Rajagopal and J. P. Schneider, *Curr. Opin. Struct. Biol.*, 2004, **14**, 480-486.
34. Y. Astier, H. Bayley and S. Howorka, *Curr. Opin. Chem. Biol.*, 2005, **9**, 576-584.
35. L. Stryer, *Biochemistry*, 4th edn., W.H. Freeman, New York, 1995.
36. J. P. Sauvage, *Acc. Chem. Res.*, 1998, **31**, 611-619.
37. M. C. Jimenez-Molero, C. Dietrich-Buchecker and J. P. Sauvage, *Chem.--Eur. J.*, 2002, **8**, 1456-1466.
38. H. Tian and Q. C. Wang, *Chem. Soc. Rev.*, 2006, **35**, 361-374.
39. M. C. Jimenez, C. Dietrich-Buchecker and J. P. Sauvage, *Angew. Chem. Int. Ed.*, 2000, **39**, 3284-3287.
40. J. P. Collin, C. Dietrich-Buchecker, P. Gavina, M. C. Jimenez-Molero and J. P. Sauvage, *Acc. Chem. Res.*, 2001, **34**, 477-487.
41. M. C. Jimenez-Molero, C. Dietrich-Buchecker and J. P. Sauvage, *Chem. Commun.*, 2003, 1613-1616.
42. M. W. Ambrogio, T. A. Pecorelli, K. Patel, N. M. Khashab, A. Trabolsi, H. A. Khatib, Y. Y. Botros, J. I. Zink and J. F. Stoddarrt, *Org. Lett.*, 2010, **12**, 3304-3307.
43. J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin, E. DeIonno, Y. Luo, B. A. Sheriff, K. Xu, Y. S. Shin, H. R. Tseng, J. F. Stoddart and J. R. Heath, *Nature*, 2007, **445**, 414-417.
44. D. H. Qu, Q. C. Wang and H. Tian, *Angew. Chem. Int. Ed.*, 2005, **44**, 5296-5299.
45. A. S. Lubbe, N. Ruangsupapichat, G. Caroli and B. L. Feringa, *J. Org. Chem.*, 2011, **76**, 8599-8610.
46. S. Saha and J. F. Stoddart, *Chem. Soc. Rev.*, 2007, **36**, 77-92.
47. X. Ma and H. Tian, *Chem. Soc. Rev.*, 2010, **39**, 70-80.
48. M. R. Panman, P. Bodis, D. J. Shaw, B. H. Bakker, A. C. Newton, E. R. Kay, A. M. Brouwer, W. J. Buma, D. A. Leigh and S. Woutersen, *Science*, 2010, **328**, 1255-1258.
49. L. Gross, F. Mohn, N. Moll, P. Liljeroth and G. Meyer, *Science*, 2009, **325**, 1110-1114.
50. R. Erni, M. D. Rossell, C. Kisielowski and U. Dahmen, *Phys. Rev. Lett.*, 2009, **102**.
51. K. W. Urban, *Science*, 2008, **321**, 506-510.
52. S. Weiss, *Science*, 1999, **283**, 1676-1683.
53. E. Barkai, Y. J. Jung and R. Silbey, *Annu. Rev. Phys. Chem.*, 2004, **55**, 457-507.
54. W. E. Moerner, *J. Phys. Chem. B*, 2002, **106**, 910-927.
55. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737-738.
56. M. Egli, V. Tereshko, M. Teplova, G. Minasov, A. Joachimiak, R. Sanishvili, C. M. Weeks, R. Miller, M. A. Maier, H. Y. An, P. D. Cook and M. Manoharan, *Biopolymers*, 1998, **48**, 234-252.
57. K. Yanagi, G. G. Prive and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 201-214.
58. S. Neidle, *Nat. Prod. Rep.*, 2001, **18**, 291-309.
59. M. D. Wang, H. Yin, R. Landick, J. Gelles and S. M. Block, *Biophys. J.*, 1997, **72**, 1335-1346.
60. B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605-608.
61. P. Yin, H. M. T. Choi, C. R. Calvert and N. A. Pierce, *Nature*, 2008, **451**, 318-U314.
62. J. Bath and A. J. Turberfield, *Nat. Nanotechnol.*, 2007, **2**, 275-284.

63. T. Peng, C. Dohno and K. Nakatani, *Chembiochem*, 2007, **8**, 483-485.
64. C. Steinhauer, R. Jungmann, T. L. Sobey, F. C. Simmel and P. Tinnefeld, *Angew. Chem. Int. Ed.*, 2009, **48**, 8870-8873.
65. H. Yao, C. Q. Yi, C. H. Tzang, J. J. Zhu and M. S. Yang, *Nanotechnology*, 2007, **18**.
66. S. Pal, Z. T. Deng, B. Q. Ding, H. Yan and Y. Liu, *Angew. Chem. Int. Ed.*, 2010, **49**, 2700-2704.
67. J. Sharma, R. Chhabra, Y. Liu, Y. G. Ke and H. Yan, *Angew. Chem. Int. Ed.*, 2006, **45**, 730-735.
68. J. Sharma, Y. G. Ke, C. X. Lin, R. Chhabra, Q. B. Wang, J. Nangreave, Y. Liu and H. Yan, *Angew. Chem. Int. Ed.*, 2008, **47**, 5157-5159.
69. Q. B. Wang, H. N. Wang, C. X. Lin, J. Sharma, S. L. Zou and Y. Liu, *Chem. Commun.*, 2010, **46**, 240-242.
70. A. J. Mastroianni, S. A. Claridge and A. P. Alivisatos, *J. Am. Chem. Soc.*, 2009, **131**, 8455-8459.
71. J. W. Zheng, P. E. Constantinou, C. Micheel, A. P. Alivisatos, R. A. Kiehl and N. C. Seeman, *Nano Lett.*, 2006, **6**, 1502-1504.
72. M. H. Caruthers, *Acc. Chem. Res.*, 1991, **24**, 278-284.
73. S. L. Beaucage and R. P. Iyer, *Tetrahedron*, 1992, **48**, 2223-2311.
74. K. B. Mullis and F. A. Faloona, *Methods Enzymol.*, 1987, **155**, 335-350.
75. R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich, *Science*, 1988, **239**, 487-491.
76. S. Cheng, C. Fockler, W. M. Barnes and R. Higuchi, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 5695-5699.
77. A. A. Henry and F. E. Romesberg, *Curr. Opin. Chem. Biol.*, 2003, **7**, 727-733.
78. G. H. Clever, C. Kaul and T. Carell, *Angew. Chem. Int. Ed.*, 2007, **46**, 6226-6236.
79. G. H. Clever and T. Carell, *Angew. Chem. Int. Ed.*, 2007, **46**, 250-253.
80. D. I. Rozkiewicz, J. Gierlich, G. A. Burley, K. Gutsmiedl, T. Carell, B. J. Ravoo and D. N. Reinhoudt, *Chembiochem*, 2007, **8**, 1997-2002.
81. J. Gierlich, K. Gutsmiedl, P. M. E. Gramlich, A. Schmidt, G. A. Burley and T. Carell, *Chem.--Eur. J.*, 2007, **13**, 9486-9494.
82. S. Jager, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum and M. Famulok, *J. Am. Chem. Soc.*, 2005, **127**, 15071-15082.
83. Y. Chen, U. Narendra, L. E. Iype, M. M. Cox and P. A. Rice, *Molecular Cell*, 2000, **6**, 885-897.
84. C. M. N. Udo Feldkamp, *Angew. Chem. Int. Ed.*, 2006, **45**, 1856-1876.
85. J. P. Zhang, Y. Liu, Y. G. Ke and H. Yan, *Nano Lett.*, 2006, **6**, 248-251.
86. C. Song, Y. Q. Chen, S. J. Xiao, L. Ba, Z. Z. Gu, Y. Pan and X. Z. You, *Chem. Mater.*, 2005, **17**, 6521-6524.
87. J. D. Le, Y. Pinto, N. C. Seeman, K. Musier-Forsyth, T. A. Taton and R. A. Kiehl, *Nano Lett.*, 2004, **4**, 2343-2347.
88. R. P. Goodman, R. M. Berry and A. J. Turberfield, *Chem. Commun.*, 2004, 1372-1373.
89. S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf and W. M. Shih, *Nature*, 2009, **459**, 414-418.
90. C. M. Erben, R. P. Goodman and A. J. Turberfield, *J. Am. Chem. Soc.*, 2007, **129**, 6992-+.
91. Y. G. Ke, S. M. Douglas, M. H. Liu, J. Sharma, A. C. Cheng, A. Leung, Y. Liu, W. M. Shih and H. Yan, *J. Am. Chem. Soc.*, 2009, **131**, 15903-15908.

92. T. Kato, R. P. Goodman, C. M. Erben, A. J. Turberfield and K. Namba, *Nano Lett.*, 2009, **9**, 2747-2750.
93. J. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd edn., Springer, New York, 2006.
94. P. Atkins and J. d. Paula, *Atkins' Physical Chemistry*, 7th edn., Oxford University Press, Oxford, 2002.
95. D. L. Andrews, *Chem. Phys.*, 1989, **135**, 195-201.
96. G. D. Scholes, *Annu. Rev. Phys. Chem.*, 2003, **54**, 57-87.
97. K. E. Sapsford, L. Berti and I. L. Medintz, *Angew. Chem. Int. Ed.*, 2006, **45**, 4562-4588.
98. J. P. Clamme and A. A. Deniz, *Chemphyschem*, 2005, **6**, 74-77.
99. E. Haustein, M. Jahnz and P. Schwille, *Chemphyschem*, 2003, **4**, 745-748.
100. S. Kawahara, T. Uchimaru and S. Murata, *Chem. Commun.*, 1999, 563-564.
101. M. Heilemann, P. Tinnefeld, G. S. Mosteiro, M. Garcia-Parajo, N. F. Van Hulst and M. Sauer, *J. Am. Chem. Soc.*, 2004, **126**, 6514-6515.
102. M. Heilemann, R. Kasper, P. Tinnefeld and M. Sauer, *J. Am. Chem. Soc.*, 2006, **128**, 16864-16875.
103. G. Sanchez-Mosteiro, E. van Dijk, J. Hernando, M. Heilemann, P. Tinnefeld, M. Sauer, F. Koberlin, M. Patting, M. Wahl, R. Erdmann, N. F. van Hulst and M. F. Garcia-Parajo, *J. Phys. Chem. B*, 2006, **110**, 26349-26353.
104. Y. Ohya, K. Yabuki, M. Hashimoto, A. Nakajima and T. Ouchi, *Bioconjugate Chem.*, 2003, **14**, 1057-1066.
105. S. Vyawahare, S. Eyal, K. D. Mathews and S. R. Quake, *Nano Lett.*, 2004, **4**, 1035-1039.
106. A. K. Tong, S. Jockusch, Z. M. Li, H. R. Zhu, D. L. Akins, N. J. Turro and J. Y. Ju, *J. Am. Chem. Soc.*, 2001, **123**, 12923-12924.
107. J. K. Hannestad, P. Sandin and B. Albinsson, *J. Am. Chem. Soc.*, 2008, **130**, 15889-15895.
108. H. Ozhalici-Unal and B. A. Armitage, *Acs Nano*, 2009, **3**, 425-433.
109. A. L. Benvin, Y. Creeger, G. W. Fisher, B. Ballou, A. S. Waggoner and B. A. Armitage, *J. Am. Chem. Soc.*, 2007, **129**, 2025-2034.
110. M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 7586-7590.
111. W. S. Wade, M. Mrksich and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 8783-8794.
112. M. Mrksich, M. E. Parks and P. B. Dervan, *J. Am. Chem. Soc.*, 1994, **116**, 7983-7988.
113. J. W. Trauger, E. E. Baird and P. B. Dervan, *Nature*, 1996, **382**, 559-561.
114. S. A. Wolfe, L. Nekludova and C. O. Pabo, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 183-212.
115. T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223-1237.
116. D. S. Wuttke, M. P. Foster, D. A. Case, J. M. Gottesfeld and P. E. Wright, *J. Mol. Biol.*, 1997, **273**, 183-206.
117. P. A. Rice, S. W. Yang, K. Mizuuchi and H. A. Nash, *Cell*, 1996, **87**, 1295-1306.
118. L. Aravind, V. Anantharaman, S. Balaji, M. M. Babu and L. M. Iyer, *Fems Microbiol Rev*, 2005, **29**, 231-262.
119. E. Fraenkel, M. A. Rould, K. A. Chambers and C. O. Pabo, *J. Mol. Biol.*, 1998, **284**, 351-361.

120. Y. C. Kim, J. H. Geiger, S. Hahn and P. B. Sigler, *Nature*, 1993, **365**, 512-520.
121. J. M. Wong and E. Bateman, *Nucleic Acids Res.*, 1994, **22**, 1890-1896.
122. A. D. Baxevanis and D. Landsman, *Nucleic Acids Res.*, 1995, **23**, 1604-1613.
123. U. M. Ohndorf, M. A. Rould, Q. He, C. O. Pabo and S. J. Lippard, *Nature*, 1999, **399**, 708-712.
124. A. H. J. Wang, G. Ughetto, G. J. Quigley and A. Rich, *Biochemistry-Us*, 1987, **26**, 1152-1163.
125. C. L. Kielkopf, K. E. Erkkila, B. P. Hudson, J. K. Barton and D. C. Rees, *Nat. Struct. Biol.*, 2000, **7**, 117-121.
126. P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani and R. Romagnoli, *Med. Res. Rev.*, 2004, **24**, 475-528.
127. J. A. Kaizerman, M. L. Gross, Y. G. Ge, S. White, W. H. Hu, J. X. Duan, E. E. Baird, K. W. Johnson, R. D. Tanaka, H. E. Moser and R. W. Burli, *J. Med. Chem.*, 2003, **46**, 3914-3929.
128. A. C. Finlay, F. A. Hochstein, B. A. Sobin and F. X. Murphy, *J. Am. Chem. Soc.*, 1951, **73**, 341-343.
129. F. Arcamone, S. Penco, P. Orezzi, V. Nicolella and A. Pirelli, *Nature*, 1964, **203**, 1064-1065.
130. J. G. Pelton and D. E. Wemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 5723-5727.
131. J. G. Pelton and D. E. Wemmer, *J. Am. Chem. Soc.*, 1990, **112**, 1393-1399.
132. P. B. Dervan and R. W. Burli, *Curr. Opin. Chem. Biol.*, 1999, **3**, 688-693.
133. J. M. Turner, S. E. Swalley, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1998, **120**, 6219-6226.
134. S. E. Swalley, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1999, **121**, 1113-1120.
135. S. Foister, M. A. Marques, R. M. Doss and P. B. Dervan, *Biorg. Med. Chem.*, 2003, **11**, 4333-4340.
136. A. Heckel and P. B. Dervan, *Chem.--Eur. J.*, 2003, **9**, 3353-3366.
137. W. Zhang, M. Minoshima and H. Sugiyama, *J. Am. Chem. Soc.*, 2006, **128**, 14905-14912.
138. W. Zhang, T. Bando and H. Sugiyama, *J. Am. Chem. Soc.*, 2006, **128**, 8766-8776.
139. M. E. Farkas, S. M. Tsai and P. B. Dervan, *Biorg. Med. Chem.*, 2007, **15**, 6927-6936.
140. C. Dose, M. E. Farkas, D. M. Chenoweth and P. B. Dervan, *J. Am. Chem. Soc.*, 2008, **130**, 6859-6866.
141. M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 1376-1380.
142. J. W. Lown, K. Krowicki, U. G. Bhat, A. Skorobogaty, B. Ward and J. C. Dabrowiak, *Biochemistry-Us*, 1986, **25**, 7408-7416.
143. C. L. Kielkopf, E. E. Baird, P. D. Dervan and D. C. Rees, *Nat. Struct. Biol.*, 1998, **5**, 104-109.
144. S. B. Singh, Ajay, D. E. Wemmer and P. A. Kollman, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 7673-7677.
145. S. White, E. E. Baird and P. B. Dervan, *Biochemistry-Us*, 1996, **35**, 12532-12537.
146. D. M. Chenoweth and P. B. Dervan, *J. Am. Chem. Soc.*, 2010, **132**, 14521-14529.

147. D. M. Herman, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1998, **120**, 1382-1391.
148. M. E. Farkas, B. C. Li, C. Dose and P. B. Dervan, *Biorg. Med. Chem. Lett.*, 2009, **19**, 3919-3923.
149. S. M. Tsai, M. E. Farkas, C. J. Chou, J. M. Gottesfeld and P. B. Dervan, *Nucleic Acids Res.*, 2007, **35**, 307-316.
150. S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird and P. B. Dervan, *Nature*, 1998, **391**, 468-471.
151. D. Renneberg and P. B. Dervan, *J. Am. Chem. Soc.*, 2003, **125**, 5707-5716.
152. C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan and D. C. Rees, *Science*, 1998, **282**, 111-115.
153. C. L. Kielkopf, R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan and D. C. Rees, *J. Mol. Biol.*, 2000, **295**, 557-567.
154. B. Wellenzohn, M. J. Loferer, M. Trieb, C. Rauch, R. H. Winger, E. Mayer and K. R. Liedl, *J. Am. Chem. Soc.*, 2003, **125**, 1088-1095.
155. J. M. Turner, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1997, **119**, 7636-7644.
156. S. E. Swalley, E. E. Baird and P. B. Dervan, *Chem.--Eur. J.*, 1997, **3**, 1600-1607.
157. A. Viger and P. B. Dervan, *Biorg. Med. Chem.*, 2006, **14**, 8539-8549.
158. N. G. Nickols, C. S. Jacobs, M. E. Farkas and P. B. Dervan, *Nucleic Acids Res.*, 2007, **35**, 363-370.
159. R. M. Doss, M. A. Marques, S. Foister, D. M. Chenoweth and P. B. Dervan, *J. Am. Chem. Soc.*, 2006, **128**, 9074-9079.
160. P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284-299.
161. D. M. Chenoweth, D. A. Harki, J. W. Phillips, C. Dose and P. B. Dervan, *J. Am. Chem. Soc.*, 2009, **131**, 7182-7188.
162. H. D. Arndt, K. E. Hauschild, D. P. Sullivan, K. Lake, P. B. Dervan and A. Z. Ansari, *J. Am. Chem. Soc.*, 2003, **125**, 13322-13323.
163. R. E. Bremer, N. R. Wurtz, J. W. Szewczyk and P. B. Dervan, *Biorg. Med. Chem.*, 2001, **9**, 2093-2103.
164. T. Best, Dissertation (Ph.D.), California Institute of Technology, 2005.
165. T. P. Best, B. S. Edelson, N. G. Nickols and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12063-12068.
166. B. S. Edelson, T. P. Best, B. Olenyuk, N. G. Nickols, R. M. Doss, S. Foister, A. Heckel and P. B. Dervan, *Nucleic Acids Res.*, 2004, **32**, 2802-2818.
167. C. F. Hsu, Dissertation (Ph.D.), California Institute of Technology, 2009.
168. M. Endo and H. Sugiyama, *Chembiochem*, 2009, **10**, 2420-2443.
169. J. D. Cohen, J. P. Sadowski and P. B. Dervan, *J. Am. Chem. Soc.*, 2008, **130**, 402-+.
170. J. D. Cohen, J. P. Sadowski and P. B. Dervan, *Angew. Chem. Int. Ed.*, 2007, **46**, 7956-7959.
171. Chayan K. N. G. R. Partha P. P. B. B. M. F. A. H. Thorsten L. Schmidt, *Angew. Chem. Int. Ed.*, 2007, **46**, 4382-4384.
172. V. C. Rucker, S. Foister, C. Melander and P. B. Dervan, *J. Am. Chem. Soc.*, 2003, **125**, 1195-1202.
173. C. F. Hsu and P. B. Dervan, *Biorg. Med. Chem. Lett.*, 2008, **18**, 5851-5855.
174. Y. N. V. Gopal and M. W. Van Dyke, *Biochemistry-Us*, 2003, **42**, 6891-6903.
175. W. Su, personal communication

176. H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem. Int. Ed.*, 2001, **40**, 2004-2021.
177. K. Sivakumar, F. Xie, B. M. Cash, S. Long, H. N. Barnhill and Q. Wang, *Org. Lett.*, 2004, **6**, 4603-4606.
178. B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org. Lett.*, 2000, **2**, 1939-1941.
179. E. Saxon, J. I. Armstrong and C. R. Bertozzi, *Org. Lett.*, 2000, **2**, 2141-2143.
180. Y. H. Pei and W. H. Moos, *Tetrahedron Lett.*, 1994, **35**, 5825-5828.
181. J. W. Bode, R. M. Fox and K. D. Baucom, *Angew. Chem. Int. Ed.*, 2006, **45**, 1248-1252.
182. N. Carrillo, E. A. Davalos, J. A. Russak and J. W. Bode, *J. Am. Chem. Soc.*, 2006, **128**, 1452-1453.
183. E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1996, **118**, 6141-6146.
184. N. R. Wurtz, J. M. Turner, E. E. Baird and P. B. Dervan, *Org. Lett.*, 2001, **3**, 1201-1203.
185. P. O. Krutzik and A. R. Chamberlin, *Biorg. Med. Chem. Lett.*, 2002, **12**, 2129-2132.
186. J. M. Belitsky, D. H. Nguyen, N. R. Wurtz and P. B. Dervan, *Biorg. Med. Chem.*, 2002, **10**, 2767-2774.
187. W. Su, S. J. Gray, R. Dondi and G. A. Burley, *Org. Lett.*, 2009, **11**, 3910-3913.
188. D. M. Chenoweth, D. A. Harki and P. B. Dervan, *J. Am. Chem. Soc.*, 2009, **131**, 7175-7181.
189. M. F. Brana, M. Moran, M. J. P. Devega, I. Pitaromero and N. Walker, *Tetrahedron*, 1995, **51**, 9127-9138.
190. G. Burley, personal communication
191. A. Nose and T. Kudo, *Chem Pharm Bull*, 1981, **29**, 1159-1161.
192. *US Pat.*, US 2683721, 1954.
193. *US Pat.*, US 2576311, 1951.
194. B. Ganem and J. O. Osby, *Chem. Rev.*, 1986, **86**, 763-780.
195. Y. Okamoto, Y. Nitta, T. Imanaka and S. Teranishi, *J. Chem. Soc., Faraday Trans. 1*, 1979, **75**, 2027-2039.
196. P. C. Maybury, R. W. Mitchell and Hawthorn.Mf, *Journal of the Chemical Society-Chemical Communications*, 1974, 534-535.
197. J. A. Schreifels, P. C. Maybury and W. E. Swartz, *J. Org. Chem.*, 1981, **46**, 1263-1269.
198. S. W. Heinzman and B. Ganem, *J. Am. Chem. Soc.*, 1982, **104**, 6801-6802.
199. J. O. Osby, S. W. Heinzman and B. Ganem, *J. Am. Chem. Soc.*, 1986, **108**, 67-72.
200. S. Caddick, D. B. Judd, A. K. D. Lewis, M. T. Reich and M. R. V. Williams, *Tetrahedron*, 2003, **59**, 5417-5423.
201. A. Nose and T. Kudo, *Chem Pharm Bull*, 1986, **34**, 3905-3909.
202. H. C. Brown and K. Ichikawa, *J. Am. Chem. Soc.*, 1961, **83**, 4372-4374.
203. I. Singh, personal communication
204. R. Kumar and J. W. Lown, *Org Biomol Chem*, 2003, **1**, 3327-3342.
205. R. Dondi, personal communication
206. G. D. Yadav and Y. B. Jadhav, *Langmuir*, 2002, **18**, 5995-6002.
207. R. B. Merrifield, *Science*, 1965, **150**, 178-185.
208. F. Albericio, *Curr. Opin. Chem. Biol.*, 2004, **8**, 211-221.
209. B. L. Bray, *Nat Rev Drug Discov*, 2003, **2**, 587-593.
210. E. Kaiser, Colescot.Rl, Bossinge.Cd and P. I. Cook, *Analytical Biochemistry*, 1970, **34**, 595-&.

211. T. Christensen, *Acta Chem. Scand., Ser. B*, 1979, **33**, 763-766.
212. M. Schnolzer, P. Alewood, A. Jones, D. Alewood and S. B. H. Kent, *Int. J. Pept. Protein Res.*, 1992, **40**, 180-193.
213. E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606-631.
214. L. A. Carpino, *J. Am. Chem. Soc.*, 1993, **115**, 4397-4398.
215. L. A. Carpino, A. Elfaham, C. A. Minor and F. Albericio, *Chem. Commun.*, 1994, 201-203.
216. L. P. Miranda and P. F. Alewood, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 1181-1186.
217. D. A. Pearson, M. Blanchette, M. L. Baker and C. A. Guindon, *Tetrahedron Lett.*, 1989, **30**, 2739-2742.
218. N. A. Sole and G. Barany, *J. Org. Chem.*, 1992, **57**, 5399-5403.
219. W. Su, M. Schuster, C. R. Bagshaw, U. Rant and G. A. Burley, *Angew. Chem. Int. Ed.*, 2011, **50**, 2712-2715.
220. C. F. Hsu, J. W. Phillips, J. W. Trauger, M. E. Farkas, J. M. Belitsky, A. Heckel, B. Z. Olenyuk, J. W. Puckett, C. C. C. Wang and P. B. Dervan, *Tetrahedron*, 2007, **63**, 6146-6151.
221. H. Eckert and B. Forster, *Angew. Chem., Int. Ed.*, 1987, **26**, 894-895.
222. E. Falb, T. Yechezkel, Y. Salitra and C. Gilon, *J. Pept. Res.*, 1999, **53**, 507-517.
223. T. L. Schmidt and A. Heckel, *Nano Lett.*, 2011, **11**, 1739-1742.
224. D. Jaramillo, Q. Liu, J. Aldrich-Wright and Y. Tor, *J. Org. Chem.*, 2004, **69**, 8151-8153.
225. L. G. Luyt, H. M. Bigott, M. J. Welch and J. A. Katzenellenbogen, *Biorg. Med. Chem.*, 2003, **11**, 4977-4989.
226. D. M. Bailey, R. E. Johnson and N. F. Albertson, *Organic Syntheses*, 1988, **50-9**, 618-619.
227. C. Schmuck, V. Bickert, M. Merschky, L. Geiger, D. Rupprecht, J. Dudaczek, P. Wich, T. Rehm and U. Machon, *Eur. J. Org. Chem.*, 2008, 324-329.
228. R. E. Bremer, J. W. Szewczyk, E. E. Baird and P. B. Dervan, *Biorg. Med. Chem.*, 2000, **8**, 1947-1955.
229. S. Brase, C. Gil, K. Knepper and V. Zimmermann, *Angew. Chem. Int. Ed.*, 2005, **44**, 5188-5240.
230. H. C. Kolb and K. B. Sharpless, *Drug Discovery Today*, 2003, **8**, 1128-1137.
231. S. K. Mamidyala and M. G. Finn, *Chem. Soc. Rev.*, 2010, **39**, 1252-1261.
232. M. Meldal and C. W. Tornoe, *Chem. Rev.*, 2008, **108**, 2952-3015.
233. J. A. Johnson, M. G. Finn, J. T. Koberstein and N. J. Turro, *Macromol. Rapid Commun.*, 2008, **29**, 1052-1072.
234. A. H. El-Sagheer and T. Brown, *Chem. Soc. Rev.*, 2010, **39**, 1388-1405.
235. P. L. Golas and K. Matyjaszewski, *Chem. Soc. Rev.*, 2010, **39**, 1338-1354.
236. V. O. Rodionov, V. V. Fokin and M. G. Finn, *Angew. Chem. Int. Ed.*, 2005, **44**, 2210-2215.
237. V. D. Bock, H. Hiemstra and J. H. van Maarseveen, *Eur. J. Org. Chem.*, 2006, 51-68.
238. J. E. Hein and V. V. Fokin, *Chem. Soc. Rev.*, 2010, **39**, 1302-1315.
239. V. O. Rodionov, S. I. Presolski, S. Gardinier, Y. H. Lim and M. G. Finn, *J. Am. Chem. Soc.*, 2007, **129**, 12696-12704.
240. V. O. Rodionov, S. I. Presolski, D. D. Diaz, V. V. Fokin and M. G. Finn, *J. Am. Chem. Soc.*, 2007, **129**, 12705-12712.
241. A. T. Poulin-Kerstien and P. B. Dervan, *J. Am. Chem. Soc.*, 2003, **125**, 15811-15821.

242. H. Tanaka, Y. Tanaka, M. Minoshima, S. Yamaguchi, S. Fuse, T. Doi, S. Kawauchi, H. Sugiyama and T. Takahashi, *Chem Commun (Camb)*, 2010, **46**, 5942-5944.
243. W. H. Binder and C. Kluger, *Curr. Org. Chem.*, 2006, **10**, 1791-1815.
244. L. Zhang, X. G. Chen, P. Xue, H. H. Y. Sun, I. D. Williams, K. B. Sharpless, V. V. Fokin and G. C. Jia, *J. Am. Chem. Soc.*, 2005, **127**, 15998-15999.
245. B. C. Boren, S. Narayan, L. K. Rasmussen, L. Zhang, H. T. Zhao, Z. Y. Lin, G. C. Jia and V. V. Fokin, *J. Am. Chem. Soc.*, 2008, **130**, 8923-8930.
246. R. B. Turner, A. D. Jarrett, P. Goebel and B. J. Mallon, 1973, **95**, 790-792.
247. F. Himo, T. Lovell, R. Hilgraf, V. V. Rostovtsev, L. Noodleman, K. B. Sharpless and V. V. Fokin, *J. Am. Chem. Soc.*, 2005, **127**, 210-216.
248. N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2004, **126**, 15046-15047.
249. J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzit, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16793-16797.
250. J. A. Codelli, J. M. Baskin, N. J. Agard and C. R. Berozzi, *J. Am. Chem. Soc.*, 2008, **130**, 11486-11493.
251. X. H. Ning, J. Guo, M. A. Wolfert and G. J. Boons, *Angew. Chem. Int. Ed.*, 2008, **47**, 2253-2255.
252. E. Lallana, E. Fernandez-Megia and R. Riguera, *J. Am. Chem. Soc.*, 2009, **131**, 5748-+.
253. S. Grecian and V. V. Fokin, *Angew. Chem. Int. Ed.*, 2008, **47**, 8285-8287.
254. V. Algay, I. Singh and F. Heaney, *Org Biomol Chem*, 2010, **8**, 391-397.
255. E. Coutouli-Argyropoulou, P. Lianis, M. Mitakou, A. Giannoulis and J. Nowak, *Tetrahedron*, 2006, **62**, 1494-1501.
256. K. S. Wankhede, V. V. Vaidya, P. S. Sarang, M. M. Salunkhe and G. K. Trivedi, *Tetrahedron Lett.*, 2008, **49**, 2069-2073.
257. Y. J. Shang and Y. G. Wang, *Tetrahedron Lett.*, 2002, **43**, 2247-2249.
258. I. Singh, J. S. Vyle and F. Heaney, *Chem. Commun.*, 2009, 3276-3278.
259. P. Aschwanden, L. Kvaerno, R. W. Geisser, F. Kleinbeck and E. M. Carreira, *Org. Lett.*, 2005, **7**, 5741-5742.
260. H. M. Lee, E. Luna, M. Hinz, J. J. Stezowski, A. S. Kiselyov and R. G. Harvey, *J. Org. Chem.*, 1995, **60**, 5604-5613.
261. L. Yi, J. Shi, S. Gao, S. Li, C. Niu and Z. Xi, *Tetrahedron Lett.*, 2009, **50**, 759-762.
262. R. Rajaganesh, P. Ravinder, V. Subramanian and T. M. Das, *Carbohydr. Res.*, 2011, **346**, 2327-2336.
263. I. Singh and F. Heaney, *Chem. Commun.*, 2011, **47**, 2706-2708.
264. M. Horiguchi and Y. Ito, *Tetrahedron*, 2007, **63**, 12286-12293.
265. A. J. Hampshire, D. A. Rusling, V. J. Broughton-Head and K. R. Fox, *Methods*, 2007, **42**, 128-140.
266. S. S. Jain and T. D. Tullius, *Nat Protoc*, 2008, **3**, 1092-1100.
267. D. J. Galas and A. Schmitz, *Nucleic Acids Res.*, 1978, **5**, 3157-3170.
268. J. W. Trauger and P. B. Dervan, in *Drug-Nucleic Acid Interactions*, Academic Press Inc, San Diego, Editon edn., 2001, vol. 340, pp. 450-466.
269. D. S. Pilch, N. Poklar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 8306-8311.
270. P. L. James, L. Le Strat, U. Ellervik, C. Bratwall, B. Norden, T. Brown and K. R. Fox, *Biophys. Chem.*, 2004, **111**, 205-212.

271. D. S. Pilch, N. Poklar, E. E. Baird, P. B. Dervan and K. J. Breslauer, *Biochemistry-Us*, 1999, **38**, 2143-2151.
272. S. E. Swalley, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1997, **119**, 6953-6961.
273. S. E. Swalley, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1996, **118**, 8198-8206.
274. D. M. Chenoweth, A. Viger and P. B. Dervan, *J. Am. Chem. Soc.*, 2007, **129**, 2216-2217.
275. R. Baliga, E. E. Baird, D. M. Herman, C. Melander, P. B. Dervan and D. M. Crothers, *Biochemistry-Us*, 2001, **40**, 3-8.