Application of Chemically Modified Oligonucleotides in Nanopore Sensing and DNA Nano – Biotechnology

Nicholas J. Mitchell

University College London

Submitted to the Department of Chemistry for the degree of Doctor of Philosophy

I, Nick Mitchell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

N. J. Mitchell

"I may not have gone where I intended to go, but I think I have ended up where I intended to be."

Douglas Adams
1952 – 2001

Acknowledgements

Firstly, I'd like to thank my supervisor, Dr. Stefan Howorka, for his guidance and advice throughout this project. A big thank-you to Drs. Helen Kinns, Vinciane Borsenberger, Jin Masania and Jonathan Richardson for their advice in all things chemistry (and otherwise) and especially for letting me moan at them over a pint. My parents Chris and Alison deserve a special acknowledgement for their unwavering love, their constant support and those occasional Sunday roasts. I should also mention all those friends outside of the lab who had little interest in my project and thankfully stopped me going on and on about it. Finally I'd like to thank Natalie for her love and her patience, particularly when three long years turned into four.

Abstract

This thesis describes how targeted chemical modification can enhance the properties of nucleic acids for use in (i) nanopore analytics and (ii) nanobiotechnology.

In nanopore analytics, individual molecules are detected as they pass a nanoscale pore to give rise to detectable blockades in ionic current. Despite progress in the sensing of a multitude of molecular species, the analytical resolution in the sensing of DNA is poor as individual bases in passing strands cannot be resolved due to the high speed of translocation. Here a new approach is presented which slows down single stranded DNA and enables the detection of multiple separate bases. Chemical tags are attached to bases, which cause a steric blockade each time a modified base passes a narrow pore. The resulting characteristic current signatures are specific for the chemical composition and the size of the tags. The unique electrical signatures can be exploited to encode sequence information as demonstrated for the discrimination between drug resistance-conferring point mutations. In addition, the generation of nucleotides with tailored properties may help develop a fast nanopore approach to size highly repetitive DNA sequences for forensic applications.

In DNA nanobiotechnology, oligonucleotides are self-assembled via hybridization to generate higher-order structures of defined geometry. Here, the functional range of DNA nanostructures is expanded by chemically modifying the constituent nucleic acids. Firstly, tetrahedron-shaped nanostructures are demonstrated to act as a scaffold to assemble a multitude of different chemical groups at tunable stoichiometry and at geometrically defined sites. The new molecular entities exhibit functional properties beneficial in biosensing and diagnostics. In addition, an approach is presented to achieve self-assembly between DNA-strands via covalently attached tags that form reversible yet tight metal chelate complexes. This chemical strategy to form supramolecular structures can potentially be extended to protein or peptide networks of interest in basic science and technology.

Table of Contents

Table of Figures

Chapter One - Figures

Chapter One - Tables

Chapter Two - Schemes

Abbreviations

AEC	Anion Exchange Chromatography
AFM	Atomic Force Microscopy
αHL	Alpha Hemolysin
AZT	Azidothymidine
BT	Biotin
Cbz	Carboxylbenzyl
CEP	Cyanoethyl diisopropylphosphoramidite
DCC	Dicyclohexanecarbodiimide
DCM	Dichloromethane
ddATP	Dideoxyadenosine Triphosphate
DIPEA	Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	Dimethoxytrityl
DNA	Deoxyribose Nucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DPDS	2,2'-dipyripyldisulfide
DS	Disulfide
DTT	Dithiolthreitol
dUTP	Deoxuridine Triphosphate
DX	Double Crossover
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDT	1,2-Ethane dithiol
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray Ionisation
FMOC	Fluorenylmethyloxycarbonyl
FPLC	Fast Protein Liquid Chromatography
HBTU	O-Benzotriazole-N,N,N’,N’-tetramethyl-uronium-hexafluoro-phosphate
HIV	Human Immunodeficiency Virus
HOAt	1-Hydroxy-7-azabenzotriazole

IMAC	Immobilized Metal Affinity Chromatography
KHz	Kiloherz
LCMS	Liquid Chromatography Mass Spectrometry
mal*bis*NTA	Maleimido-*bis*-Nitrilotriacetic Acid
MALDI MS	Matrix Assisted Laser Desorbtion Ionisation Mass Spectroscopy
MHz	Megaherz
miRNA	Micro Ribose Nucleic Acid
MRFS	Molecular Force Recognition Spectroscopy
mRNA	Messenger Ribose Nucleic Acid
MWCO	Molecular Weight Cut Off
NHS	N-Hydroxysuccinimide
NIH	National Institute of Health
NLS	Nuclear Localizing Signal
NMR	Nuclear Magnetic Resonance
NTA	Nitrilotriacetic Acid
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PEG	Polyethylene gycol
PNK	Polynucleotide kinase
POC	Peptide-Oligonucleotide Conjugate
PS	Polystyrene
PTFE	Polytetrafluoroethylene
RISC	RNA-Induced Silencing Complex
RNA	Ribose Nucleic Acid
RP-HPLC	Reverse Phase High Performace Liquid Chromatography
RT	Room Temperature
SEC	Size Exclusion Chromatography
SIAB	*N*-Succinimidyl-(4-iodoacetyl)-aminobenzoate
siRNA	Small Interfering Ribose Nucleic Acid
SMPB	Succinimidyl-4-(p-maleimidophenyl)butyrate
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism
SPDP	*N* -Succinimidyl 3-(2-pyridyldithio)-propionate

SPPS	Solid Phase Peptide Synthesis
TBE	Tris-Borate-EDTA
TBTU	O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate
^{t}Bu	*tert* Butyl
TEA	Triethylamine
TEM	Tunnelling Electron Microscopy
TES	Triethyl silane
TFA	Trifluoroacetic acid
THAP	2',4',6'-Trihydroxyacetophenone
tRNA	Transfer Ribose Nucleic Acid
TX	Triple Crossover
XPS	X-Ray Photoelectron Spectroscopy

Chapter One

Introduction

1.1 Deoxyribose Nucleic Acid (DNA)

The biopolymer deoxyribose nucleic acid (DNA) is at the basis of the central dogma in molecular biology.[1] It underpins all processes that occur within the cell, carrying the genetic blueprint for life and with it the information that determines our traits as both a species and as individuals. It is a molecule that has fascinated both the scientific and non-scientific communities since its discovery in 1869, the image of the double helix has become an icon of popular science in the twentieth century. DNA continues to be at the forefront of medical diagnostics, healthcare and forensics. As well as playing an important role in current medical advances, it is presumed that several future technologies improving the diagnosis and treatment of disease will be based around this remarkable molecule.

The timeline for the discovery of DNA stretches over 100 years; the molecule was first discovered by the Swiss biochemist Johann Friedrich Miescher in 1869.[2] He extracted a phosphorus-containing compound found in the nucleus of white blood cells of human pus that he obtained from bandages of wounded soldiers. He named

this material nuclein due to its position of residence inside the cell. In 1919, Phoebus Levene identified phosphate, deoxyribose and the nucleobases as the chemical constituents of the material.[3] He suggested that the molecule is comprised of a string of monomer units linked together through the phosphate groups. In 1937 William Astbury produced the first X-ray diffraction patterns of DNA proving that it possessed a repeating structure.[4] The theory that DNA carried the genetic blueprint for life was first demonstrated by Frederick Griffith when, in 1928, he transferred the traits of the 'smooth' form of the bacteria *Streptococcus Pneumococcus* to the 'rough' by mixing the killed smooth form with live rough.[5] The suggestions put forward by this work were later proven when Alfred Hershey and Martha Chase showed that DNA is the genetic material of the T2 phage.[6] With the publication of their seminal paper in Nature in 1953[7] Watson and Crick announced a molecular structure of what was thereafter known as deoxyribose nucleic acid (DNA) using X-ray diffraction data produced by Rosalind Franklin and Raymond Gosling. Later X-ray analysis of the compound proved this theory to be correct and with it emerged the new field of molecular biology.

1.1.1 DNA in Healthcare

DNA is central to the processes of the cell, carrying the information required to synthesize the proteins essential for life. The manipulation and application of DNA therefore plays an important role in the biomolecular sciences. By using DNA as the blueprint and production line for proteins one can learn more about the intricate workings of the cell. Using DNA as a molecular tool, scientists have been able to interject when these processes go wrong, enabling diagnosis and treatment of disease. By understanding the genomes from biological pathogens such as viruses and bacteria, it is possible to identify and neutralize these potentially life-threatening organisms. Modern sequencing techniques allowed the genomes of pathogens such as the H5N1 and H1N1 avian and swine flu viruses to be decoded quickly allowing vaccines to be prepared thereby helping to avoid, or at least mitigate the effect of, global pandemics.

1.1.2 Chemical Modification Tailor the Properties of DNA for Molecular Biology

Advances in both the solid phase synthesis of oligonucleotides and the enzymatic incorporation of nucleotides into a template have allowed researchers to synthesize DNA carrying non-biogenic bases. Via incorporation of bases that carry designed functional groups, scientists are able to probe the nature of DNA and to enhance the properties of the molecule to suit a particular application.

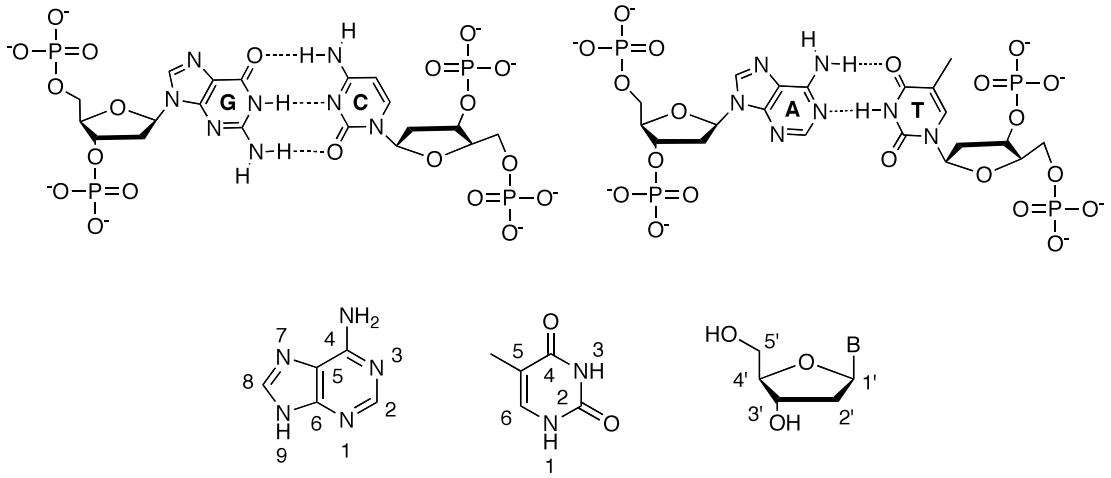


Figure 1.1. Chemical structures of the G-C and A-T Watson and Crick base pairs and labeling of the purines, pyrimidines and deoxyribose sugar units.

The structure of the nucleic acids fulfills several important biochemical and genetic tasks. The functional groups on both the purine and pyrimidine bases allow for hydrogen bonding to their corresponding complementary base, as indicated (Figure 1.1). This specificity and the formation of a double helical structure underpins the role of DNA as an efficient carrier of genetic information as well as the semi-conservative nature of the replication process. Furthermore, the 3' and 5' hydroxyl groups on the deoxyribose sugar are attached to phosphate groups to form the phospodiester backbone of the polymer. Any man-made chemical modification of DNA should retain these structural/functional properties of DNA if the biological function of the molecule is to remain unaffected.

It is advantageous in some cases, however, to functionalize nucleotides at these positions. The development of the antiretroviral drug AZT (Azidothymine,

Figure 1.2)[8], for example, disrupts the action of the reverse transcriptase RNA polymerase enzyme that helps to incorporate the HIV viral genome into a DNA of the host cell. Omission of the 3' hydroxyl group – replaced instead with the azide moiety – causes termination of the transcription process of viral RNA. This dramatically slows down the propagation of the virus in the host.

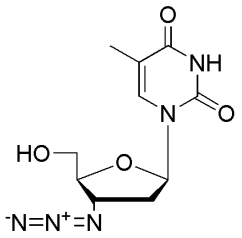


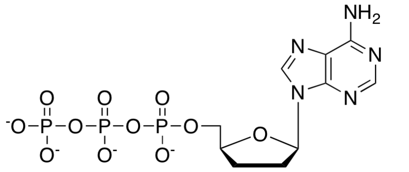
Figure 1.2. Azidothymine (AZT).

If it is not the intention to disrupt the process of DNA replication then chemical modification is usually conducted at position 5 in the pyrimidines and position 7 of the purines (Figure 1.1). Chemical groups attached to these sites are exposed to the major groove in the duplex and do not interfere with either the enzymatic incorporation of the nucleotide into a strand or the base pairing when duplexes are formed.[9] If modified nucleosides are to be used as biological agents then they must first be converted into their corresponding triphosphate to allow enzymatic incorporation of the monomer into a template strand. In the lab the enzyme polymerase is used, as it is in nature, to link the triphosphate to the 3' terminus of the DNA strand. Different polymerases can be more promiscuous than others and therefore allow a wider range of groups attached to the base to be tolerated. Researchers also have the option of incorporating a modified base into a DNA template via solid phase synthesis. These options give scientists a broader tool-kit for the application of non-biogenic DNA to modern scientific problems.

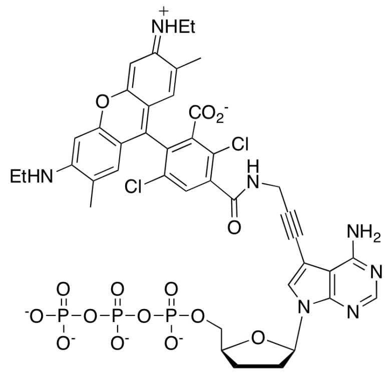
1.1.3 Non-Biogenic Nucleotides

Examples of the modification of DNA to enhance its scope of application are ubiquitous in molecular biology. Historically an important method of sequencing DNA is the so-called chain termination methods developed by Sanger *et al.*[10] This

procedure is based on the use of 2',3'-dideoxy analogues (ddNTP) of natural triphosphates (Figure 1.3). Given the triphosphate group at 5' position of the ribose sugar unit, these ddNTPs can be incorporated into a DNA strand by the 5'->3' activity of DNA polymerase enzymes. However, since the ddNTPs do not carry a hydroxyl group at the 3' position of the sugar they cannot be linked to further nucleotides via a phosphodiester bond. This results in termination of the strand at this position. The DNA strand to be sequenced and associated reagents and compounds are split into four aliquots. To each sample one of the ddNTP's is added in slight excess to the other dNTP's. Heating and cooling cycles anneal the primer to the template and facilitate chain extension using the polymerase enzyme. Incorporation of the ddNTP in the sample causes chain termination leading to a truncated strand. Incorporation of the dNTP analogue, however, allows the extension to progress. This process continues until the polymerase reaches the end of the strand. Once the sequencing is complete the four samples are denatured and analyzed on a poly-acrylamide denaturing gel.



1



2

Figure 1.3. (1) ddATP utilized in Sanger sequencing; (2) ddA-PA-5dR6G nucleotide used for dye – terminator sequencing diagnostics.[11]

The truncated strands present in each sample separate by size on the gel, the DNA can be visualized in bands due to the ^{32}P labeling of either the primers or nucleotides. By reading the bands off by order of size from each of the four samples the sequence of the strand can be deduced (Figure 1.4).

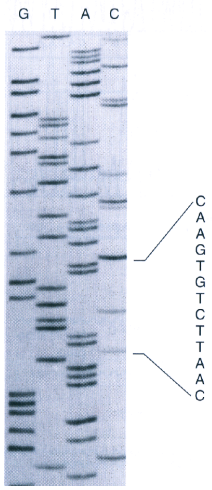


Figure 1.4. An example of a Sanger chain termination sequencing gel[12].

These ddNTP were originally synthesized[13,14,15,16] but became commercially available as the technique gained popularity. Since Sanger *et al.* published these results the chain termination sequencing became the method of choice. Modern scientific advancements have led to fluorescence techniques replacing ^{32}P radiolabeling of the extended strands. First primers carrying fluorescent tags[17] and then ddNTP's[18] were developed (Figure 1.3) and employed to terminate the chain extension. The synthesis of the four ddNTP's each carrying a fluorophore that emits at a different wavelength has allowed sequencing to take place in one single reaction vessel. Each truncated strand carries a fluorophore representing one of the four bases at its 3' terminus. The strands are again separated by size usually using capillary electrophoresis and the particular emission of the tagged strand allows identification of the terminating base. In this way the sequence of a strand can be deduced from a one-pot reaction and set up as a fully automated system.

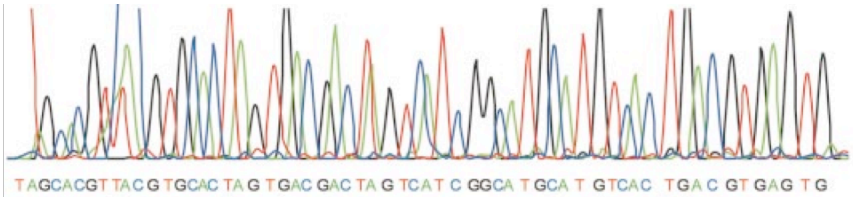


Figure 1.5. The sequencing output from fluorescently labeled chain termination sequencing[19].

This research has paved the way for such scientific milestones as the sequencing of the complete human genome in 2003.[20] Since Sanger based sequencing is only viable for fairly short reads, (4000 – 7000 base pairs) a method known as shotgun sequencing[21,22] was used for whole genome sequencing. In this method the genome is broken up randomly and sequenced using the dye – terminator method. The fragments are then computationally pieced back together to form the completed genome. Such developments have revolutionized modern medicine and science, each one relying on the synthesis and application of non – biogenic nucleotides.

1.1.4 Antisense Therapeutics

A vitally important application of DNA is in the regulation of gene expression - the control of potentially pathogenic protein translation allowing treatment of both genetic (heritable) disorders and various forms of cancer.[23] This type of therapy allows treatment, not of the symptoms but of the causes of the disease itself.

Protein biosynthesis (i.e. transcription and translation) relies on the section of DNA that carries the appropriate gene initially being unzipped by the protein DNA Helicase. The enzyme RNA polymerase then reads out the non-coding DNA strand of the gene, resulting in an RNA copy of the coding sequence i.e. the sense strand, termed messenger RNA (mRNA). This strand diffuses out of the nucleus and into the cytoplasm. It then enters the Ribosome and each trinucleotide 'codon' carried by the mRNA sequence binds to a specific amino acid linked transfer RNA (tRNA) component. As the Ribosome reads the mRNA strand the specific amino acids that are bound to the tRNA molecules link together to form a nascent polypeptide chain. Once the translation is complete the polypeptide folds into a tertiary structure forming the active protein.

DNA can be used to interfere with this process of protein biosynthesis. By synthesizing oligonucleotides that are complementary in sequence to the mRNA strand the oligo can bind to the mRNA preventing it from entering the ribosome and effectively silencing translation of the gene (a process known as steric blocking). Although this method has been shown to silence protein production it has its drawbacks. Since the antisense strand will bind in a 1:1 stochiometry with the mRNA it must be present in at least the equivalent molarity to that of the mRNA. It is

difficult to ascertain the amount of mRNA released and so an excess of the antisense strand must be employed. This requires, in biological terms at least, a large amount of the DNA to be delivered.

An alternative method of gene silencing is to exploit the cells own natural regulatory processes. When a protein is being over expressed or if a cell needs to down regulate gene expression it uses a process called RNA interference.[24] This involves two types of small RNA molecules, either double stranded small interfering RNA (siRNA) or single stranded micro RNA (miRNA). These molecules are processed from non – coding RNA (i.e. RNA which is not translated into protein) transcribed from genes carried in our DNA. Both strands are then processed by a variety of enzymes to produce the active RNA components. Both siRNA and miRNA effect gene knockdown via activation of a protein complex called RISC (RNA induced silencing complex). The 'guide strand' of the miRNA or siRNA is incorporated into this complex and is complementary to the target mRNA. Once it hybridizes to its target the *Argonaute* protein (the catalytic component of the RISC complex) is activated and degrades the mRNA. The siRNA differs to that of miRNA in that the latter demonstrates incomplete base pairing to the mRNA target. This allows the miRNA a wider target base whereas the siRNA's are more specific – degrading just the one mRNA target strand. In this way the cell can regulate gene expression via an effective catalytic mechanism.

These processes can be used to efficiently silence targeted genes and therefore decrease the output of potentially pathogenic proteins. It is also a very important method in the investigation of biological processes. The ability to 'knock-out' specific genes is vital in helping researchers gain a better understanding of complex biological processes. Transfection of viral plasmids carrying the siRNA duplex or introduction of small miRNA single strands made synthetically can target specific mRNA strands and silence the corresponding gene. However, RNA is liable to enzymatic degradation, therefore short oligonucleotides can be used in place of their RNA counterparts to increase serum and cytoplasm stability. The DNA can be further stabilized via the synthesis of a phosphothioate backbone. This reduces its liability towards DNA degradation enzymes (Figure 1.6).

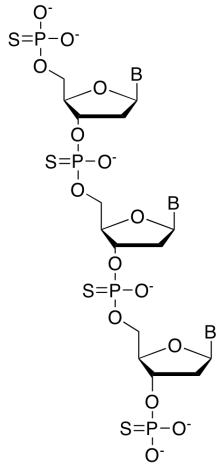


Figure 1.6. The phosphothioate backbone of an oligonucleotide synthesized via solid phase.

Antisense Therapy could be potentially very effective; however, the first hurdle in this type of therapy is the introduction of the DNA into the cell. The net negative charge of the oligonucleotide poses a number of issues relating to cell delivery. Due to its large net negative charge the oligonucleotide cannot diffuse through the lipid bilayer and acting alone will not be recognized by the cell surface proteins. Therefore to effect cell delivery the oligonucleotides must carry a separate component, one that will either allow the compound to pass though the bilayer or encourage recognition by the proteins on the surface. Since these biological processes involve proteins it seems natural to assume that they would also make good agents for cell delivery. However, proteins are fairly large components to attach to an oligo strand. Peptides carrying the specific amino acid sequences that possess the relevant biological qualities required are much more feasible as cell delivery agents.

1.1.5 Peptide – Oligonucleotide Conjugates

Due to their use in biological signaling peptides are ideal tags to assist in transmembrane penetration and localization to the sub-cellular compartments where the oligos are most effective. The therapeutic efficacy of the DNA can be greatly enhanced by attaching peptides with biologically relevant amino acid sequences. It is generally believed that oligonucleotides are transported into cells via an endocytic

mechanism and so components that may enhance this process will increase cellular uptake and therefore efficacy of the therapeutic DNA.

Peptides can help improve DNA delivery in two ways; firstly by mitigating the net charge of the molecule. Since the lipid bilayer does not allow charged species to diffuse through it, the DNA must be neutralized in a complex before introduction to a cell. Poly-*L*-lysine containing peptides, for example, will bind electrostatically to the negatively charged backbone of the oligonucleotide to effectively cancel out its net charge (Table 1.1). Although the DNA is able to enter the cell as a neutral complex, it does so inefficiently. The cellular uptake is therefore increased, secondly, by using special cell-penetrating or cell-targeting peptides.[25]

Function	Target	Sequence
Cell surface binding	Fc receptor	TQPREEQYNSTFRV[26]
	IGF-1 receptor	D-GCSKAPKLPAALC[27]
	Beta-endorphin receptor	YGGFLRRG[28]
Membrane fusion	Viral fusion	AVGAIGALFLGFLGAAG[29]
Localizing peptides	ER localizing	YGEEDTSEKDEL[30]
	Nuclear localizing	CNSAAFEDLRVLS[31]
DNA binding	Cationic peptides	KKAAKKACAKKAAKK[32]

Table 1.1. Examples of biologically active peptide sequences.

In the latter method, the peptide-DNA conjugate is targeted to cell surface proteins. When the peptide moiety is recognized by a given receptor, the whole conjugate is likely to be taken up by the cell. Peptides that localize the DNA cargo in the nucleus i.e. Nuclear Localization Signals (NLS), have also been used to good effect. The most successful strategies harness a combination of approaches i.e. cell surface binding peptides, nuclear localizing peptides, as well as covalently attached lipids to assist membrane fusion.[33]

The peptide-oligonucleotide conjugates (POCs) are usually prepared using either a complete 'in-line' solid phase synthesis or via a convergent strategy in which DNA and peptide are synthesized separately and then conjugated. The first procedure

attempts to achieve the convenient all-encompassing solid phase approach available for modern peptide and oligonucleotide synthesis.

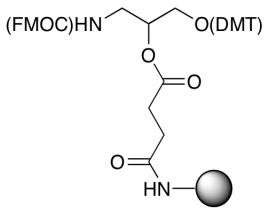


Figure 1.7. Duel peptide and oligonucleotide solid phase linker.[34]

A duel linker was developed by M. Antopolsky *et al.*[34,35,36] to facilitate the synthesis of the POC via solid phase (Figure 1.7). The linker contains a standard fluorenylmethyloxycarbonyl (Fmoc) protected amine for initiating peptide synthesis and a dimethoxy trityl (DMT) protected hydroxyl group for the DNA synthesis. The conventionally used chemistries for DNA and peptide synthesis, however, are incompatible. The usually harsh final acidic deprotection of the peptide from the resin via TFA may result in depurination of the oligonucleotide while the milder acidic deprotection of the DMT groups at the 5' hydroxyl of the oligos may effect unwanted deprotection of the protected peptide R groups. In addition, it was also postulated that the ammonium cleavage of the oligo may result in residue racemization or hydrolysis of the peptide, however, this assumption has since been proved as unfounded.[37, 38]

Despite these problems, the solid phase method has been used successfully to generate POCs comprised of poly-dT and poly-Ala components.[39] The problems often encountered with the combined solid-phase route could be avoided since no R group amino acid protection was needed. To develop a successful in-line solid phase method capable of preparing more complicated conjugates alternative protecting groups and resin linkers must be employed, thereby restricting the range of accessible sequences. Due to the limitations of this method many researchers have chosen to employ a convergent strategy to prepare their POCs.

In the convergent approach, the peptide and oligonucleotide components are synthesized separately. Using appropriate functionality the two compounds are conjugated together post solid-phase. To be of synthetic utility, the intended conjugating reaction must firstly be able to propagate in an aqueous solvent as the

DNA is buffered to prevent acidic depurination. In addition, the reaction has to be chemoselective to prevent non-specific reactions with the deprotected amino acid groups. The most commonly used functional group is usually a thiol since its reaction range tends to be specific to the sulfur group. Usually, a thiol is incorporated into the peptide via the inclusion of a cysteine residue. Consequently, the complementary electrophilic group capable of reacting with the thiol is attached to a non-biogenic functionality (Figure 1.8) on the oligonucleotide. A common route is the use of heterobifunctional linkers (Figure 1.9)[40,41] containing groups such malemides[42], iodo acetamides[43], and orthopyridyl disulfides to enable reaction with the peptide under mild conditions.[33]

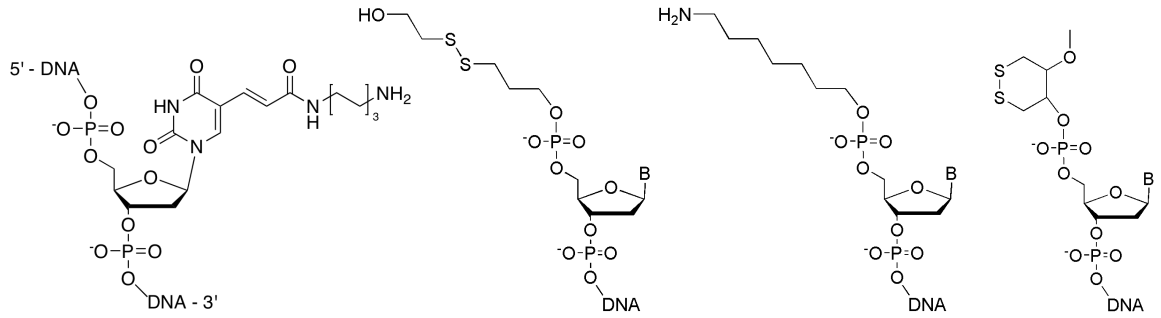


Figure 1.8. Examples of chemically modified nucleotides for solid phase inclusion.

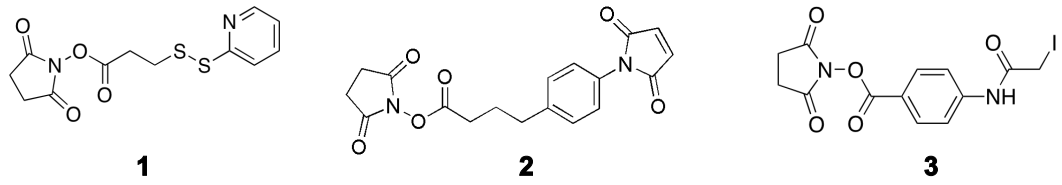


Figure 1.9. Examples of heterobifunctional linkers; (1) *N* -Succinimidyl 3-(2-pyridyldithio)-propionate (SPDP); (2) Succinimidyl-4-(p-maleimidophenyl)buterate (SMPB); (3) *N*-Succinimidyl (4-iodoacetyl)aminobenzoate (SIAB).

The heterobifunctional crosslinkers also usually feature the amine-reactive group N-hydroxysuccinimide for coupling to amine-modified nucleotides (Figure 1.9). Alternative chemistry such as oxime formation,[44] native ligation,[45] Click[46] and Diels Alder[47] have also been employed. Amide coupling methods using water-soluble coupling reagents such as EDC have been successful previously[48] but due to the likelihood of multiple amine groups being present on the peptide moiety this strategy

is not always viable.

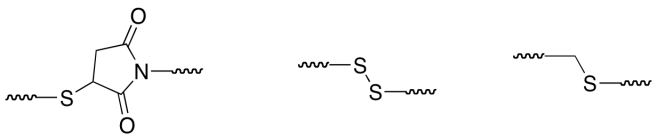


Figure 1.10. Linkages formed from mild thiol-specific reactions; (1) thioether bridge formed by reaction between maleimide and thiol group; (2) disulfide formed via disulfide exchange; (3) thioether formed via thiol attack on Iodo acetimide.

Since the modified base and cysteine residues can be placed at any point along the oligo or peptide sequence the conjugate can be synthesized with an internal peptide (or oligonucleotide) at no extra effort. Because of these benefits a convergent strategy is currently the method of choice for the synthesis of peptide-oligonucleotide conjugates.

Once synthesized the conjugate is usually purified by Reverse Phase High Performance Liquid Chromatography (RP-HPLC) or Anion Exchange Chromatography (AEC). Poly-acrylamide electrophoresis (PAGE) or size exclusion chromatography (SEC) can also be used. Given its large net negative charge, the oligo usually dictates the behavior of the conjugate during chromatography while the success of the purification depends on the peptide moieties' ability to add sufficient charge or alter the lipophilic nature of the compound enough to achieve analytical resolution between the modified and non-modified oligonucleotides.

1.1.6 Chemical Modification for Nanoscale Devices

Employment of non-biogenic nucleotides has enabled researchers to impart functions and interactions in DNA other than that of Watson and Crick base pairing. Due to the nanometer dimensions of a DNA duplex and its inherent physical properties i.e. its ability to self assemble and its relatively high persistence length, DNA is a suitable material for the preparation of nano-scale assemblies. The application of DNA in nanotechnology has therefore been explored with great interest over the last few decades.

An example of this research is the use of DNA in the preparation of nanoelectronic devices. It has been suggested that DNA could be turned into a nano-scale wire by increasing the conductivity of a DNA duplex. One of the advantages of DNA in this area is the ability to self-assemble in a repeating pattern and its versatile means of synthesis (i.e. via enzymatic or solid phase means). To endow the duplex with electrical properties, the biogenic bases are to be modified to carry metallic ions able to impart the appropriate properties to the duplex. A number of different bases have been synthesized and incorporated for this purpose.[49, 50, 51, 52]

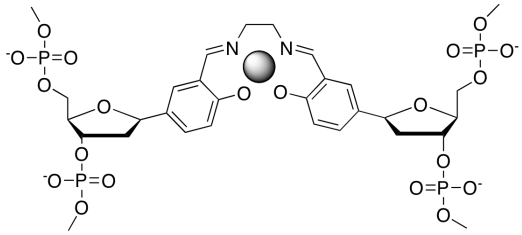


Figure 1.11. Nucleotides modified with M(II) complexing bases for nano-wire construction.

The nucleosides shown in Figure 1.11 have been equipped with the ability to conjugate to either Cu(II) or Hg(II) metal ions. This functionality has replaced the *exo* groups on the bases that allow specific base pairing. Instead of the hydrogen bonded Watson and Crick base pairing, the duplex is now stabilized using metal ions. This allows a duplex to self–assemble via the incorporation of alternate metal ions running through the central axis of the helix (Figure 1.12).

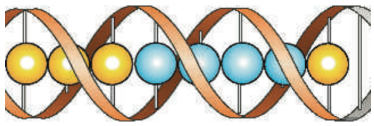


Figure 1.12. Metalized DNA duplex via incorporation of non – biogenic metal binding nucleotides.[52]

Research by Carell *et al.*[53] has utilized alkyne-modified nucleotides to allow efficient 'Click' reactions post incorporation. This allows the DNA template to be functionalized with groups that can be metalized with Ag ions as a base for the

development of wires. Further work has investigated the incorporation of bulky porphyrin groups[54] into oligonucleotides, which allows efficient resonance energy transfer between chromophores and serves to stabilize the formation of the duplex. Chelation of metal ions within the porphyrin chromophores would again enable the generation of nano-scale electronic devices. Alternative functionalization such as multiple spin labeled bases for EPR detection[55] and modification with multiple functional groups[56,57,58] have also been investigated to further enhance the scope of application of nucleotides and DNA.

1.1.7 Areas not yet been Exploited by Chemical Modification

Two areas of biotechnological interest that have not yet been fully exploited using chemical modifications of nucleic acids are: (1) the sensing of DNA strands with nanopore recordings and (2) the functionalization of DNA nanostructures. Both are emerging areas of biotechnology and will likely benefit from chemical modification of DNA.

1.2 Nanopore Sensing

The advancement of modern technologies has given rise to the emergence of single-molecule techniques. Most conventional analytical methods rely on the observation of average changes in the bulk solution, for example, the increase in fluorescence in a majority of soluble dye molecules. However, with advances in the sensitivity of fluorescence detection schemes and a drop in the price of equipment, researchers are now able to follow changes at the single-molecule level. This development has revolutionized analytics, especially in the field of molecular biology. The ability to detect single molecules of DNA for example, allows minute amounts of material to be analyzed without the cost, time and accuracy issues of PCR amplification. In 2004 the National Institute of Health (NIH) set the challenge of $1000 as the cost of complete human genome sequencing. This would allow personalized genomics, which is predicted to revolutionize personalized medicine. Among the single-molecule

methods, the two most advanced are based on fluorescence. The first technique was developed at Helicos Biosciences (http://www.helicosbio.com) and termed 'sequencing by synthesis[59] while SMRT sequencing (single molecule real-time sequencing)[60] technology was developed at Pacific Bioscience (http://www.pacificbiosciences.com). Whilst different, both techniques involve the incorporation of fluorescently tagged nucleotides into the DNA template to be sequenced and the detection of the incorporation at the single molecule level.

Nanopore recording is an alternative single-molecule technique that also has the potential to reach this $1000 benchmark. It is an electrical analytical technique in which an electrical field drives charged species through a biological or inorganic nanopore.[61] As the translocating molecules displace electrolyte ions flowing through the pore, the measurement of the passing ionic current can enable identification of the molecules. The technique is able to detect individual molecules since (i) only one pore is used and (ii) the dimensions of the pore are chosen to only enable the passage of one molecule at a time.

1.2.1 Nanoscale Protein Pores

A commonly used biological pore is that of the heptameric protein α-Hemolysin. The αHL protein from the gram-positive bacterium *Staphylococcus Aureus* is an exotoxin which causes the lysis of red blood cells. Bacteria use these toxins to obtain nutrients, such as iron, from host cells. The structure of this heptameric protein pore has been characterized to 1.9 Å resolution using X-ray crystallography.[62] The protein has a cap and a membrane inserted stem region. The inner vestibule in the cap region has an inner diameter of 42 Å, while a transmembrane β-barrel stem measures 14 Å at the tightest constriction positioned at the boundary between the cap and the stem region (Figure 1.13).

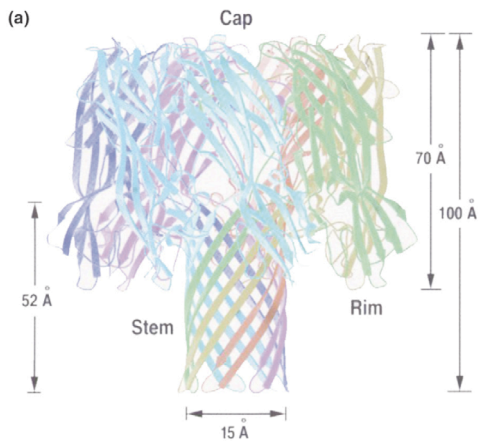


Figure 1.13. Ribbon representation of the protein pore α – hemolysin.[63]

The nanopore apparatus is comprised of a lipid bilayer deposited over a μm-sized hole in a teflon sheet. This section divides two buffer-filled compartments connected by an electrical circuit. A voltage is induced and the protein is introduced to one compartment. Once the αHL pore has inserted into the bilayer, a current is detected and the compartments can be replaced with fresh buffer to flush out excess protein pores. The simple electrical set-up enables researchers to follow the translocation of molecules across a single pore.

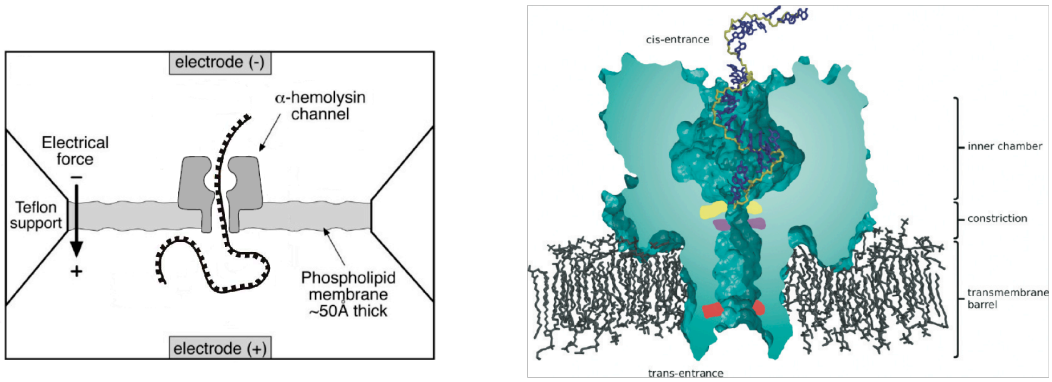


Figure 1.14. (1) Schematic of the Nanopore set up[64]; (2) molecular model of a DNA strand translocating through a nanopore[65].

This single-molecule approach has been exploited to analyze a number of molecular species such as toxins[66], chemical reactions, drug molecules[67] and protein

analytes.[68] However it is the analysis of DNA strands where the technique has gained the most interest. Due to the size of the proteins aperture at its most constricted (approximately 1.4 nm) it is possible to translocate single strands of DNA/RNA through (the approximate diameter of a DNA strand is 0.9 – 1.2 nm). It was noted by Kasianowicz *et al.*[61] and others[69] that although it was possible to discriminate between different homo-polynucleotides due to the different characteristic current blockades, it was not however, possible to detect individual bases. The DNA translocates too quickly for information about the sequence to be determined. A number of developments have since been made. Specific DNA sequences have been identified by tethering an oligo within the pore and translocating the complementary strand to observe specific binding.[70] Further research has shown that single point mutations can be detected in static double stranded DNA.[71,72] It was observed that the structural differences between the purines and pyrimidines are sufficient to alter the current characteristically in a static strand. However, it has so far not been possible to identify individual bases on a translocating strand.

One potential approach to this issue of translocation speed has been to fit a 'molecular adaptor' composed of a cyclodextrin into the β-barrel of the protein pore.[73] This further decreases the diameter of the constriction allowing more subtle structural differences to be identified, yet not at single-base resolution. The approach using the molecular adapter can also be exploited for a related approach, which does not thread DNA through the pore but is rather based on individual nucleotides.[74] The proposed strategy uses an exonuclease, which cleaves the phosphodiester backbone of the DNA strands allowing individual nucleotides to enter the pore. With the molecular cyclodextrin adaptor fitted (Figure 1.15), it has been demonstrated that individual bases can be resolved based on the different current blockades. While the step of attaching a fully functional enzyme to the pore awaits experimental implementation, the successful detection of nucleotides is nevertheless a large step towards a whole genome sequencing method using the nanopore technique. Due to the speed of the process it is postulated that a complete genome could be analyzed in a matter of days using parallel nanopore devices. Due to the low cost of materials, and since this is a single – molecule technique it is expected that this method may be able to reach the benchmark of whole genome sequencing at under $1000 set by the National Institue of Health (NIH) in 2004.

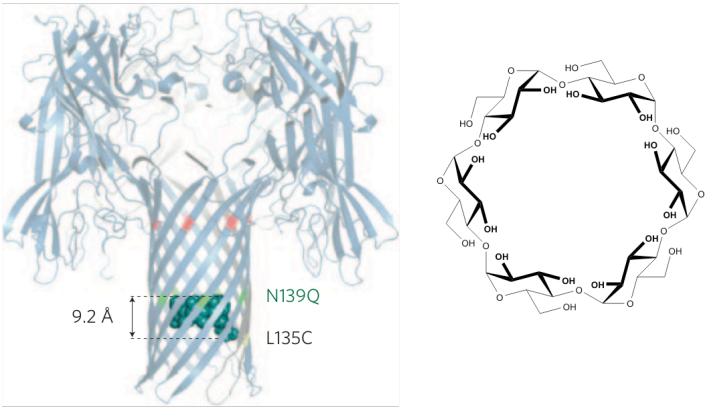


Figure 1.15. The α-hemolysin pore carrying a cyclodexdrin filter.[74]

1.2.2 Chemical Modification of DNA to Increase the Resolution of Nanopore Recording

The approaches detailed above tackle the identification of individual nucleotides via alteration of the protein pore itself. An alternative approach is to instead modify the translocating DNA. By attaching various bulky or charged groups to the individual bases of the strand it may be possible to engineer each nucleotide to produce a different current modulation. Given the requirement to chemically label the strands, it is envisaged that this method would not necessarily be useful for whole genome sequencing but rather be applicable in point-of-care genetic diagnostics i.e. testing at or near the site of patient care. Due to its minimal technical requirements, the nanopore apparatus can be miniaturized. For example, it may be possible use chemical tagging of DNA to analyze forensic DNA samples using portable equipment.

A central hypothesis is that attaching chemical tags to separate bases would increase the cross-sectional diameter of the DNA and hence slow down translocation each time a modified base passes the narrow pore constriction.

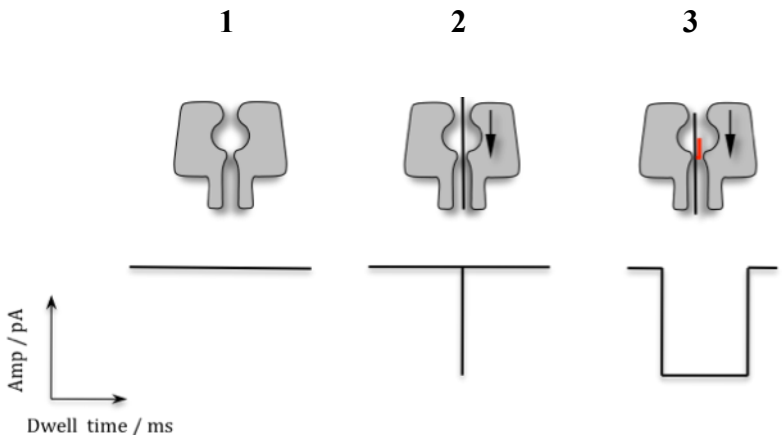


Figure 1.16. Schematic translocation of unmodified and modified DNA. The open channel current (1) is interrupted via translocation of a DNA strand (2) leading to a fast event. The tagging of the DNA strands should slow down the dwell time of the strand through the pore (3).

In this way it should be possible to identify the tagged DNA from untagged using the characteristic current modulations observed. Further research should enable tagging of multiple bases and identification using this technique. Depending on the ability of the chosen tags to resolve the translocating bases it may be viable to detect multiple repeating bases such as trinucleotide repeats found in forensically important DNA samples.

1.3 DNA nanotechnology

A second area yet to be fully exploited using chemical modification is that of higher-order DNA nanostructures. This is a rapidly expanding research topic within the area of nanotechnology. Its nanoscale size and inherent biophysical properties such as sequence-specific self-assembly, high persistence length of approx. 50 nm and manipulation via enzymes make DNA an attractive building block for nanoscale structures.[75,76] Indeed, using the complementary nature of Watson and Crick base pairing, DNA strands have been sequence-specifically assembled into artificial higher-order nanoscale structures[77,78] that are of use in biotemplating,[79,80] biocomputing,[81] sensing,[82] and biophysical studies.[83] As well as 'inert' structures, hybridization of chemically modified strands into DNA duplexes has been utilized to

bring chemical groups into defined contact to, for example, enhance their reaction.[84] These complex and impressive two and three dimensional higher order structures could be further functionalized by the incorporation of chemically modified groups. This would allow researchers to widen the scope of application of these nano-scale structures further still.

1.3.1 Higher Order DNA Based Nanoscale Structures

As early as 1982 Nadrian Seeman published a paper postulating the use of DNA for the construction of 3D nanoscale objects.[85] He recognised the potential this molecule had due to the specificity of its Watson and Crick base pairings. Since a DNA strand will only hybridise to its complement this provides a way to programme structural formation via self-assembly. Seeman suggested using 'junctions' from which 3 – 8 double helices may propagate forming the basis of these complex structures. Since this proposal a large variety of incredibly complex structures have be formed by clever design of the DNA sequences. Although complex in design the formation of these structures simply requires the heating and cooling of the relevant DNA strands. Compared to the synthesis of non-biological nanostructures this method is incredibly economical. By utilizing the properties inherent to the storage of genetic information researchers are able to easily form complex architecture.

1.3.2 Two-Dimensional DNA Based Arrays

The application of DNA to the formation of templated arrays has attracted a great deal of interest. These complexes could be used for molecular electronics, or patterned surfaces to which proteins, for example, could potentially be attached. A successful method for the formation of these arrays is in the application of double[86,87] (DX) and triple[88] (TX) crossover 'tiles' (Figure 1.17). These tiles consist of two or three double helices that link by crossover of one of their strands to another helix. In this way planer, rigid tiles are formed that carry sticky ends to link tiles together.

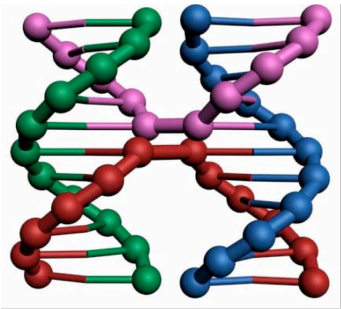


Figure 1.17. Schematic of a DX crossover tile.[89]

A number of arrays have been generated in this way; moreover tubular structures[90] can be formed via the folding of these sheets. Functional arrays that perform computation[81] and mechanical[91] actions for example have also been investigated along with sheets that are able to conjugate proteins and nanoparticles.[83,92] The junctions specified in Seemans 1982 paper have also been utilised extensively, for example in the formation of lattices that can be functionalized with proteins and metalized with silver to form highly conductive nanowires[93] (Figure 1.18).

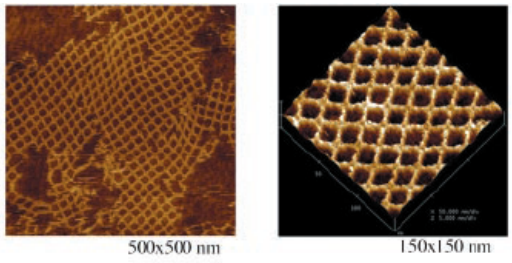


Figure 1.18. DNA based arrays based on DNA junctions able to form nano-wires.[93]

1.3.3 Three-Dimensional DNA Based Structures

Various three-dimensional structures have been prepared by exploiting the base pairing specificity of DNA. By designing oligonucleotide strands to carry sequences complementary to sections of the partnering oligos complex structures can be formed such as tetrahedrons[94,95], cubes[96] and octahedral structures.[97] Previous research has

also demonstrated that a selection of these structures are capable of encapsulating nanoparticles for possible application in drug delivery.[98,99]

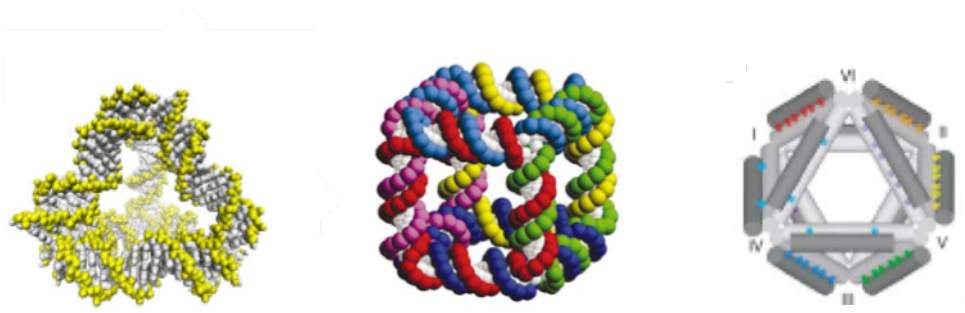


Figure 1.19. Examples of DNA based structures; (1) tetrahedron[94]; (2) cube[96]; (3) octahedron.[97]

Visualisation of these structures is of importance for characterisation – electrophoresis provides a good indication of correct formation however direct visualization using such high-resolution techniques as AFM and TEM is ideal. As well as the design and formation of these structures, researchers have also generated impressive high-resolution images (Figure 1.20).

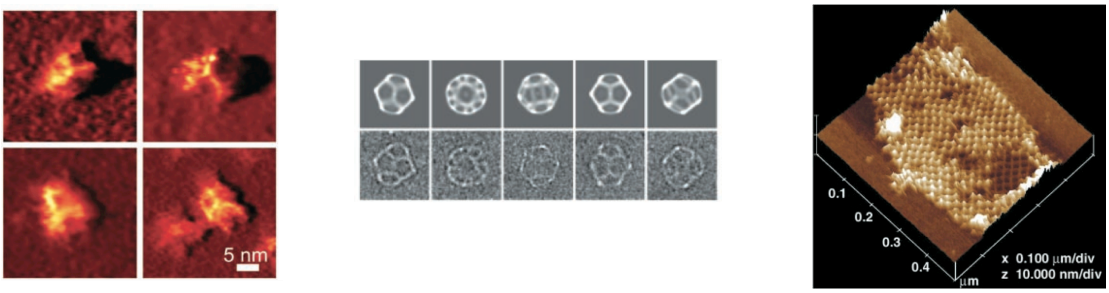


Figure 1.20. AFM and TEM visualization of (1) DNA tet[95]; (2) Octahedral geometries[97]; (3) DNA lattice.[93]

Chemical functionalization has been explored in this area to impart designed functionality onto these structures. Examples of the functionalization of DNA nano-structures via chemical modification include; attachment of fluorophore-tagged oligonucleotide into a DNA ring[100,101] and the addition of a linker enabling two nano-scale rings to be joined.[102]

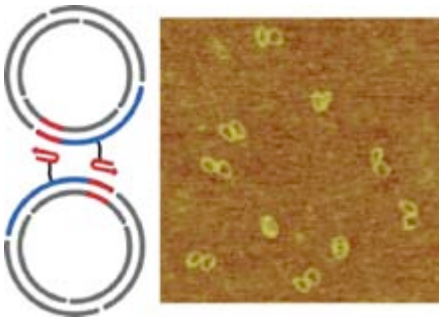


Figure 1.21. DNA based nano-circles joined by a polyamide 'anchor' enabling formation of double-ring structures.[102]

1.6 Aims of Thesis

This thesis is concerned with the application of chemically modified DNA in emerging areas of biotechnological research, specifically; Nanopore analytics and DNA-based nano-scale structures.

The first aim of this thesis is to improve single-molecule DNA sensing using nanopores. Nanopore recording is a relatively new technique, which senses charged compounds that are electrophoretically driven through a nano-scale pore. As no bulky optical equipment is needed, nanopore analytics can be miniaturized and potentially used for point-of-care applications. However, the sensing of important analyte-type DNA is hampered by poor analytical resolution. Individual bases cannot, currently, be detected as the DNA strands translocate too quickly through the pore for any sequence information to be read. Chemical modification could help enhance the resolution of this single-molecule technique. By attaching bulky chemical tags to the DNA strands it may be possible to slow down its translocation. This should enable identification of those strands carrying a tag from those that remain unmodified. The modification of multiple bases should allow identification of multiple tags and therefore sensing of specific bases, which has so far not been achieved using this technology. Strategies to attain the aim of improving nanopore analytics via chemistry are described in Chapter 2.

The second aim of the thesis is to enhance DNA-nanobiotechnology using chemically modified oligonucleotides. Due to its interesting physical properties, DNA

has been used widely as a building block to form higher-order structures. Most of the new architectures have been built by utilizing the base-specific hybridization of complementary DNA strands. Whilst these constructs are a fascinating example of the versatility of DNA, functionalization by attachment of chemical tags would create a new area within DNA-nanobiotechnology. For example, chemical modification may help impart new functions, namely the ability to form DNA structures by a chemical trigger. It is also conceivable that DNA structures could enhance chemistry by providing a nanoscale scaffold onto which different functional groups are placed. The results of this endeavor are described in Chapters 3 and 4.

Chapter Two

Single Molecule Sensing

2.1 Summary

In nanopore analytics, individual DNA strands are electrophoretically driven through a nanopore. Despite progress in this general area the analytical resolution in the sensing of DNA strands is poor as individual bases in translocating DNA strands cannot be resolved. This chapter describes how nanopore analytics can be enhanced by chemically tagging DNA strands. In particular, it is demonstrated that short peptide tags attached to internal base positions can help discriminate individual bases in translocating synthetic DNA strands. Furthermore, the chemical tagging strategy can be applied in a biological context to discriminate between single nucleotide polymorphisms (SNP), as shown using template-directed DNA polymerization with peptide-tagged nucleotides. Finally, adamantane-tagged nucleotides are synthesized to lay the foundations for a fast point-of-care method to detect highly repetitive DNA strands of forensic importance.

2.2 Introduction

Nanopore recording is an attractive electrical analytical technique in which individual molecules block a nanometer-scale pore leading to detectable modulations in ionic current. The single-molecule approach has been exploited to analyse toxins[66], chemical reactions/drug molecules[67], proteins[68] and single point mutations in double stranded DNA.[71,72] Despite progress in the sensing of isolated nucleotides[73,74] and a single base position in static DNA strands, it has so far not been possible to detect multiple bases in an individual translocating strand. One of the main technical hurdles towards this aim is the high speed with which the voltage-driven DNA strands pass through the pore leading to insufficient analytical resolution. In order to address this shortcoming, a new approach was developed to slow down single stranded DNA and thereby enable the detection of multiple separate bases.

This approach consisted of attaching chemical tags to individual bases in order to increase their size in comparison to the non-modified bases. It was anticipated that the chemical tags would facilitate the detection of the specific bases. As the first of three aims, it was investigated whether the chemical tags cause a steric blockade each time a modified base passes a narrow pore, and whether chemical tags of different composition and size give rise to different characteristic current signatures. This first aim was experimentally pursued by covalently attaching short oligopeptides to a synthetic oligonucleotide. After establishing the proof-of-principle with a synthetic system, the second aim was to show that tagging of bases can help infer biomedically relevant sequence information from DNA templates. The viability of the tagging approach was tested with peptide-modified nucleotide triphosphates in combination with template-directed DNA polymerisation. The third aim was to develop the tagging approach into a sensing technique that can compete with state-of-the art technologies and advance the general field of bioanalysis. In particular, the chemical foundations were to be created to enable the electrical sizing of repetitive DNA sequences that are of relevance in the identification of forensic samples. This final aim was investigated using nucleotides which carry the small and compact adamantane tag instead of the longer peptide tag.

2.3 Materials and Methods

2.3.1 Reagents

All reagents were purchased from Sigma-Aldrich Co. Ltd. unless otherwise stated, and used without further purification. All solvents were of commercial anhydrous quality and used as received. Ethanol refers to absolute ethanol (> 99.7%) and brine refers to a saturated solution of sodium chloride. For all moisture sensitive reactions, glassware was dried in an oven at 120 °C and an argon atmosphere used.

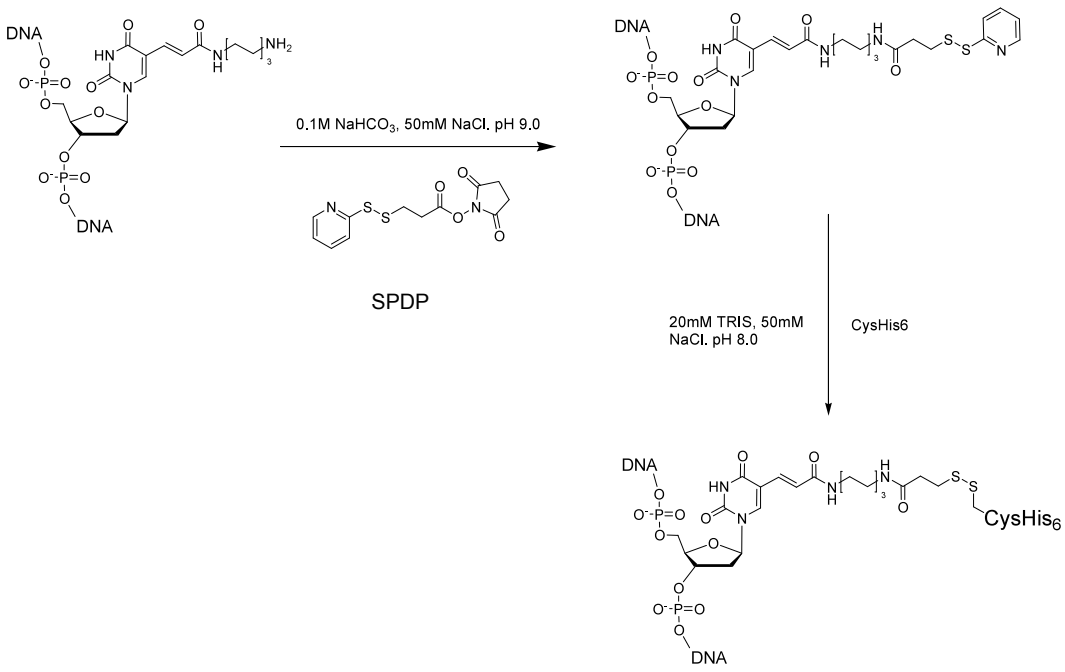
Oligonucleotides were purchased from Integrated DNA Technologies (Leuven, Belgium; www.idtdna.com), all amino acids and resins purchased from NovaBioChem (EMD bioscience, Gibbstown, NJ, USA; www.novabiochem.com). Peptides were synthesized on a Syro automated machine (see section 2.3.10 for details). Anion Exchange (AEC), Size Exclusion (SEC) and Immobilised Metal Affinity (IMAC) were all performed on an AktaPurifier 10 FPLC system (GE Healthcare). HPLC was performed on a Varian ProStar system with a Model 210 solvent delivery module and a Model 320 UV detector using Varian analytical, semi – prep and prep columns. Thin Layer Chromatography (TLC) was performed on aluminium backed Sigma-Aldrich TLC plates with F_{254} fluorescent indicator. Visualisation was done by absorption of UV light or by staining the plates with phosphomolybdic acid (PMA) solution [PMA hydride (12 g), conc. sulphuric acid (10 mL), ethanol (250 mL)] or potassium permanganate solution [$KMnO_4$ (1.25 g), Na_2CO_3 (6.25 g), water (250 mL)]. Normal phase flash chromatography was carried out using silica gel (43 – 60 μm) supplied by Merck. Flash chromatography on alumina was carried out using alumina activated from Brockman activity I to III (via addition of 6% v/w of water to the dry Alumina powder).

MALDI MS was performed on a Waters MALDI MicroMX machine using 2',4',6'-trihydroxyacetophenone (THAP) as the matrix (1 mg/mL in methanol) and ammonium acetate or citrate as the co-matrix. LC/MS was performed on a Waters Acquity uPLC SQD using HPLC grade water and Acetonitrile (both with 0.1% formic acid) as the solvents. NMR (^{1}H and ^{13}C) was performed on both 300 and 500 MHz AMX Bruker Spectrometers (as stated). The chemical shifts (δ) were given in units of

ppm relative to tetramethylsilane (TMS), where δ (TMS) = 0 ppm. Coupling constants (*J*) were measured in Hertz using the Mestrec® analysis program (Mestrelab Research, Santiago de Compostela, Spain). Multiplicities for ^{1}H coupling are shown as s (singlet), d (doublet), t (triplet), m (multiplet), or a combination of the above. Deuterated chloroform ($CDCl_3$), dimethylsulfoxide (d_6-DMSO) and methanol (CD_3OD) were used as solvents (as stated) for all NMR analysis.

2.3.2 Preparation of peptide-tagged DNA strands for Aims 1 and 2

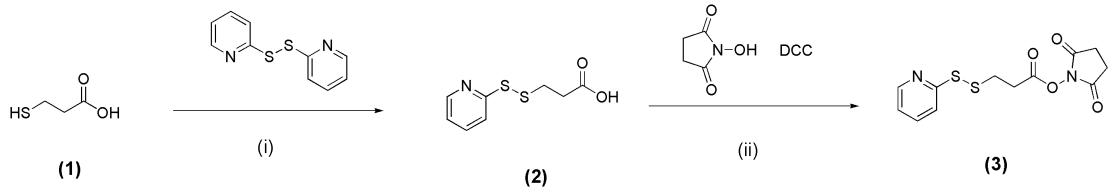
Oligopeptides of different length and composition were covalently attached to one or two internal base positions of a synthetic oligonucleotide. Scheme 2.1 summarises the two-step procedure to prepare a single-modified peptide-DNA conjugate.



Scheme 2.1. Conjugation of peptide tags to an internal base. The initial modification of non-biogenic amine-carrying thymine with the heterobifunctional crosslinker SPDP and subsequent attachment of cysteine containing peptide sequence.

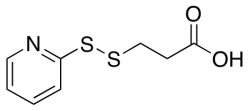
Using the heterobifunctional linker *N*-Succinimidyl 3-(2-pyridyldithio)-propanoate (SPDP; synthesis in Scheme 2.2) an internal thymine carrying a non-biogenic amine linker was coupled to an activated pyridyl disulfide group. This

moiety is particularly liable to attack from thiols and therefore introduction of a cysteine containing peptide allows formation of the conjugate.



Scheme 2.2. Synthesis of SPDP; (i) DPDS, mercapto-propionic acid, DCM, RT; (ii) DCC, *N*-hydroxysuccinimide, DCM, RT.

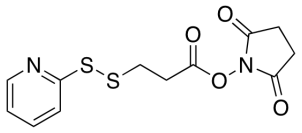
2.3.3 3-(2-Pyridyldithio)-propanoic acid (2)



(2)

2'-2'-Dipyridyldisulfide (DPDS, 10.98 g, 0.0498 mol) was dissolved in 80 mL anhydrous ethanol. Acetic acid (1.5 mL) and 3-mercapto-propanoic acid (2.64 g, 2.16 ml, 0.0249 mol) were added and the solution left stirring at room temperature for 2 hours. After this time the solvent was removed under reduced pressure to yield a viscous yellow oil, this was placed under high vacuum to remove all traces of acetic acid. The crude product was purified using a basic Al_2O_3 column (Brockman activity III) with dichloromethane/ethanol (3/2) as the eluent. Once the yellow band corresponding to the thione by-product had eluted from the column, 4 ml of acetic acid per 100 ml solvent was added to elute the desired product. Fractions containing this compound were pooled and the solvent removed under reduced pressure. The resulting viscous oil was dried on a high vacuum line to yield the title compound **2** (4.1 g, 0.019 mol, 77% yield). Calculated mass ($C_8H_9NO_2S_2$) 215.01; observed mass (ESI^+) m/z – 216.04 (MH^+), 238.03 ($M + Na^+$). ^1H NMR (300 MHz; d_6-DMSO) δ/ppm 2.59 (2H, t, J = 7.0), 2.98 (2H, t, J = 7.0), 7.23 (1H, ddd, J = 7.1, 4.8, 1.2), 7.75 (1H, d, J = 8.0), 7.8 (1H, td, J = 7.9, 1.4), 8.4 (1H, d, J = 4.9), 12.43 (1H, broad s). ^{13}C NMR (75 MHz, d_6-DMSO) δ/ppm 33.6, 33.7, 119.2, 121.1, 137.7, 149.5, 159.0, 172.7.

2.3.4 *N*-Succinimidyl 3-(2-pyridyldithio)-propanoate (3)



(3)

3-(2-Pyridyldithio)-propanoic acid **2** (3.45 g, 0.016 mol) in anhydrous DCM was cooled to 0°C in an ice bath. *N*-hydroxysuccinimde (2.03 g, 0.0176 mol) was added followed by DCC (3.64 g, 0.0176 mol), both in anhydrous DCM. The solution was brought back up to room temperature and left stirring for 3 hours After the reaction had reached completion (followed by TLC) the solution was filtered under vacuum from the urea precipitate and the solvent removed from the filtrate under reduced pressure to yield a crude yellow oil. This crude material was purified via silica gel column chromatography using dichloromethane/methanol (50/2) as the eluent. The fractions containing the product were combined and the solvent removed under reduced pressure to yield a colourless oil that crystallized overnight at -20°C (3.93 g, 0.0126 mol, 79% yield). Calculated mass ($C_{12}H_{12}N_2O_4S_2$) 312.36; observed mass (ESI^+) m/z – 313.35 (MH^+), 335.36 ($M + Na^+$). ^1H NMR (300 MHz; d_6-DMSO) δ/ppm 2.81 (4H, s), 3.12 (4H, m, AB_{sys}), 7.25 (1H, ddd, J = 6.4, 4.8, 1.6), 7.76 (1H, d, J = 7.1), 7.82 (1H, td, J = 7.5, 1.6), 8.48 (1H, d, J = 4.9). ^{13}C NMR (75 MHz; d_6-DMSO) δ/ppm 25.4, 30.2, 32.8, 119.5, 121.4, 137.8, 149.7, 158.5, 167.4, 170.0.

2.3.5 Solid Phase Peptide Synthesis (SPPS)

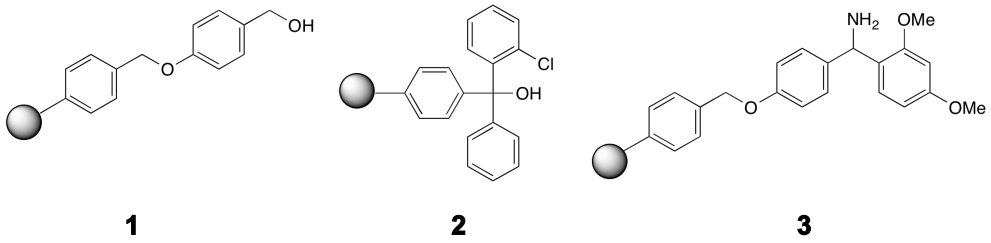
This section will describe in detail the synthesis of peptides via Fmoc solid phase peptide synthesis (SPPS) using an automated synthesizer. To rationalize the reagents and conditions used during the preparation, the general technique will be briefly introduced.

Since the development of the Merrifield resin[103] modern peptide synthesis has been carried out via a solid phase peptide synthesis (SPPS) method, making the synthesis of peptides of up to around 50 amino acids routine. Modern peptide synthesizers can make multiple peptides in parallel so a large number of compounds may be prepared overnight. SPPS involves the step-wise attachment of amino acid

residues to an appropriately modified polystyrene (PS) resin. The PS beads are held in a small inert reaction vessel (a PTFE syringe, for example) that contains a frit and is attached to a vacuum system. This allows the beads to sit at the bottom of the syringe whilst solutions containing the reagents are deposited via an automated system and allowed to react. Once the reaction is complete, the syringe is put under vacuum to evacuate the solution, leaving the beads in the vessel. In this way the peptide can be prepared via addition and evacuation of sequential reagents, a procedure know as flow-through synthesis. The chemistry involved is repetitive and orthogonal, allowing the process to be simplified and conducted robotically. Scales in the hundreds of mg of each peptide are usually prepared however industrial processes can produce g – kg of material. Once the synthesis is complete the peptide can be cleaved from the resin and all protecting groups removed in a single step to yield the target compound. Depending on the success of the synthesis, which usually relies on the specific sequence being attempted, the peptide may not require any further purification.

2.3.6 Resins and Linkers

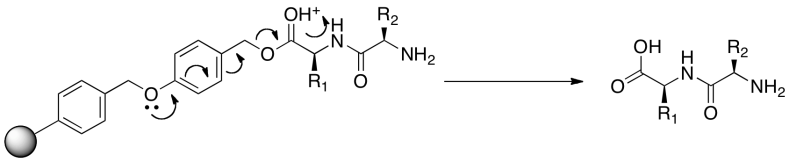
The resins used for SPPS are usually 1% divinylbenzene cross-linked polystyrene (PS) based and packed as porous beads. Modern resins can be functionalized with PEG (polyethylene glycol) to increase the characteristics of the resin i.e. hydrophilicity. These resins must then be functionalized with 'linker' compounds which allow attachment of the amino acid residues forming the peptide chain. The choice of linker dictates the ease and conditions with which the peptide is cleaved from the resin. The most commonly used linker/resin combination tends to be the Wang resin (a hydroxymethyl benzene linker attached to a PS resin) but chlorotrityl and Rink amide are also regularly used (Figure 2.1).



1 **2** **3**

Figure 2.1. Examples of common linkers and their cleavage conditions; (1) Wang (strong acid); (2) Chlorotrityl (weak acid); (3) Rink amide (allows cleavage of peptide with inclusion of an amine group at its Carboxy terminus).

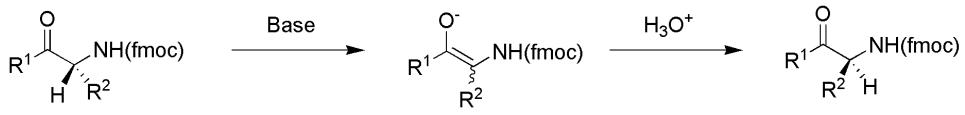
As the coupling steps that make up the peptide compound are base-initiated, the cleavage of the peptide from the linker tends to be acid labile. Depending on the linker used, different conditions will effect cleavage. In the case of the Wang resin, a strong acidic solution such as 95% TFA in DCM is employed. This condition causes protonation of the C-terminal ester and cleavage from the linker via electron resonance (Scheme 2.3). Alternative linkers, such as chlorotrityl, require milder acidic conditions allowing the peptide to be cleaved without deprotection of the side chain groups.



Scheme 2.3. Acid-initiated cleavage of a peptide from a Wang linker.

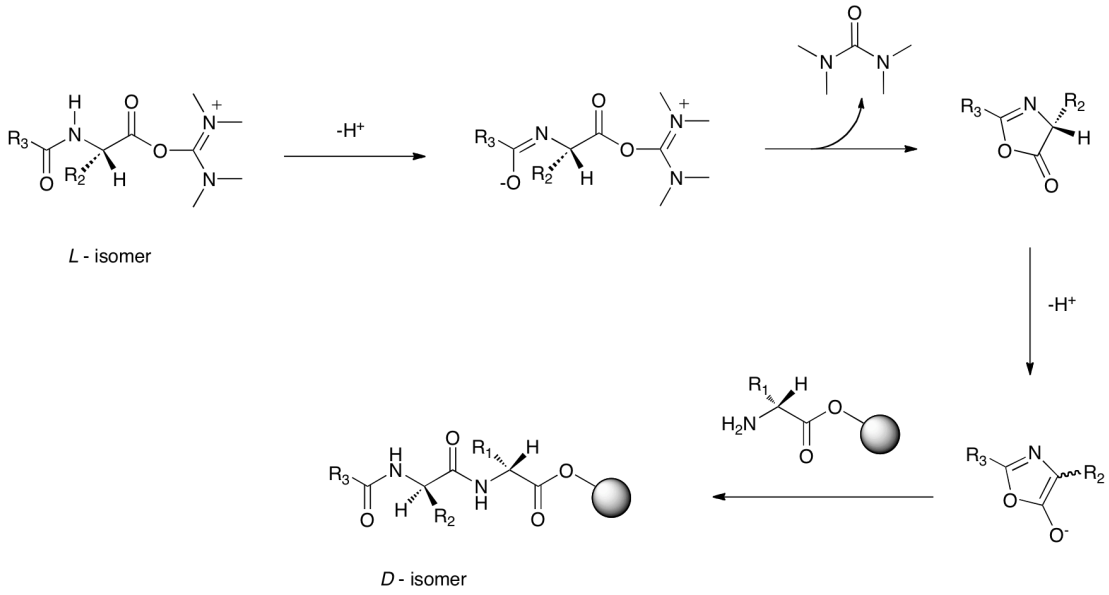
The first amino acid residue must be attached to the linker via an esterification coupling. This initial attachment step is very important for several reasons. Firstly, the yield of the peptide depends on the degree of functionalization of the resin i.e. the loading of the resin. Secondly, the remaining hydroxyl sites that do not react may be acylated later on in the synthesis leading to C-terminally truncated peptides. Due to the harshness of this initial esterification, required to achieve optimal loading, enantiomerisation can be an issue. The strong bases involved ensure an efficient coupling of the hydroxyl to the activated amino acid but may deprotonate the alpha carbon of the incoming amino acids in solution. This will lead to the enolate product

which, when picking up another proton, may give rise to the undesired enantiomer (Scheme 2.4).



Scheme 2.4. Enantiomerisation of the first amino acid residue due to deprotonation of the alpha carbon and formation of the enolate intermediate leading to racemization.

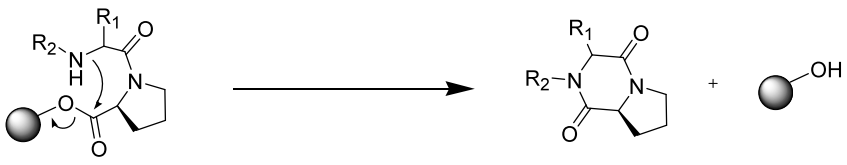
This problem is especially prevalent when attaching an initial cysteine or histindine, as the side chain groups on these amino acids are able to stabilize the enolate and thereby facilitate deprotonation and subsequent racemisation. Racemisation can also occur during the synthesis, activation of the C-terminus of the peptide chain can be followed by the deprotonation of an amide and the subsequent formation of an oxazole (Scheme 2.5). This in turn encourages enolate formation which may lead to the alternative, undesired enantiomer. This is most likely to occur during convergent coupling of peptide fragments.



Scheme 2.5. Oxazole formation and subsequent racemisation.

Premature cleavage from the resin is also an issue during solid phase synthesis. Sequences containing proline as the first residue may present problems due

to the ease of which diketopiperazine is formed. Due to its positioning, attack from the terminal amine of the residue coupled to the proline (Scheme 2.6) to the ester linking the peptide to the resin is possible. This results in the cleavage of these initial two residues leaving the hydroxyl exposed for potential alkylation further on in the synthesis.



Scheme 2.6. Diketopiperazine formation leading to cleavage of the peptide chain from the resin.

Due to these issues, the resin is usually purchased with the linker and first residue already attached. In addition, mild bases are used during the synthesis. The residues are also coupled one amino acid at a time to minimize enatiomerization via ozaxolone formation.

2.3.7 FMOC SPPS

The development of Fmoc peptide synthesis[104] has allowed large peptide compounds to be prepared under mild condition. The solid-phase synthesis of peptides relies on orthogonal chemistry. Base-initiated reactions lead to the formation of the peptide while acidic conditions effect its deprotection and cleavage. Each amino acid is purchased (or synthesized) protected at its N-terminus by a fluorenylmethoxycarbonyl (Fmoc) group (Figure 2.2). If necessary, the functional side chain group of the amino acid is also protected by a moiety that is usually acid labile (i.e. *tert*-butyl ester or trityl groups). However, many of these protecting groups are available to suit a specific type of amino acid R group and its own deprotection conditions. It is therefore possible to plan the synthesis of a peptide so selected R groups remain protected even during the cleavage conditions. This is especially useful when employing non-biogenic amino acids.

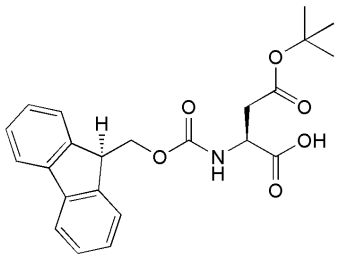
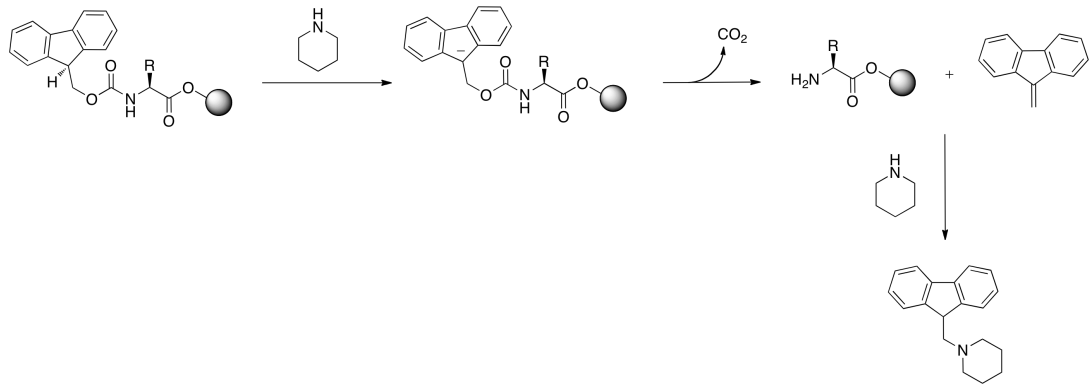


Figure 2.2. An example of a glutamic amino acid carrying a *tert*-butyl protecting group on its carboxy R group and an fmoc group at its amino terminus.

As the Fmoc group is base-labile, the coupling steps can proceed without cleavage of the R protecting groups or the ester linking the peptide to the resin. The deprotection of the Fmoc group is initiated via a solution of 20% piperidine, which deprotonates the acidic chiral proton on the Fmoc group. This leads to the resonance-stabilized anion, which is able to cleave from the peptide chain resulting in the evolution of $CO_{2(g)}$ and an alkene which may subsequently undergo reaction with the piperidine base (Scheme 2.7).

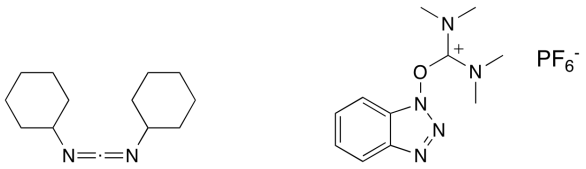


Scheme 2.7. Mechanism for FMOC removal; deprotonation of the acidic chiral proton shown leads to the evolution of CO_2 and release of the free amine. The resulting fluorenyl alkene is attacked by any free piperidine to prevent side reactions.

2.3.8 Coupling Step

Once the Fmoc group on the first residue has been deprotected, the next amino acid can be attached. This incoming residue must first be activated. This is achieved by reacting the unprotected C-terminus carboxyl group with a coupling reagent to

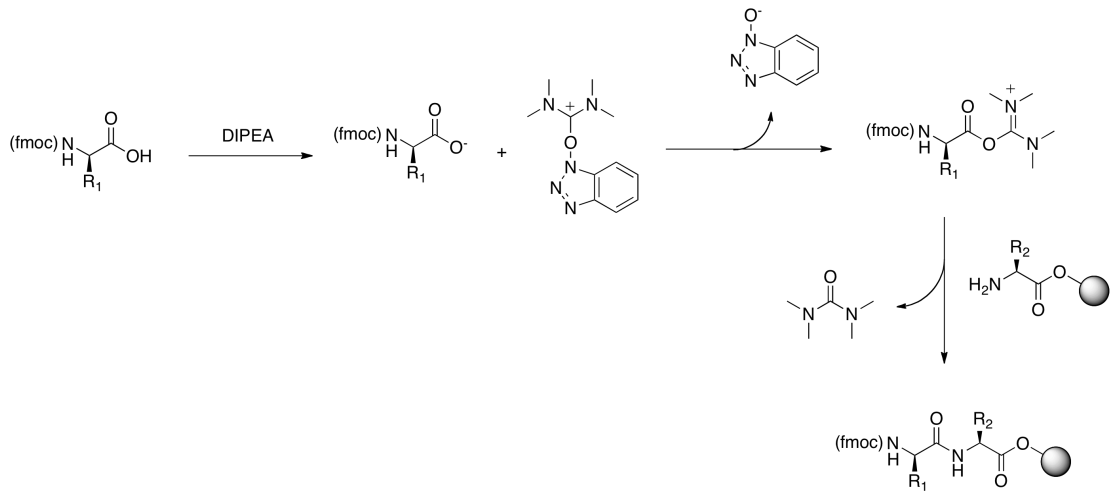
generate the ester *in situ.* Carbodiimide based compounds (dicyclohexylcarbodiimide – DCC, for example) were the first coupling agents to be used, and are still very effective. Due to their increased yield, the alternative aminium/phosphonium salts are now more widely utilized (e.g. HBTU, see Figure 2.3).



1 **2**

Figure 2.3. Examples of two commonly used coupling agents; (1) Dicyclohexylcarbodiimide; (2) HBTU (*O*-Benzotriazol-1-yl-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate).

In the case of the phosphonium salt-based coupling agent (Scheme 2.6), the reaction is initiated via the deprotonation of the amino acid carboxyl group via a tertiary base such as DIPEA (*N*,*N*-diisopropylethylamine). The carboxylate then attacks the carbocation of the activating group kicking out HOBt forming the activated ester. At this point the Fmoc-deprotected amine can attack the activated ester to create the stable amide bond.



Scheme 2.8. HBTU coupling of an amino acid residue to the resin linked peptide chain; Deprotonation of the carboxylic acid of the incoming amino acid allows it to attack the coupling agent forming an activated ester *in situ.* The free amine terminus of the immobilized peptide chain can then attack this ester to displace the urea by-product and form of the new amide bond.

The yield of each coupling step is especially important in this type of solid phase synthesis. It is the deciding factor in the total length of peptide strand the technique is able to prepare. For example, if each step goes to 99% completion then a 25-mer will be synthesized in 77% yield. Any decrease in yield per step will dramatically decrease the yield of the final peptide and increase the amount of unwanted truncated by-products. As the yield of this step is very important, additional compounds may be added to achieve high yield. 1-hydroxy-7-azabenzotriazole (HOAt), for example, has been shown to improve yields. This group acts to encourage attack on the activated ester by hydrogen bonding with the N-terminus amine of the resin-bound peptide. This brings the groups into close proximity allowing for a more efficient reaction.

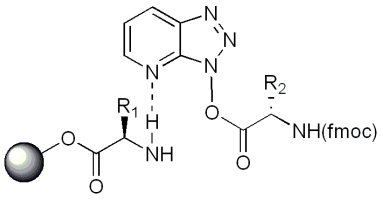


Figure 2.4. Hydrogen bonding of HOAt to the peptide chain bringing its terminal amine closer to the activated ester and keeping both moieties static relative to one another increasing the efficiency of the coupling reaction.

2.3.9 Peptide Cleavage and Work-Up

The cycling of the synthetic procedure involves the FMOC group of a newly coupled residue first being taken off with 20% piperidine and the next amino acid activated and coupled. This process continues until the desired number of residues have been coupled to the resin. The beads are then washed with the appropriate solvents to remove any reagents still remaining in the resin meshwork. The solid phase is then treated with a cleavage solution (95% TFA in DCM for the Wang linker) including scavenger agents such as triethyl silane (TES) and compounds such as ethylene dithiol (EDT). These reagents react with any carbocations (arising, for example, from the deprotection of *tert* - Butyl groups) and ensure that thiol groups (e.g. cysteines) remain reduced. After cleavage from the resin, the peptide is precipitated in diethyl ether and analyzed via RP-HPLC and ESI-MS. If there are no truncated sequences, the peptide may be pure enough to use without need for further HPLC purification.

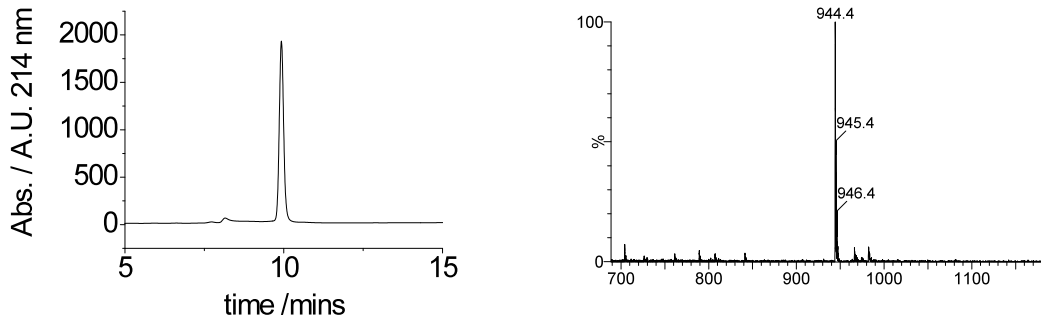


Figure 2.5. RP-HPLC (left) and ESI MS (right) examples of a $CysHis_6$ peptide – see experimental methods section 2.3.10 for details.

2.3.10 General SPPS Procedure

All peptides employed for the chemical tagging of DNA were synthesized using standard Fmoc solid-phase peptide synthesis on a Syro automated system. Pre-loaded Wang resin was used with standard HBTU/DIPEA coupling chemistry. All amino acids were purchased as N-protected Fmoc and appropriately protected R group compounds. Each coupling step was conducted for 40 mins and each Fmoc deprotection allowed to proceed for 10 min in 20% piperidine. Four equivalents of amino acids were used. After completion of the synthesis, peptides were deprotected and cleaved from the resin by incubation in TFA/TES/EDT/H_2O (2500:150:75:150 μL) for 3 hours. The collected peptide solution was mixed with diethyl ether and allowed to precipitate at -20°C. The suspension was centrifuged at 4000 rpm, and the pellet was re-dissolved in diethyl ether. This purification process was repeated three times after which the precipitate was dissolved in deionized H_2O or 0.1% TFA containing HPLC grade water if necessary, frozen in liquid N_2 and freeze-dried overnight. The peptides were analyzed and if necessary purified via reversed phase HPLC using a Varian ProStar system with a Model 210 solvent delivery module and a Model 320 UV detector. The preparative purification was performed using a Varian column (100 x 21.2 mm, 5 μm beads, flow rate of 10 mL/min) loaded with 100 μL aliquots of a 20 mg/mL solution of peptide dissolved in 0.1% TFA containing water. The mobile phase was a linear gradient beginning with 98:2 water (0.1% TFA) / acetonitrile (0.1% TFA) to 40:60 water (0.1% TFA) / acetonitrile (0.1% TFA) over 30 min, this gradient was adapted when necessary. The analysis of the chromatograms was conducted using Star Chromatography Workstation software Version 1.9.3.2. Fractions containing the correct peak were pooled, the solvent removed under reduced pressure to approximately 2 mL, and the solution freeze-dried overnight. The resulting white powder was analyzed via HPLC using a Gemini column (250 x 4.6 mm, 5 μm beads, flow rate of 1 mL/min) using the solvent system and gradient detailed above. ESI-MS analysis was performed on a Waters Acquity Ultra Performance LC/MS machine in positive ion mode using the same solvent system and gradient as the analytical HPLC procedure detailed above (0.6 ml/min, 0.1% formic acid used in place of TFA).

2.3.11 Coupling of Cysteine-Terminated Peptides to SPDP-Activated DNA

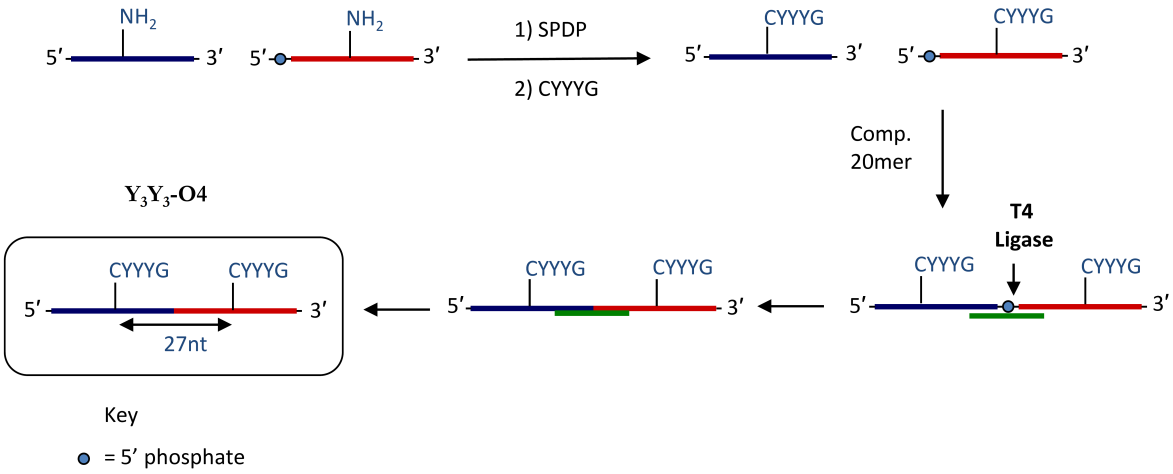
For oligonucleotide activation, a 1 mM solution of oligonucleotide (25 µL, 25 nmol) in 10 mM Tris, 1 mM EDTA, pH 8.0 was diluted with 0.1 M $NaHCO_3$, 50 mM NaCl, pH 9.0 (65 µL). The mixture was combined with a 250 mM solution of SPDP in DMSO (10 µL, 2.5 µmol) leading to a white precipitate, which immediately re-dissolved upon agitation. The reaction was incubated at room temperature for 30 mins. Unreacted SPDP was removed by size exclusion chromatography using an AKTA purification system. A Sephadex column with a volume of 5 mL was loaded with 100 µL of the reaction mixture at a flow rate of 5 mL/min. The mobile phase was buffer A (20 mM Tris, 50 mM NaCl, pH 8.0), and the DNA containing fraction was collected using an isocratic gradient.

A cysteine-terminated peptide was covalently coupled to SPDP-activated DNA oligonucleotides. Peptide (5 µmol) was dissolved in distilled H_2O containing 0.1 % TFA (100 µL, final concentration of 50 mM). An aliquot (10 µL, 0.5 µmol) was combined with an aqueous solution of purified SPDP-activated DNA (2 mL, 25 nmol). In the case of R_7C_1-O1, 50% v/v of formamide was added to the buffer containing the SPDP-activated oligonucleotide and the peptide to improve the solubility of the conjugate. The solutions for all peptide oligonucleotide mixtures were agitated on a shaker at room temperature for 2 hours followed by purification via anion exchange chromatography on an AKTA purifier system using a Resource Q column (volume 1 mL). A linear gradient was employed beginning with buffer B (20 mM Tris, pH 8.0) to 60:40 buffer B / buffer C (buffer B plus 2 M NaCl) over 45 min at a flow rate of 1 mL/min. Product-containing fractions were desalted and concentrated to a volume of 200 µL using ultrafiltration spin-columns (Amicon, 5 mL, 5KDa MW cut-off).

The concentration of the purified conjugates were determined using UV/vis spectroscopy (Cary Eclipse, Varian), and their purity assessed using analytical AEC. The purity of the His_6-tagged conjugate was also controlled via immobilized metal affinity chromatography (IMAC) on an AKTA system with a Ni-NTA containing His-Trap column (1 mL/min). The conjugate was loaded onto the column using ten column volumes of buffer D (0.1 M Na_2HPO4, 50 mM NaCl, pH 8.0) and eluted using a step gradient to buffer E (buffer D plus 1 M imidazole). The mass of conjugates were determined via MALDI MS analysis (Section 2.4.3)

2.3.12 Ligation of peptide-modified oligonucleotides

To prepare an oligonucleotide carrying two peptide tags, a convergent method was applied (Scheme 2.9). Two single-modified 27-mers were initially prepared, one containing a phosphorylated 5'-terminus. Using the enzyme T4 Liagase and a complementary bridging 20-mer strand, the duplex was ligated to form a continuous strand with 54-nt in length.



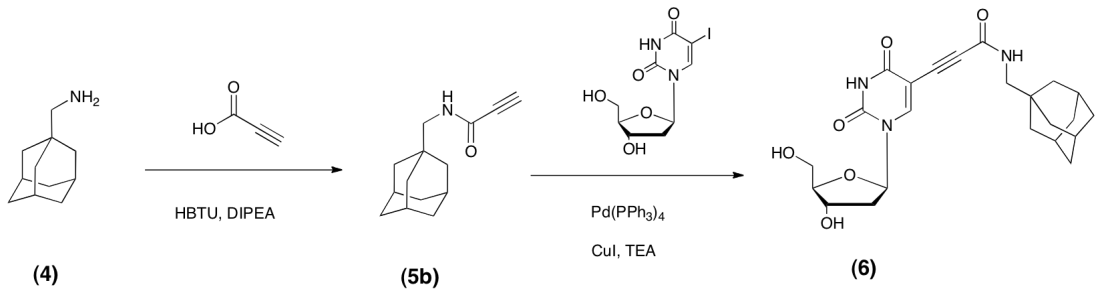
Scheme 2.9. Convergent synthesis of a double modified 54-mer oligonucleotide via ligation of two single-modified 27-mers using T4 Ligase and a complementary bridging 20mer oligonucleotide.

The synthesis of Tyr_3/Tyr_3-O4 (Table 2.1) was performed with 27-mer oligonucleotides O4A of sequence 5' – ACA TTC CTA ACA **T**CA CTA ACT AAT CTT -3' and O4B of sequence 5'-/5Phos/TCA ACT CGA TAC G**T**A CCT ATC AAT CTA -3'. The two internal amine-modified thymine bases (underlined) were activated with SPDP and coupled to $Gly_1Tyr_3Cys_1$ as described previously. The chemical analysis of resulting POCs $Gly_1Tyr_3Cys_1$-O4A and $Gly_1Tyr_3Cys_1$-O4B was conducted via anion exchange chromatography and MALDI MS. To ligate the two POCs, a bridging oligonucleotide O4C was added. O4C with the sequence 5'-A TCG AGT TGA AAG ATT AGT T-3' was complementary to the 3' end of O4A and to the 5' end of O4B. Equimolar amounts of $Gly_1Tyr_3Cys_1$-O4A, $Gly_1Tyr_3Cys_1$-O4B, and O4C (12.5 nmol) were diluted in 50 mM Tris, 10 mM $MgCl_2$, 1 mM ATP, pH 7.5 (370 µL). The mixture was heated in a PCR block to 95 ºC for 2 min and cooled to room

temperature. The duplex was ligated by adding T4 ligase (16 μL, 6400 units) followed by incubation at room temperature for 16 hours. The resulting 54-mer POC Tyr_3/Tyr_3-O4 was separated from bridging O4C using denaturing 15% PAGE. The peptide-modified DNA strands were used to achieve the first two aims, namely the proof-of-principle demonstration that peptide tags can slow down DNA translocation through a nanopore, and the application of the technology to discriminate between single nucleotide polymorphism (SNP).

2.3.13 Preparation of Adamantane-Tagged Nucleosides for Aim 3.

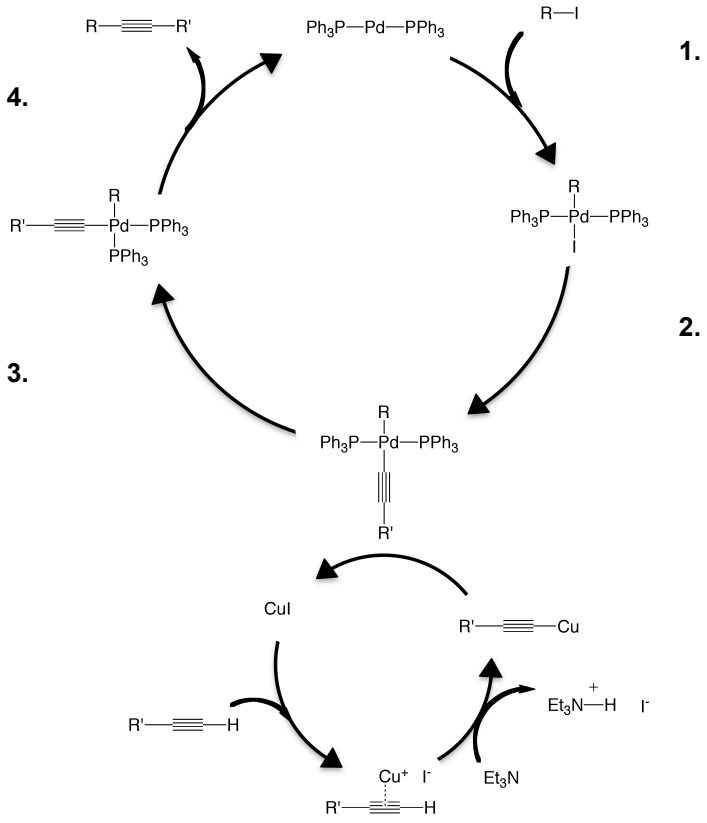
The third aim in this section of the thesis is concerned with the incorporation of a smaller, more versatile tag into an oligonucleotide for sizing highly repetitive DNA sequences of forensic importance. The chemical foundations towards this aim were laid by preparing an adamantane-derivatized nucleotide. The synthesis of the nucleoside form of the analogue is outlined in Scheme 2.10 and encompasses the Sonogashira-mediated coupling of an appropriately modified adamantane-acetylene compound **5b** with 5-iodo-2'-deoxyuridine. In addition to this compound a maleimide modified adamantane compound **(5a)** was also synthesized to allow coupling onto an appropriately modified oligonucleotide.



Scheme 2.10. Synthesis of 5-adamantane-2'-deoxyuridine **(6)**; HBTU mediated coupling of an appropriate acetylene with adamantane carrying a primary amine. Sonogashira coupling of this compound **5b** with 5-iodo-2'-deoxyuridine yields the desired nucleotide **6**.

Sonogashira coupling[105] allows an alkyne to be attached to an alkyl halide (Scheme 2.11); it is an efficient, versatile and high yielding reaction that must be conducted under an inert atmosphere to prevent oxidation of the palladium. The

mechanism initially involves the Pd(0) carrying two triphenyl phosphine ligands. The Pd metal oxidatively inserts into the carbon – halide bond of the starting compound (oxidizing to Pd(II) in the process). Meanwhile the terminal proton of the alkyne is replaced by the Cu metal of CuI via association of the metal first with the triple bond and with the help of triethyl amine removal and replacement of the hydrogen. This organo-metallic then displaces the halide on the Pd complex.

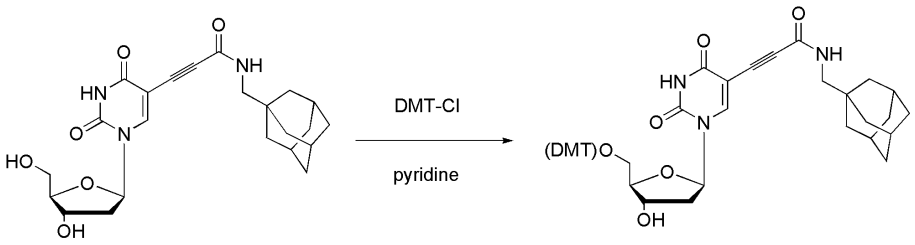


Scheme 2.11. Pd(II) catalyzed Sonogashira reaction; (1) oxidative insertion; (2) trans metallation; (3) trans – cis isomerization; (4) reductive elimination.

The ligands then orientate so the reacting compounds are *cis* to one another allowing reductive elimination to produce the desired compound and eliminating the starting Pd(0) catalyst. Since the Pd complex is prone to oxidation all solutions must be thoroughly degassed with Argon and the reaction must be conducted under an inert atmosphere throughout.

The resulting compound **6** isolated from the Sonogashira coupling can then be phosphorylated to give the triphosphate nucleotide analogue for enzymatic

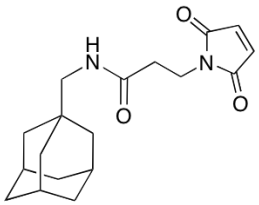
incorporation into a DNA template. Alternatively the corresponding phosphoramidite can be prepared via protection of the 5' hydroxyl group with dimethoxytrityl (DMT) (Scheme 2.12) and subsequent phosphitylation of the 3' hydroxyl group using cyanoethyl diisopropylphosphoramidite (CEP).



(6) (7)

Scheme 2.12. Dimethoxytrityl protection of the 5' hydroxyl group.

2.3.14 Maleimide propyl amino adamantane (5a)

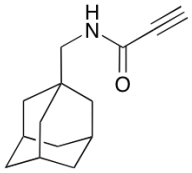


(5a)

To 338 mg (2 mmol) of *N*-maleoyl-β-alanine in anhydrous DCM (10 mL) was added 412 mg (2 mmol) of DCC also in anhydrous DCM. To this solution, 321.5 μL (300 mg, 1.82 mmol) of 1-adamantane methyl amine was added, and the reaction left stirring under an atmosphere of argon at RT overnight. After this period, the solvent was removed under reduced pressure, the crude was dissolved in neat methanol and reduced onto silica. The crude was purified via flash chromatography using a solvent system of 2% methanol in DCM. The fractions containing the desired product were pooled, and the solvent removed under reduced pressure to yield the title compound **5a** (356 mg, 64.5% yield). ^{1}H NMR (500 MHz; d_6-DMSO) δ/ppm 1.35 (6H, s), 1.6 (6H, dd, J = 12.0), 1.9 (3H, m), 2.35 (2H, t, J = 7.3), 2.68 (2H, d, J = 6.3), 3.6 (2H, t, J

= 7.3), 7.0 (2H, s), 7.7 (1H, t, J = 6.0). ^{13}C NMR (125 MHz; d_6-DMSO) δ/ppm 24.5, 25.3, 27.7, 33.3, 33.5, 33.9, 34.4, 36.5, 47.5, 50.1, 134.6, 156.6, 169.5, 170.7.

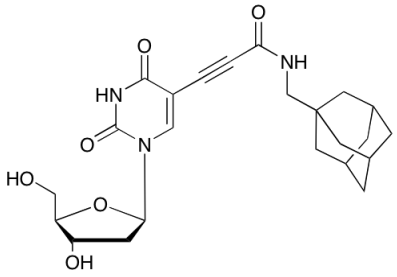
2.3.15 Propiolic (adamantane-1-yl methyl) amide (5b)



(5b)

To 2.137 g (6.65 mmol) of TBTU in 30 mL anhydrous DCM was added 0.466 g (6.65 mmol) of propiolic acid (also in DCM). The solution was stirred in ice for the duration of the reaction. 2.107 mL (12.1 mmol) of DIPEA was added to the reaction, followed by the dropwise addition of 1 g (6.05 mmol) of 1-aminomethyladamantane in anhydrous DCM. The solution was stirred on ice for 4 hours. During this period, the solution turned bright yellow. The solvent was then removed under reduced pressure and residue was purified via column chromatography using a solvent system of 1:4 ethyl acetate:hexane, The fractions containing the desired product were pooled and the solvent removed under reduced pressure to yield the titled compound **5b** (0.74 g 56.3% yield). ^{1}H NMR (500 MHz; d_6-DMSO) δ/ppm 1.4 (6H, s), 1.6 (6H, dd, J = 12.3), 1.9 (3H, m), 2.8 (2H, d, J = 6.4), 4.1 (1H, s), 8.6 (1H, t, J = 6.4). ^{13}C NMR (125 MHz; d_6-DMSO) δ/ppm 13.9, 24.3, 25.0, 25.0, 25.4, 30.3, 31.6, 33.2, 49.5, 52.8, 71.2, 153.2, 167.7.

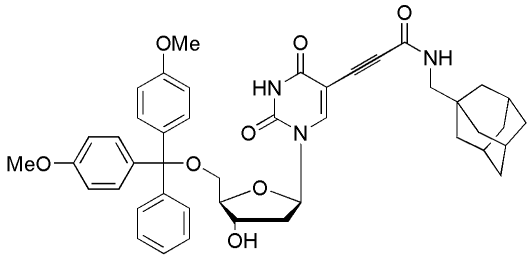
2.3.16 5-Adamantane-2'-deoxyuridine



(6)

5-Iodo-2'-deoxyuridine (0.677 g, 1.91 mmol), palladium tetrakis triphenyl phosphine (221 mg, 0.191 mmol) and copper iodide (72.8 mg, 0.382 mmol) were placed in a dried round bottom flask. The flask was purged and flushed with argon, followed by the addition of 30 mL of degassed anhydrous DMF and a degassed solution of TEA (0.968 g, 1.33 mL, 9.56 mmol) in 10 mL of degassed DMF. 0.829 g (3.82 mmol) of propiolic (adamantane-1-ylmethyl) amide (**5b**) dissolved in 20 mL of degassed anhydrous DMF was then added dropwise, and the reaction allowed to proceed at RT under argon. After 22 h, the solvent was removed under high vacuum to yield a crude brown oil. This crude product was diluted in 1:1 DCM:MeOH, reduced onto silica and purified via flash chromatography using a 5% MeOH in DCM solvent system. The fractions containing the desired product were pooled, and the solvent removed under reduced pressure to yield the titled compound **6** (0.72 g, 85.4% yield). Calculated mass ($C_{23}H_{29}N_3O_6$) 443.5 m/z; observed mass (ESI^-) *m/z* – 442.6 (MH^-). ^1H NMR (500 MHz; d_6-DMSO) δ/ppm 1.4 (6H, m), 1.6 (6H, dd, J = 12.0), 1.9 (3H, m), 2.1 (2H, dd, 5.0), 2.8 (2H, d, J = 6.6), 3.6 (2H, m), 3.8 (1H, dd, J = 3.5), 4.2 (1H, m), 5.1 (1H, t, J = 4.7), 5.2 (1H, d, J = 4.1), 6.1 (1H, t, J = 6.6), 8.4 (1H, s), 8.5 (1H, t, J = 6.6), 11.7 (1H, bs). ^{13}C NMR (125 MHz; d_6-DMSO) δ/ppm 8.8, 27.7, 34.0, 36.5, 45.8, 48.6, 50.4, 60.9, 70.0, 77.2, 85.1, 87.0, 87.7, 96.0, 146.0, 149.3, 152.6, 161.2.

2.3.17 5-Adamantane-5'-DMT-2'-deoxyuridine (7)



(7)

The adamantane-modified nucleoside **6** (0.5 g, 1.128 mmol) was co-evaporated with anhydrous pyridine three times before being re-dissolved in 10 mL of pyridine under argon. DMT-Cl (0.42 g, 1.24 mmol) dissolved in argon-flushed pyridine was added drop-wise via a syringe. The reaction was allowed to proceed at RT under argon overnight. A further 0.1 equivalents of DMT-Cl was added (38.3 mg in 1 mL of anhydrous pyridine). After a further 1h, the solvent was removed under high vacuum and co-evaporated with toluene three times. The crude oil was dissolved in DCM and washed with a saturated solution of $NaHCO_3$ and then brine. The organic phase was dried over $MgSO_4$ and the solvent was removed under reduced pressure, followed by co-evaporation with toluene (3 x). The crude was then purified via flash chromatography using a solvent system of 5:94:1 MeOH:DCM:pyridine. The fractions containing the desired product were pooled and the solvent removed via reduced vacuum. The compound was co-evaporated with toluene three times to yield the titled compound **7** (0.183 g, 21.8%). Calculated mass ($C_{44}H_{47}N_3O_8$) 745.8 m/z; observed mass (ESI^-) 744.6 m/z (MH^-). ^1H NMR (500 MHz; d_6-DMSO) δ/ppm 1.4 (6H, m), 1.6 (6H, dd, J = 11.6), 1.9 (3H, m), 2.2 (2H, m), 2.8 (2H, d, J = 6.6), 3.2 (2H, m), 3.7 (6H, s), 3.9 (1H, m), 4.2 (1H, m), 5.3 (1H, d, J = 4.6), 6.1 (1H, t, J = 6.6), 7.2 (13H, m), 8.0 (1H, s), 8.3 (1H, t, J = 6.6), 11.8 (1H, bs). ^{13}C NMR (125 MHz; d_6-DMSO) δ/ppm 21.0, 27.7, 34.0, 36.4, 50.4, 55.0, 63.8, 70.5, 76.7, 85.7, 85.8, 86.0, 87.1, 96.2, 113.2, 125.3, 126.7, 127.6, 127.9, 128.2, 128.9, 129.6, 129.8, 135.2, 135.7, 137.3, 144.7, 145.7, 149.2, 152.5, 158.1, 158.1, 161.2.

2.3.18 Nanopore recordings

Single-channel current recordings were performed by using a planar lipid bilayer apparatus as described.[106] Briefly, a bilayer of 1,2- diphytanoyl-sn-glycerophosphocholine (Avanti Polar Lipids) was formed on an aperture (80 μm in diameter) in a Teflon septum (Goodfellow Corporation, Malvern, PA) separating the cis and trans chambers of the apparatus. Each compartment contained 2 M KCl, 50 mM Tris-HCl, pH 8.0 unless otherwise stated. Gel-purified heptameric αHL protein (final concentration 0.01– 0.1 ng/ml) was added to the cis compartment, and the electrolyte in the cis chamber was stirred until a single channel inserted into the bilayer. Transmembrane currents were recorded at a holding potential of +100 mV (with the cis side grounded) by using a patch–clamp amplifier (Axopatch 200B, Axon Instruments, Union City, CA). For analysis, currents were low-pass filtered at 10 kHz or 30 kHz and sampled at 50 kHz or 100 kHz, respectively, by computer with a Digidata 1200 A/D converter (Axon Instruments), as described.[107]

2.4 Sensing of Chemically Modified DNA using Nanopore Recordings

To enhance the resolution of nanopore recording individual bases were chemically modified with tags carrying steric bulk and/or charge. By adding these components to the DNA, the energy needed to thread the strand through the pore was expected to enlarge. If successful, the strategy should result in the lengthening of the translocation time. Furthermore, it was also anticipated that different chemical tags would give rise to characteristic current modulations. In this way, identification of differently modified bases via the characteristic signatures should be possible.

Peptides are ideal for this purpose due to their convenient synthesis via solid phase and due to the ease of with which their molecular characteristics can be altered by employing alternative amino acid sequences.

2.4.1 Peptide – Oligonucleotide Conjugates

To test the hypothesis that peptide tags slows the translocation of DNA through the nanopore, a 27-mer oligonucleotide was modified with a variety of peptides. The chemical structure of one conjugate, focusing on the modified base, is shown in Figure 2.6. The cysteine-terminated peptide is attached via a heterobifunctional SPDP crosslinker to a non-biogenic thymine base which carries a primary amine group linked to position 5.

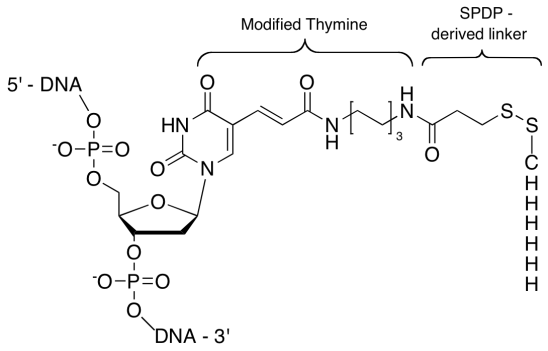


Figure 2.6. Chemical structure of an oligonucleotide carrying a peptide modification at an internal base position.

2.4.2 Convergent Synthesis

Given the chemical incompatibilities of the synthesis for peptides (acid-based) and oligonucleotides (alkali-based) previously discussed, a convergent approach to the conjugate synthesis was pursued. In this strategy, the oligonucleotide and peptide are synthesised and purified separately and, using complementary functionalization on each component, reacted to form the conjugate.

For successful conjugation of the peptide-modified DNA strands, a mild and high yielding chemistry based on the heterobifuctional crosslinker *N*-succinimidyl 3-(2-pyridyldithio)-propanoate (SPDP) was used. This linker carries an *N*-hydroxysuccinimide activated ester that allows attack from the primary amine of the modified oligo at room temp. in a buffered solution (pH 9.0) in under 30 min. SPDP was coupled to the amine-modified DNA oligonucleotide in a buffered solution, and

excess reagent was removed by size exclusion chromatography. Subsequently, the orthopyridyl disulfide group of the cross-linker was reacted with the thiol side chain group of a cysteine bearing peptide such as NH_2-Cys-His_6-COOH at room temp. within 1h.

This strategy was found to be successful. The final peptide-conjugated oligonucleotides were isolated at very high purity (> 95% see Figure 2.7) as determined with anion exchange chromatography without severe loss of material (approx. 50% yield). The ease of purification depended on the charge and hydrophobic steric bulk of the peptide tag, leading to some conjugates being easier to purify than others. For example the Arg_7 POCs eluted a few minutes before their unmodified counterparts on AEC due to the seven positive charges on the backbone of the peptide effectively neutralizing seven of the negative charges on the DNA. This allows the oligo to elute faster than the unmodified oligo and allows more effective purification. The opposite for the poly-Glycine peptides is observed, due to a lack of bulk and charge the modified oligonucleotides are hard to purify for the unmodified strands.

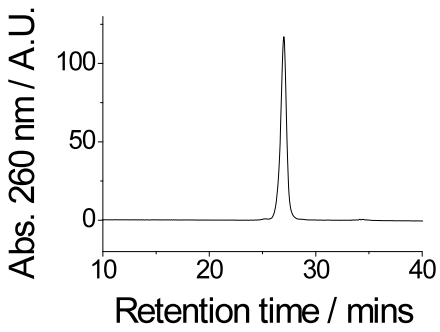


Figure 2.7. Representative anion – exchange chromatograph for $CysTyr_3$ – O1 POC (27mer).

2.4.3 Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI MS)

Due to the high mass and chemical complexity of the conjugates, NMR was not a viable method for their chemical characterization. Firstly, a relatively large amount of material is needed even when running a ^1H NMR experiment on a powerful machine (i.e. > 600 MHz) and it is expensive to synthesize a large amount of modified

oligonucleotide. For the nanopore experiments described here only 25 nmol (or around 0.2 mg) of oligonucleotide was modified (and around 50% of the final product isolated). Whereas only picomole amounts of material are needed for electrophoresis and mass spec. The complex spectrum resulting from an NMR experiment would also take time and expertise to elucidate. Considering these issues NMR is clearly not the ideal characterization technique. Mass Spectrometry is in fact the method of choice alongside analytical chromatographic methods for determination of successful conjugation. MALDI-MS was used in place of electrospray since it has a higher mass range than and importantly is more tolerant to salts. Prior to spotting the DNA onto the MS-plate, the samples were desalted using ZipTips© to avoid the unfavorable neutralization of the overall negatively charged peptide-DNA conjugates by sodium ions. DNA was eluted using NH_4OAc buffer as ammonium does not bind as tightly to the DNA backbone as sodium. The MS-matrix 2,3,4-trihydroxyacetophenone (THAP) was found to give the best results. Large, complex biological molecules can be accurately characterised using this technique. Figure 2.8 shows a representative MS-spectrum for the 37-mer DNA (O3) strand carrying two Cys_1Tyr_3 peptide tags at over 13,100 Da in mass. The peaks for the molecular ion (M^{1-}) and the M^{2-} ion were observed.

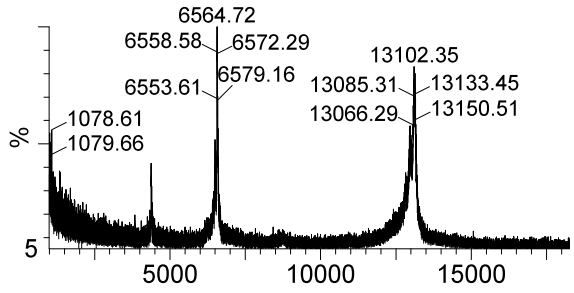


Figure 2.8; MALDI spectrograph of a peptide – oligonucleotide conjugate, M^{1-} and M^{2-} peaks observed (calculated mass; 13105 $g.mol^{-1}$).

The resolution of the peak depends largely on the desalting procedure and concentration of the sample; the more thorough the desalting and the higher the conc. the more accurate and well defined the peak. Taking an average of 100 scans an error of approx. 0.2% is common; therefore this technique is not as accurate as its ES counterpart. However, with proper desalting and by taking the average of a smaller data set errors well below this number can be achieved, right down to a mass M – 1

i.e. the accurate mass of the M^{1-} molecular ion. The size of the molecule also effects the resolution, oligos containing in excess of 30 residues become harder to resolve than those with fewer.

2.4.4 Summary of Synthesized Conjugates

Using the procedures detailed above, a variety of peptide-oligonucleotide conjugates (POC) were successfully prepared (Table 2.1). The conjugates were either singly- or double-modified. The tags for the single-modified conjugates comprised peptides of different composition (glycine, histidine, arginine, tyrosine) and length (2, 4, or 6 residues in addition to a single cysteine). The double-modified conjugates carried two $Gly_1Tyr_3Cys_1$ tags on DNA oligonucleotides of different length.

Oligo-nucleotide		Peptide				POC		
Name [a]	Length [nt]	Sequence [b]	Mass [m/z][c] Expected	Mass [m/z][c] Observed	HPLC retention time [min][d]	Mass [m/z][e] Expected	Mass [m/z][e] Observed	AEC retention time [min] / purity [%][f]
O1	27	n.a.	n.a.	n.a.	n.a.	8380.6	8379.4	25.5 / 99
O1	27	H_6C_1	944.00	944.36	9.7[i]	9411.2	9410.8	24.1 / 96
O1-term.	27	$G_3H_6C_1$	1253.00	627.82[g]	15.0[j]	9746.6	9759.03	25.4 / 75
O1	27	H_4C_1	669.72	670.50	5.2	9136.1	9134.6	24.5 / 93
O1	27	H_2C_1	395.44	396.33	4.2	8862.6	8862.4	24.5 / 99
O1	27	G_6C_1	463.47	486.41[h]	3.9	8930.8	8930.5	24.1 / 98
O1	27	G_4C_1	349.36	372.24[h]	3.8	8816.2	8815.2	24.2 / 98
O1	27	G_2C_1	235.26	258.31[h]	3.7	8702.3	8703.8	24.1 / 99
O1	27	$G_1R_7C_1$	1271.52	636.55[g]	13.1[k]	9738.6	9737.9	21.3 / 99
O1	27	$G_1R_5C_1$	959.15	480.48[g]	12.5	9426.2	- [m]	21.8 / 90
O1	27	$G_1R_3C_1$	646.80	647.49	11.5	9113.9	- [m]	22.5 / 96
O1	27	$G_1Y_3C_1$	667.73	668.18	13.3[l]	9134.6	9135.6	27.4 / 96
O2	27	$G_1Y_3C_1$	667.73	668.18	13.3[l]	9149.6	9148.1	27.4 / 99
O3	37	$2xG_1Y_3C_1$	667.73	668.18	13.3[l]	13105	13102	30.4 / 99
O5	37	$2xG_1Y_3C_1$	667.73	668.18	13.3[l]	13098	13098	29.0 / 99.3
O4	54	$2xG_1Y_3C_1$	667.73	668.18	13.3[l]	18160	18161	28.8 / 96.8

Table 2.1; Chemical Characterization of Peptides and Peptide-Oligonucleotide Conjugates (POCs)

n.a.; not applicable; [a] For sequences of oligonucleotides see table 2.2.[b] The peptide sequence is from the C to the N terminus. [c] Peptide masses were obtained in the ESI+ mode. Peptides were dissolved in HPLC grade water containing 0.1% TFA. [d] Unless otherwise stated, retention times are for HPLC analysis using C_{18} column (250 mm x 4.6 mm) and mobile phases 0.1% TFA in water and 0.1% TFA in acetonitrile with a gradient of 5 % to 20 % acetonitrile/TFA over 10 min at a flow rate of 1 mL/min. [e] Oligonucleotide masses were obtained with MALDI–TOF using THAP as matrix and ammonium acetate as co-matrix. [f] Anion exchange chromatography (AEC) was performed using a Resource Q column (1 mL) and buffer B (20 mM Tris, pH 8.0) and buffer C (buffer B plus 2 M NaCl) with a linear gradient from 0 to 40 % B over 45 min at a flow rate of 1 mL/min. [g] M2+ ion. [h] Peptide mass for the monosodium ion; [i] Gradient of 3 % to 10 % acetonitrile/TFA over 20 min.; [j] Gradient of 2 % to 50 % acetonitrile/TFA over 20 min.; [k] Gradient of 2 % to 60 % acetonitrile/TFA over 20 min.; [l] Gradient of 2 % to 98 % acetonitrile/TFA over 20 min. [m] Entire sample required for nanopore recording therefore successful conjugation inferred from shift in AEC retention time.

Oligo.	Sequence
O1	5'-ACA TTC CTA AGT TCT GAA ACA TTA CAG-3'
O2	5'-TCA ACT CGA TAC GTA CCT ATC AAT CTA-3'
O3	5'-Phosph-TCA ACT CGA TAC GTA CCT ATC AAT CTA GTA CCT ATC A-3'
O4	5' – ACA TTC CTA ACA TCA CTA ACT AAT CTT TCA ACT CGA TAC GTA CCT ATC AAT CTA – 3'
O5	5'-Phosph-TCA ACT CGA TAC GTA CCT ATC TAT CTA GTA CCT ATC A – 3'

Table 2.2; Oligonucleotide sequences used; underlined **T** indicates amino linker modification.

2.5 Nanopore Analysis of Tagged Oligonucleotides

Once the appropriate peptide modified oligos had been prepared, the first aim of this chapter was addressed; to use the tags to slow the translocation of DNA through a biological nanopore. It was also intended to identify the tags as they passed the inner constriction of the pore via their characteristic current modulations.

2.5.1 Nanopore Traces

As a voltage is applied across a bilayer containing the biological nanopore the resulting current flow can be recorded and monitored via the *pClamp* program. This software, usually used by electophysiologists, can analyze the details of the traces and deposit the data into a spreadsheet for further graphical analysis. A typical trace usually lasts up to 30 mins (see Figure 2.9) after this time the lipid bilayer tends to become comprimised due to the effect of the induced voltage. This period, however, is long enough to obtain thousands of individual events to give statistically relevant data. It is also short enough to utilise in a point-of-care fashion i.e. for swift analysis of samples.

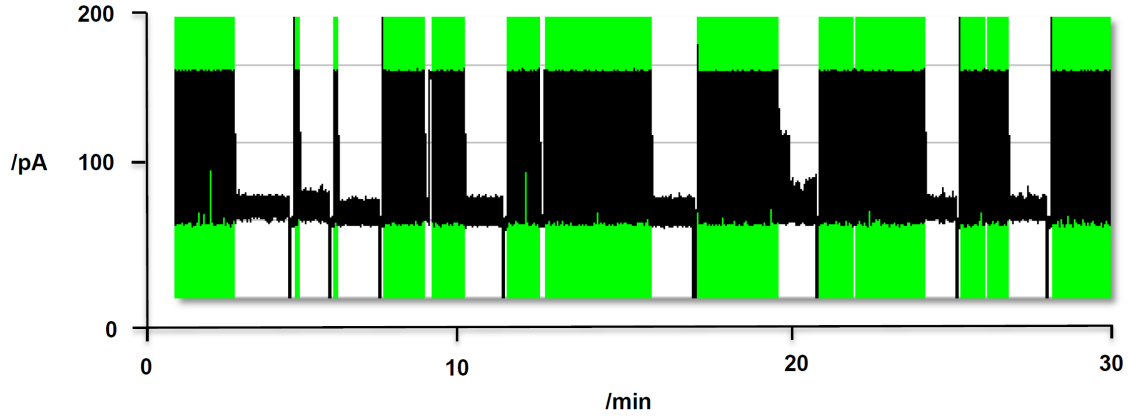


Figure 2.9. The digital output from a 30 minute nanopore recording of a POC. The black sections represent the current traces.

The open channel current is termed τ_{on}, once a charged species is introduced into the *cis* compartment and begins to translocate through the pore we observe sharp decreases in current termed τ_{off}. It is the duration and amplitude of these τ_{off} events that we are interested in analyzing. The pClamp program compresses the τ_{on} duration i.e. the open channel current between events (seen in green in Figure 2.9 & 2.10) to decrease the size of the data file. Typically there are a few translocations every second, this frequency changes with increasing voltage and differs across translocating species.

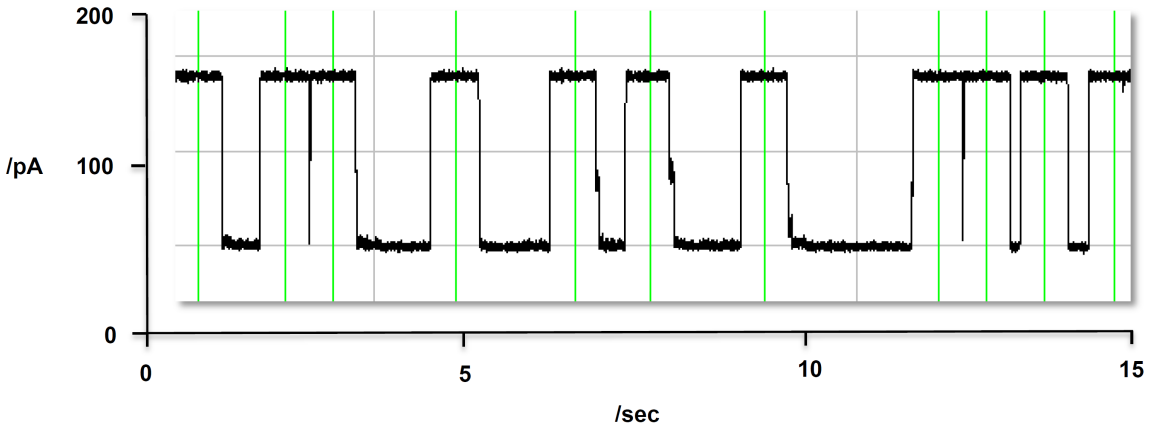


Figure 2.10. A section of the trace magnified, each current blockade represents a single molecule of DNA passing through the pore, the τ_{on} sections are compressed (green lines) to allow for faster processing.

Each event is characterized via its dwell time value (τ_{off} / ms) and its amplitude (A_h; meaning high amplitude event) (see Figure 2.11). The latter value is processed in pA and normalized so as to represent a decimal value of the open current (1 being 100% of the current and therefore a blocked pore). Depending on how the software calculates this value it may give the residual current I_{res} of the event in which case $[I_{open} - I_{res}] / I_{open} = A_h$. Alternatively it may represent the amplitude of an event as a negative value in pA, setting the open current (200 pA) as 0 pA. In this case A_h is calculated via $I_{event} / I_{open} = A_h$. Either way the value will be normalized to the open current to allow comparisons between tags even at varying voltages.

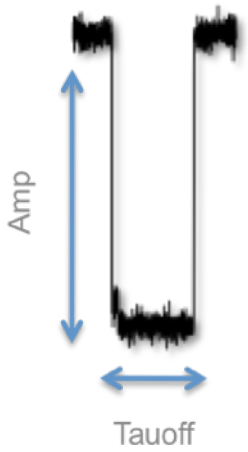


Figure 2.11. Example of a current modulation arising from the translocation of a peptide tagged oligonucleotide.

When translocating ssDNA through a nanopore the current modulations can become relatively complicated. The strand may self – associate causing it to linger in the inner vestibule of the protein. This leads to a current flow of approx. 50% of the open pore current as the strand restricts the flow of ions but cannot yet fully enter the inner constriction. The oligo may then fully translocate, blocking the pore completely and leading to a stepped event. Alternatively the strand may escape the pore leading to an incomplete translocation. These complications are exacerbated by the attachment of peptide tags. For instance tags such as poly-glutamic acid carry a negative charge (at pH 8.0) on each residue. The tag, although flexible, is held perpendicular to the DNA backbone. This could result in the peptide threading into the pore first leading to a blockage and a long noisy event. Due to these complexities the traces are manually analyzed using the ClampFit program to prevent inconsistent

results. The current level and duration of each event is measured individually using the program, which saves the data in the form of a spreadsheet. By analyzing separate traces covering at least 1000 individual events we can calculate the average translocation and pore blockage values for each tagged oligo. It is time consuming (considering the traces of a number of DNA strands can be recorded in a single day) however it is the best way to guarantee that the data accurately reflect the characteristics of the tags.

To ensure that the data is statistically valid, and that no miscellaneous pore blockages affect the results, averages are not taken. The data is plotted and fit to an appropriate trend. The plotted amplitude data always gives a bell curve due to the fact that the majority of events are of a similar current. Any erroneous events with uncharacteristically high (complete blockages) or low (incomplete translocations) amplitude will make up only a minority. A Guassian curve is fitted to the plot (see Figure 2.12) and the value at the apex of the curve taken as the statistical average blockage for that set of results. The dwell time data (τ_{off}) is plotted against 'number of events', which gives an exponential decay due to the majority of the current modulations possessing dwell times in the ms range. The few anomalous complete blockages that are registered account for the very few longer events. So taking an average of these data points would allow these uncharacteristically long events to adversely effect the data set and so the τ_{off} value is taken from an exponential fit of the plot. The $\log_{10}$ of the y – axis is taken to ensure the closest fit and this value taken as the τ_{off} value. Error is calculated through the standard deviation (SD) of the analyzed data via the spreadsheet. The data for each of the oligonucleotide strands under investigation is comprised of at least three separate recordings to ensure its validity. Finally the τ_{on} data, i.e. the duration between events, can be used to calculate the frequency of occurrence. τ_{on} is recorded in ms therefore the inverse of this value ($1 / \tau_{on}$) gives us the frequency of events per ms (ms^{-1}).

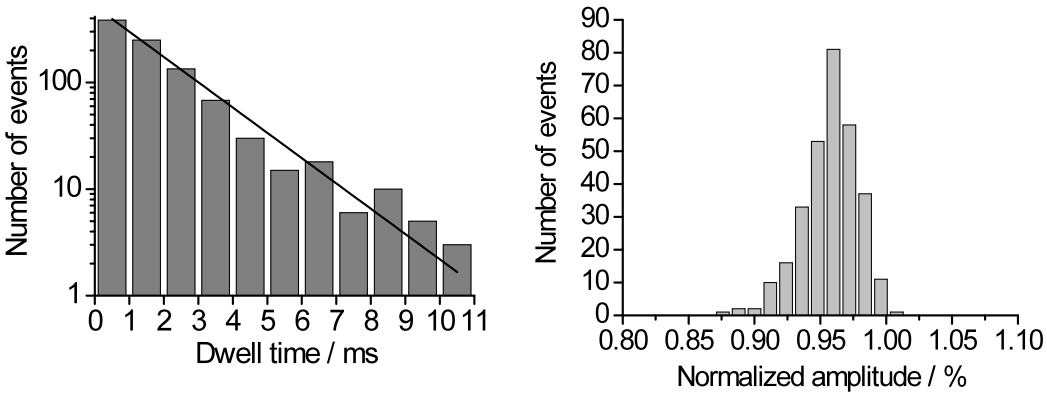


Figure 2.12. τ_{off} (left) and amplitude (right) plots for DNA strand translocation.

Using the collated data we can build up a description of the characteristics for each tag. By comparing the dwell time and pore blockage of each of the POCs we can select those that may be easily differentiated from; (1) the unmodified strands and (2) from one another.

2.5.2 Translocation of Unmodified 27mer Oligonucleotides

The pore recording is conducted in a buffered solution of 2M KCl, at voltages of 100 – 180 mV. In the absence of DNA, the wild-type αHL pore yielded a conductance of 1900 ± 120 pS (n = 4; n, number of independent recordings) when a positive potential was applied at the *trans* side. Once an oligonucleotide translocates through the pore the current flow is momentarily blocked. By attaching tags to internal bases on these strands both the dwell time and the amplitude should increase characteristically for each tag. Attaching a tag on the terminus of the strand should have no overall effect on the translocation as the tag at this position is not increasing the bulk of the strand (Figure 2.13).

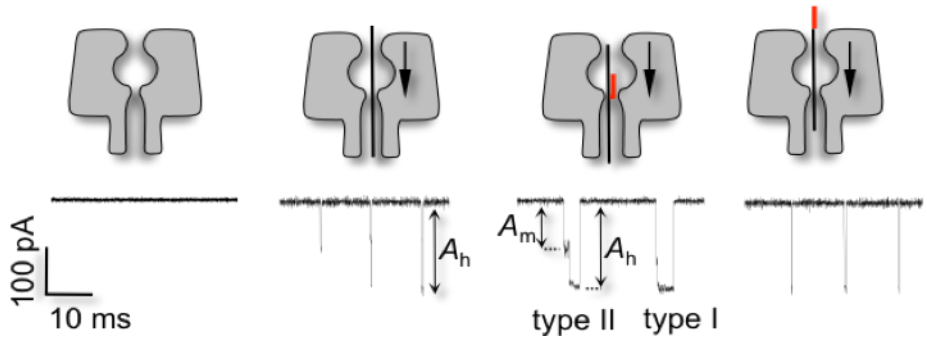


Figure 2.13. Schematic illustration of oligonucleotide translocation through a nanopore. The first trace shows the open channel current. The second illustrates the current blockages observed when an un-tagged oligonucleotide translocates through. The third shows the two types of events common in translocating tagged oligonucleotide strands; Type I - complete translocation & Type II - a stepped event resulting from first incomplete and then complete translocation of the strand. The fourth illustration shows the current events typical for an oligonucleotide tagged at its terminus.

The addition of an unmodified oligonucleotide O1 (27mer; see Table 2.2 for oligonucleotide sequences) to the *cis* side of the pore led to short high-amplitude blockades. Over 1000 individual events were analyzed; the high-amplitude events were characterized by an average amplitude of 91.7 ± 1.1 % relative to the open pore current, and a duration, τ_{off}, of 0.18 ± 0.06 ms (n = 3). The short events represent the fast translocation of individual strands from the *cis* to the *trans* side of the pore (Figure 2.13). The recordings also displayed blockades with 50% amplitude, which were not pursued further as they most likely represent the reversible threading of a strand into and the escape from the *cis* opening rather than the complete translocation to the *trans* side.

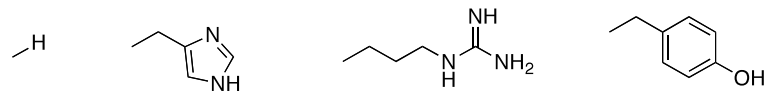
Figure 2.14. Event modulations characteristic of unmodified 27mer oligos (O1).

A τ_{off} value of 0.18 ms for a 27mer oligonucleotide confirms the previously discussed issues of translocation speed. No information regarding sequence or even

length can be derived from this value. Therefore, to enhance the resolution of this technique chemical tags must be employed to slow the speed of translocation and allow identification of the DNA strands. To test the viability of this hypothesis peptides shall initially be used as tags.

2.5.3 Utilizing Peptides as Tags

When considering the sequence of the peptide tags to be employed, the characteristics needed to effect a change in the current modulations when translocation the oligonucleotide through the pore, must first be considered. Charge and sterics will be the most prominent features that will alter the translocation events. Therefore bulky, aromatic substituents such as histidine, tyrosine, phenylalanine, tryptophan etc would be ideal. Charged groups that are able to interact with the residues situated around the inner constriction and neutralize the negative charges on the DNA backbone will also slow the translocation, such as arginine, lysine, glutamic acid and aspartic acid etc. All peptides will be synthesized with a cysteine residue at the N – terminus (for chemoselective attack at the activated disulphide group on the DNA) and a glycine at the C – terminus (due to use of a pre-loaded Gly Wang resin during the solid phase synthesis).



G **H** **R** **Y**

Figure 2.15. Examples of aromatic and charged amino acid R groups; Glycine (G), Histindine (H), Arginine (R) and Tyrosine (Y).

A range of peptides will initially be synthesized and investigated, starting with those carrying very little steric bulk. By gradually increasing the bulk of each peptide via an increase in the number of residues and adding amino acids that carry bulky R groups it will be possible to monitor the effect of sterics on the dwell time of these tagged oligonucleotide. Of the 20 natural amino acids available a number contain bulky aromatic substituents and positive or negative charge that would be ideal for

this application. The option of synthesizing non – biogenic amino acids and incorporating them into a peptide via solid phase synthesis is also available. This will, however, be time consuming and due to the range of natural residues available, not considered necessary.

The nanopore recordings will be conducted just above physiological pH at around 8.0. Therefore residues such as glutamic and aspartic acid with a pKa of around 4 will carry a negative charge at this pH. Lysine and arginine with pKa's of 10.5 and 12 respectively will carry a positive charge. Since the voltage applied will pull negative charge through the pore, a negatively charged tag, as it is adding to the overall net charge of the molecule, may in fact pull the oligonucleotide through faster. However, as the tag, although flexible, is situated perpendicular to the main DNA backbone it may well be pulled into the pore first. This could cause the oligonucleotide to block the pore, requiring the voltage to be switched. Alternatively a positive charged tag will cancel out an equal number of negative charges on the backbone of the strand. This should lead to a decrease in the force a potential difference is able to exert onto the strand thus slowing it down (similar to decreasing the voltage of the system). Charges introduced via a tag may also associate with the same or opposite charges carried by the residues surrounding the constriction of the pore. The resulting attractive and repulsive forces may have a marked effect on the dwell time of the oligonucleotide.

2.5.4 Glycine Series

The initial aim of the project was to confirm that a single peptide tag is capable of retarding strand translocation. Oligonucleotide strand O1, of 27 bases in length, was modified with a series of Gly containing peptides ($CysGly_x$; x = 2, 4, 6) at an internal base. The resulting peptide-DNA conjugates G_xC_1-O1 were analyzed via nanopore recordings. Due to the relatively small size and neutral charge of the tags these oligos have very similar retention times as the unmodified oligos on AEC making purification a challenge. MALDI MS shows mainly a modified peak, although unmodified strands may well be present as it is unlikely that the reaction is 100% effective.

Translocation of the $Gly_{2\text{-}6}$ POCs (1000 – 1500 events, 3 traces analyzed) gave a τ_{off} value of 0.53 – 0.56 ms (± 0.4 – 0.6) (Table 2.3) giving an increase of a little over 3 fold. The pore blockage (accounting for error) is unchanged from the unmodified oligonucleotide (approx. 91%). This is to be expected since no charge has been added and the sterics are not changed significantly hence the slight increase in τ_{off}. Therefore peptides carrying charged groups and/or bulkier substituents were analyzed.

2.5.5 Histidine series

Once again three POCs were synthesised each carrying a peptide with 2, 4 or 6 Histidine residues. Two event types were observed during the translocation of these oligos; Type I is the usual event seen previously, the current is reduced by > 90% and returns to the open 200 pA level once the strand is through the pore. The second Type II events are stepped; the current is reduced to approx. 50% of the open level before increasing to 90% then back to the open current once the strand is through. This modulation is most likely caused by the oligonucleotide being held momentarily in the inner vestibule of the protein. This causes the flow of ions to be restricted before the rest of the strand threads through to complete the translocation.

Figure 2.16. Current modulation characteristic of the His_6 tagged oligonucleotide; variations of Type I and II seen.

The first oligonucleotide carrying a H_2C-O1 tag translocated with an average dwell time of 0.82 ± 0.16 ms with a pore blockage of 0.93 ± 0.7 %. This is close to a 5-fold increase in dwell time and an increase in 2% pore blockage over the unmodified strand. Translocation of the H_4C-O1 tagged strand led to an increase in dwell time to 1.57 ± 0.29 ms with a pore blockage of 96.0 ± 0.6 %. Finally the H_6C_1-

O1 modified strand was analyzed, this yielded typical events with a τ_{off} duration of 1.83 ± 0.26 ms and a high amplitude blockage of 96.8 ± 0.5 %.

The type II events previously discussed were more prevalent during the H_6C_1-O1 translocation. In these cases the event started with a mid-amplitude level, A_m, of 56.6 ± 2.6 % with a duration, $\tau_{off\text{-}m}$, of 1.34 ± 0.36 ms. This medium level was followed by a high-amplitude blockade, A_h, of 97.4 ± 0.9 % and a duration, $\tau_{off\text{-}h}$, of 1.96 ± 0.48 ms. As discussed the mid-amplitude blockade of type II events possibly stems from mis-folded strands which reside in the internal cavity reducing the current flow by 50% before eventually threading into the inner constriction. Due to the uncertainty in the assignment of the mid-level blockade, the focus was shifted onto the more clearly defined type I events, which only exhibited high-amplitude blockades, for the remainder of this investigation. A comparison of type I events from H_6C_1-O1 with unmodified DNA shows that the histidine tag slowed down translocation by a factor of 10 and increased the current amplitude by 5 %.

Several lines of evidence support the notion that the current blockades of the histidine-modified strand are caused by the steric hindrance encountered when a wide peptide-DNA segment passes the narrow inner constriction. Firstly, histidine tags with six, four and two residues led to correspondingly shorter high-amplitude blockades in type I events (Table 2.3, H_xC-O1, x = 6, 4, 2). This implies that the peptide is elongated and aligned parallel to the DNA strand while being translocated. Tags composed of the less bulky glycine residues did not exhibit the same length-dependence indicating that this small amino acid does not reach the critical size threshold required for continually slowing down the DNA. Secondly, the 27mer strand O1 carrying a H_6C_1 tag at a terminal rather than an internal position did not greatly retard DNA passage as shown by a short event time of 0.23 ± 0.10 ms. The absence of a major retardation is attributed to the fact that the peptide can sequentially pass the pore after the DNA strand without the formation of a bulky peptide-DNA segment. The peptides tags are certainly the molecular reason for the increase in translocation time and may exert their effect by either hindered diffusion or an increase in friction mediated by steric, electrostatic, polar, and/or hydrophobic interactions.

Translocating molecules must overcome inherent energetic and entropic barriers to successfully thread through the pore. The diameter of the inner constriction

of the αHL nanopore is approx 1.4 nm. The diameter of a single DNA strand is slightly less than 1 nm; therefore unmodified strands will not encounter much steric clash as they thread through the pore. By attaching bulky groups such as histidine to increase the diameter of the strands to approaching that of 1.4 nm we are increasing the likelihood of the strand encountering the sides of the protein. As the potential difference (Voltage) across the lipid bilayer forces the tagged oligonucleotide through the pore it will encounter friction as it pushes past the amino acid residues around the constriction. The larger the diameter of the strand the more energy will be needed to ensure it can pass though. If the tag is too big no amount of energy (Voltage) will allow the oligonucleotide to pass, as it would require the oligo to break through the covalent bonds of the actual protein to fit through and overcome the steric clash. The translocating oligonucleotide must also overcome an entropic barrier. To enter the inner constriction of the protein pore the strand must transition form a disordered to a semi-ordered state. The Voltage of the system is sufficient to overcome this barrier and force the strands through the pore. Entropic considerations are independent of chemical tags and so investigation of this variable is outside the remit of this thesis.

As well as being aromatic and therefore possessing more steric bulk than the glycine series, the histidine moieties also have a pKa of approx. 6. Therefore by altering the pH of the buffer used in the nanopore recordings it is possible to protonate the backbone of the poly-His peptides. This positive charge may influence the translocation of the oligo.

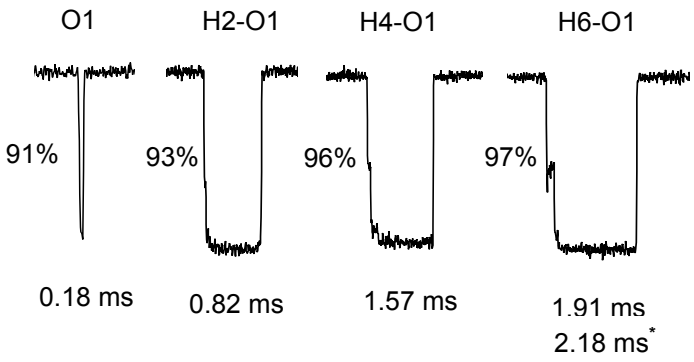


Figure 2.17. The O1 trace shows an actual current event observed from the translocation of unmodified oligonucleotide strands. The others illustrate strands tagged with 2, 4 & 6 histidine residues. Each τ_{off} reading was obtained at pH 8.0 except for the indicated (*) result which was recorded at pH 6.4.

To investigate the effect positive charge had on the DNA translocation, nanopore recordings of the H_6C_1-O1 POC were conducted at a pH of 6.4 (approaching the pKa for histidine). At this pH a greater proportion of the tag will be positively charged and an increase in the translocation dwell time from 1.91 ms to 2.18 ms is seen. Comparing this result with the tranlocation of the same POC at pH 8.0 it is clear that the extra positive charge slows the passage of the oligonucleotide through the pore (Figure 2.17). This can be explained by either the interaction of this charge with other charged amino acids that encircle the inner constriction of the pore. Alternatively, His tagged oligos at this pH carrying positive charge may effectively cancel out the negative charges on the phosphodiester backbone of the DNA. Fewer negative charges on the molecule would equate to less force effected on the strand by the Voltage, thus slowing down the translocation. This effect will be discussed in greater detail in the following section.

2.5.6 Arginine series

The next series of peptides to be investigated was an arginine set to demonstrate that strand retardation is a general feature of bulky/charged amino acids and not only restricted to histidines. An additional aim was to identify tags that give rise to current signatures distinguishable from the potentially useful histidine blockades.

The first peptide, R_7C_1-O1, was investigated; in the nanopore analysis, type I events of R_7C_1-O1 exhibited a duration of 25 ± 5 ms and an amplitude of 98.9 ± 0.6 %. (Table 2.3). This blockade is much higher and longer than that of H_6C_1-O1. The more pronounced results of R_7C_1-O1 are possibly due to the charge carried by the peptide. Since the residues will carry a positive charge at pH 8.0 the tag may be more likely to fold back onto the DNA backbone resulting in a bulkier DNA – peptide segment. This may also account for the lack of noise associated with these events compared to H_6C-O1 traces (Figure 2.18). The histidine R groups are planer in geometry whereas the guanidinium groups of the arginine will possesses a bulkier sphere of influence. The residual positive charge of these groups will also cancel out the same number of negative charges on the backbone, thus lessening the effect the potential difference has on the species and leading to a retardation of the translocation (as seen in the translocation of H_6C-O1 at pH 6.4).

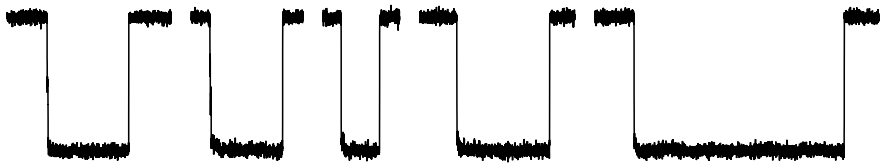


Figure 2.18. Current modulations for the R_7C_1-O1 POCs

Due to the pronounced results obtained from the translocation of the R_7 oligo we moved onto the investigation of smaller poly-Arg peptides using oligos conjugated to R_3 and R_5 peptides. During Nanopore recording of the R_5 POC an average dwell time of 7 ms and a pore blockage of 98% was observed. The smaller R_3 tag gave rise to a 1.9 ms τ_{off} value with an A_h of 98%. These results again support our hypothesis that the length, sterics and charge are all involved in the translocation process.

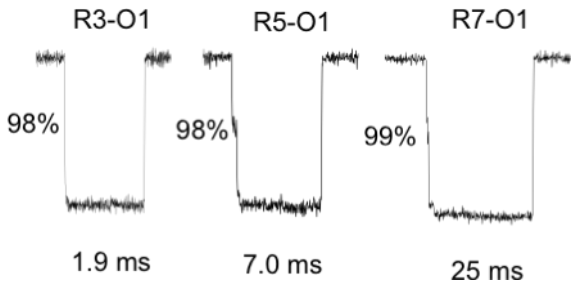


Figure 2.19. Summary of the arginine series current modulations.

The pronounced events associated with the Arg_{3-7} series of POCs provide a good system for the modeling of tagged DNA translocation. By plotting the data it was observed that the dwell time has an exponential dependence on the number of residues that the tag carries (Figure 2.20). Since the current modulations are dominated by the tag (i.e. the pore blockage and dwell time are far greater than those of the untagged strands) – these tags are observed to greatly exaggerate the processes of strand translocation. Compared to the other tags investigated the Arg series also appears to exaggerate the effects of a peptide tag. It is therefore possible to build up a realistic model system of tagged strand translocation based on these physical results.

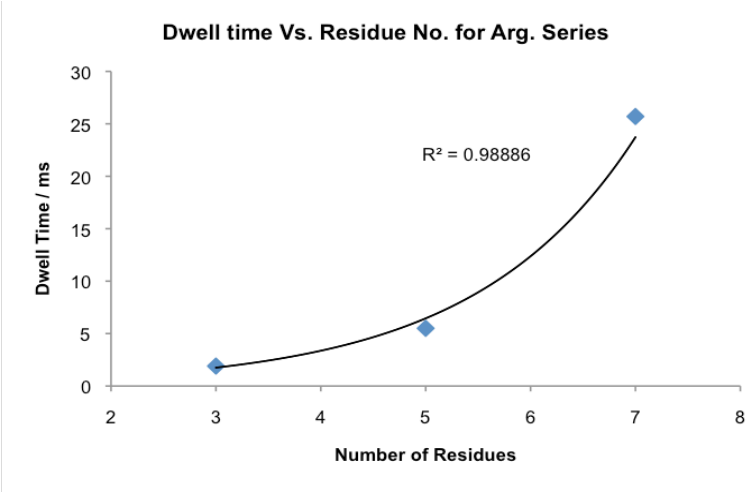


Figure 2.20. Plot of dwell time vs residue number for the Arg series; data fitted with an exponential trend.

Previous research has addressed the translocation of single stranded DNA through a nanopore.[108] However, since the results described in this thesis are, to the best of our knowledge, the first instance of tagged DNA being used in this context, the physical data set obtained from the arginine POC experiments could be used to build a model describing the events observed.

The mathematical modeling of this translocation data is outside the remit of this investigation.[109] It may be beneficial however, to postulate in more detail as to the effect the variables of sterics and charge have on the mechanism of DNA translocation. As described previously, the peptide tags increase the steric bulk of the oligonucleotide; this causes the molecule to require more potential energy to force it through the pore during translocation. The larger the strand the more steric interaction it will have with the residues around the pore and therefore the more energy it will require to overcome these steric interactions. However, in the case of the arginine tags, positive charge is also introduced to the strand.

The translocation of DNA is driven by the creation of a potential difference (Voltage) across the lipid bilayer. This potential difference causes an electrical potential gradient to be set up across the bilayer. This Voltage exerts force on all charged species causing ions to flow through the pore in the direction of the gradient thus creating a current, which is measured using the nanopore apparatus. The more negative charge a molecule carries, the more force the potential difference of the system can apply to this molecule. Therefore a species of net neutral charge will not

feel the potential gradient and therefore will not translocate through the pore. By increasing either the net negative charge of the compound or the Voltage of the circuit we can increase the force exerted on each charged molecule and therefore effectively increase the speed of translocation. Conversely, by decreasing the Voltage or decreasing the amount of negative charge on each strand we can slow the speed of translocation. Therefore it may be fair to presume that neutralizing a portion of the negatively charged DNA backbone with the positive charges on the peptide tag would decrease the force acting on this strand. This would slow the progress of the oligonucleotide through the pore. However, the dramatic retardation of the Arg_7 tagged strands cannot be explained simply by charge reduction alone. Removing negative charge thus decreasing the net charge of the molecule does not have such a pronounced effect. For example a 20mer oligonucleotide has seven fewer negative charges than that of a 27mer. However the difference in τ_{off} for these two strands is very slight. Therefore to build up an accurate model of the translocation of chemically tagged DNA strands we must marry the idea of steric hindrance and charge reduction to give us a realistic representation of the data. Since both these factors must contribute to the dramatic current modulation observed using the Arg_7 tags.

2.5.7 Tyrosine series

The fourth peptide series to be investigated was that of a poly-tyrosine peptide; initially the synthesis of a peptide carrying multiple phenylalanines was attempted. This peptide was, however, too insoluble to utilise. Tyrosine, which carries a phenol group, is more soluble in a buffered solution and so the peptide Y_3C_1 was prepared.

Translocation of Y_3C_1-O1 through the nanopore led to type I events with two observed current levels (Figure 2.22, inset, dotted lines). The first level at an A_h of 92.4 ± 0.6 % is similar to the blockade amplitude of unmodified DNA. It is therefore very likely that this level stems from a DNA strand threaded into the inner constriction but kept from passing the β-barrel because the bulky peptide has not yet entered the narrow pore region.

Figure 2.21. Examples of current modulations from Tyr_3 tagged oligonucleotides.

The second A_h level at 97.8 ± 0.5 % is ascribed to a state where the peptide-DNA segment has entered the inner constriction and translocates the β-barrel. The step signature was observed for 60 % of type I Y_3C_1-O1 events. In the remaining events, the transition between the two current levels resembled a slope (Figure 2.22, second inset).

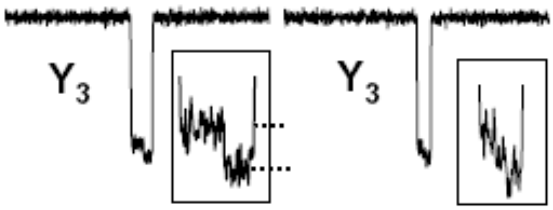


Figure 2.22. Two typical events present in the Tyr3 tagged oligonucleotide traces; stepped (left) and sloped (right).

This sloped effect could reflect a peptide-DNA segment being gradually, rather than abruptly, pulled into the inner constriction. Importantly, the step-like blocking effect of Y_3C_1 was independent of the DNA since same event characteristics were also seen for Y_3C_1-O2 with a different sequence (see Table 2.2 for sequence information and Table 2.3 for translocation data; Y_3C_1-O1 /step vs Y_3C_1-O2 /step).

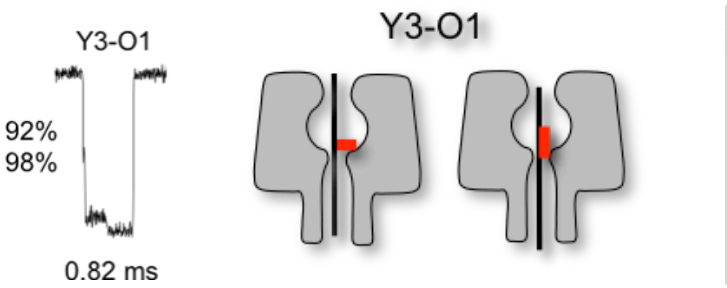


Figure 2.23. Details of the stepped events seen in the Y3-O1 traces and a schematic of the translocation of this tag.

Modified DNA	A_h [%][b]	$\tau_{off\text{-}h}$ [ms][c]
O1[d]	91.7 ± 1.1	0.18 ± 0.06
H_6C_1-O1	96.8 ± 0.5	1.91 ± 0.23
H_4C_1-O1	96.0 ± 0.6	1.57 ± 0.29
H_2C_1-O1	93.0 ± 0.7	0.82 ± 0.16
G_6C_1-O1	91.4 ± 0.6	0.56 ± 0.15
G_4C_1-O1	92.5 ± 0.5	0.55 ± 0.12
G_2C_1-O1	92.7 ± 0.4	0.53 ± 0.13
H_6C-O1-term.[d]	93.9 ± 0.7	0.23 ± 0.10
H_6C_1-O1 / pH 6.4	96.9 ± 0.7	2.18 ± 0.42
R_7C_1-O1	98.9 ± 0.6	25 ± 5
Y_3C_1-O1 /step	92.4 ± 0.6 / 97.8 ± 0.5	0.46 ± 0.15/ 0.35 ± 0.13
Y_3C_1-O1 /slope	94.8 ± 0.7	1.00 ± 0.22
Y_3C_1-O2 /step	91.8 ± 0.6 / 98.9 ± 0.7	0.43 ± 0.12 / 0.39 ± 0.08
Y_3C_1-O2 /slope	96.7 ± 0.8	0.97 ± 0.13

Table 2.3; Summary of POC translocation and pore blockage; [a] The recordings were conducted at 2 M KCl, 20 mM Tris, pH 8.0, filtered at 10 kHz and sampled at 50 kHz unless stated otherwise. The number of events analyzed for each type of DNA was between 1500 and 2000. n = 3. [b] The relative amplitude was calculated using $A = (I_{OC} - I_{BC}) / I_{OC}$, where I_{OC} and I_{BC} are the conductance levels from the open and blocked channel, respectively. I_{OC} and I_{BC} were derived using all-point histograms. [c] The average duration represents the mono-exponential fit of the dwell-time histogram. [d] Filtered at 30 kHz and sampled at 100 kHz.

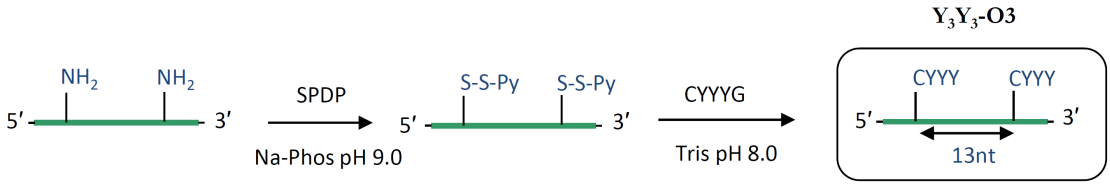
2.6 Sensing of Multiple Bases

DNA strands with two separate chemical tags were tested to illustrate that tags act independently and give rise to corresponding distinct current modulations. Two oligonucleotides sequences were modified with peptide tags, a 37mer and a 54mer (Table 2.1), both prepared using alternative procedures. Sensing of these multiple bases using nanopores allowed identification of the individual tags along the strand via characteristic current modulations.

2.6.1 Synthesis of Oligonucleotides Containing Two Peptide Tags

To prepare strands carrying two peptide tags the same procedure used for the synthesis of a singularly tagged oligonucleotides was applied to a slightly longer

37mer template carrying 2 non-biogenic thymines (see Table 2.2 for sequence). A 37mer with two internal thymines carrying primary amine linkages was modified with the crosslinker SPDP and then conjugated to two CysTyr$_3$Gly peptides as previously described.



Scheme 2.13. Schematic of double modified oligonucleotide preparation; activation of the non-biogenic primary amine bases with the heterobifunctional linker SPDP followed by conjugation with a cysteine containing peptide.

This procedure was successful and after purification via AEC the 37mer carrying 2x CysTyr3 peptides was characterized using MALDI MS.

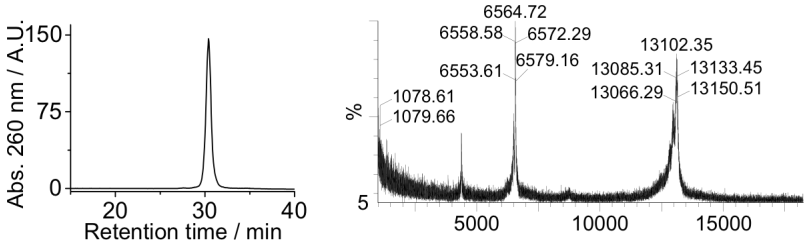


Figure 2.24. AEC and MALDI of double modified 37mer; calc mass 13,105 Da; observed 13,102 Da.

The same method was then applied for the preparation of a 54mer carrying two non-biogenic thymines; this reaction was unsuccessful. It is assumed that due to the length of the strand sections of the oligonucleotide are prone to self-hybridization preventing the reaction from taking place. A doubly modified 54mer would also be a challenge to purify as the peptide tags had little effect on the retention time through an AEC column. Therefore, in this case it was more effective to first modify two separate, smaller oligonucleotides and then ligate them using appropriate enzymes (Scheme 2.9). Two 25mers carrying the internal amine modified thymine were

conjugated as detailed previously with the $CysTyr_3Gly$ peptide (see Table 2.2 for sequence information). The second oligonucleotide carried a phosphate group at its 5' terminus allowing enzymatic ligation. Both strands were hybridized to a 20mer oligonucleotide that was complementary to the last 10 bases on the first strand and the first 10 bases on the second allowing it to bridge the two by forming a duplex bringing the two termini into close proximity. Using a T4 Ligase enzyme the two oligonucleotides were then ligated to form one continuous strand carrying two peptide tags.

The resulting oligonucleotide was purified via denaturing gel electrophoresis (made up in 8M urea), the denaturing conditions of the gel resulted in the comp. 20mer strand separating and migrating at a faster rate than the 54mer (Figure 2.26, lower band). The gel was then visualized under a UV lamp and the section containing the 54 mer strand cut from the gel. The cut out portion of the gel was then crushed and incubated in 5M ammonium acetate solution to elute the oligonucleotide. The slurry was filtered through a QIAgen column and desalted to yield the doubly modified 54mer single strand (see gel 2 of Figure 2.26). A second 54mer carrying two tyrosine tags was also synthesized with alternative positioning of the peptide modifications (see Table 2.2). Both oligonucleotides (including the 25mer intermediates, Figure 2.26) were characterized using AEC and MALDI to confirm successful modification.

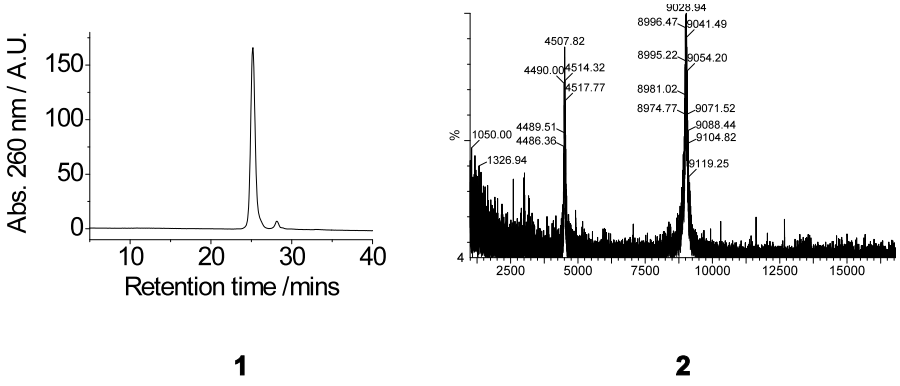


Figure 2.25. (1) AEC and (2) MALDI analysis of the $GlyTyr_3Cys$-O4A intermediate 25mer. Purity 96.2%, calc mass – 9029.6 m/z; observed mass – 9028.9 m/z

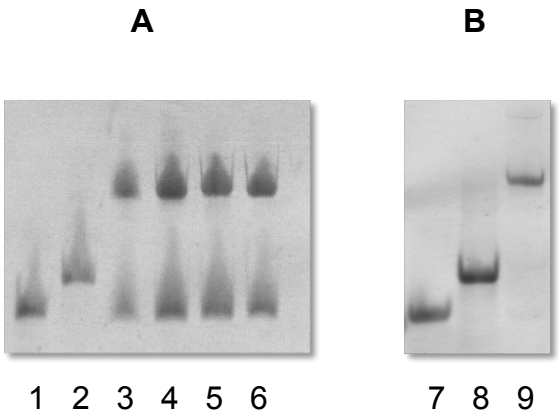


Figure 2.26. (A) 8M Urea PAGE of; (1) complementary 20mer; (2) Y3 modified 25mer; (3-6) 2x Y3 modified 54mer; (B) Native PAGE of; (7) complementary 20mer; (8) Y3 modified 25mer; (9) 2x Y3 modified 54mer.

2.6.2. Nanopore Analysis of Double Modified Oligos

The first strand to be analyzed was a 37-mer (O3) carrying Y_3/Y_3 tags in which the peptides are tethered to two modified bases separated by 13 nucleotides. Similar to a single modified Y_3C_1-O1 strand, double modified DNA gave rise to unresolved slope events as well as fully resolved step-like events (Figure 2.27).

Figure 2.27. Examples of the current modulations resolving two individual tags from a 37mer oligo modified with two Tyr_3 tags,

In the latter events, the blockade amplitude fluctuates twice between two levels sequentially from event segments 1 to 4 (Figure 2.28, event segments numbered red). The average current levels for segments 1 and 3, and 2 and 4 were 92 and 99%, respectively (Table 2.4).

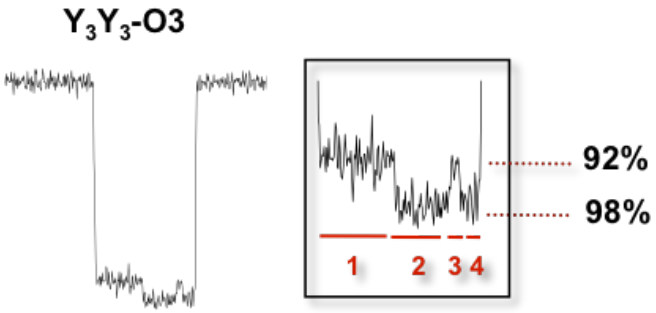


Figure 2.28. Example of a current modulation identifying 2 separate peptide tags.

This step-like signature is in line with expectations for two Y_3C_1 peptides; one peptide is known to cause a blockade step from 92 % to 98 % (Table 2.3, Y_3C_1-1 /step). The signature of Y_3/Y_3-O3 in Figure 2.28 strongly suggests that the current alterations reflect the sequential pulling of a DNA strand through the pore as illustrated schematically in Figure 2.29.

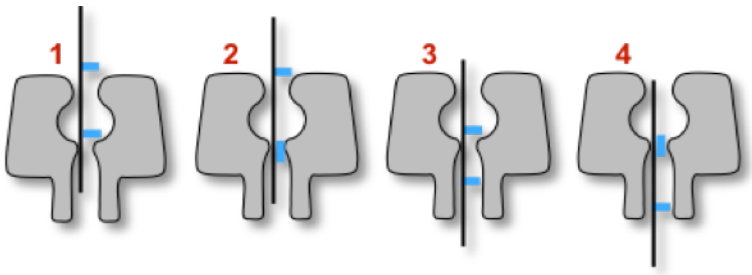


Figure 2.29. Schematic representation of double modified oligos translocating through a nanopore; (1) first tag of translocating strand meets the inner constriction leading to a 92% blockage of the current; (2) the tag is pulled through to give a 98% blockage; (3) the second tag meets the inner constriction shown by the brief return to a 92% blockage; (4) it is then fully pulled though to give a second 98% blockage.

These events consist of 4 sections; the first has a pore blockage of 92%, it is assumed that (similar to the singular Tyr_3 events) this initial reduced current level is due to the first peptide tag reaching the inner constriction of the pore. Once the tag folds up onto the DNA backbone the strand is allowed to pass through the pore almost blocking it entirely; hence a reduction of the current level to 98% of its open state. Once this tag passes through the constriction the current level raises back to 92%

reflecting the unmodified bases between the peptide tags. As the second tag hits the inner constriction the current level again drops to 98%.

This interpretation of the stepped events as the sequential pulling of both tags through the pore is supported by the finding that the 54-mer DNA strand Y_3/Y_3-O4 with a separation of 27 nt between tags showed a similar set of current modulations. The two current levels were identical within experimental error to the values observed for Y_3/Y_3-O3 (Table 2.4). The duration of event segment 3 was, however, longer for Y_3/Y_3-O4 than for Y_3/Y_3-O3.

2.6.3 Analysis of the Space Between Peptide Tags

Since the third section of the event reflects the bases in-between the peptide tags the duration of this section should be representative of this gap. In Table 2.4 we compare the information gathered by manually analysing the events associated with each of the three doubly tagged strands.

Segments	1	2	3	4
O3 A_h [%][b]	92.2 ± 1.1	99.7 ± 0.7	92.3 ± 0.9	99.7 ± 0.6
O3 $\tau_{off\text{-}h}$ [ms][c]	3.43 ±0.67	0.80 ±0.22	0.26 ±0.08	0.61 ±0.15
O4 A_h [%][b	92.8 ± 1.1	98.2 ± 0.8	91.1 ± 1.5	98 ± 0.7
O4 $\tau_{off\text{-}h}$ [ms][c]	1.34 ±0.75	0.64 ±0.26	0.67 ±0.21	0.74 ±0.37

Table 2.4. τ_{off} and amp data summary for 37 and 54 mer double modified oligos; [a] The recordings were conducted at 2 M KCl, 20 mM Tris, pH 8.0, filtered at 10 kHz and sampled at 50 kHz. [b] and [c] are defined as in Table 2.3.

We observe that the spacing for the O3 oligo, that of 13nt, gives us a translocation section 3 lasting 0.26 ± 0.08 ms. Extending this gap to 27nt for the O4 oligonucleotide increases the dwell time of this section to 0.67 ± 0.21 ms. A third strand (O5) was prepared; a 37mer with the same sequence as O4 carrying the Tyr tags positioned 7nt from one another. The double modified current modulation observed for oligonucleotides O3 and O4 was not seen using this strand. This result was to be expected since the $CysTyr_3Gly$ peptide tag is long enough to bridge the 7nt gap between the tags (approx. 2 nm). Therefore, in this case, the two modified bases

cannot be resolved. The percentage of stepped events as well as the quality of the current steps was far lower than for Y_3/Y_3-O3 and Y_3/Y_3-O4 (42% seen for O3, 57% for O4 and only 5% for O5). This reduced resolution agrees with molecular models showing that the tag with an extended length of 2.8 nm bridges the gap between two tagged bases separated by 7nt or 2.2 nm.

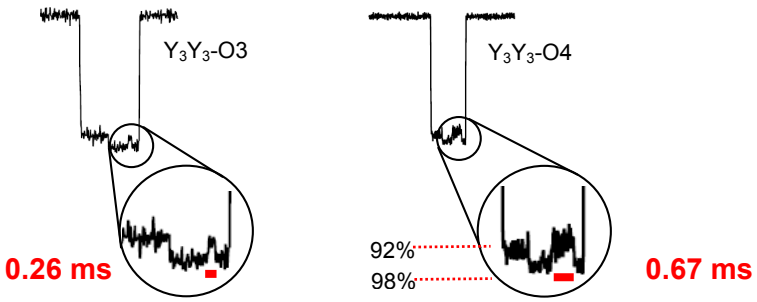


Figure 2.30. The τ_{off} value representing the spacing in-between the tags increases from 0.26 ms to 0.67 ms with a gap increase from 13 to 27 nt.

Although the increase in τ_{off} for section 3 reflects the longer gap between the tags this is most likely a qualitative value. Although a relative increase in τ_{off} indicates an increased gap between the tags this value this most likely cannot be used to identify the number of bases with any accuracy.

2.7 Using Chemically Modified Nucleotides to Tag Biologically Relevant SNPs

After establishing that chemical tags can slow the translocation of DNA through a biological nanopore, thereby fulfilling the first aim of this thesis, the next step was to apply the tagging technology in a biologically relevant context. This second aim was addressed by tagging bases that represent single nucleotide polymorphisms (SNPs) within important sections of DNA.

2.7.1 Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide Polymorphisms (SNPs) occur throughout the genome and reflect mistakes that are made when DNA is erroneously copied during cell division. As the majority of our DNA exists as regions that do not appear to contain genes many SNPs usually do not have any biological effect. However, if these errors occur in a sequence of DNA that regulates gene expression or encodes for a protein, then SNPs may have an impact on the cell and organism. For example, SNPs falling within the codons for a protein may cause amino-acid substitutions that can alter the folding, structure, and function of an enzyme, and thereby change the organism's ability to react to a drug. The sensing of SNPs not only helps predict the response of humans to drugs but is also of interest in microbiology where SNPs may cause resistance against antibiotics or antivirals. Knowing the genetic make-up of infectious strains could inform about specific resistances and help tailor the selection of drugs to improve the efficacy of treatment.

2.7.2 Sensing of SNPs

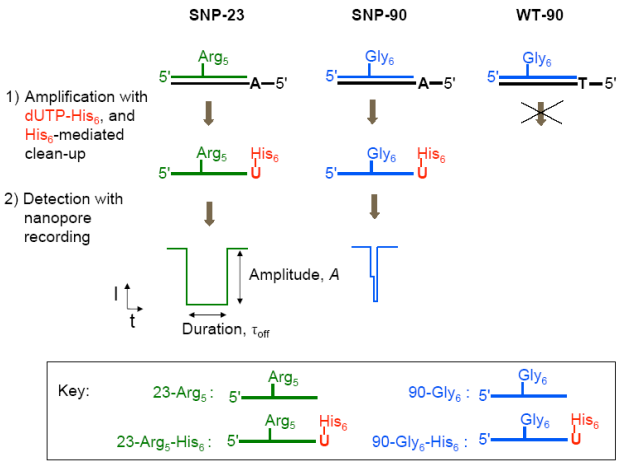


Figure 2.31. Schematic of SNP detection; tagging of SNP and IMAC purification using His-dNTP and nanopore identification of primer tag using nanopores.

For the nanopore sensing of biologically relevant sequences, two SNPs from the protease gene of the human immunodeficiency virus (HIV) were selected. Codons 23 and 90 confer resistance against the HIV drug Nelfinavir. The principle of the SNP

sensing strategy is based on the use of peptide-tagged oligonucleotides and nucleotides to encode sequence information for nanopore read-out. The two-step strategy is summarized in Figure 2.31. First, two oligonucleotide probe strands that are complementary to a sequence stretch upstream the SNP site are hybridized to the SNP-containing template. Each primer has attached to it a peptide tag, which is either a Gly_6 or Arg_5 peptide. Importantly, the two tags are known from the previously reported results to give rise to two different current signatures. The two tags have therefore the potential to encode sequence information based on different current signatures. As illustrated in Figure 2.31, the primers are hybridized to templates that either do or do not carry the target SNPs. Using a polymerase enzyme, only the primers for the SNP-carrying templates are extended with the nucleotide analogue dUTP-His_6 while the non-SNP template does not result in primer extension. The optionally incorporated nucleotide analogue carries a His_6 tag attached to position 5 of the thymine base. The analogue dUTP-His_6 (Figure 2.32) was previously prepared by Dr. Vinciane Borsenberger. Importantly, the His_6 tag helps to purify the extended primers using immobilized metal affinity chromatography (IMAC). By contrast, in the absence of an SNP, the probe strand will not be retained on the IMAC column. The mere presence of the strand after purification of the solution is therefore an indication of the SNP (Figure 2.33). The successful primer extension and the His_6-tag purification of the probe strands were confirmed by gel electrophoresis (Figure 2.34; experiments conducted by Dr. Borsenberger).

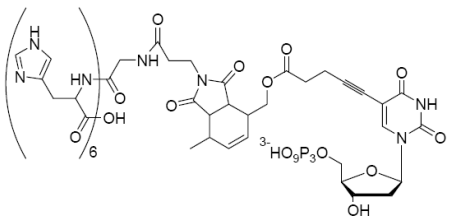


Figure 2.32. His_6 modified dUTP

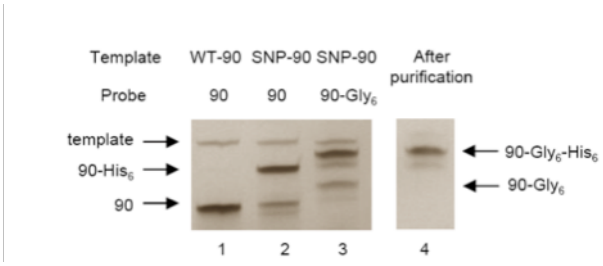


Figure 2.33. SNP-dependent enzymatic extension of the peptide-modified DNA probe strand with dUTP-His_6; (1) the wild type 90 oligonucleotide; (2) the SNP-90 strand showing extension *via* an untagged oligonucleotide due to the presence of the SNP; (3) extention of the SNP-90 strand using the Gly_6 tagged nucleotide; (4) the strand in lane 3 post IMAC purification.

Once the extended DNA template was isolated it was subjected to nanopore recording (recording conducted by S. Howorka, analysis by N. Mitchell). Since the His_6 tag is at the terminus of the oligo, it has little effect on the translocation. It is the characteristics of the tag attached to the probe strand that determine the current modulations. The characteristics for the Gly_6 and Arg_5 tags are, as concluded in the previous section, very different. Indeed, the current blockades of the Gly_6 and Arg_5 modified probe strands are strikingly different (Figure 2.34) in their average amplitudes and durations (Table 2.5). The Gly_6 peptide resulted in an oligo translocation value of τ_{off} = 0.76 ms and an A_h of 91%. While the Arg_5 tag gave a τ_{off} value of 9.33 ms and an A_h value of 99%. These two data sets can be easily differentiated via the plotted data allowing the two SNPs to readily be identified (see Figure 2.35).

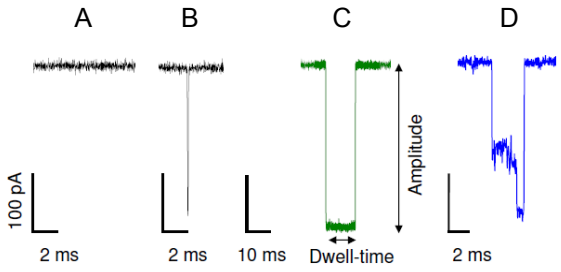


Figure 2.34. Examples of current modulations resulting from Gly_6 and Arg_5 modified oligos; (A) Open-pore current trace; (B) current blockage resulting from translocation of unmodified oligonucleotide; (C) current blockage resulting from translocation of Gly_6 tagged oligonucleotide; (D) current blockage resulting from translocation of Arg_5 oligonucleotide.

Probe strand	τ_{off} [ms][a]	A [%][a,b]
23-R_5-H_6	9.33	99
90-G_6-H_6	0.76	91
	1.21 (13%) / 8.07 (87 %)[c]	59 (80%) / 72 (20%)[c]

[a] Obtained as described in section2.3.18. The values are the average from three independent recordings each for 23-Arg_5-His_6 and 90-Gly_6-His_6, and a recording in which both modified DNA probes were present. [b] Amplitude, *A*, was normalized to the conductance of the open unblocked channel. [c] τ_{off} and *A* values for mid-level events of 90-G_6-H_6. Two values each are provided to reflect the two components in the respective distributions. The relative proportion of the components is given in brackets.

Table 2.5. Results of the translocation of the peptide modified extended DNA templates.

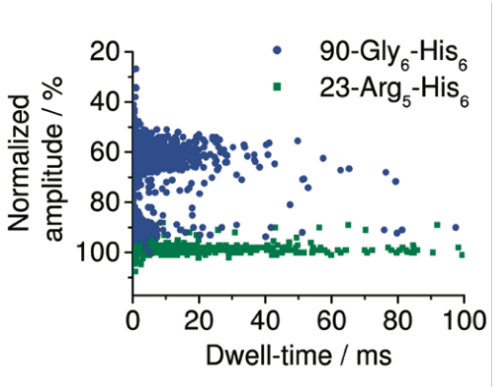


Figure 2.35. A scatter plot with 90-Gly_6-His_6 and 23-Arg_5-His_6 differentiated in blue and green respectively.

2.8 Sensing Repeating Bases Using Alternative Tags

The first preceding section of this chapter introduced peptides as useful chemical tags for the sensing of bases along a translocating DNA strand (aim 1). The nanopore approach was then applied in a biological context by using peptide-modified nucleotides and oligonucleotides to identify SNPs on biologically relevant DNA templates (aim 2). In the final aim, a series of nucleotides analogues were to be synthesized to help progress and develop the tagging approach into a technique that is able to compete with state-of-the-art DNA sensing strategies. As an attractive target analyte, the nucleotide patterns in highly repetitive DNA sequence of forensic

importance were chosen. The short nucleotide repeats (tri, tetra, or penta-repeats) occur in different regions of the genome, and the number of repeats in each region is characteristic for each person. While repeat sequences are routinely analyzed using conventional sequencing, it would be of benefit to quickly size the number of repeats within samples or suspects at the crime scene. The proposed nanopore strategy can potentially deliver a fast turn-around time.

The proposed approach to size repetitive DNA strands is illustrated in the Figure 2.36. DNA strands containing multiple trinucleotide repeats are used as a template to direct the enzymatic synthesis of a DNA copy. The DNA polymerization is conducted using a nucleotide analogue that carries a small chemical tag, The resulting DNA copy carries a defined number of tags that mirror the number of repeats in the template. The DNA copy is then analyzed using nanopore recordings. The threading of the tagged strands through the pore is expected to create a characteristic comb-like blockade pattern in which the number of blockades reflects the number of tags in the translocating strand.

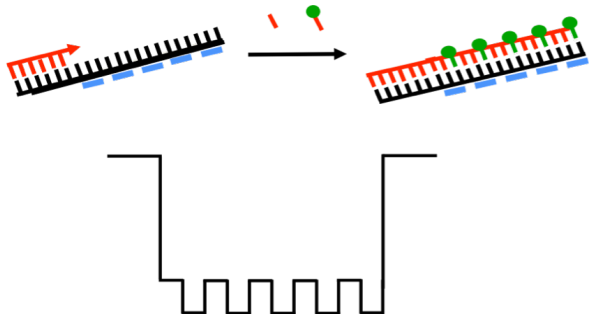


Figure 2.36. Comb-like current modulations; PCR incorporation of tagged and untagged nucleotides into a DNA template and the electrical sensing of multiple chemical tags along the strand using nanopore techniques.

The proposed sensing strategy is based on and inspired by two previous experimental results, namely the revelation that synthetic DNA strands with two $GlyTyr_3Cys$ peptide tags give rise to a two-tooth comb-like current signature, and the experimental demonstration that tagged DNA strands can be generated via enzymatic polymerization using a nucleotide analogue. The proposed strategy for trinucleotide repeats cannot, however, use the same tags and nucleotide analogue given their unsuitable dimensions. In particular, the tyrosine-containing peptide tag is too long to

be used for e.g. a pentanucleotide sequence, as a distance of two nucleotide positions separated by four bases, is shorter than the size of the extended peptide tag. Indeed, problems were encountered in section 2.6.3 when two Tyr_3-tags separated by 7 nt could not be resolved in nanopore recordings. Based on the known dimensions of DNA repeats as well as the nanopore dimensions, a different chemical tag was designed.

2.8.1 Adamantane Tag

Adamantane ($C_{10}H_{16}$) is a potentially ideal tag; it has dimensions of approx 5 Å cubed which will increase the diameter of a single stranded oligo to approx 1.4 nm, the size of the inner constriction of the nanopore. This should be large enough to temporarily block the pore to a sufficient extent. If a short linker is used between the tag and the base there should also be no issues with the masking of downstream bases.

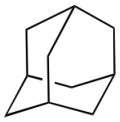


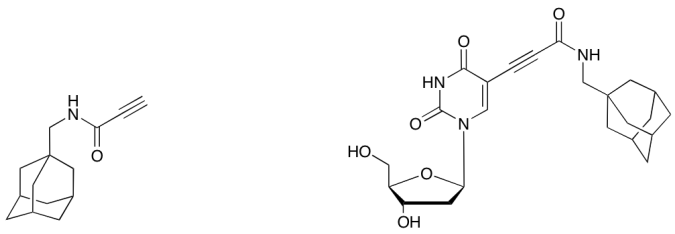
Figure 2.37. Chemical structure of adamantane.

A model system based on a synthetic DNA strand was first investigated to confirm that the adamantane would be appropriate. The coupling of aminomethyl adamantane to maleoyl-β-lanine (see section 2.3.14) enabled conjugation of the adamantane tag to the 27-mer oligonucleotide which had been previously used in the peptide-conjugation experiments. This oligonucleotide was first activated using the cross-linker SPDP as detailed previously and the newly installed disulfide group reduced using DTT. The maleimide modified adamantane compound **5a** could then be conjugated onto the strand. The nanopore analysis of the resulting adamantane tagged 27-mer was conducted and analyzed. Both type I and II events were observed in the recorded trace. The type I events gave a τ_{off} value of 0.64 ms and an average A_h value of 96.0 ± 1.9 %. These blockade characteristics are ideal as they alter the translocation of the modified oligonucleotide sufficiently so as to distinguish from unmodified strands. The linker between adamantane and the DNA is, however, 2 nm and therefore

too long for the sensing approach. Another shortcoming of the strategy to modify commercially available DNA oligonucleotides are the few non-biogenic groups that are available, limiting the possible chemistry options. The strands can also only be obtained above a few hundred μg at high cost. This means that synthesis of individual nucleosides from small molecule starting materials is certainly a more viable option.

2.8.2 Synthesis of the Adamantane-Tagged Nucleoside

An appropriate target nucleotide analogue carrying the adamantane tag via a suitably short linker is illustrated in Figure 2.38; compound **7**.



(5b) **(7)**

Figure 2.38. (1) alkyne modified adamantane; (2) 5-adamantane-2'-deoxyuridine.

This compound can potentially be synthesized from the starting nucleoside 5-iodo-2'-deoxyuridine (see Figure 2.38). This uracil/thymine starting material has several advantages. It is cheap compared to the possible alternative purine base analogues. In addition, it is also the least complicated to modify due to the absence of any *exo* amines (unlike the cytosine base). Finally, the iodo-functionality can be used to introduce a variety of groups at the 5 position of the base at high yields using the well-established Pd(II) cross-coupling chemistry.

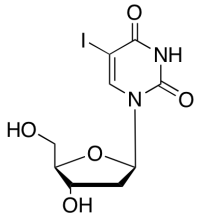


Figure 2.39. 5-Iodo-2'-deoxyuridine

The target nucleoside **(7)** was synthesized by coupling the acetylene-modified adamantane (as described in Sections 2.3.15 – 16) to a 5-iodo-2'-deoxyuridine nucleoside via the Pd(II)-catalyzed Sonogashira reaction. This chemical transformation was successful; the desired product was isolated in 85% yield. This nucleoside is an important intermediate towards the synthesis of the corresponding phosphoramidite and triphosphate derivatives to generate adamantane-tagged DNA strands via solid phase chemical synthesis and enzymatic polymerization, respectively.

2.8.3 Synthesis of Phosphoramidites for the Generation of Tagged Oligonucleotide Strands via Solid-Phase

Compound 7 is an important intermediate for the preparation of the dimethoxytrityl (DMT)-protected phosphoramidite derivative (Figure 2.40) via reaction of the 5' hydroxyl group with dimethoxytrityl-chloride and the subsequent modification of the 3' hydroxyl group with cyanoethoxy-*N,N*-diisopropyl-chlorophosphine (CEP). The protected phosphoramidite is the stepping-stone towards the preparation of synthetic oligonucleotides carrying the adamantane tag.

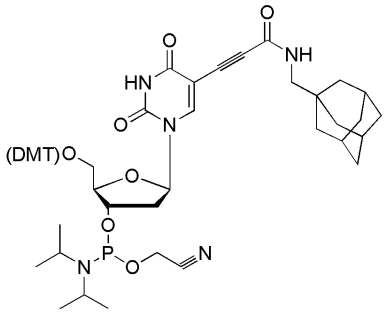
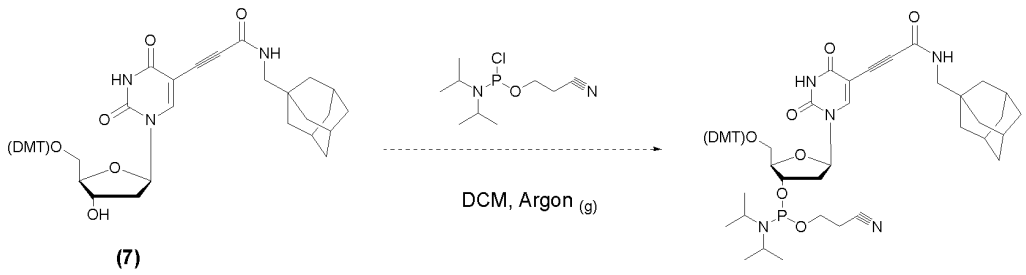


Figure 2.40. Chemical structure of proposed Thymine derived phosphoramidite.

The DMT-protected nucleoside was successfully synthesized and isolated albeit in a low 26% yield. This value may possibly be due to the steric clash of the bulky adamantane and DMT groups, caused by the inflexible linker at the 5′ position. A possible solution to address this (apart from utilizing a longer linker) could be the use of an S_N2 catalyst such as DMAP[110] to achieve higher yields.

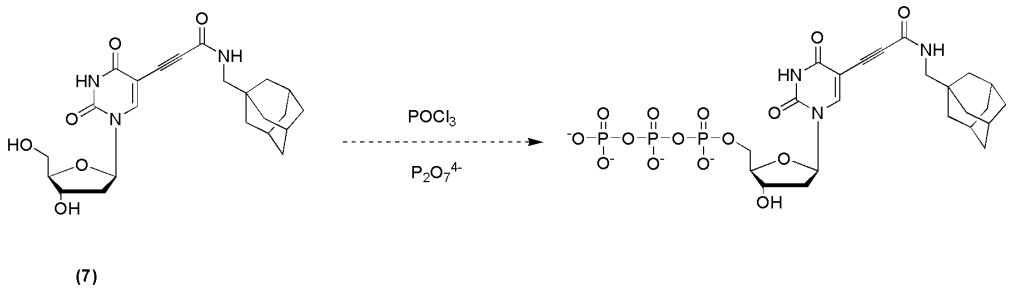


Scheme 2.14. Proposed synthesis of adamantane containing phosphoramidite.

The DMT-protected nucleoside was then subjected to the phosphitylation reaction with CEP. TLC analysis established that the probable desired product was being formed but then degrading almost immediately. Most likely, the degradation is due to the oxidation of the phosphine. This result was surprising, as several precautions had been taken to avoid this undesired degradation. The phosphitylation reaction had been conducted under inert atmosphere and all solutions had been purchased in anhydrous form and thoroughly degassed beforehand. Due to time constraints the reaction could not be repeated to resolve this issue.

2.8.4 Enzymatic Incorporation of Modified Nucleotides into a DNA Template

Compound 7 also represents a starting point for the synthesis of the triphosphate analogue (Scheme 2.15), which is proposed to be used for the generation of adamantane-tagged DNA via template-directed enzymatic polymerization. It is known that chemical tags at position 5 of deoxyuridine triphosphate analogues can be successfully accepted as substrates by DNA polymerase enzymes.[111]



Scheme 2.15. Proposed synthesis of 5-adamantane-5'-triphosphate-2'-deoxyuridine.

An adapted version of the Yoshikawa procedure[112,113] was used to phosphorylate compound 7. The 5' hydroxyl group was first modified with $POCl_3$ in trimethyl phosphate at -20 °C for 20 min to generate an activated phosphorodichloridate species (not shown). This activated compound is then reacted with pyrophosphate to form the desired triphosphate. The reaction was conducted following the literature procedure but resulted in the mono-phosphorylated by-product (as indicated by LC/MS); the triphosphate could not be isolated. Due to time constraints, the reaction could unfortunately not be repeated to improve the results.

2.9 Conclusions

In this study, a new nanopore-based strategy has been developed to enable the detection of separate bases in DNA strands. Using a model system of DNA oligonucleotides modified with peptides, chemical tags have been used to demonstrate that tags attached to the bases are capable of causing characteristic current signatures for strands translocating through nanopores. The proof-of-principle experiments show that blockade duration, amplitude and signature can be tuned by changing the length, charge and size of the tags. The current modulations are independent of the surrounding DNA sequence, and two tags on a strand retain their characteristic signatures opening up the possibility of attaching multiple tags to DNA. This work therefore demonstrated for the first time that (i) pore recordings can detect one or two separate bases in translocating individual DNA strands, and that (ii) chemically modified DNA can be used to infer base-specific information.

While the experiments were performed with synthetic oligonucleotides, the approach could potentially be applied to sense DNA from biological samples. This research illustrated that peptide tags can be incorporated into copied DNA strands using chemically modified nucleotides and sequence-specific primer extension. This approach was suitable to sense the presence or absence of single-nucleotide polymorphisms by incorporating and detecting a modified base only if the target mutation is present. With further improvements on the tags, -e.g. decreasing the size of the linker and the length of the tags - it may also be possible to reduce the nucleotide distance between the tags and thereby detect multiple bases in biologically

relevant DNA strand. For example, the highly repetitive trinucleotide repeat regions in forensically important sections of our DNA could be sized by labelling the same base in all repeats. Adamantane was identified as a viable tag for this purpose as initial nanopore experiments give encouraging results. A uracil nucleoside analogue was synthesized carrying this tag and reactions towards the preparation of both the phosphoramidite and triphosphate were investigated.

The extension of the technology towards sequencing by measuring the ionic current modulation for each base would be very difficult to achieve due to the small distance between neighbouring bases. This does, however, not limit its potential, because the concept of slowing down DNA via chemical tags is novel and can be applied to various related nanopore approaches. These include fluorescence- or ionic current-based sequencing of DNA-derived design polymers where the spacing between individual bases has been increased, or strategies which detect tunnelling current.

The general approach of using chemically modified bases is especially relevant for DNA sensing with solid-state nanopores. These pores exhibit very high mechanical stability, which makes them ideally suited for rugged electrical sensor devices. Despite progress in their fabrication, solid-state nanopores cannot be engineered to the same atomic precision as the protein pore αHL. For example, inorganic pores are usually not narrow enough to discriminate between single and double stranded DNA thereby posing constraints on their ability to detect DNA via a hybridization approach. The use of chemically modified DNA strands can address this limitation by tuning the cross-sectional diameter of ssDNA to existing pore dimensions rather than matching the pore dimensions to the size of the DNA strand.

In summary, our strategy represents a paradigm shift as it uses the chemical modification of the analyte rather than pore engineering to expand and enhance the sensing repertoire of nanopores. The concept can be applied to other bioanalytes.

The results described in this chapter have been published in Angewandte Chemie International Edition (Nick Mitchell & Stefan Howorka; 2008; 47; 5565 – 5568) and Journal of the American Chemical Society (Vinciane Borsenberger, Nick Mitchell, Stefan Howorka; 2009; 131; 7530 - 7531). As second author of the second of these publications my involvement was limited to the analysis and interpretation of the nanopore traces. Synthesis of the His_6-dUTP and the polymerase extension assays

were conducted by Dr. V. Borsenberger, and all nanopore recording was conducted by Dr. Stefan Howorka.

Chapter Three

DNA based Nanoscale Tetrahedrons as Scaffold for the Spatial and Stoichiometric Positioning of Chemical Groups

3.1 Summary

In DNA nanobiotechnology, oligonucleotides are self-assembled via hybridization to generate higher-order structures of defined geometry. This chapter describes how the functional range of DNA nanostructures can be expanded by chemically modifying the constituent nucleic acids. DNA-based tetrahedron-shaped nanostructures were equipped with different chemical groups. As their position within the DNA was pre-determined, the nanostructure acted as a scaffold to arrange the different groups at tunable stoichiometry and at geometrically defined sites. The resulting chemically tagged molecular entities exhibited functional properties beneficial in biosensing and diagnostics. The new strategy for assembling chemical groups at the nanoscale may be expanded to endow other DNA structures with rationally designed functions.

3.2 Introduction

DNA can be a versatile material in nanobiotechnology and chemical biology. Due to its high persistence length and sequence-specific assembly via Watson Crick base pairing, DNA strands can be used to assemble artificial higher-order structures.[114] These architectures can be applied in areas such as biotemplating[79,80], bio-computing[81], sensing[82], and biophysical studies[83]. In chemical biology, hybridization of chemically modified strands into DNA duplexes has been utilized to bring chemical groups into defined contact to e.g. enhance their reaction.[84]

Here a new approach is described which merges DNA-based nanostructures and the chemical modification of DNA. The new approach exploits DNA-based tetrahedrons as molecular scaffold to position functional chemical groups at defined geometry. DNA tetrahedra are nanostructures with edges composed of double stranded DNA[94,95] (Figure 3.1).

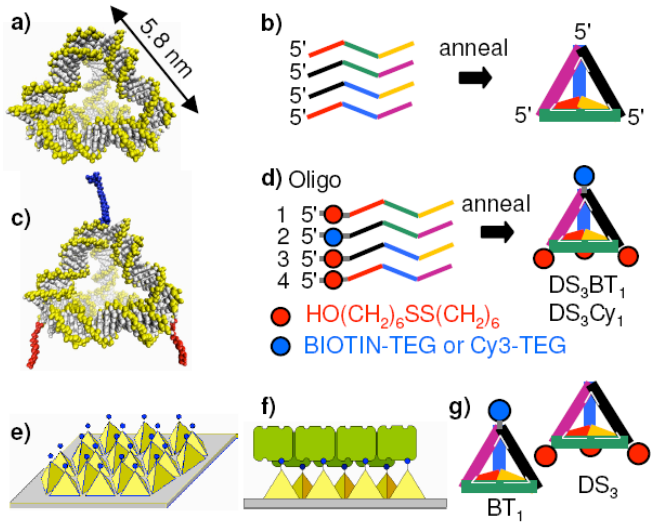


Figure 3.1. Molecular model of a DNA tetrahedron (A); DNA tetrahedra functionalized via biotin (blue) and disulphide (red) groups (C); Annealling procedure (B+D+G); Functionalisation of surfaces (E+F).

Tetrahedra are obtained by annealing four single stranded DNA oligonucleotides (ssDNA), the 55-nt strands used here give rise to tetrahedron edges of 5.8 nm in length (Figure 3.1.A). Once assembled, the free 5' and 3' termini of the ssDNA are positioned at the vertices of the tetrahedron (Figure 3.1.B). This structural

characteristic was exploited by placing biotin (BT) and disulfide (DS) groups at the four vertices (Figure 3.1.C) by using four ssDNAs carrying the chemical modifications at their 5' ends (Figure 3.1.D). By positioning three DS-groups and one BT group at the vertices, the rationally designed structures were expected to exhibit desirable functional properties. In particular, the tetrahedra were anticipated to (i) bind via three thiol legs with high affinity onto gold substrates, and - due to their oriented binding - (ii) present the BT moiety at a defined surface-distance exposed to the ambient (Figure 3.1.E) capable of capturing streptavidin at high density (Figure 3.1.F). The experimental strategy to analyze the DNA structures encompassed gel electrophoresis of solubilized tetrahedrons, and the surface characterization of DNA-tetrahedron films with X-ray photoelectron spectroscopy (XPS), atomic force microscopy (AFM), and fluorescence microscopy, along with spectrometric measurements on the binding of tetrahedrons onto gold particles.

3.3 Materials and Methods

3.3.1 Formation of a DNA Tetrahedral Structure

DNA tetrahedra were generated using four DNA oligonucleotides 55 nt in length. The sequences of the DNA strands were derived from the oligonucleotides used by Goodman *et al.*[94,95] and are summarized in Table 3.1. The four oligonucleotides were termed Oligo-1, Oligo-2, Oligo-3, and Oligo-4 and carried chemical modifications including a disulfide (DS), a biotin (BT), or a Cy3 group at the 5'-end, depending on which type of DNA tetrahedron was generated. Five tetrahedra were obtained using this procedure: DS_3BT_1, DS_3, BT_1, DS_3Cy_1, and DS_2Cy_1. Each contained disulfide groups at up to three vertices, and an optional biotin or Cy3 group at the forth vertex.

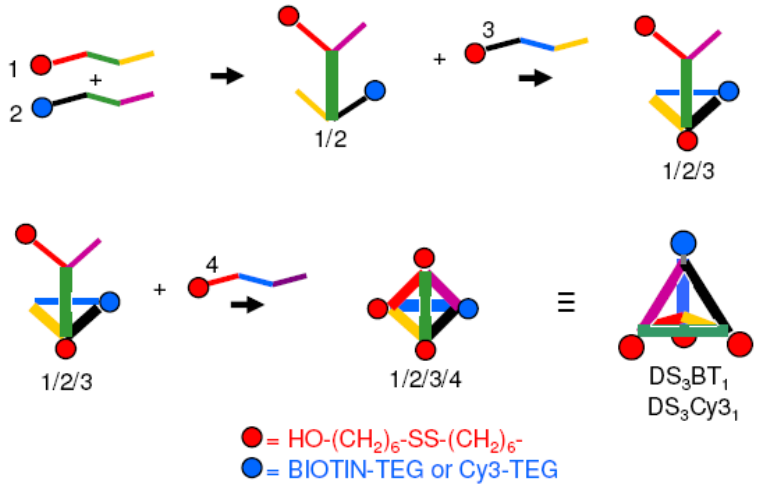


Figure 3.2. Step-wise hybridization of the oligonucleotides comprising the DNA tetrahedral.

Oligonucleotides Oligo-1 to Oligo-4 (see Table 3.1), were dissolved in TE buffer (10mM Tris, 1mM EDTA, pH 8.0) to a conc. of 100µM. 1 nmol of each strand (10 µL) was combined with 20 µL of 0.1M Tris, 0.1M NaCl, 1mM EDTA, pH 8.0 and heated to 95°C in a PCR block for 2 min before immediately being cooled on ice. A 12.5% PAGE gel was prepared, along with the appropriate controls (i.e. single strands and hybridisation products omitting one or two strands, see Figure 3.3) and the gel was run in a 4 °C cabinet. To prepare structures carrying disulfide, biotin and Cy3 modifications the appropriate oligonucleotides carrying 5' modifications were substituted for the unmodified strands.

DNA Oligo	Sequence and Optional Modification
Oligo-1	5'-ACA TTC CTA AGT CTG AAA CAT TAC AGC TTG CTA CAC GAG AAG AGC CGC CAT AGT A-3'
Oligo-2	5'-TAT CAC CAG GCA GTT GAC AGT GTA GCA AGC TGT AAT AGA TGC GAG GGT CCA ATA C-3'
Oligo-3	5'-TCA ACT GCC TGG TGA TAA AAC GAC ACT ACG TGG GAA TCT ACT ATG GCG GCT CTT C-3'
Oligo-4	5'-TTC AGA CTT AGG AAT GTG CTT CCC ACG TAG TGT CGT TTG TAT TGG ACC CTC GCA T-3'
Oligo-1-DS	Sequence of Oligo-1 carrying a disulfide group via a hexamethylene (C6) linker at the 5' end
Oligo-2-BT	Sequence of Oligo-2 carrying a biotin group via a tri(ethylene glycol) (TEG) linker at the 5' end
Oligo-2-Cy3	Sequence of Oligo-2 carrying a Cy3 group via a TEG linker at the 5' end
Oligo-3-DS	Sequence of Oligo-3 carrying a disulfide group via a C6 linker at the 5' end
Oligo-4-DS	Sequence of Oligo-4 carrying a disulfide group via a C6 linker at the 5' end

Table 3.1. Names, Chemical Modifications, and Sequences of DNA Oligonucleotides Oligo-1 to Oligo-4.

Tet.	# of Disulfide Groups	# of Biotin or Cy3 Tags	Oligonucleotides Used for the Preparation of Tetrahedra
DS_3BT_1	3	1	Oligo-1-DS, Oligo-2-BT, Oligo-3-DS, Oligo-4-DS
DS_3	3	0	Oligo-1-DS, Oligo-2, Oligo-3-DS, Oligo-4-DS
BT_1	0	1	Oligo-1, Oligo-2-BT, Oligo-3, Oligo-4
DS_3Cy_1	3	1	Oligo-1-DS, Oligo-2-Cy3, Oligo-3-DS, Oligo-4-DS
DS_2Cy_1	2	1	Oligo-1, Oligo-2-Cy3, Oligo-3-DS, Oligo-4-DS

Table 3.2. Names and Composition of DNA Tetrahedra

3.3.2 Streptavidin Conjugation

Prior to analyzing biotin-tagged DNA tetrahedron DS_3BT_1 by the coupling to the protein streptavidin, a simple biotin-modified ssDNA oligonucleotide was tested to confirm its ability to form a streptavidin-biotin bridge. Oligo-2-BT (200 pmol) of 55 nt length carrying a biotin moiety was added to a 10 μL aliquot of streptavidin (1mg / ml). After incubation at room temp. for 10 mins the sample was run on a 15% TBE polyacrylamide gel in addition to two controls consisting of Oligo-2-BT without streptavidin protein, and the non-biontinylated Oligo-2 plus streptavidin. To confirm conjugation of BT modified structures the DNA tetrahedra was formed via the procedure detailed in section 3.3.1 using three 5'-disulfide modified oligonucleotides Oligo-1-DS, Oligo-3-DS, Oligo-,4-DS and the 5'-biotin-modified oligonucleotide Oligo-2-BT. Samples containing biotin-modified tetrahedron were mixed with 10 μL of a 1 mg/ml solution of steptavidin and allowed to incubate at RT for 10 mins. The samples were then run on a 12.5% PAGE.

3.3.3 Au Nanoparticle Functionalization using DNA Tetrahedrons

To 500 μL of a Au particle solution (4.5×10^{10} particles/ml, 48.9 nm in diameter; 67pM; Ted Pella Inc.) 5 nmol of 5'-thiolated DNA (29 nt; extinction coefficient = 290 900 dm^3/mole.cm) was added and left overnight. The solution was then diluted up to 1 mL with sodium phosphate buffer (5 mM Na_2HPO_4/NaH_2PO_4, 0.5 M NaCl, pH 7.0) and left overnight. The solution was concentrated slowly at 30 °C in a vacuum centrifuge to 100 μL, resulting in a red precipitate. The particles were re-suspended in

Na_2HPO_4/NaH_2PO_4 buffer, followed by centrifugation, removal of the supernatant, resuspension of the particles in phosphate buffer, and centrifugation. The supernatants from two washing procedures were combined and concentrated, and the amount of DNA was determined via UV/Vis specroscopy.

This procedure was repeated for an unmodified 29-nt control strand and for a tetrahedron DS_3 carrying three thiol groups at the apex positions. To release the particle-bound DNA material, 100 µL of a 100 mM aqueous solution of DTT was added to Au-DNA solution in 300 uL of buffer (10 mM phosphate buffer, 0.3 M NaCl, pH 7.0). Aliquots were taken at specific time intervals and scanned using a UV/Vis spectrophotometer from 800 – 200 nm.

3.3.4 XPS Analysis

XPS analysis was conducted for gold surfaces cleaned with an ion beam and after incubation with DS_3BT_1. The surface analysis was performed using an ESCALAB 200i-XL spectrometer operated using the large area-XL magnetic lens mode and an AlKa monochromated X-ray source operated with a 1 mm spot size. The take-off angle was maintained at 0° to the surface normal. Survey spectra were collected at a pass energy of 100 eV and detailed spectra at 20 eV pass energy. Removal of Shirley backgrounds and Gaussian peak fitting was performed using CasaXPS (Casa Software Ltd., U.K.).

3.3.5 AFM Analysis

Atomically flat template-stripped gold (TSG) was prepared as described by Hegner *et al.*[115] and used within two weeks of preparation. Gold surfaces were incubated with 230 nM solutions of DNA tetrahedra DS_3BT_1, DS_3 or BT_1 in buffer TNE (50 mM Tris, 3 M NaCl, 1 mM EDTA, pH 7.5) for one hour at room temperature. The surfaces were then rinsed with PBS buffer (100 mM NaCl, 20 mM NaH_2PO_4, pH 7.5). For the optional coating with streptavidin, DS_3BT_1 and DS_3 surfaces were incubated with a solution of streptavidin (0.1 mg/ml) in buffer PBS-SDS, composed of PBS supplemented with 0.1 % (w/v) sodium dodecylsulfate. After incubation for 1h, the surfaces were rinsed in PBS-SDS. Holes were scratched into the surface in contact

mode AFM using a cantilever (Sharpened MicroLever tips, Veeco Instruments S.A, Dourdan Cedex, France) with a nominal spring constant of 0.03 N/m. The force set point was 9V and scanning at a speed of 10 lines per second was repeated 4 times. All images were acquired at a speed of 1 line per second using the same cantilever in contact mode. Bare and DNA-coated TSG were imaged in PBS, while surfaces incubated with streptavidin were analyzed in PBS-SDS.

3.3.6 Fluorescence Measurements on the Affinity of Gold-Bound DNA Tetrahedra

Glass substrates bearing gold squares with a size of 50 x 50 µm and a height of 40 nm were prepared using standard photolithography. The substrates were then incubated with 500 nM solutions of DS_3Cy_1, DS_2Cy_1 and single stranded DNA oligonucleotide DS-DNA-Cy in TNM buffer (10 mM Tris, 2 M NaCl, 20 mM $MgCl_2$, pH 8.0). Oligonucleotide DS-DNA-Cy has the sequence 5'- CTAGACCGGTACAGATGC GTTCGAA-3' and carries a disulfide group via a hexamethylene (C6) linker at its 5' end, and a Cy3-C6 modification at its 3' end. After incubation for 3 h, slides were washed with TNM buffer for 15 min. The desorption of thiolated DNA tetrahedra or DS-DNA-Cy3 was enhanced by incubating the samples with TNM buffer supplemented with 10 mM dithiothreitol (DTT) for time periods ranging from two minutes to up to one day. Afterwards, the surfaces were rinsed in DTT-free TNM buffer and kept under liquid for fluorescence readout. A glass cover slip was placed over the sample to facilitate the fluorescence analysis. Fluorescence images were obtained with an in-house developed fluorescence scanning device[116,117] which is based on an inverted epifluorescence microscope (Axiovert 200, Zeiss, Oberkochen, Germany). For the measurements, a 100x objective (Zeiss, a Plan-FLUAR 100x/1.45) was used. Samples were mounted upside-down on a scanning stage (Märzhäuser, Wetzlar-Steindorf, Germany) and illuminated with a diode-pumped solid-state laser with an emission line of 532 nm (Millennia Iis, Spectra Physics, Irvine, CA). Images were taken with a Photometrics CoolSnap HQ digital camera (Roper Scientific, Trenton, NJ) (1392 x 1040-element CCD; pixel pitch, 6.45 x 6.45µm; 12-bit; QE, 0.6) using a time delayed integration mode. Image processing and analysis of the images were performed with V++ (Roper Scientific).

3.4 Formation and Analysis of DNA Tetrahedra

The formation of DNA tetrahedron DS_3BT_1 by annealing four DNA oligonucleotides Oligo-1-DS, Oligo-3-DS, Oligo-4-DS, Oligo-2-BT was analyzed via polyacrylamide gel electrophoresis. The four single stranded oligonucleotides migrated as single bands (lanes 1 – 4 figure 3.3). Lane 5 contains all four strands hybridized together resulting in a major high running band. The band of tetrahedron DS_3BT_1 migrated higher due to its increased mass and bulk. The next few lanes are controls, each the product of annealing two or three of the oligonucleotides. In these cases, hybridization to smaller, and faster running by-products are seen but the high running tetrahedron band is missing.

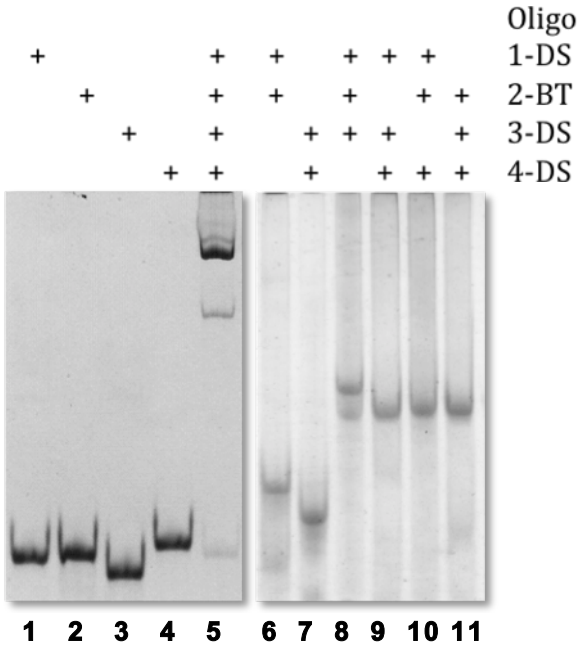


Figure 3.3. PAGE analysis on the formation of DNA tetrahedron DS_3BT_1. Lanes 1 – 4; single strands; lane 5, DNA tetrahedron obtained by mixing all four oligonucleotides; lane 6 – 11, omission of one or two strands from the annealing mixture.

3.4.1 Biotin-Functionalized Tetrahedra

The molecular recognition of the biotin moiety at the apex of DNA tetrahedron DS_3BT_1 was investigated using a gel-shift assay with streptavidin.

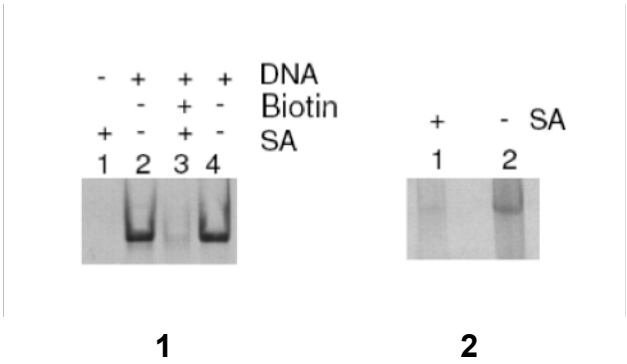


Figure 3.4. PAGE gels showing the results of the biotin pull down assay; (1) conjugation of a biotin – carrying ssDNA with Streptavidin (lane 3) versus the unmodified strand (lane 4); (2) conjugation of the biotin modified tet (lane 1) relative to the unmodified structure.

Prior to analyzing DNA tetrahedrons, the assay was tested with biotinylated single stranded DNA. As would be expected, incubation of streptavidin with the biotinylated DNA created a large complex, which did not enter the gel meshwork (Figure 3.4.1, lane 3) while non-biotinylated DNA in the presence of streptavidin migrated as a single band (Figure 3.4.2, lane 4). The same assay was repeated using biotin-modified tetrahedron DS_3BT_1. Incubation of this structure with the streptavidin solution led to a disappearance of the DNA band on the gel suggesting successful conjugation (Figure 3.4.2).

3.5 Functionalization of Surfaces with DNA Tetrahedra

The DNA tetrahedra carrying three disulfide groups were first coupled to Au nanoparticles to provide evidence that the binding is mediated via the formation of disulfide-gold bonds. A film composed of the thiol-functionalized tetrahedra was then formed on Au surfaces and characterized via XPS, AFM and fluorescence microscopy. The specific aim of the latter two characterizations was to demonstrate that the three anchoring disulfide points would bind the tetrahedron in a directed orientation and with high affinity to the gold surface.

3.5.1 Au Nanoparticles

Au nanoparticles were used to confirm that disulfide-tagged DNA tetrahedra bind specifically via the thiol group to the gold surface. The specificity was tested by comparing the binding of disulfide-modified with non-modified DNA strands, and by releasing gold-bound thiol-modified DNA via incubation in a reducing enviroment.[118]

The disulfide modified DNA tetrahedron DS_3 and 5'-modified ssDNA were annealed to the Au particles and concentrated (as detailed in section 3.3.3). At the chosen high nanoparticle and salt concentration, unmodified gold particles usually precipitate but the DNA-modified particles are held in suspension due to the electrostatic repulsion of the negatively charged DNA coating. Suspended particles absorb at 530 nm giving rise to a red colour. Breaking the bond between thiolated DNA and gold via reducing agent DTT results in the aggregation of the particles, and a concomitant change in absorbance, turning the suspension from red to light purple. Since the DNA tetrahedrons carry three disulfide legs per structure they should resist reduction for longer than that of their single stranded counterparts.

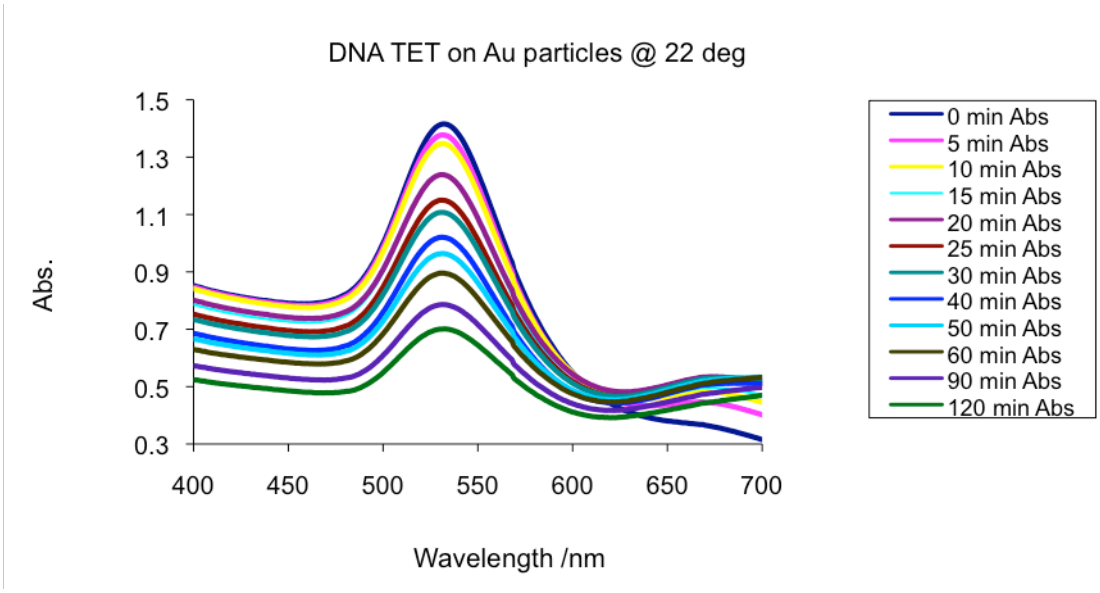


Figure 3.5. Absorbance curves of the tetrahedra-modified Au particles in the presence of DTT for increasing incubation times.

A time-course experiment of the reduction was conducted by taking aliquots of the DNA/Au in a DTT solution at increasing intervals (Figure 3.5). A decrease in absorbance at 530 nm is clearly seen over the course of the 2 h. This experiment was

repeated using a single stranded 55-mer carrying a 5'-modified disulfide group. The decrease in the absorbance at 530 nm for both the single strand and tetrahedron was then normalized to the maximum and minimum absorbance range (Figure 3.6).

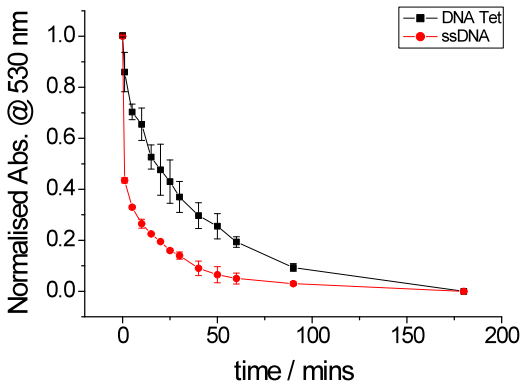


Figure 3.6. Plot of the decreasing absorbance at 530 nm of the Au particle solutions over 2h.

A slower decline of absorbance at 530 nm is clearly seen for the tetrahedron sample compared to that of the ssDNA indicating it is indeed more stable towards cleavage. The difference in the initial decay of the graphs was approximately 5 fold, however a greater increase in stability was expected due to the three anchor points of the structure over a single strand. This lowered stability may be attributed to the reduction of NaCl concentration in the solution during the annealing procedure. The salt concentration must be initially high to bind the DNA; during this procedure the NaCl conc. is increased from 0.5 M to over 1 M as the volume of the solution is reduced. However, the particles must then be re-suspended in buffer – if the conc. of NaCl is too high the absorbance of the DNA at 260 nm is masked and the absorbance at 530 nm interfered with. Therefore the salt levels are decreased to get an accurate read; this however will also decrease the T_m of the structures and thus weaken the binding of the tetrahedron to the particles, hence a lower than expected stability. This hypothesis can be tested by repeating the experiment once the structures are bound to a Au surface substrate for AFM analysis. Using the fluorescence of a Cy3 group at the apex of the structure we will not encounter the same issues that have arisen in this assay.

3.5.2 Characterization of DNA-Tetrahedron Films on Au Surfaces

The ability of DS_3BT_1 to bind onto gold surfaces was assessed using X-ray photoelectron spectroscopy.[119] In XPS, a surface is irradiated with X-rays of defined energy, and the kinetic energy of the escaping electrons is measured. By calculating the energy difference, information relating to the elemental composition and electronic state of the elements can be ascertained.

Planar gold supports were coated with DNA-tetrahedron DS_3BT_1 and subjected to XPS analysis to confirm the presence of a DNA film. Figure 3.8A shows the survey spectra of cleaned bare and DS_3BT_1-treated gold surfaces. The bare gold surface displays major peaks at 84, 87, 335, 353, and 546 eV which are characteristic for $Au4f_{7/2}$, $Au4f_{5/2}$, $Au4d_{5/2}$, $Au4d_{3/2}$ and $Au4p_{3/2}$, respectively. Upon treatment with DNA- tetrahedron DS_3BT_1, the Au signals were reduced while peaks for P2p (135 eV), C1s (285 eV), N1s (400 eV), and O1s (530 eV) appeared. These peaks are consistent with the formation of a DNA film. The changes in the elemental composition of the Au surfaces upon DNA treatment are summarized in Table 3.3. Additionally, small quantities of Na, S and Zn were observed, again in agreement with the formation of a DNA film. XPS analysis also showed the attenuation of the Au4f signal as well as the appearance of peaks for C1s, N1s, O1s, and P2p at relative ratios expected for DNA.

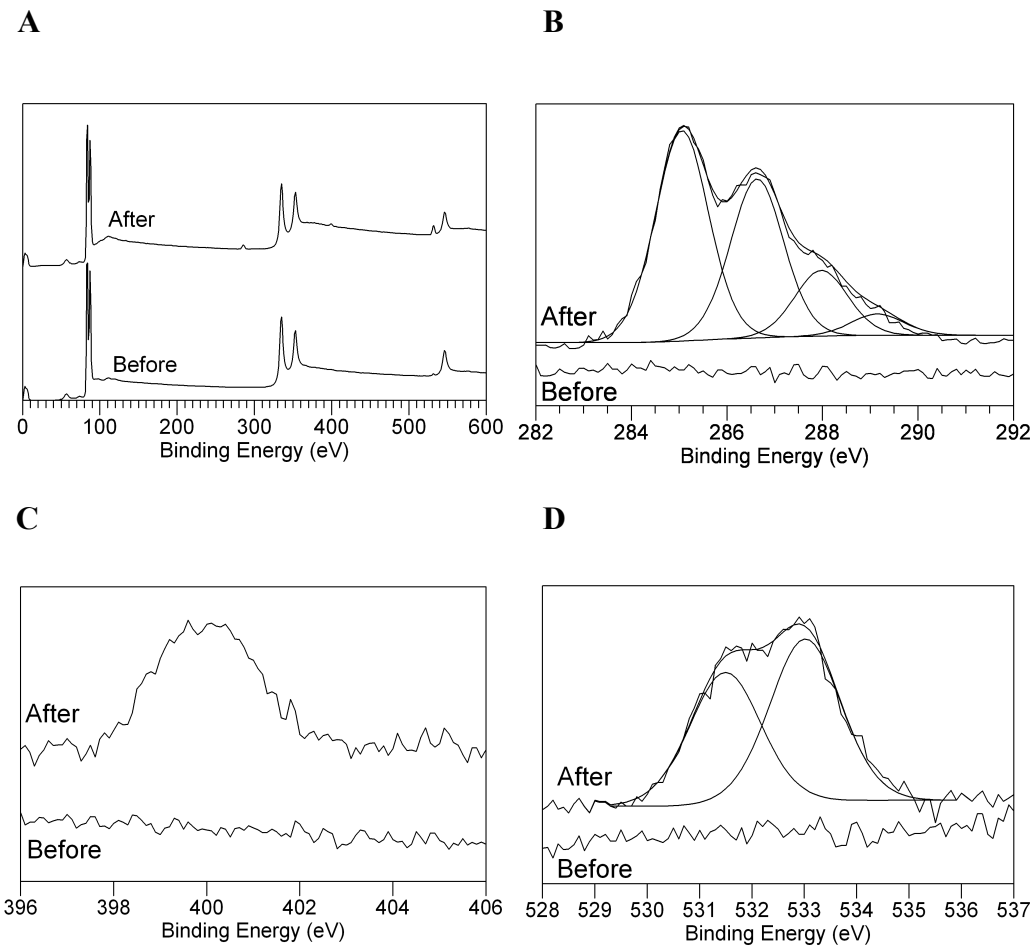


Figure 3.7. XPS analysis of Au surfaces coated with a films of DNA tetrahedra. (A) Survey spectra of Au substrates before and after incubation with DNA tetrahedron DS_3BT_1. High resolution XPS spectra of (B) C1s, (C) N1s, and (D) O1s before and after treatment with DS_3BT_1.

	Au	C	N	Na	O	P	S	Zn
Bare Au	96.6	0.8	0.2	0	0.1	1.8	0.3	0.2
Au / DS_3BT_1	19.3	44.8	10.7	0.6	20.7	2.5	0.9	0.4

Table 3.3. Full chemical quantification of bare Au surfaces and Au surface coated with tetrahedron DS_3BT_1 as determined by XPS.

The detailed XPS spectra for C1s, N1s, and O1s of the uncoated and DNA-treated surfaces (Figure 3.8B to Figure 3.8D, respectively) were subjected to a detailed analysis to derive elemental ratios for the organic film. The analysis of the XPS region for C1s and O1s (Figure 3.8B) was performed by fitting Gaussian peaks

to the spectra. Four separate peaks were required in the deconvolution procedure for C1s to achieve a random residual. The four peaks were at 285.0, 286.6, 288.0 and 289.2 eV with relative intensities of 46, 35, 14 and 5%, respectively (Figure 3.8B). The peak energies and relative intensities are similar to the detailed C1s spectrum for an oligothymidine strand of 25 nt in length. As the DNA tetrahedron DS_3BT_1 did, however, contain all four bases rather than just T, a simple scheme was utilized to predict the spectra of our DNA sample. This scheme accounted for changes in the energies of the carbon peaks caused by the complex structure of and the conjugation within the heteroaromatic base rings, as well as the electron-withdrawing phosphate groups in the deoxyribose. For example, a shift of +1.5 eV was applied for each C-O bond and a shift of +0.4 eV for each C-N bond. This procedure yielded nine unique binding energy peaks. To simplify the analysis and to allow a direct comparison to the experimental data, theoretically derived peaks of similar binding energy separated by less than 0.4 eV were combined to yield four peaks. The energies of the four peaks were 285.2, 286.3, 288.4 and 288.8 eV with estimated peak area ratios of 36, 43, 18 and 3%, respectively. These calculated data compare to the experimental peaks at 285.0, 286.6, 288.0 and 289.2 eV with relative intensities of 46, 35, 14 and 5%, respectively. The agreement between the peak positions is good considering the complexity of the actual DNA structure. Hydrocarbon contamination is the likely cause of the slightly elevated levels of the 285.0 eV peak in the measured spectra. The ratio of the two peak areas at 288.0 eV and 286.6 eV is 0.41, which is in excellent agreement to the expected value of 0.41 for DNA. The difference of 2% for the energy peak at 289.2 eV is reasonable considering the low intensity of this signal. The detailed N1s region displayed a strong peak at 400 eV for the DNA surface but no signal prior to the treatment (Figure 3.8C). The small asymmetry of the peak is due to the multiplicity of binding environments in DNA bases. Similarly, upon addition of DNA the spectra for the expanded oxygen region (Figure 3.8D) showed two peaks at 531.6 and 533.1 eV with relative peak intensities of 43 and 57%. The energies and relative areas of the two peaks are in excellent agreement with a report on the high-resolution XPS analysis of DNA films with values of 531.7 eV and 533.3 eV, and peak area ratios of 44 and 56%. Table 3.4 summarizes the experimental and the theoretical ratios of the elemental signals. For example, the experimental O/N ratio with 1.9 is identical to the calculated ratio. Similarly, the determined O/P ratio of 8.3 is close to the expected value of 7.0. The ratios C/N, C/O, C/P, and C/S were

corrected for the presence of a carbon contamination and yielded reasonable to good agreement with the predictions. Taken together, the XPS spectrum provides strong support for the presence of a DNA film on the Au surface.

Elemental Ratio	C/N	C/O	C/P	O/N	O/P
Experimental	4.2	2.1	17.9	1.9	8.3
Theoretical	2.6	1.4	9.8	1.9	7.0

Table 3.4. Elemental ratios of DS_3BT_1 films determined from XPS analysis compared to calculated ratios expected for DNA.

Atomic Force Microscopy (AFM) was used to determine the extent of the DNA tetrahedra binding and its ability to conjugate to streptavidin whilst maintaining its adherence to the gold.[120] A DS_3BT_1 film was formed and incubated with streptavidin. Scratching of the surface revealed a profile step of 7 nm (Figure 3.8, panel 3), which is consistent with the formation of a monomolecular layer of streptavidin (size of protein, 4x8x7 nm) on top of a DNA film. The elevated, bright features around the hole (Figure 3.8, panel 3) represent streptavidin and DNA which had been removed from the film. In line with a specific biotin-streptavidin interaction, BT-free tetrahedron DS_3 did not capture streptavidin as indicated by the smaller step size (Figure 3.8, panel 4).

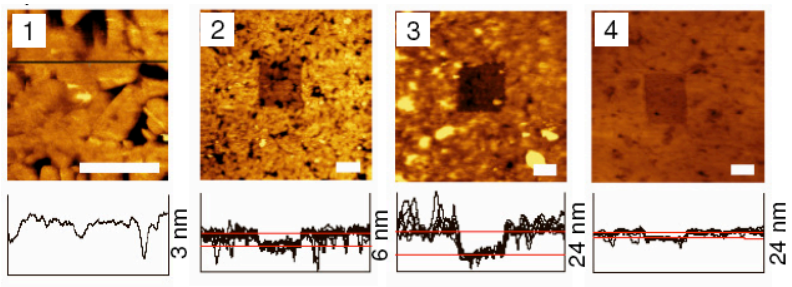


Figure 3.8. AFM analysis of Au surfaces coated in DNA tet layer and Streptavidin protein layer; (1) bare surface; (2) DNA tet layer; (3) DNA/Protein layer; (4) DNA tet minus biotin layer.

As would be predicted for oriented and specific binding, tetrahedron DS_3BT_1 yielded a monomolecular film in AFM analysis. For the topographic measurements, template-stripped gold (TSG) was used. TSG exhibited flat (rms noise, 0.3 nm) facets

of 100 to 1000 nm in size, which were separated by 2-10 nm deep trenches (Figure 3.8, panel 1). By comparison, gold surfaces incubated with DS_3BT_1 displayed an increased rms noise of 0.5 nm. In addition, a hole with a depth of 1.5 nm could be scratched into DS_3BT_1 surfaces (Figure 3.8, panel 2). This value is lower than the height of 4.5 nm of the DNA tetrahedron because soft biological material can be compressed in contact mode AFM. In agreement with the disulfide-specific tethering, gold surfaces incubated with DS-free tetrahedron BT_1 did not form films according to AFM analysis.

After confirming the oriented binding of DS_3BT_1, the affinity of the tetrahedra to the gold substrate was determined. Due to the anticipated multivalent enhancement of individual gold-thiol interactions, the DNA structures with three DS legs should bind tighter to the surface than constructs with two or a single leg. The differential affinity was experimentally determined by first forming films composed of fluorescence-labeled tetrahedra DS_3Cy_1 or DS_2Cy_1, or single-stranded oligonucleotide DS-DNA-Cy, and then monitoring their time-dependent desorption using fluorescence microscopy. Desorption was enhanced by incubating the DNA films with 10 mM dithiothreitol (DTT) which displaces thiolated DNA from the gold surface. The kinetics of fluorescence decrease are displayed in Figure 3.9.

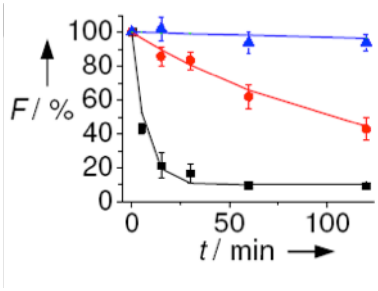


Figure 3.9. Plot of DNA release from DNA surface; fluorescence of Cy3 moiety plotted against time; blue triangles – DNA tetrahedra equipped with three disulphide 'legs'; red circles – DNA tetrahedra with two disulphide legs; black squares – single stranded oligonucleotide with only one disulphide group.

During an observation window of 120 min, DS_3Cy_1 was only slightly removed by 5% (Figure 3.9, blue triangles) while the coverage of DS_2Cy_1 with two thiol legs fell linearly to below 50% of the starting value (Figure 3.9, red circles). DNA with a single DS group exhibited the lowest affinity with the signal exponentially decaying

to 10% after 30 min (Figure 3.9, black squares). The residual constant signal is due to autofluorescence of gold and background fluorescence of desorbed and solvated DNA strands. Comparative analysis of the initial rates of desorption revealed that DS_3Cy_1 with three legs bound with a 5000-fold higher affinity to gold than monothiolated DNA confirming our previous suspicions of sample denaturation during the Au nanoparticle investigation.

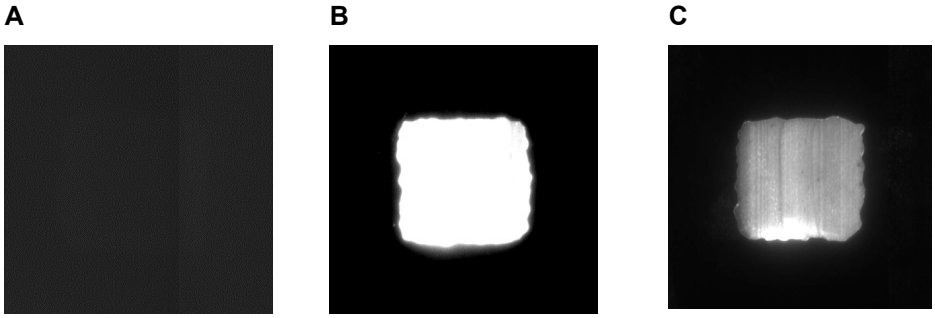


Figure 3.10. Fluorescence microscopic analysis on the binding of DS_2Cy3_1 onto gold squares and the DTT-mediated release of the fluorescence-labeled tetrahedra from the gold surface. Fluorescence images show glass substrates with a 50 x 50 µm gold square without DNA film (A) and with a freshly formed layer of DS_2Cy_1 bound to the gold surface (B). (C) displays a DS_2Cy_1-coated gold square after 60 min incubation with 10 mM DTT.

3.6 Addressing Individual Biotinylated DNA Tetrahedra with Scanning Probe Techniques

DNA tetrahedra were investigated with two scanning probe techniques to confirm that individual DNA tetrahedra DS_3BT_1 were binding in an oriented fashion to the gold surface thereby presenting the biotin moiety to the ambient. Two single-molecule scanning probe techniques were used: atomic force microscopy (AFM) and molecular recognition force spectroscopy (MRFS). AFM imaging of individual molecules detects topographic height changes when a cantilever is scanned over a sample surface. By comparison, force microscopy utilizes cantilevers whose tips are chemically modified with receptors. The tethered receptor can recognise a surface-bound ligand leading to changes in the amplitude of an oscillating AFM tip.

3.6.1 Molecular Recognition Force Spectroscopy (MRFS)

For the investigation of the biotin-carrying DNA tetrahedron DS_3BT_1 by recognition force spectroscopy, an AFM tip was coated with streptavidin proteins (Figure 3.11). The tip carrying primary amine groups was first reacted with the heterobifunctional linker NHS-PEG-aldehyde. The aldehyde group was then coupled to a streptavidin protein via reductive amination of surface accessible lysine groups.

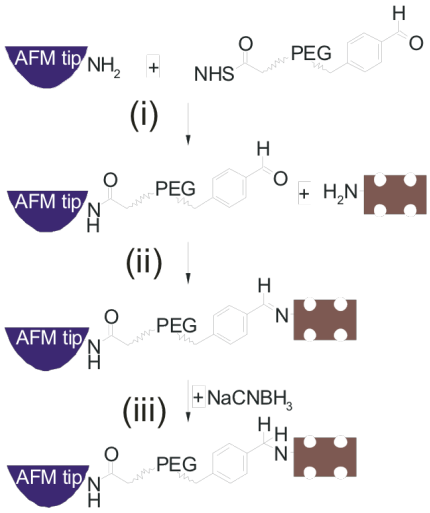


Figure 3.11. Functionalization of an AFM tip: (i) Attachment of an aldehyde linker via attack of an amine-functionalized tip to an activated ester linker; (ii+iii) attachment of a streptavidin protein via reductive amination.

3.6.2 Application of MRFS and AFM to Analyze DNA Tetrahedrons

MRFS was first used to characterize DS_3BT_1 tetrahedra on a Au surface. The MRFS experiment is illustrated in Figure 3.12. As the oscillating AFM tip passes over one of the biotin-modified DNA-tetrahedrons, streptavidin attached to the tip binds to the biotin moiety. As the oscillating tip is retracted, it is held back by the surface-tethered biotin-streptavidin bond. As the weakest component of the system, the non-covalent association of the biotin group and the protein ruptures once the tip is retracted far enough. The force at which this occurs is characteristic to the biotin-streptavidin

complex and can be used to investigate the binding affinities of the substrate on the surface.

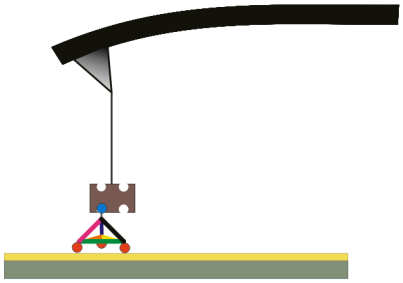


Figure 3.12. Identification of the DNA structure via molecular recognition force spectroscopy (MRFS).

Figure 3.13 (left) shows the binding and rupture of the complexes visualized in Figure 3.12. After repeating the binding and captured cycles 1000 times, the rupture forces were displayed in a histogram to reveal a typical force of 50 pN which is characteristic for biotin-streptavidin (Figure 3.13, right). Controls which block the specific molecular interaction also reduce the probability of recognition in the histograms (Figure 3.13, right).

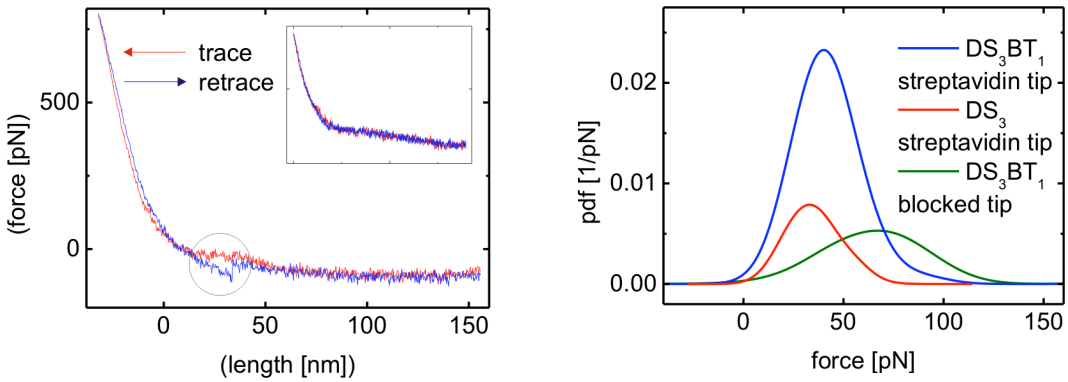


Figure 3.13. MRFS data (left) traces showing the conjugation and release of protein-modified AFM tip and (right) probability density function of the rupture force for different conditions.

The ability of biotin-functionalized tetrahedra to bind individual streptavidin molecules was also proven by AFM imaging. First, DS_3BT_1 tetrahedra were immobilized on a gold surface via their disulfide legs (Figure 3.14A, top). The resulting surface density of DS_3BT_1 was lower than in the previous unbinding experiments in order to permit the unambiguous identification of individual tetrahedra (Figure 3.14A, bottom). Full and dotted circles mark individual DNA tetrahedra with a height of 2 nm and a diameter between 10-15 nm. The measured height is lower than the nominal height of the DNA tetrahedra due to the compression by the AFM tip. In the second step, streptavidin was added to form DS_3BT_1-streptavidin complexes (Fig. 3.14B, top). Given that the same surface area was imaged, the enlarged topographical signals (Figure 3B, bottom, dashed circles) could be assigned to individual tetrahedron-streptavidin complexes. Only ~8% of all tetrahedra did not bind streptavidin (solid circles) whereas non-specific binding of streptavidin occurred even less frequently (dashed circle).

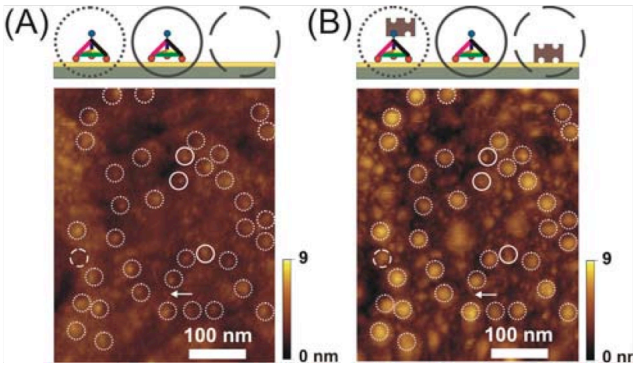


Figure 3.14. AFM of individual DNA tetrahedral molecules on a Au surface (A) and imaging of protein conjugation (B).

3.7 Conclusions

This chapter presented a new strategy, which exploits DNA-nanostructures as scaffolds to combine different chemical groups at defined geometrical distance and tunable stoichiometry. The rationally designed structures exhibit functional properties, which may be exploited for the immobilization of DNA or proteins on gold for biosensing, diagnostics, and cell biological research. Given the great variety of DNA

nanostructures, the approach for chemical enhancement has the potential to be extended to other nanostructures and different bio-chemical tags. For example, it is envisioned that the structures may help prepare new templates for chemical reactions, create functional building blocks for defined multimeric enzyme complexes, or build labeling reagents carrying tunable numbers of tags.

It may also be possible use thiolated DNA tetrahedrons as spacers to immobilize cargo DNA onto surfaces in biochips. Compared to highly densely packed ssDNA, DNA on tetrahedrons would be spaced at greater intervals thereby improving steric accessibility and hybridization with an analyte DNA strand of complementary sequence. Furthermore, the defined position of the DNA cargo at the top of the tetrahedron and distant to the substrate surface would also improve accessibility compared with the non-defined orientation of ssDNA resulting from the point-like attachment via single thiol group.

The work described in this chapter has been published in Angewandte Chemie Int. Ed (Nick Mitchell, Robert Schlapak, Markus Kastner, David Armitage, Wojciech Chrzanowski, Johannes Riener, Peter Hinterdorfer, Andreas Ebner and Stefan Howorka; 2009; 48; 525 – 527). My contribution to this publication included investigation of DNA tet formation and design/inclusion of chemically modified oligos. I carried out the original Au nanoparticle investigation and prepared all samples subjected to surface analysis. The AFM analysis was conducted by Andreas Ebner and the X-ray photoelectron spectroscopy by Wojciech Chrzanowski. All other authors gave assistance with either the experimental work or the composition of the publication.

Chapter Four

Chemical tags mediate the self-assembly of DNA strands into supramolecular structures

4.1 Summary

The following chapter shows that chemically tagged DNA strands can form defined higher-order structures via the formation of Ni(II)-mediated chelate bridges. Tags that are capable of chelating Ni(II) ions were synthesized and conjugated to DNA duplexes. The selective interaction between duplexes via the metal-chelate bridges was proven using an assay capable of physically isolating the higher-order structures from the monomeric duplexes. In addition, chemically tagged helices were synthesized to self—assemble into a 2D array, as visualized by AFM analysis. Finally, chemically tagged DNA rings were synthesized with the aim of forming tubular stacks of rings held together via the chelate bridges.

4.2 Introduction

Self-assembly is an important functional property of biological biopolymers, as illustrated by naturally occurring higher-order structures of proteins and DNA.[121,122,123] The ability to engineer self-assembly into monomeric peptides or proteins via rational design is an emerging topic in supramolecular chemistry, nanobiotechnology and synthetic biology due to the potential applications in basic science and technology.[124]

Existing approaches towards imparting self-assembly use the engineering of protein surfaces,[125] the bottom-up design of linear peptides,[126] or selection of complementary DNA sequences.[127] An alternative approach, which achieves self-assembly independent of the re-engineering of the structural framework of the subunits, can be conceived. The new approach would use chemical tags that are covalently attached to the bio-polymer subunits to form reversible yet tight metal chelate complexes resulting in designed supramolecular biopolymer structures.

Here the new, generic approach, is tested by the creation of nanoarrays of DNA duplexes. Our strategy to impart self-assembly onto biomolecules exploited the two chemical tags *bis*-nitrilotriacetic acid (*bis*NTA) and a His_6 peptide, which are able to form a metal chelate complex (Figure 4.1). In the bisNTA-Ni^{2+}-His_6 complex, two nitrilo-triacetic acid (NTA) groups are bonded via two Ni^{2+} ions to the imidazole moieties of a hexahistidine peptide.

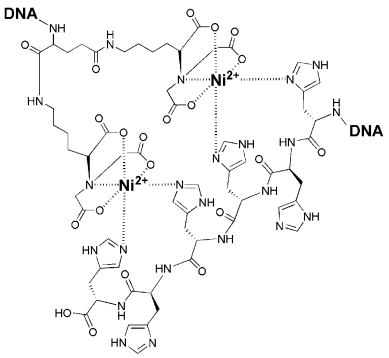


Figure 4.1. DNA bound via *bis*NTA – Ni(II) – His_6 complex

The ability to reversibly bind DNA duplexes using designed chemical tags could have interesting potential applications in biotechnology. By utilizing tags that

are able to form reversible metal chelates it should be possible to bind either proteins or DNA to one another or to a solid support, for example, without the need to alter their specific sequence or structure. Importantly, this allows the biopolymers to retain their biological function. The NTA-Ni(II)-His_6 complex is already used to good effect in Immobilized Metal Affinity Chromatography (IMAC) where NTA groups are immobilized onto a solid support and loaded with Ni(II) ions. Proteins that have His_6 sequences encoded into their terminus can be retained on the column due to the formation of the above complex and eluted via competitive binding by flooding the system with excess imidazole.

The *bis*NTA tag has been developed to bind with a K_d of 10^8 M^{-1} to the His_6 peptide.[128] In addition to high affinity, a linkage via bisNTA-Ni^{2+}-His_6 features several other beneficial properties useful in the design of supramolecular structures. These include reversibility and controllability by adding or removing Ni^{2+}, and directional binding as a *bis*NTA tag can only interact with His_6. Reflecting these advantages, our aim was to attach *bis*NTA and His_6 tags to single base positions within DNA duplexes. Given the position of the tags at opposite sites of the duplex, they are expected to endow DNA with the ability to form parallel aligned nanoarrays (Figure 4.2). The array formation process is triggered by Ni^{2+} while a metal chelating agent such as $EDTA^{4-}$ (ethylenediamine tetraacetic acid) can revert the organisation and make the system regress back into a disordered state.

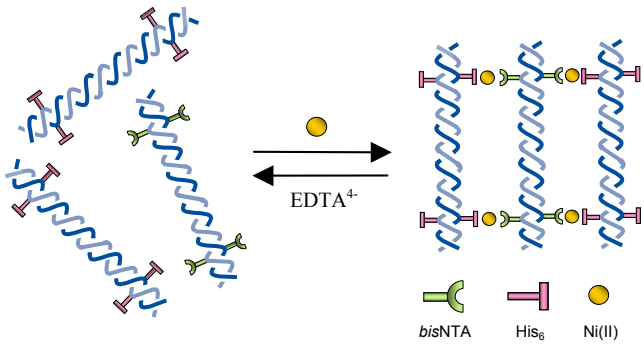


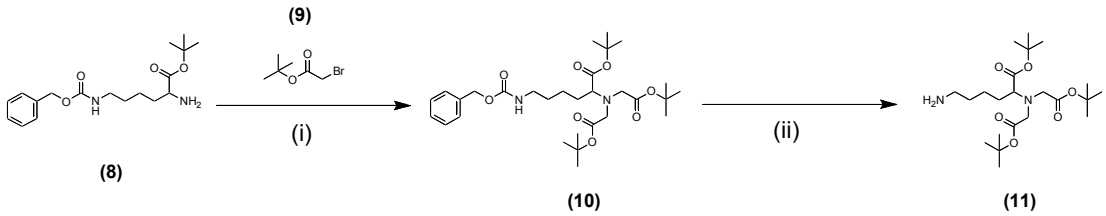
Figure 4.2. Schematic diagram of Ni(II) mediated array formation via binding of chemically tagged duplexes.

The *bis*NTA and $CysHis_6$ peptide components were initially synthesized and conjugated to oligonucleotides that contained non-biogenic thymine bases carrying an amine linker. These chemically modified oligonucleotides were used in an assay to

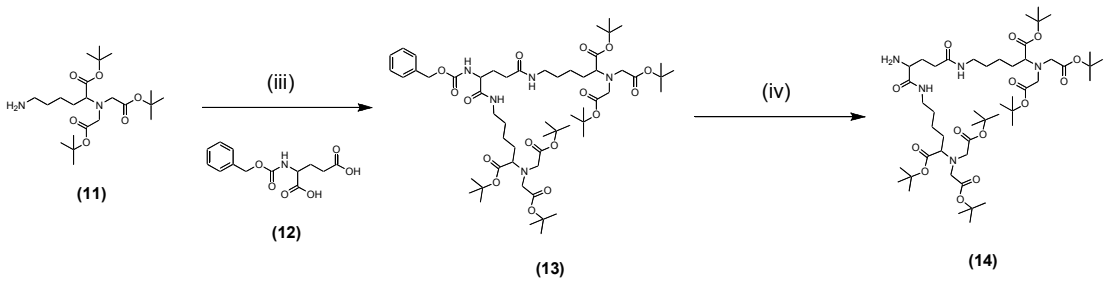
detect the formation of a Ni(II)-mediated dimer between a *bis*NTA and a His_6-tagged duplex.

4.3 Materials and Methods

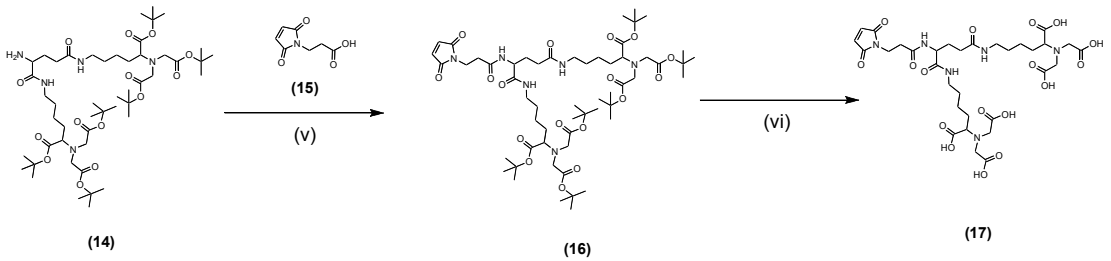
The *bis*NTA compound carrying a maleimide functionality (compound **17**) was synthesized via a procedure adapted from Lata et al.[128]



Scheme 4.1. Synthesis of ^{t}Bu protected mono-nitrilotriacetic acid; (i) Cbz/tBu protected Lysine, tert butyl bromo acetate, DIPEA, DMF, reflux; (ii) 10% Pd/C, $H_{2(g)}$, MeOH, RT.

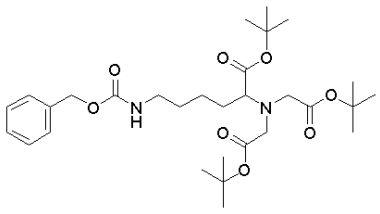


Scheme 4.2; Synthesis of tBu protected *bis*NTA; (iii) mono-NTA, Cbz protected Glutamic Acid, TBTU, DIPEA, DCM; (iv) 10% Pd/C, $H_{2(g)}$, MeOH, RT.



Scheme 4.3; Synthesis of mal–*bis*NTA; (v) tBu protected bis NTA, Maloeyl – beta – alanine, TBTU, DIPEA, DCM; (vi) 50% TFA/DCM.

4.3.1 Synthesis of Z-*bis*NTA (OtBu) (10)

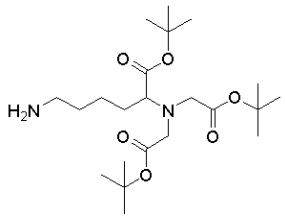


(10)

6-benzyloxycarbonylamino-2-(bis-tert-butoxycarbonylmethyl-amino)-hexanoic acid **(10)** was synthesized via sequential addition of *tert* butyl bromoacetate (**9**; 7.91 mL; 53.6 mmol) and diisopropylethylamine (DIPEA) (11.7 mL; 67 mmol) to a solution of N^{ε}-benzyloxycarbonyl-*L*-lysine *tert*-butyl ester (**8**; 5 g; 13.4 mmol) in 80 mL anhydrous DMF. The reaction was stirred magnetically at 55 °C overnight. After this period the solvent was removed under reduced pressure. Petroleum ether : ethyl acetate (3:1) was added to the crude oil causing precipitation of the bromide salt of DIPEA. The slurry was filtered, and the precipitate washed multiple times with the petroleum ether : ethyl acetate solution until the light brown coloration had been removed. The crude was then purified via flash column chromatography using the solvent system petroleum ether : ethyl acetate (5:1). The fractions containing the desired product were pooled, the solvent removed under reduced pressure and the resulting oil dried on a high vacuum line to yield the titled compound **10** as a yellow oil (6.97 g, 12.35 mmol, 92 %). Calculated mass ($C_{30}H_{48}N_2O_8$) 564.34 g.mol^{-1}; observed mass (ESI^+) 565.46 (MH^+) & 587.50 ($M + Na^+$). ^1H NMR (300 MHz; $CDCl_3$) δ/ppm 1.41 (18H, s), 1.44 (9H, s), 1.52 (4H, m), 1.62 (2H, m), 3.18 (2H, m),

3.29 (1H, t, J = 6.9), 3.37 (4H, dd, J = 17.1, 4.8), 5.07 (2H, s), 5.10 (1H, m), 7.33 (5H, m); ^{13}C NMR (75 MHz; $CDCl_3$) δ/ppm 23.0, 28.1, 28.2, 29.2, 30.1, 40.8, 53.9, 65.1, 66.4, 80.7, 81.1, 128.0, 128.1, 128.4, 136.8, 156.5, 170.7, 172.4.

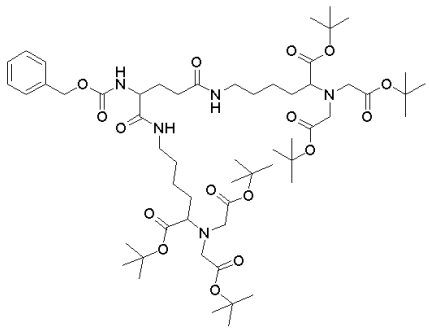
4.3.2 Deprotection of Z-*bis*NTA (O^{t}Bu) yielding 11



(11)

6-amino-2-(bis-tert-butoxycarbonylmethyl-amino)-hexanoic acid tert-butyl ester **(11)** was synthesized by dissolving compound **10** (2.5g; 4.43 mmol) in 100 mL of anhydrous methanol, 10% Pd/C (50 mg) was added, and the resulting solution was purged with $Ar_{(g)}$. The reaction was left stirring vigorously for 5 hours under an atmosphere of $H_{2(g)}$ at room temperature. After this period the solution was filtered through celite, and the solvent was removed under reduced pressure. The resulting light yellow oil was dried on a high vacuum line to yield the titled compound **11** (1.82 g; 4.24 mmol; 96%). Calculated mass ($C_{22}H_{42}N_2O_6$) 430.30 g.mol^{-1}; observed mass (ESI$^+$) 431.63 (MH^+). ^{1}H NMR (300 MHz; $CDCl_3$) δ/ppm 1.42 (18H, s), 1.44 (9H, s), 1.47 (4H, m), 1.64 (2H, m), 1.75 (2H, m), 2.67 (2H, t, J = 5.9), 3.29 (1H, t, J = 7.5), 3.38 (4H, dd, J = 17.1, 7.5); ^{13}C NMR (75 MHz; $CDCl_3$) δ/ppm 23.2, 28.1, 28.2, 30.5, 33.2, 41.9, 53.8, 65.3, 80.7, 81.0, 170.7, 172.4.

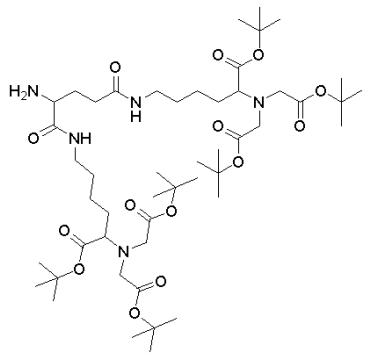
4.3.3 Synthesis of Z-*bis*NTA (O^{t}Bu) (13)



(13)

6-{4-benzyloxycarbonylamino-4-[5-(bis-tert-butoxycarbonylmethyl-amino)-5-tert-butoxycarbonyl-pentylcarbamoyl] butyryl amino}-2-(bis-tert-butoxycarbonylmethyl-amino)-hexanoic acid tert-butyl ester was synthesized by dissolving compound **11** (1.3 g; 3.02 mmol) in anhydrous DCM (50 mL) followed by addition of compound **12** (386.5 mg; 1.374 mmol), TBTU (970 mg; 3.02 mmol) and DIPEA (956 μL; 5.49 mmol). The slurry was purged with $Ar_{(g)}$ and allowed to stir at RT under $Ar_{(g)}$ overnight. After this period the solvent was removed under reduced pressure to yield a slurry. The crude product was dissolved in DCM and washed with water and brine. The organic phase was dried over anhydrous magnesium sulfate, and the solvent removed under reduced pressure. The crude was purified via column chromatography using a solvent system of petroleum ether : ethyl acetate (4:1). All fractions containing the product were pooled and the solvent removed under reduced pressure to yield the titled compound **13** as a white solid (1.2 g; 1.09 mmol; 79%). Calculated mass ($C_{57}H_{95}N_5O_{16}$) 1105.68 g.mol^{-1}; observed mass (ESI$^+$) 1106.86 (MH^+), 1128.85 ($M + Na^+$). ^{1}H NMR (300 MHz; $CDCl_3$) δ/ppm 1.43 (36H, s), 1.45 (18H, s), 1.50 (8H, m), 1.64 (4H, m), 1.97 - 2.08 (2H, m), 2.27 - 2.39 (2H, m), 3.19 - 3.34 (6H, m), 3.38 (8H, dd, J = 17.1, 6.4), 4.17 (1H, q, J = 7.0), 5.08 (2H, s), 6.37 (1H, d, J = 6.4), 6.57 (1H, at), 7.05 (1H, at), 7.33 (5H, m). ^{13}C NMR (75 MHz; $CDCl_3$) δ/ppm 22.8, 23.1, 28.8, 29.3, 29.7, 30.0, 32.4, 39.3, 53.7, 54.0, 54.4, 64.9, 66.7, 80.8, 80.9, 81.1, 81.2, 128.98, 128.0, 128.5, 170.7, 170.8, 171.3, 172.3, 173.0.

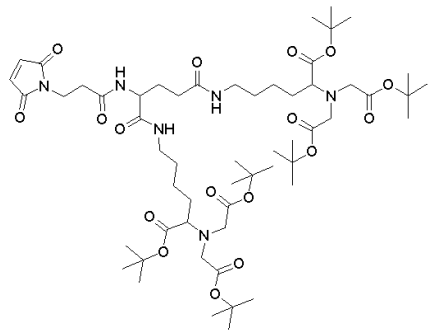
4.3.4 Deprotection of Z-*bis*NTA (O^{t}Bu) yielding 14



(14)

6-{4-amino-4-[5-(bis-tert-butoxycarbonylmethyl-amino)-5-tert-butoxycarbonyl-pent-ylcarbamoyl]-butyrylamino}-2-(bis-tert-butoxycarbonylmethyl-amino)-hexanoic acid tert-butyl ester was synthesized by dissolving compound **13** (1.08 g; 0.98 mmol) in 50 mL of anhydrous methanol, 10% Pd/C (20mg) was added, and the resulting solution was purged under $Ar_{(g)}$ and left stirring vigorously for 5 hours under an atmosphere of $H_{2(g)}$ at RT. Analysis using LC/MS showed the presence of *tert* butyl groups, therefore a further 100 mg of 10% Pd/C added to the solution. The flask was evacuated and purged with Argon and then with a fresh $H_{2(g)}$ balloon. After a further 2h the solution was filtered through celite, and the solvent removed under reduced pressure. The product was dried under high vacuum to yield the titled compound **14** as a white foam (0.866 g; 0.92 mM; 94%). Calculated mass ($C_{50}H_{93}N_5O_{14}$) 971.64 g.mol^{-1}; observed mass (ESI$^+$) 972.70 (MH^+) g.mol^{-1}, 994.60 (MH^+) g.mol^{-1}. ^{1}H NMR (300 MHz; d_6-DMSO) δ/ppm 1.42 (36H, s), 1.44 (18H, s), 1.51 (8H, m), 1.63 (4H, m), 1.8 – 2.0 (2H, m), 2.2 – 2.39 (2H, m), 3.21 – 3.31 (6H, m), 3.43 (9H, ddd, J = 3.2, 2.7), 3.46 (1H, s), 6.70 (1H, t. J = 9.3, 4.8), 7.48 (1H, t, J = 10.7, 5.4). ^{13}C NMR (75 MHz; d_6-DMSO) δ/ppm 23.3, 28.1, 28.2, 28.5, 30.2, 39.0, 53.8, 53.9, 80.7, 80.8, 81.1, 170.7, 170.8, 172.4, 173.0.

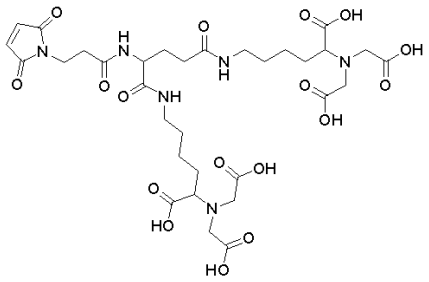
4.3.5 Synthesis of mal-*bis*NTA (O^{t}Bu) (16)



(16)

2-(bis-tert-butoxycarbonylmethyl-amino)-6-{4-[5-(bis-tert-butoxycarbonylmethyl-amino)-5-tert-butoxycarbonyl-pentylcarbamoyl]-4-[*N*-(2-carboxyethyl)maleimide]-butyrylamino}-hexanoic acid tert-butyl ester **16** was synthesized by dissolving compound **15** (0.862 g, 0.888 mmol) in anhydrous DCM (15 mL) and adding a solution of 0.165 g (0.98 mmol) *N*–maleoyl–β–alanine (in DCM) followed by 0.315 g of TBTU (0.98 mmol) and DIPEA (308.6 μL. 0.23 g, 1.78 mmol). The reaction was purged with $Ar_{(g)}$ and stirred overnight at RT under an $Ar_{(g)}$ atmosphere. Complete conversion to the desired product was monitored via LC/MS. Once the reaction had reached completion the solvent was removed under reduced pressure, and the crude solid purified via column chromatography using a solvent system of 4% methanol in DCM. The fractions containing the desired product were pooled, and the solvent removed under reduced pressure to yield the titled compound as a white solid/gum (0.97 g; 0.863 mM; 97%). Calculated mass ($C_{56}H_{94}N_6O_{17}$) 1123.6754 g.mol^{-1}; observed mass (ESI$^+$) 1123.6810 (MH^+) g.mol^{-1}. ^{1}H NMR (300 MHz; d_6-DMSO) δ/ppm 1.42 (36H, s), 1.44 (18H, s), 1.50 – 2.55 (11H m), 3.17 – 3.32 (8H, m), 3.44 (12H, m), 3.82 (3H, t, J = 6.4), 4.31 (1H q, J = 5.9), 6.70 (2H, s), 6.71 (1H, t, J = 5.4), 7.2 (1H, t, J = 5.4); 7.7 (1H, d, J = 6.4). ^{13}C NMR (75 MHz; d_6-DMSO) δ/ppm 28.1, 28.2, 38.6, 39.4, 53.7, 54.0, 80.8, 81.0, 81.2, 134.3, 170.5, 170.6, 170.7, 171.0, 173.5.

4.3.6 Synthesis of mal-*bis*NTA (17)



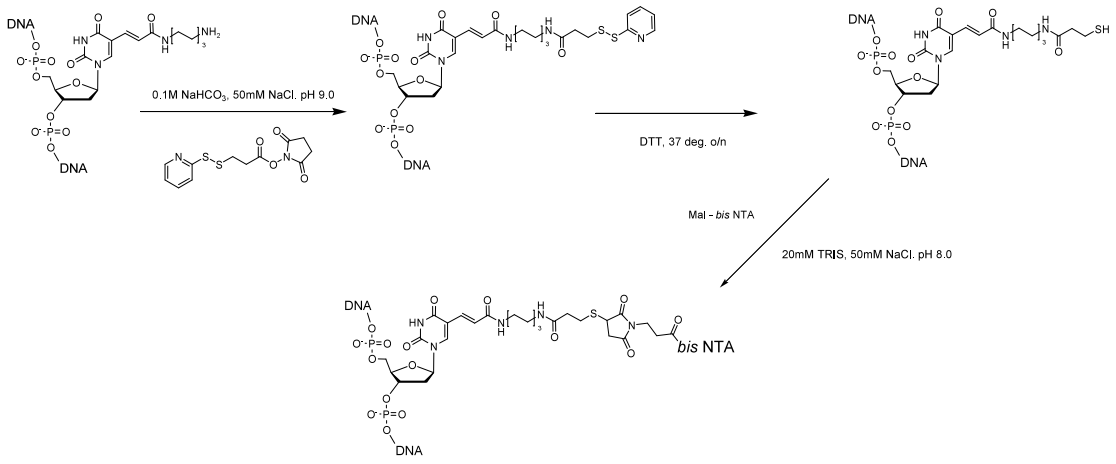
(17)

2-(bis-carboxymethyl-amino)-6-{4-[5-(bis-carboxymethyl-amino)-5-carboxy-pentyl carbamoyl]-4-[*N*-(2-carboxyethyl)male- imide]-butyrylamino}-hexanoic acid **17** was synthesized by dissolving compound **16** (200 mg) in 5 mL anhydrous DCM and 5 mL TFA. After stirring the solution magnetically for 5 hrs the solvent was removed under reduced pressure. The resulting oil was purified via preparative RP-HPLC using a (A) water (0.1% TFA) : (B) acetonitrile (0.1% TFA) gradient of 2 – 30% B over 20mins. The fractions containing the desired peak were pooled, the solvent removed under reduced pressure and the resulting product dried under high vacuum to yield the titled compound as a white solid (98.8 mg, 65.5%). Calculated mass ($C_{32}H_{46}N_6O_{17}$) 786.29 g.mol^{-1}; observed mass (ESI$^+$) 787.50 (MH^+) g.mol^{-1}, 809.50 ($M+Na^+$) g.mol^{-1}. ^{1}H NMR (500 MHz; d_6-DMSO) δ/ppm 1.36 (8H, m), 1.55 (2H, m), 1.63 (3H, m), 1.80 (1H, m), 2.01 (2H, m), 2.38 (2H, t, J = 7.0), 2.98 (4H, m), 3.40 (2H, td, J = 7.3, 3.5), 3.57 (10H, m), 4.09 (1H, td, J = 8.2, 5.7), 6.97 (2H, s), 7.74 (1H, t, J = 5.7), 7.80 (1H, t, J = 5.7), 8.06 (1H, d, J = 7.9), 12.29 (6H, bs); ^{13}C NMR (125 MHz; d_6-DMSO) δ/ppm 23.2, 26.9, 27.0, 27.6, 28.2, 32.3, 34.6, 34.7, 38.5, 39.0, 39.1, 53.8, 54.6, 67.4, 134.8, 169.6, 169.7, 171.1, 172.9, 173.3, 173.8, 175.0

4.3.7 Coupling of Mal-*bis*NTA to SPDP-Activated DNA

Compound **17** was conjugated to an oligonucleotide which contained an internal thymine base carrying a non-biogenic amino linker. The conjugation is outlined in Scheme 4.4. Briefly, the oligonucleotide was first activated using the heterobifunctional cross-linker SPDP and then reduced using a DTT solution to

generate a free thiol group, which was subsequently reacted with mal-*bis*NTA **17** to yield the DNA-conjugate.



Scheme 4.4. Conjugation of the mal-*bis*NTA compound (17) to an oligonucleotide carrying an internal non-biogenic thymine.

25 nmol of oligonucleotide O1 carrying an internal amine modification (1 mM; 25 µL; 5′ – TCT CTA AAA AA**T** ATA TAA AAA – 3′, amine modified base bold and underlined) (IDT-DNA, Coralville, IO) was diluted in 65 µL of 0.1 M $NaHCO_3$, 50 mM NaCl, pH 9.0 and mixed with 10 µL of a 0.25 M DMSO solution of *N*-succinimidyl 3-(2-pyridyldithio)-propanoate (SPDP). After incubation the solution at RT for 30 min, unreacted SPDP was removed by size exclusion chromatography using an AKTA purification system, a Sephadex column with a volume of 5 mL and a sample loop of 100 µL. The mobile phase was buffer A (20 mM Tris, 50 mM NaCl, pH 8.0). An isocratic gradient at a flow rate of 5 mL/min was applied to collect the DNA containing fractions. The eluent was mixed with 85 µL of a 0.3 M aqueous solution of dithiothreitol (DTT) and incubated at 37 °C overnight. Excess DTT was removed by size exclusion chromatography using the same conditions as detailed above. The volume of the eluent was reduced to approx. 500 µL using spin columns with a molecular weight cut-off (MWCO) of 3 kD. The reduced oligo solution was mixed with 12.5 µL of a 0.05 M aqueous solution of the mal-*bis*NTA **17** and incubated under agitation at RT for 30 min. Anion exchange chromatography (AEC) was conducted on an AKTA purifier system using a Resource Q column with a volume of 1 mL. A linear gradient was employed beginning with buffer B (20 mM

Tris, pH 8.0) to 60:40 buffer B / buffer C (buffer B plus 2 M NaCl) over 45 min at a flow rate of 1 mL/min. The oligonucleotide was then purified using a 15 – 30 % gradient buffer B / buffer C over 100 column volumes (CV) and the fractions containing the desired peak collected and combined. The solution was desalted and concentrated through a 5 ml Amicon desalting column (cat # UFC800296) with a MWCO of 3 kD. The purity of the conjugate was confirmed via analytical AEC (product eluted after 21.3 min at 96% purity), and the concentration was determined using UV/vis spectroscopy (Cary Eclipse, Varian). For analysis via MALDI MS, the oligonucleotide was first desalted via aspiration through C_{18} ZipTip pipette tips (Millipore, cat # ZTC18S096); the tips were washed with both 50% acetonitrile in distilled water and 0.1 M triethylammonium bicarbonate pH 8.0 buffer. The oligonucleotide sample was aspirated through the tips, which were then washed using distilled water followed by the 0.1M TEAB buffer. Finally the strands were eluted using the 50% acetonitrile solution and the desalted DNA solution was mixed with matrix solution containing 2',4',6'-trihydroxyacetophenone (THAP) (4 μL, 50 mg/mL in methanol) and 1 μL of the co-matrix ammonium acetate (0.1 M). The mixture was spotted onto the MALDI plate, and the sample was allowed to crystallize before being analyzed via MALDI MS using a Waters Micro MX machine; calculated mass - 7521.5 m/z, observed mass - 7521.8 m/z.

4.3.8 Generation of the His_6-Modified Oligonucleotide

The scheme to generate a His_6-modified oligonucleotide comprised the activation of an oligonucleotide carrying an internal amine by reaction with heterobifunctional cross-linker SPDP, followed by the coupling of the thiol-reactive portion of the tethered cross-linker with the peptide $CysHis_6$.

25 nmol of O1 oligonucleotide carrying an internal amine modification (1 mM; 25 μL; 5' – TCT CTA AAA AA**T** ATA TAA AAA – 3') (IDT-DNA, Coralville, IO) was activated with SPDP and purified from excess reagent as described in section 4.3.7. The eluent from the chromatographic purification was mixed with 10 μL of a 50 mM solution of the peptide $CysHis_6$ in 0.1% TFA/H_2O. The mixture was agitated at RT for 2 hours followed by purification via anion exchange chromatography on an AKTA purifier system using a Resource Q column as described in section 4.3.7.

Fractions containing the conjugate O1-His_6 were desalted and concentrated to a volume of 200 µL using ultrafiltration spin-columns (Amicon, 5 mL, MWCO of 3 kDa). The concentration of the purified conjugates was determined using UV/vis spectroscopy (Cary Eclipse, Varian), and its purity was assessed using analytical AEC as described in section 4.3.7. The purity of the His_6-tagged conjugate was also controlled via immobilized metal affinity chromatography (IMAC) on an AKTA system with a Ni-NTA containing His-Trap column (1 mL/min). The conjugate was loaded onto the column using ten column volumes of buffer D (0.1 M Na_2HPO4, 50 mM NaCl, pH 8.0) and eluted using a step gradient to buffer E (buffer D plus 1 M imidazole). The mass of conjugate O1-H was determined via MALDI MS analysis; calculated mass – 7678.5 m/z; observed mass – 7678.8 m/z.

4.3.9 Formation of DNA Duplexes for the Metal-Chelate Pull-Down Assay

Five DNA duplexes were prepared for the metal-chelate binding assay (Figure 4.3). Duplex 42 and duplex 42-H (the latter carrying two His_6 tags) both have a length of 42 bp. The remaining three duplexes had a length of 63 bp and comprised unmodified duplex 63, duplex 63B carrying a terminal biotin residue, and duplex 63B-N equipped with three *bis*NTA tags and a biotin tag. The duplexes were obtained via hybridization of corresponding single stranded DNA oligonucleotides as described in Figure 4.3 For example, duplex 42 was obtained by association of 42-mer O2 and two 21-mers O1. It is noted that the two His_6 tags in 42-H and the three *bis*NTA tags in 63B-N face the same side within the duplex as the tags were separated by 21 nt or two full helical turns in the double stranded DNA. The sequences of the oligonucleotides used for the formation of the five duplexes are summarized in Table 1.

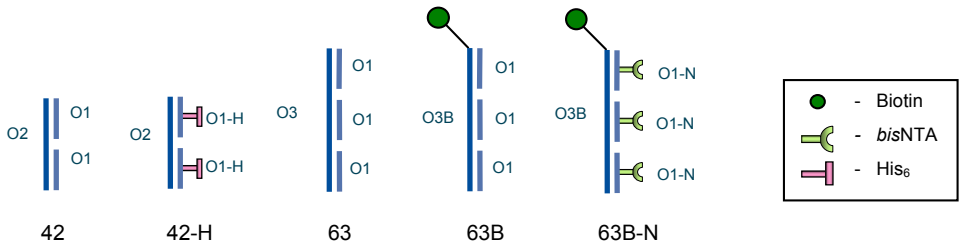


Figure 4.3. Chemically tagged oligonucleotides used for the pull-down experiments.

Oligo	Sequence/Modification
O1	5' – /5Phos/TCTCTAAAAAA**T**ATATAAAAA – 3'
O1-H	Sequence of O1 carrying a His_6 group at the internal amine-modified T
O1-N	Sequence of O1 carrying a $bisNTA_6$ group at the internal amine-modified T
O2	5' – TTTTTATATATTTTTTAGAGATTTTTATATATTTTTTAGAGA – 3'
O3	5' – TTTTTATATATTTTTTAGAGATTTTTATATATTTTTTAGAGATTTTTATATATTTTTTAGAGA – 3'
O3B	Sequence of O3 carrying a biotin group via a tri(ethylene glycol) (TEG) linker at the 5' end

Table 4.1. Summary of the DNA sequences of the oligos used for the metal-chelate assay.

For the formation of duplexes 42 and 42-H, 100 pmol of 42 nt DNA oligonucleotide O2 and either 200 pmol 21 nt oligonucleotide O1 or 200 pmol O1-H were dissolved in 5 µL 20 mM Tris, 50 mM $MgCl_2$, pH 8.0. For the generation of the 63 bp duplexes, 100 pmol of DNA oligonucleotides 63 and 63B and 300 pmol of oligos O1 or O1-N were mixed, as specified in Figure 4.3. The mixtures were thermally heated to 95 ºC for 2 min then cooled back down to 23 ºC. The samples were analyzed via non-denaturing polyacrylamide gel electrophoresis (PAGE) in TAE buffer followed by straining with Stains-all to confirm that the hybridization had gone to completion.

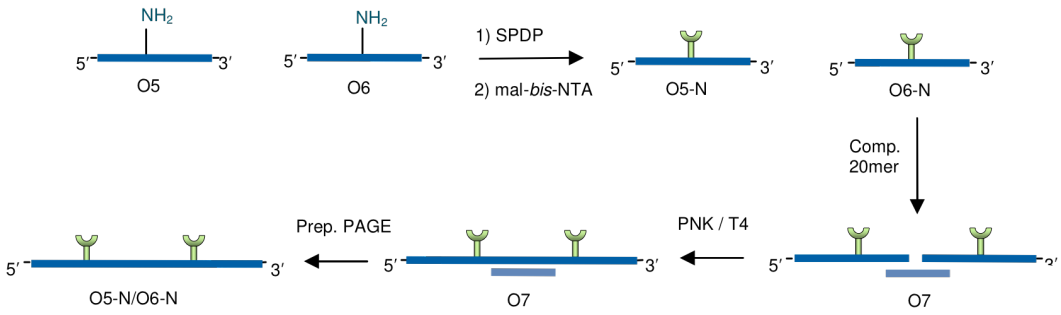
4.3.10. Metal-Chelate Pull-Down Assay

The metal chelate binding assay was conducted to test whether two *bis*NTA-Ni-His_6 metal chelate bonds can mediate the binding of two DNA duplexes. 100 pmol of the duplexes 63B-N and 42-H were combined with 0.25 µL of 2.5 M Ni^{2+} and incubated for 10 mins. Similarly, negative controls 1 and 2 were prepared by mixing 63 and 42, and 63B and 42, respectively, with Ni^{2+}. The binding of duplex 42-H to biotin-tagged 63B-N was probed using a pull-down assay with streptavidin-coated magnetic beads in combination with gel electrophoresis as a read-out. By analyzing the supernatant, the appearance or disappearance of a DNA band was used as an indicator for non-binding or binding of DNA to the beads, respectively. The pull-down assay was conducted by combining duplexes 42 & 63, 42 & 63B, and 42-H & 63N-N with 10 µL of as a suspension of streptavidin-coated magnetic beads (M-0444 Sigma) in PBS. The samples were agitated at RT for 10 min, mixed with 3 µL of 50% glycerol and analyzed via native PAGE in TAE followed by staining with ethidium-bromide.

4.3.11 Formation of Double Modified 50mers and Hybridization to Form Quad-Modified 50bp Duplexes

Using previously detailed procedures as described in section 4.3.7/4.3.8, the *bis*NTA-modified oligonucleotides O5-N, O6-N, O8-N and O9-N and His_6-tagged strands O5-H, O6-H, O8-H, and O9-H were synthesized. Briefly 25-mer oligonucleotides O5, O6, O8 and O9 (sequences in Table 4.2) were activated with SPDP and coupled to mal-bisNTA, similarly, O5, O6, O8 and O9 were coupled to the His_6 tag. The purity and chemical identity of the *bis*NTA and His_6-modified 25-mer strands was confirmed via AEC and MALDI-MS, respectively. The results of the chemical analysis are summarized in Table 4.3.

In order to form 50-mer strands O5-N/O6-N and O8-N/O9-N each carrying two *bis*NA tags, and the same sequences carrying His_6 tags, 25-mers O5-N and O6-N, and O7- N and O10-N were ligated using T4 Ligase (Scheme 4.5).



Scheme 4.5. Generation of double modified 50mers via ligation of singularly modified 25mer strands.

The enzymatic ligation was conducted with the bridging oligonucleotides O7 and O10 for O5-N/O6-N, and O8-N/O9-N, respectively. The bridging oligonucleotide O7 (sequence in Table 4.2) is complementary to 10 bases at the 3' end of oligonucleotide O5-N and 10 bases at the 5' end of oligonucleotide O6-N. Similarly, bridging oligonucleotides O10 is complementary to 10 bases at the 3' end of O8-N and 10 bases at the 5' end of O9-N. The ligation mix contained 6 nmol each of O5-N and O6-N (as an example), 50 µL of 10x T4 Ligase buffer (500 mM Tris, 100 mM $MgCl_2$, 10 mM ATP, pH 7.4) and made up to a final volume of 500 µL using

Ultrapure water (Fluka 95289). The 5'-ends of the oligonucleotides were phosphorylated by adding 5 μL of PNK (10,000 U/mL; NEB MO201S), and incubating the mixture for 1 h at 37 °C. It is noted that prior to the phosphorylation reaction, any DTT in the enzyme stock was deactivated by adding a 5 mM aqueous solution of SPDP (the activated NHS group of this compound had been hydrolyzed by incubation in water for 20 min). After the phosphorylation, 6 nmol (6 μL; 1mM) of the bridging oligonucleotide O7 was added. The DNA solution was heated to 95 °C and allowed to cool to RT in the PCR block to 23 °C before 10 μL of T4 Ligase (NEB MO202S; 400, 000 U/mL) was added and left to ligate the oligonucleotides at 16 °C overnight. It is again noted that prior to the ligation reaction, any DTT in the enzyme stock was deactivated by adding a 5 mM aqueous solution of SPDP. A similar procedure of phosophorylation and ligation was used to prepare O8-N/O9-N, and for His_6-tagged 50-mer strands O5-H/O6-H and O8-H/O9-H.

DNA strands O5-N/O9-N, O8-N/O9-N, O5-H/O9-H, O8-H/O9-H, were purified from the bridging oligonucleotides using preparative 15% PAGE in 8 M Urea and TBE. The DNA samples were prepared by reducing their volume using a spin column with a MWCO of 3 kDa. 50% volume of 8M urea was added, and the solutions were briefly heated to 95 °C. After cooling, glycerol was added and the samples were loaded onto the preparative gels (0.75 mm thickness, slot width 0.5 cm). After completing the gel electrophoretic separation, the DNA bands were visualized via UV-shadowing. The higher migrating 50 mer bands were cut out and crushed in 0.5 M ammonium acetate, frozen at -80°C and left to agitate at room temp. over-night. The solubilized DNA was separated from the crushed gel slurry by filtration using QIAgen shredder filters (79654). The remaining slurry was washed once with ammonium acetate. The combined solutions were then reduced in volume and their concentration was determined via UV/Vis. Analytical denaturing PAGE analysis of gel-purified DNA strands O5-N/O9-N and O8-N/O9-N demonstrates the successful ligation into 50-mer oligonucleotides as well as the removal of bridging oligonucleotides.

Oligo	Sequence	Modification
O5	5' – TCGAACTCCCTCAGCACAA**T**CAGTA – 3'	His_6/bisNTA
O6	5' – CCTATCAATCACCA**T**ATCTCAATAC – 3'	His_6/bisNTA
O7	5' – GATTGATAGGTACTGATTGT – 3'	n/a
O8	5' – GTATTGAGATATGG**T**GATTGATAGG – 3'	His_6/bisNTA
O9	5' – TACTGATTG**T**GCTGAGGGAGTTCGA – 3'	His_6/bisNTA
O10	5' – ACAATCAGTACCTATCAATC – 3'	n/a

Table 4.2. DNA sequences of the oligos used for the formation of DNA arrays.

Oligonucleotide	AEC [min]	Purity [%]	Mass expected [m/z]	Mass found [m/z]
O5-H	23.8	90.10	8739.2	8717.2
O5-N	24.4	82.00	8583.5	8564.4
O6-H	24.2	91.60	8673.2	8629.6
O6-N	24.4	92.64	8517.5	8496.4
O8-H	23.4	86.82	9015.3	9039.9
O8-N	23.6	89.90	8859.6	8811.4
O9-H	24.0	86.16	8952.3	8958.1
O9-N	24.0	83.3	8796.6	8793.5

Table 4.3. Summary of AEC and MALDI data for the modified oligonucleotides.

4.3.12 Melting Point Measurements (T_M)

DNA duplexes 50-N and 50-H were obtained by hybridization of single stranded DNA strands. 100 pmol of gel-purified 50-mers O5-N/O9-N and O8-N/O9-N, and O5-H/O9-H and O8-H/O9-H were diluted in 20 mM Tris, 20 mM $MgCl_2$, pH 8.0 to a final volume of 80 µL, and heated in a PCR block to 95 ºC for 2 min followed by annealing at 22 ºC for 5 min. Gel electrophoresis was conducted to confirm the formation of the duplex. The T_m measurements were made on a Cary UV/Vis system by denaturing and re-annealing using temperature ramps from 25 – 100 ºC and 100 – 25 ºC, respectively, at a ramp speed of 2 ºC min^{-1}. The melting curves for modified duplexes 50-N and 50-H, and for unmodified duplex 50. The T_m values for the chemically modified duplexes were 4°C lower than the unmodified duplex, suggesting that the tags only slightly affected the thermal stability of the duplex.

4.3.13 AFM Analysis

DNA nanoarrays were prepared by diluting 10 μL of a 2.5 μM mixture of 50-N and 50-H (1:1) into 300 μL of a freshly prepared buffer F (33 mM $NiCl_2$, 33 mM Tris pH 8). The mixture was immediately deposited onto mica surfaces. After incubation for 30 min, the surface was rinsed with a total of 2 mL of buffer F. For acquiring images in MAC Mode, a magnetically driven dynamic force microscope (Agilent Technologies, Chandler, AZ) and magnetically coated MacLevers with a nominal spring constant of 0.292 N/m (Agilent Technologies) were used. During imaging, the vertical position of the cantilever was adjusted by using an electronic feedback loop in order to keep the amplitude of the oscillating tip constant. A scanning frequency of 1.5 lines/sec was employed. To achieve gentle non-destructive imaging, the peak-to-peak oscillation amplitude was set to 5 nm at 15 kHz oscillation frequency. Images were acquired in 1 mM Ni^{2+}, 10 mM Tris pH 8.0.

4.3.14 Formation of Nanoscale DNA Rings

Rings with a diameter of approximately 12 nm were formed by ligating three highly curved duplexes α, β, and γ. The three duplexes had terminal single stranded overhangs to facilitate their ligation into a ring. The three duplexes were obtained by hybridization of two oligonucleotides each, duplex α carried a single biotin residue.

100 pmol of the two α strands (see Table 4.4 for sequence information) were combined in the buffer containing 20 mM Tris, 50 mM $MgCl_2$, pH 8.0. The mixture was heated to 95 °C for 2 min and then rapidly cooled to RT. The same procedure was conducted for the β and γ strands. After the hybridization, duplexes α and β were combined with 8.5 μL of 10 x T4 Ligase buffer and diluted up to 85 μL with Ultrapure water. 0.5 μL of T4 Ligase was added, and the mixtures were incubated at RT for 30 min. The γ duplex was then added along with 1.5 μL of the 10 x ligase buffer and a further 0.5 μL of T4 Ligase. This solution was again allowed to incubate at room temp for 30 min. 25 μL of the solution was analyzed by a 15% TBE PAGE.

To investigate the ability of the DNA rings to bind to streptavidin, samples of the duplex constructs α, ligated α/β, and α/β/γ were incubated with a 1 mg/ml solution of streptavidin at RT and analyzed using 15% TBE PAGE.

Name	Sequence
α	5'/5Phos/TCTCTAAAAAATATATAAAAATCTCTAAAAAATATATAAAAA/iBiodT/CTCTAAAAAATAT-3'
	5' – /5Phos/TTTATATATTTTTTAGAGATTTTTATATATTTTTTAGAGATTTTTATATATTTTT -3'
β	5' - /5Phos/ATAAAAATCTCTAAAAAATATATAAAAATCTCTAAAAAATATATAAAAATCTCTAA – 3'
	5'-/5Phos/ATTTTTTAGAGATTTTTATATATTTTTTAGAGATTTTTATATATTTTTTAGAGATT-3'
γ	5'/5Phos/AAAATATATAAAAA/iBiodT/CTCTAATAATTACTATCATGTCTGCGAAAAATATATAAAAA-3'
	5'- /5Phos/AGAGATTTTTATATATTTTTCGCAGACATGATAGTAATTATTAGAGATTTTTATAT -3'

Table 4.4. Sequences of the oligonucleotides used to form the DNA rings.

4.4 Reversible Ni(II) Mediated Binding of DNA Helices.

A proof-of-principle experiment was conducted to demonstrate that two chemically tagged duplexes could interact via metal ion complexation. A pull-down assay based on the binding of biotinylated (BT) duplexes, carrying the appropriate NTA tags, with Streptavidin magnetic beads was used. Conjugation of the BT duplexes with the beads and the Ni(II) binding of the NTA groups to a smaller duplex carrying His tags clearly illustrated that the whole complex could be bound and removed from solution.

4.4.1 Metal-Chelate Assay on the Formation of a Dimer of Duplexes

The duplexes, 42-H and 63B-N, used for this experiment are shown in Figure 4.4A. The first duplex was 42 bp long and carried two His_6 tags, while the second, 63 bp-long duplex was equipped with three *bis*NTA groups. Based on rational design, His_6 and *bis*NTA tags were positioned on one side of the duplexes in order to facilitate the creation of heterodimers.

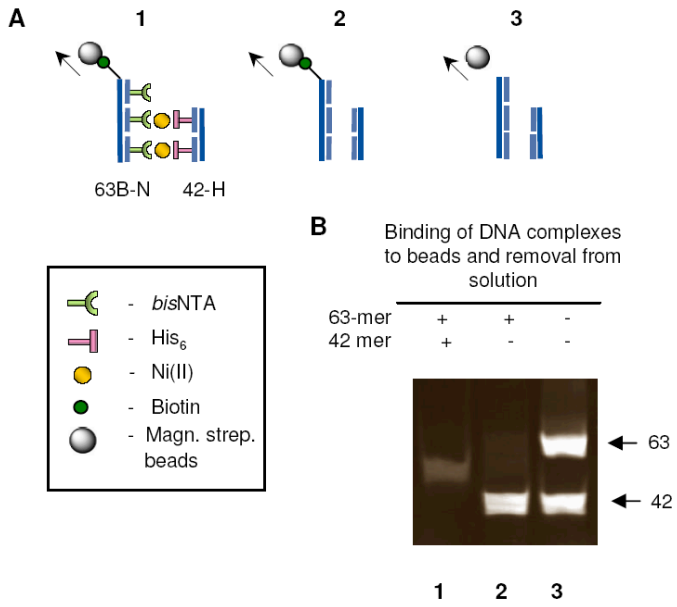
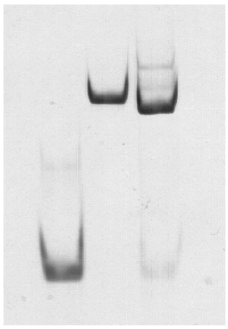


Figure 4.4. Metal-chelate pull-down assay; the successful binding of DNA duplexes is demonstrated by the removal of a BT-63mer duplex conjugated to streptavidin-coated magnetic beads.

Duplex 42-H was generated by hybridizing a 42 nt long single stranded DNA oligonucleotide to two 21-mer oligonucleotide O1-H carrying a His_6 tags. Similarly, 63B-N was prepared by association of a 63 nt long template carrying a 5' tri(ethylene glycol)-linked biotin group and three 21-mer oligonucleotides O1-N carrying a *bis*NTA tag (see section 4.3.9). The successful synthesis of the single-stranded bioconjugates O1-H and O1-N was confirmed by ion exchange chromatography and mass spectrometry (Table 4.3). Further analysis by gel electrophoresis provided evidence that duplexes 42-H and 63B-N had successfully formed (Figure 4.5).



1 2 3

Figure 4.5. 15% TBE PAGE analysis of (1) single stranded 21-mer O1-N; (2) the single stranded 63-mer used to form (3) the 63 bp duplex by mixing with a 3-fold molar excess of O1-N.

The proof-of-principle experiment illustrating dimer formation is schematically summarized in Figure 4.4A and is based on a heterodimer forming between the *bis*NTA and His_6-tagged DNA in the presence of Ni(II). In this assay, the duplex carrying the biotin and NTA tags (63B-N) is bound to streptavidin-coated magnetic beads. Addition of the 42 bp duplex carrying the His_6 tags in the presence of Ni(II) should lead to the formation of the Ni-chelate complex and the binding of 42-H to the beads. Removal of the magnetic beads should therefore deplete the solution of 42-H, and gel electrophoresis of the cleared supernatant can be used to follow the depletion of 42-H.

The metal-chelate assay was performed by first confirming the ability of streptavidin to capture the biotinylated DNA. As shown by gel electrophoresis analysis of the cleared supernatant, non-biotinylated duplex 63 migrated as a defined band (Figure 4.4B, lane 3), while biotinylated 63B was depleted due to capture by the beads (Figure 4.4B, lane 2). Gel electrophoretic analysis also demonstrated that the 63-mer and 42-mer duplex do not interact in the absence of *bis*NTA and His_6 tags because the non-tagged 42 duplex migrated in the gel (Figure 4.4B, lane 2). By contrast, when the fully tagged duplexes 63B-N and 42-H were incubated with beads in the presence of Ni^{2+}, the intensity of the 42-H band was reduced by over 95% (Figure 4.4B, lane 1) suggesting successful formation of a heterodimer via two *bis*NTA-Ni^{2+}-His_6 complexes (Figure 4.4A – 1).

4.5 Formation of 2D Arrays

The convincing results of the pull-down assay prompted the construction of larger supramolecular structure as illustrated in Figure 4.6. For the design of the nanoarray, two different DNA duplexes 50-N and 50-H were generated. The two 50-mers contain either four *bis*NTA or four His_6 tags, respectively (Figure 4.6). The duplexes were obtained from smaller units in a multi-step procedure. This method involved coupling the chemical tags to 25 nt-long single stranded DNA oligonucleotides, followed by the ligation of each two 25-mers to 50-mer strands carrying two chemical tags and the final hybridization of two complementary strands to yield duplexes with four tags each.

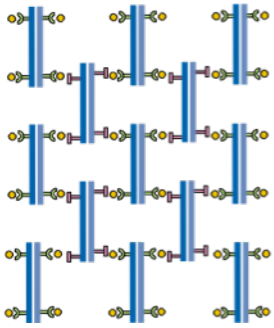


Figure 4.6. Schematic of a DNA nanoarray formed via Ni(II) chelation of 50 bp duplexes carrying either four His_6 or four *bis*NTA tags.

4.5.1 Design of Arrays and DNA Sequences

To ensure that a 2D array is able to form the tags within one duplex must be directly opposite one another (Fig. 4.6). If any of the tags are placed at angles other than 180° from one another then the assembly of multiple duplexes would not be planar but an undesired warped higher-order structure.

One helical turn of a DNA duplex occurs approx. every 10 bases, therefore the internal chemical modification within a DNA oligonucleotide was designed to be placed 20 bases apart. To improve the accuracy in determining the optimal base position, an X-ray structure of a DNA helix composed of two complementary strands of the sequences 5'-CGCAAATTTGCG-3' and 5'-CGCAAATTTGCG-3' was used as

a model (Figure 4.7). Inspection of the duplex revealed that the 5 methyl group on one thymine base is 180° from the methyl group on another thymine four bases downstream on the complementary strand. Given that the chemical tags are to be linked to position 5 of the pyrimidine base, the pattern of two thymines separated by four bases was used as the basis for the design of the 50-mer nanoarray strands (Table 4.5).

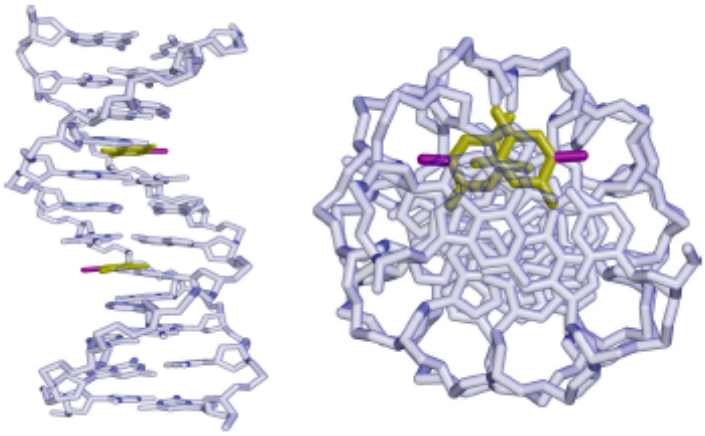


Figure 4.7. Molecular model of a double helix illustrating a thymine on each strand in yellow with the methyl group at the 5 position in magenta. If tags are positioned in place of these methyl groups they should protrude at 180° from one another.

5' – TCG AAC TCC CTC AGC ACA A**T**C AGT ACC TAT CAA TCA CCA **T**AT CTC AAT AC – 3'
3' – AGC TTG AGG GAG TCG **T**GT TAG TCA TGG ATA GTT AG**T** GGT ATA GAG TTA TG – 5

Table 4.5. The sequence of the designed 50 mers used for the DNA nanoarray. The position of the thymines carrying an internal amine modification are highlighted.

4.5.2 Formation of Double Modified 50-mers for the DNA Nanoarray

The 50 bp duplexes 50-N and 50-H, carrying four *bis*NTA or His_6 modifications, respectively (Figure 4.8), were generated via a convergent procedure as previously discussed (4.3.11). The *bis*NTA compound and $CysHis_6$ peptide were first synthesised separately and conjugated to 25-mer oligonucleotides yielding His_6–tagged O5-H, O6-H, O8-H and O9-H, and *bis*NTA-tagged O5-N, O6-N, O7-N and O8-N. Chemical

analysis by MS and AEC confirmed the successful coupling of the tags to the single-stranded DNA (Table 4.5). Each two of the modified oligonucleotides were then ligated into a single stranded 50-mer using a bridging oligonucleotide O7 and O10 as shown in Fig. 4.8. The enzyme-mediated ligation products carrying two chemical tags were purified via denaturing PAGE.

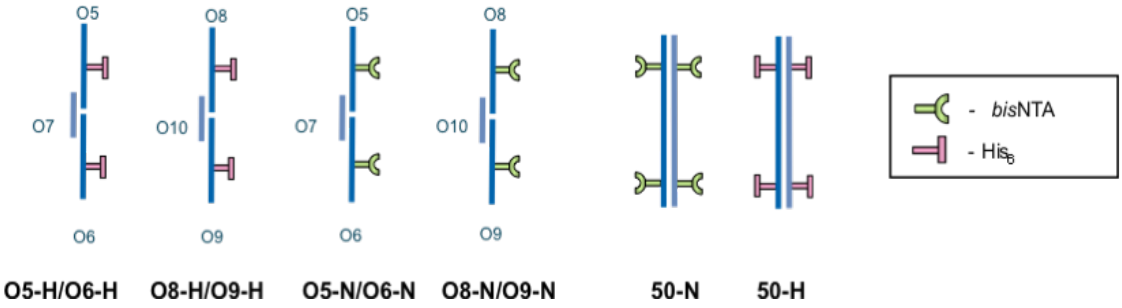


Figure 4.8. Nomenclature of oligonucleotides used for the formation of DNA arrays. The sequences are provided in Table 4.4.

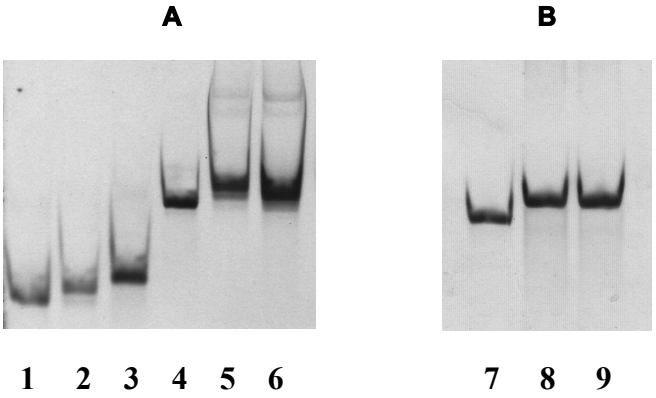


Figure 4.9. (A) Gel illustrating 25mer modification and 50 mer ligation; Lane 1) unmodified 25mer; Lane 2) His_6 modified 25mer; Lane 3) *bis* NTA modified 25mer; Lane 4) unmodified 50mer; Lane 5) 2x His6 modified 50mer; Lane 6) 2x bisNTA modified 50mer. Gel (B); 50bp hybridisation; Lane 1) 50 mer A; Lane 2) 50 mer B; Lane 3) 50bp duplex.

As double modified strands O5-H/O6-H and O8-H/O9-H (Fig. 4.8) are complementary, hybridization of the two 50-mers was used to generate His_6-tagged target duplex 50-H. Similarly, the *bis*NTA-tagged target duplex 50-N was obtained by hybridization of O5-N/O6-N and O8-N/O9-N. Their formation was followed by gel electrophoresis and thermal denaturation profiles. A change in the thermal stability

can be expected as the modifications are placed within internal positions of the duplex. The melting temperature (T_m) derived from the melting profiles (Fig. 4.10) of both modified duplexes was slightly lower (75 ºC) than the T_m of unmodified duplex (79 ºC) (Table 4.6). This illustrates successful duplex formation and suggests that the tags only weakly affected the stability of the duplex.

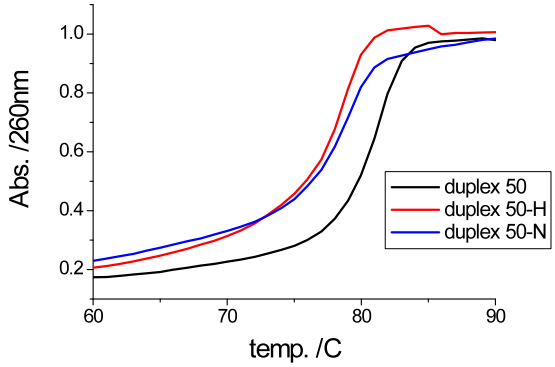


Figure 4.10. Melting profiles of unmodified 50-mer (black); His_6-modified 50-H (red); and *bis*NTA-modified 50-N (blue).

Duplex	Tm [ºC]
50	79.6 ± 0.3
50-H	75.3 ± 0.0
50-N	75.6 ± 0.6

Table 4.6. T_M data for the modified 50-mers.

4.6 Formation and AFM Analysis of DNA-Based Arrays

The formation of the DNA nanoarray was confirmed by dynamic force atomic force microscopy (AFM). DNA arrays were generated on mica surfaces by incubating a 1:1 mixture of 50-H and 50-N duplexes in Ni^{2+} containing buffer followed by removal of unbound DNA by gently rinsing with buffer. The incubated mica surface (Figure 4.11) showed ~ 100 nm long strands which were aligned parallel into a regular 2D sheet. The height of the supramolecular structures (Figure 4.11 inset) was 1.35 ± 0.15 nm, similar to the diameter of DNA duplexes in other AFM studies. Furthermore, the

distance between the parallel DNA strands was 3.0 ± 0.2 nm. Importantly, no nanostructures were observed in the absence of Ni^{2+} or with DNA duplexes lacking the chemical tags. Taken together, these data strongly support the interpretation of the parallel structures as a nanoarray composed of interlinked DNA duplexes as schematically illustrated in Figure 4.6.

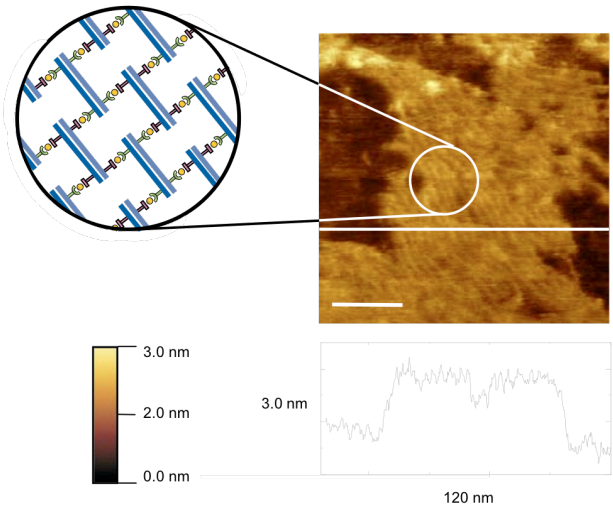


Figure 4.11. AFM image of DNA based arrays.

4.7 Formation of Circular DNA Rings

The successful use of the *bis*NTA–Ni(II)–His_6 complex in DNA-nanobiotechnology prompted us to investigate whether the complex can be exploited to create other DNA-based supramolecular structures from simple building blocks. One such block is a DNA loop or ring. It is known that biological DNA can have a highly curved secondary structure which is important for the packaging of genetic material into viral capsides or around proteins such as histones. Previous research has utilized this natural design principle to form DNA loops from smaller highly curved duplexes.[100,101,102] It is conceivable that chemical functionalization of the duplexes could help create additional structures. By chemically modifying oligonucleotides with the *bis*NTA and His_6 tags and forming the ring structures, individual units could have the ability to bind via Ni(II)-mediated complexation leading to the formation of stacks of RNA rings (Figure 4.12).

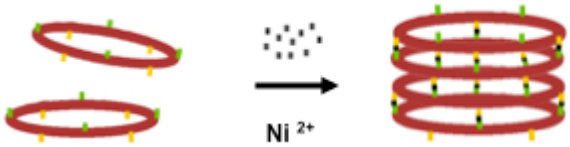


Figure 4.12. Formation of DNA rings with tags able to bind reversibly in the presence of Ni(II) ions.

4.7.1 Formation of DNA Rings

Initially it was attempted to form the DNA circles using 21 bp duplex monomers. It has been previously reported[100] that these 21 bp units can be annealed to form rings of varying diameters. However, no structures could be formed using these sequences. Alternative duplexes[101] were utilized successfully to form the rings. Three 63 bp duplexes (α, β, and γ) were formed by hybridization of two complementary single stranded DNA oligonucleotides (see Table 4.2 for sequence information). The α duplex (seen in lane 2 in Figure 4.14) was then hybridized to the β duplex via designed sticky ends and followed by formation of a covalent linked using T4 Ligase reaction. The ligation was partly successful as inferred from a higher migrating DNA band (Fig. 4.13, lane 3, α/β). Addition of the third duplex γ and ligase reaction led to the formation of the even higher migrating band, likely composed of covalently linked α/β/γ (Fig. 4.13, lane 4).

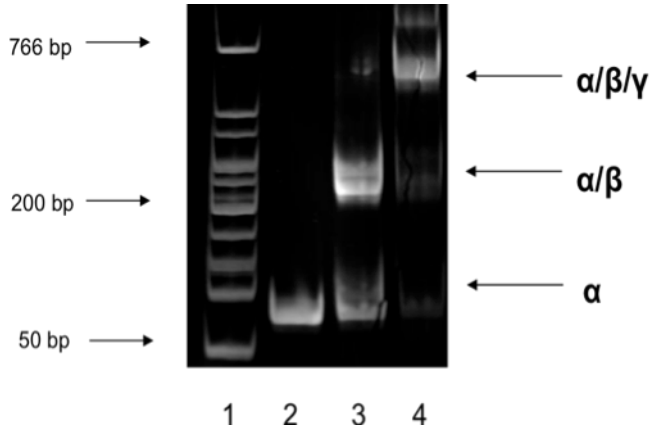


Figure 4.13. Stepwise formation of DNA rings; (1) DNA ladder; (2) alpha duplex; (3) alpha-beta ligated duplex; (4) DNA loop.

It is noted that not all of the α and β duplexes seem to initially hybridize as an excess of 63 bp duplex is seen in lane 3. The lower excess band and mid α/β duplex band, however, disappear upon ligation with γ in lane 4. This is possibly due to the formation of longer by-products that are seen in the well of the gel. In the gel illustrated in figure 4.13, a good yield of the ring product was observed. This was not repeatable when the conditions and samples were exactly replicated. Several parameters were varied to improve the yield, such as dilution of the sample to avoid intermolecular ligation, favoring intramolecular, and a decrease in the amount of γ as a slight excess of the γ duplex may have caused the undesired formation of the long by-products to form. However, none of these approaches proved successful.

4.7.2 AFM Analysis

The DNA ring samples were analyzed by a collaborating AFM specialist[129] with the aim to confirm the formation of DNA rings in the low-to-average yielding preparations. Indeed, mostly long linear strands were observed (Figure 4.14). However, fully circularized structures were also present (Figure 4.14; inset) indicating at least partially successful formation of the rings. The images of the covalently closed and open structures are currently being analyzed and the DNA observed characterized using their curvature. As the curvature occurs for DNA molecules of specific sequence, a comparison between circular and linear structures can shed led light on the influence of sequence variation and understand the biophysical basis of circular DNA found in nucleosomes and replication sequences.

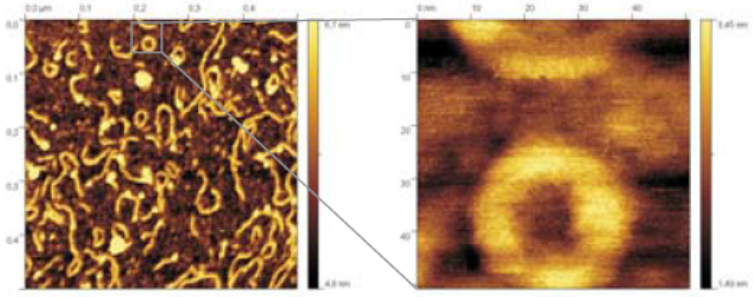


Figure 4.14. AFM images of DNA ring as well as non-ligated loops and ligated concatemers adsorbed on mica.

4.8 Possible Future Applications of Metal-Chelate Bridges

In conclusion, following rational design the covalent attachment of His_6 and *bis*NTA tags was shown to endow DNA duplexes with the ability to self-assemble via reversible chelate bonds. This work represents a novel way to form nanoscale structures which do not require the redesign of underlying structural framework of the constituent monomeric units. We expect that the new approach can be extended to protein arrays with defined functional properties. In general, the ability to engineer self-assembly into monomeric peptides or proteins is an emerging topic in supramolecular chemistry, nanobiotechnology and synthetic biology due to the potential applications of the protein and peptide networks in basic science and technology. By covalently attaching a small tag onto a single position within a biomolecule, this new approach has the potential to address the shortcomings of existing approaches towards imparting self-assembly. In particular, while current approaches use the re-engineering of whole protein surfaces or the bottom-up design of linear peptides, the new generic approach is independent of reengineering the structural framework of the subunits.

Publications from PhD Research

Chemical Tags Facilitate the Sensing of Individual DNA Strands with Nanopores; Nick Mitchell, Stefan Howorka; *Angew. Chem. Int. Ed.;* **2008**; 47; 5565 – 5568.

A DNA Nanostructure for the Functional Assembly of Chemical Groups with Tunable Stoichiometry and Defined Nanoscale Geometry; Nick Mitchell, Robert Schlapak, Markus Kastner, David Armitage, Wojciech Chrzanowski, Johannes Riener, Peter Hinterdorfer, Andreas Ebner, Stefan Howorka; *Angew. Chem. Int. Ed.;* **2009**; 48; 525 – 527.

Chemically Labeled Nucleotides and Oligonucleotides Encode DNA for Sensing with Nanopores; Vinciane Borsenberger, Nick Mitchell and Stefan Howorka; *J. Am. Chem. Soc.*; ***2009;*** 131 (22); 7530 – 7531.

Single-Molecule Sensing

DOI: 10.1002/anie.200800183

Chemical Tags Facilitate the Sensing of Individual DNA Strands with Nanopores**

*Nick Mitchell and Stefan Howorka**

Nanopore recording is an electrical analytical technique in which individual molecules block a nanometer-scale pore and cause detectable modulations in ionic current.[1–5] The single-molecule approach has been exploited to analyze proteins, toxins, metal ions, drug molecules, and single-point mutations in double-stranded DNA.[6–11] Despite progress in the sensing of isolated nucleotides[12] and single base positions in static DNA strands,[13] it has, so far, not been possible to detect multiple bases in an individual strand. One of the main technical hurdles towards this aim is the high speed at which the voltage-driven DNA strands pass through the pore; this leads to insufficient analytical resolution. Herein, we present a new approach that slows down single-stranded DNA (ssDNA) and enables the detection of multiple separate bases. We show that chemical tags attached to bases cause a steric blockade each time a modified base passes through a narrow pore. The resulting characteristic current signatures are specific for the chemical composition and the size of the tags. Our approach for detection with modified DNA is independent of pore engineering and can potentially be applied to a wide range of solid-state nanopores to extend their sensing repertoire.[2,3,14] This is a novel strategy because the detection is facilitated by the chemical modification of the analyte molecules rather than the engineering of pores.

This new approach for the base-specific identification of DNA was tested with the nonengineered version of the protein pore α-hemolysin (αHL; Figure 1A). The αHL pore has been widely used in the past for the sensing of unmodified RNA and DNA strands.[1,15] It therefore constitutes a good reference point for sensing with chemically modified DNA. αHL is of defined architecture with a lumen diameter of 1.3 nm at the narrow inner constriction and of 2 nm in the β barrel at the trans side of the pore (Figure 1A).[16] ssDNA

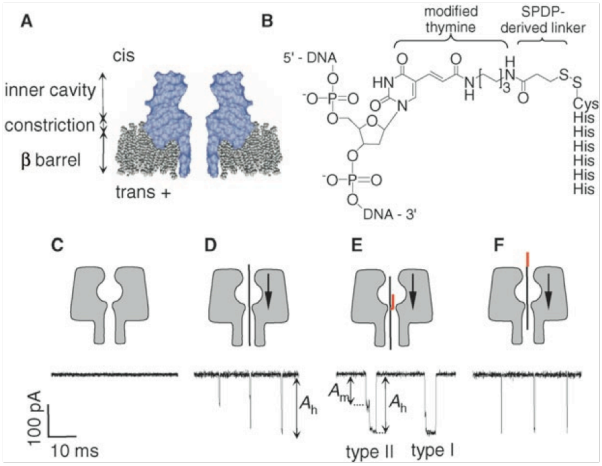


Figure 1. A) The α-hemolysin (αHL) pore embedded in a lipid bilayer; "+": positive potential. B) The chemical linkage between DNA and the peptide within construct H_6C_1–O1. C) Schematic representation of the αHL pore and a representative single-channel current trace. D) Events caused by oligonucleotide O1 without a peptide tag. E) Trace for the translocation of H_6C_1–O1. F) Events for H_6C_1–O1-term. (carrying the H_6C_1 tag at a terminal rather than an internal position). The traces were obtained from recordings in 2 M KCl and 20 mM tris(hydroxymethyl)aminomethane (Tris, pH 8.0), filtered and sampled at 10 and 50 kHz, respectively.

with an average cross-sectional diameter of 0.9–1.2 nm[17] is known to pass the inner constriction.[1,18] We speculated that chemical tags attached to separate bases would increase the cross-sectional diameter of the DNA and hence slow down translocation each time a modified base passes the narrow pore constriction. To test the approach, a model system consisting of commercially available DNA oligonucleotides and tags composed of peptides was used. Peptides were selected because their size, length, charge, and hydrophobicity can be easily tuned in a modular fashion to optimize pore blockades.

We first confirmed that a single peptide tag is capable of retarding strand translocation. DNA-strand oligonucleotide O1, of 27 bases in length, was modified with the hexahistidine tag H_6C_1 at an internal base (Figure 1B; see the Supporting Information). The resulting peptide–DNA conjugate, H_6C_1–O1, was analyzed in nanopore recordings. In the absence of DNA, the wild-type αHL pore yielded a conductance of (1900 ± 120) pS (number of independent recordings $n = 4$) when a positive potential was applied at the trans side (Figure 1C and A). The addition of unmodified oligonucleotide O1 to the cis side of the pore (Figure 1A) led to short high-amplitude events (Figure 1D). These were characterized by an average amplitude of (91.7 ± 1.1)% relative to the

[*] N. Mitchell, Dr. S. Howorka
Department of Chemistry
University College London
20 Gordon Street, London WC1H OAJ (UK)
Fax: (+44) 20-7679-7463
E-mail: s.howorka@ucl.ac.uk

[**] Funded by the Engineering and Physical Sciences Research Council and University College London Chemistry. We thank A. B. Tabor for use of a peptide synthesizer, H. Bayley for providing recording equipment, and H. Martin for assistance with molecular modeling. N.M. holds a Scholarship from the UCL Graduate School.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200800183. It contains details of the synthesis of peptide-modified DNA strands, details of nanopore recordings, representative current events, additional data on the event analysis, and a discussion on the combined use of chemically modified DNA and solid-state nanopores.

open-pore current and had a duration, τ_{off}, of (0.18 ± 0.06) ms ($n = 3$). The short events represent the fast translocation of individual strands from the cis to the trans side of the pore (Figure 1D).[1,18] The recordings also indicated blockages with 50% amplitude, which were not pursued further as they likely represent the reversible threading of a strand into and the escape from the cis opening rather than complete translocation to the trans side.

When modified DNA strand H_6C_1–O1 was analyzed, events of two different types were observed. Type I events (Figure 1E) had a high amplitude, A_h, of (96.8 ± 0.5)% with an average duration, τ_{off-h}, of (1.83 ± 0.26) ms. Due to the nature of this very defined blockage, type I events certainly represent complete pore translocation. By comparison, type II events (Figure 1E) started with a medium amplitude, A_m, of (56.6 ± 2.6)% with a duration, τ_{off-m}, of (1.34 ± 0.36) ms. This medium level was followed (Figure 1E) by a high-amplitude blockage of (97.4 ± 0.9)% with a duration of (1.96 ± 0.48) ms. The medium-amplitude blockage of type II events possibly stems from misfolded strands that reside in the internal cavity but eventually thread into the inner constriction. Due to the uncertainty in the assignment of the medium-level blockage, we focused our further investigations on the more clearly defined type I events, which only exhibited high-amplitude blockages. A comparison of type I events from H_6C_1–O1 with the results obtained with unmodified DNA shows that the histidine tag slowed down translocation by a factor of 10 and increased the current amplitude by 5%.

Several lines of evidence support the notion that the current blockages with the histidine-modified strand are caused by the steric hindrance encountered when a wide peptide–DNA segment passes the narrow inner constriction (Figure 1E). First, histidine tags with six, four, or two residues led to correspondingly shorter high-amplitude blockages in type I events (Table 1, H_xC_1–O1, $x = 6$, 4, or 2; for type II events, see the Supporting Information). This implies that the peptide is elongated and aligned parallel to the DNA strand while being translocated. Second, tags composed of less-bulky glycine residues did not exhibit the same length dependence, which indicates that the smaller amino acid does not reach the critical size threshold required for continually slowing down DNA (type I events: Table 1, G_xC_1–O1, $x = 6$, 4, or 2; for type II events, see the Supporting Information). Third, oligonucleotide O1 carrying an H_6C_1 tag at a terminal rather than an internal position did not greatly retard the DNA passage, as shown by the short event time of (0.23 ± 0.10) ms (Figure 1F; Table 1, H_6C_1–O1-term.). The absence of major retardation is attributed to the fact that the peptide can sequentially pass through the pore after the DNA strand, without the formation of a bulky peptide–DNA segment.[19] The peptide tags are certainly the molecular reason for the retardation and may exert their effect by either hindered diffusion or an increase in friction[20,21] mediated by steric, electrostatic, polar, and/or hydrophobic interactions.

Additional peptides were examined to demonstrate that strand retardation is a general feature of bulky amino acids and not only restricted to histidine residues. An additional aim was to identify tags that give rise to current signatures that are distinguishable from the histidine blockages. Two different peptides were investigated. The first peptide, R_7C_1, was composed of seven arginine residues. In the nanopore analysis, type I events of R_7C_1–O1 (Figure 2A) exhibited a duration of (25 ± 5) ms and an amplitude of (98.9 ± 0.6)% (Table 1, R_7C_1–O1). This blockage has higher amplitude and is longer than that of H_6C_1–O1 (Table 1). The more pronounced blockade of R_7C_1–O1 is possibly due to the longer amino acid side chain of arginine compared to that of histidine (Figure 2E) or to the folding back of the positively charged arginine onto the negatively charged DNA backbone to generate a compact and bulkier DNA–peptide segment. The second peptide investigated was Y_3C_1. Tyrosine has an uncharged aromatic side chain (Figure 2E). Translocation of

Table 1: Characteristics of type I translocation events of DNA carrying a single chemical tag.[a]

Modified DNA	A_h [%][b]	τ_{off-h} [ms][c]
O1[d]	91.7 ± 1.1	0.18 ± 0.06
H_6C_1–O1	96.8 ± 0.5	1.83 ± 0.26
H_4C_1–O1	96.0 ± 0.6	1.57 ± 0.29
H_2C_1–O1	93.0 ± 0.7	0.82 ± 0.16
G_6C_1–O1	91.4 ± 0.6	0.56 ± 0.15
G_4C_1–O1	92.5 ± 0.5	0.55 ± 0.12
G_2C_1–O1	92.7 ± 0.4	0.53 ± 0.13
H_6C–O1-term.[d]	93.9 ± 0.7	0.23 ± 0.10
H_6C_1–O1/pH 6.4	96.9 ± 0.7	2.18 ± 0.42
R_7C_1–O1	98.9 ± 0.6	25 ± 5
Y_3C_1–O1/step	$92.4 \pm 0.6/97.8 \pm 0.5$	$0.46 \pm 0.15/0.35 \pm 0.13$
Y_3C_1–O1/slope	94.8 ± 0.7	1.00 ± 0.22
Y_3C_1–O2/step	$91.8 \pm 0.6/98.9 \pm 0.7$	$0.43 \pm 0.12/0.39 \pm 0.08$
Y_3C_1–O2/slope	96.7 ± 0.8	0.97 ± 0.13

[a] The recordings were conducted in 2 M KCl and 20 mM Tris (pH 8.0), filtered at 10 kHz, and sampled at 50 kHz unless stated otherwise. The number of events analyzed for each type of DNA was between 1500 and 2000. $n = 3$. [b] The relative amplitude was calculated by using $A = (I_{OC} - I_{BC})/I_{OC}$, in which I_{OC} and I_{BC} are the conductance levels from the open and blocked channel, respectively. I_{OC} and I_{BC} were derived by using all-point histograms. [c] The average duration represents the mono-exponential fit of the dwell-time histogram. [d] Filtered at 30 kHz and sampled at 100 kHz.

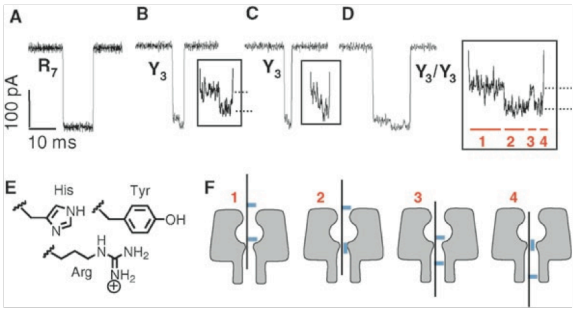


Figure 2. Representative nanopore translocation events for A) R_7C_1–O1, B) Y_3C_1–O1 exhibiting a current step, C) Y_3C_1–O1 exhibiting a current slope, and D) Y_3/Y_3–O3. The insets display magnified views of the high-amplitude regions of the events. E) Amino acid side chains of histidine, arginine, and tyrosine. F) Schematic representation to account for the current signature of the Y_3/Y_3–O3 events in (D).

Y_3C_1–O1 led to type I events with two current levels (Figure 2B, inset, dotted lines). The first level at (92.4 ± 0.6) % (Figure 2B, inset, top dotted line) is similar to the blockage amplitude of unmodified DNA. It is therefore very likely that this level stems from a DNA strand that is threaded into the inner constriction but is kept from passing through the β barrel because the bulky peptide has not yet entered the narrow pore region. The second level at (97.8 ± 0.5) % (Figure 2B, inset, bottom dotted line) is ascribed to a state in which the peptide–DNA segment has entered the inner constriction and translocates the β barrel. The step signature was observed for 60% of type I Y_3C_1–O1 events. In the remaining events, the transition between the two current levels resembled a slope (Figure 2C, inset). This could reflect a peptide–DNA segment that is being gradually, rather than abruptly, pulled into the inner constriction. Importantly, the step-like blocking effect of Y_3C_1 was independent of the DNA sequence around the modified base because the same event characteristics were also seen for Y_3C_1–O2 with a different oligonucleotide sequence (Table 1; Y_3C_1–O1/step versus Y_3C_1–O2/step).

DNA strands with two separate chemical tags were tested to prove that tags act independently and give rise to correspondingly distinct current modulations. The first strand was a 37-mer Y_3/Y_3–O3 in which 2 Y_3C_1 peptides are tethered to 2 modified bases separated by 13 nucleotides. Similar to the results with the single-modified Y_3C_1–O1 strand, double-modified DNA gave rise to unresolved slope events (see Supporting Information) as well as fully resolved step-like events (Figure 2D). In the latter events, the blockage amplitude fluctuates twice between 2 levels, sequentially from event segments 1–4 (Figure 2D, event segments numbered in red). The average current levels for segments 1 and 3 and for segments 2 and 4 are approximately 92 and 99%, respectively (Table 2). The step-like signature is in line with the expectations for two Y_3C_1 peptides because one peptide is known to cause a blockage step from approximately 92% up to 98% (Table 1, Y_3C_1–O1/step). The signature of Y_3/Y_3–O3 in Figure 2D strongly suggests that the current alterations reflect the sequential pulling of a DNA strand through the pore, as illustrated schematically in Figure 2F (red numbers correspond to the segments in Figure 2D).

Table 2: Characteristics of type I translocation events of Y_3/Y_3–O3 and Y_3/Y_3–O4 carrying tags separated by 13 and 27 nucleotides, respectively.[a]

Segments	1	2	3	4
O3 A_h [%][b]	92.2 ± 1.1	99.7 ± 0.7	92.3 ± 0.9	99.7 ± 0.6
O3 τ_{off-h} [ms][b]	3.43 ± 0.67	0.80 ± 0.22	0.26 ± 0.08	0.61 ± 0.15
O4 A_h [%][b]	92.8 ± 1.1	98.2 ± 0.8	91.1 ± 1.5	98 ± 0.7
O4 τ_{off-h} [ms][b]	1.34 ± 0.75	0.64 ± 0.26	0.67 ± 0.21	0.74 ± 0.37

[a] The recordings were conducted in 2 M KCl and 20 mM Tris (pH 8.0), filtered at 10 kHz, and sampled at 50 kHz. [b] Defined as in Table 1.

The interpretation of the stepped events as sequential pulling is supported by the finding that the 54-mer DNA strand Y_3/Y_3–O4, with a separation of 27 nucleotides between the peptides, showed similar current modulations (see the Supporting Information). The two current levels were identical within experimental error to the values observed for Y_3/Y_3–O3 (Table 2). The duration of event segment 3 was, however, longer for Y_3/Y_3–O4 than for Y_3/Y_3–O3 (Table 2). This positive correlation between duration and tag separation indicates that a longer DNA strand takes more time to pass the pore. The 37-mer Y_3/Y_3–O5, with a separation of 7 nucleotides between the peptides, also displayed the step behavior. The percentage of stepped events, as well as the quality of the current steps, was lower for Y_3/Y_3–O5 than for Y_3/Y_3–O3 and Y_3/Y_3–O4 (see the Supporting Information). This reduced resolution agrees with molecular models showing that the tag with an extended length of 2.8 nm bridges the gap between the two tagged bases separated by 7 nucleotides or 2.2 nm.

In this study, we have developed a new nanopore-based strategy to enable the detection of separate bases in DNA strands. By using a model system of DNA oligonucleotides modified with peptides, we have demonstrated that chemical tags attached to bases are capable of causing characteristic current signatures for strands translocating through nanopores. The proof-of-principle experiments show that the blockage duration, amplitude, and signature can be tuned by changing the length, charge, and size of the tags. The current modulations are independent of the surrounding DNA sequence and two tags on a strand retain their characteristic signatures, which opens up the possibility of attaching multiple tags to DNA. To the best of our knowledge, this is the first time 1) that pore recordings have detected one or two separate bases in translocating individual DNA strands and 2) that chemically modified DNA has been used to infer base-specific information.

While our experiments have been performed with synthetic oligonucleotides, the approach can potentially be applied to sense DNA from biological samples. For example, peptide tags can be incorporated into copied DNA strands by using chemically modified nucleotides and sequence-specific primer extension. This approach would be suitable to sense the presence or absence of single-nucleotide polymorphisms by incorporating and detecting a modified base only if the target mutation is present. With further improvements in the tags, such as, a decrease in the size of the linker and the length of the tags, it will also be possible to reduce the nucleotide distance between the tags and thereby detect multiple bases in biologically relevant DNA strands. For example, the highly repetitive DNA regions in trinucleotide-expansion-disease genes could be sized by labeling the same base in all of the repeats.[22] The extension of the technology towards sequencing by measuring the ionic-current modulation for each base[1] would be very difficult to achieve due to the small distance between neighboring bases. This does not, however, limit the potential of this method, because the concept of slowing down DNA through chemical tags is new and can be applied to various related nanopore approaches. These include fluorescence- or ionic-current-based sequencing of DNA-derived designed polymers in which the spacing between individual bases has been increased[23] (see the Supporting Information) or strategies that detect tunneling current.[24]

The general approach of using chemically modified bases is especially relevant for DNA sensing with solid-state nanopores. These pores exhibit very high mechanic stability, which makes them ideally suited for rugged electrical sensor devices. Despite progress in their fabrication, solid-state nanopores cannot be engineered to the same atomic precision as the protein pore αHL. For example, inorganic pores are usually not narrow enough to discriminate between single- and double-stranded DNA; this thereby places constraints on their ability to detect DNA through a hybridization approach.[9] The use of chemically modified DNA strands can address this limitation by tuning the cross-sectional diameter of ssDNA to existing pore dimensions rather than matching the pore dimensions to the size of the DNA strand. A more detailed discussion on the combined use of chemically modified DNA and solid-state nanopores can be found in the Supporting Information.

In summary, our strategy represents a new approach because it uses the chemical modification of the analyte, rather than pore engineering, to expand and enhance the sensing repertoire of nanopores. The concept can be applied to other bioanalytes.

Received: January 14, 2008
Published online: June 13, 2008

Keywords: biosensors · DNA · nanopores · peptides · single-molecule studies

[1] J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Deamer, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13770.
[2] C. Dekker, *Nat. Nanotechnol.* **2007**, *2*, 209.
[3] C. R. Martin, Z. S. Siwy, *Science* **2007**, *317*, 331.
[4] M. Rhee, M. A. Burns, *Trends Biotechnol.* **2007**, *25*, 174.
[5] H. Bayley, *Curr. Opin. Biotechnol.* **2006**, *10*, 628.
[6] S. Howorka, S. Cheley, H. Bayley, *Nat. Biotechnol.* **2001**, *19*, 636.
[7] W. Vercoutere, S. Winters-Hilt, H. Olsen, D. Deamer, D. Haussler, M. Akeson, *Nat. Biotechnol.* **2001**, *19*, 248.
[8] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, D. Branton, *Phys. Rev. Lett.* **2003**, *90*, 238101.
[9] J. Mathe, H. Visram, V. Viasnoff, Y. Rabin, A. Meller, *Biophys. J.* **2004**, *87*, 3205.
[10] J. Nakane, M. Wiggin, A. Marziali, *Biophys. J.* **2004**, *87*, 615.
[11] Q. Zhao, G. Sigalov, V. Dimitrov, B. Dorvel, U. Mirsaidov, S. Sligar, A. Aksimentiev, G. Timp, *Nano Lett.* **2007**, *7*, 1680.
[12] Y. Astier, O. Braha, H. Bayley, *J. Am. Chem. Soc.* **2006**, *128*, 1705.
[13] N. Ashkenasy, J. Sanchez-Quesada, H. Bayley, M. R. Ghadiri, *Angew. Chem.* **2005**, *117*, 1425; *Angew. Chem. Int. Ed.* **2005**, *44*, 1401.
[14] M. Gershow, J. A. Golovchenko, *Nat. Nanotechnol.* **2007**, *2*, 775.
[15] A. Meller, L. Nivon, D. Branton, *Phys. Rev. Lett.* **2001**, *86*, 3435.
[16] L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, J. E. Gouaux, *Science* **1996**, *274*, 1859.
[17] C. R. Cantor, P. R. Schimmel, *Biophysical chemistry. Part III. The behavior of biological macromolecules*, W. H. Freeman, New York, **1980**.
[18] T. Z. Butler, J. H. Gundlach, M. A. Troll, *Biophys. J.* **2005**, *90*, 190; the inclusion of the amino linker on the internal thymine base had no effect on either the event amplitude or translocation duration.
[19] The fast translocation of the terminally modified DNA compared to that of the internally modified DNA also suggests that electrostatic factors are not a major factor in strand retardation. This would also be expected from the low ionization state of the imidazole side chains at pH 8.0 (pK_a 6.0). The passage of internally modified H_6C_1–O1 was, however, slowed down in a buffer with a pH value of 6.4, with a slight increase from 1.91 to 2.18 ms for type I events (Table 1, H_6C_1-O1/pH 6.0).
[20] J. Mathe, A. Aksimentiev, D. R. Nelson, K. Schulten, A. Meller, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12377.
[21] I. A. Kathawalla, J. L. Anderson, J. S. Lindsey, *Macromolecules* **1989**, *22*, 1215.
[22] R. Schlapak, H. Kinns, C. Wechselberger, J. Hesse, S. Howorka, *ChemPhysChem* **2007**, *8*, 1618.
[23] http://www.lingvitae.com.
[24] N. Blow, *Nat. Methods* **2008**, *5*, 267.

Nano-biotechnology

DOI: 10.1002/anie.200804264

A DNA Nanostructure for the Functional Assembly of Chemical Groups with Tunable Stoichiometry and Defined Nanoscale Geometry**

*Nick Mitchell, Robert Schlapak, Markus Kastner, David Armitage, Wojciech Chrzanowski, Johannes Riener, Peter Hinterdorfer, Andreas Ebner, and Stefan Howorka**

DNA is a versatile material in nano-biotechnology[1] and chemical biology. DNA strands can be sequence-specifically assembled into artificial higher-order nanoscale structures[2,3] that are useful in, for example, biotemplating,[4,5] biocomputing,[6] sensing,[7] and biophysical studies.[8] In chemical biology, hybridization of chemically modified strands into DNA duplexes has been utilized to bring chemical groups into defined contact to, for example, enhance their reaction.[9] Herein we describe a new approach which merges DNA-based nanostructures and the chemical modification of DNA. We show that tetrahedron-shaped nanostructures can act as scaffolds to assemble a multitude of different chemical groups at tunable stoichiometry and at geometrically defined sites. The resulting molecular entities exhibit functional properties beneficial in biosensing and diagnostics. Our new strategy for assembling chemical groups at the nanoscale may be expanded to endow other DNA structures with rationally designed functions.

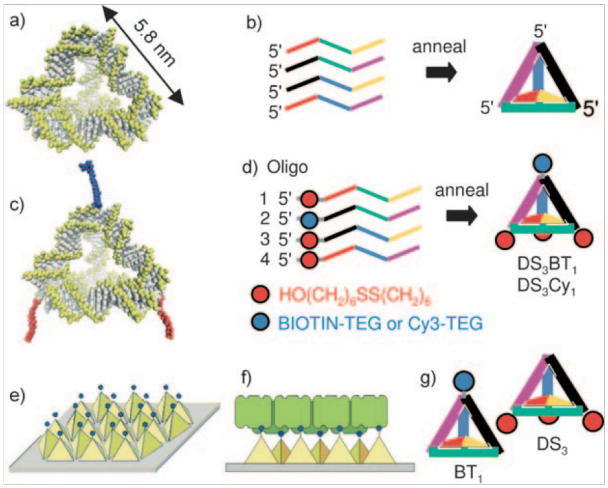


Figure 1. Structure and formation of DNA tetrahedra. a) Molecular model of a DNA tetrahedron formed by b) annealing of four ssDNA each carrying three out of six different, color-coded sequence blocks. The 5′ termini of the ssDNA oligonucleotides are positioned at the four vertices. c) Model of tetrahedron DS_3BT_1 with three DS and one BT tag formed by d) annealing four 5′-modified ssDNA oligonucleotides. Similarly, tetrahedron DS_3Cy_1 carrying the Cy3 fluorophore instead of BT was also generated. The BT and Cy3 tags are attached through a flexible tri(ethylene glycol) (TEG) linker to the 5′ ends and vertices. e) DS_3BT_1 or DS_3Cy_1 bind through their DS legs in an oriented fashion to a gold leaving the BT or Cy3 tag exposed to the ambient. f) A monomolecular film of DS_3BT_1 can capture a layer of streptavidin. g) Tetrahedra BT_1 and DS_3 that carry one BT or three DS groups, respectively.

DNA tetrahedra, as intoduced by Goodman et al, are nanostructures with edges composed of double-stranded DNA (dsDNA; Figure 1a).[10,11] Tetrahedra are obtained by annealing four single-stranded DNA oligonucleotides (ssDNA; Figure 1b); the 55 nucleotide strands used in this study give rise to tetrahedron edges of 5.8 nm in length (Figure 1a).[10] Once assembled, the free 5′ and 3′ termini of the ssDNA are positioned at the vertices of the tetrahedron (Figure 1b). We exploited this structural characteristic and placed a combination of biotin (BT) and disulfide (DS) groups at the four vertices (Figure 1c) by using four ssDNA oligonucleotides that carry the chemical modifications at their 5′ ends (Figure 1d). By placing three DS groups and one BT group at the vertices, the rationally designed structures were expected to exhibit desirable functional properties. In particular, the tetrahedra were anticipated to (1) bind through three

[*] N. Mitchell, Dr. S. Howorka
Department of Chemistry, University College London
20 Gordon Street, London WC1H OAJ (UK)
Fax: (+44) 20-7679-7463
E-mail: s.howorka@ucl.ac.uk
R. Schlapak, Dr. S. Howorka
Center for Biomedical Nanotechnology, Upper Austrian Research
Linz, 4020 (Austria)
M. Kastner, J. Riener, Prof. P. Hinterdorfer, Dr. A. Ebner
Institute for Biophysics, Johannes Kepler University
Linz, 4040 (Austria)
Dr. D. Armitage
Leicester School of Pharmacy, De Montfort University
Leicester, LE1 9BH (UK)
Dr. W. Chrzanowski[+]
Biomaterials and Tissue Engineering, Eastman Dental Institute,
University College London, London, WC1X 8LD (UK)
[+] Present address: Department of Mechanical Engineering
University of Glasgow, Glasgow, G12 8QQ (UK)

[**] Funded by the Austrian Science Foundation (project N00104-NAN), the European Fund for Regional Development (EFRE), and the government of Upper Austria. N.M. holds a PhD studentship from the Department of Chemistry, UCL and a scholarship from the Postgraduate School, UCL. We thank Hugh Martin for the preparation of Figure 1a and 1c.

Experimental procedures and results of the analysis of DNA tetrahedron films by XPS and fluorescence microscopy are provided in the Supporting Information for this article, which is available on the WWW under http://dx.doi.org/10.1002/anie.200804264.

thiol legs with high affinity onto gold substrates and (2) because of their oriented binding, present the BT moiety at a defined surface distance so that the tag is exposed to the ambient (Figure 1e) and is able to capture streptavidin at high density (Figure 1f).

Tetrahedron DS_3BT_1 (with three DS and one BT tags) was generated by annealing four ssDNA oligonucleotides (Figure 1d). The formation of DS_3BT_1 was confirmed by monitoring the position of the DNA bands in polyacrylamide gel electrophoresis. In comparison to the fast-migrating ssDNA bands (Figure 2a, lanes 1–4), hybridization product DS_3BT_1 was at the top of the lane (Figure 2a, lane 5) because the

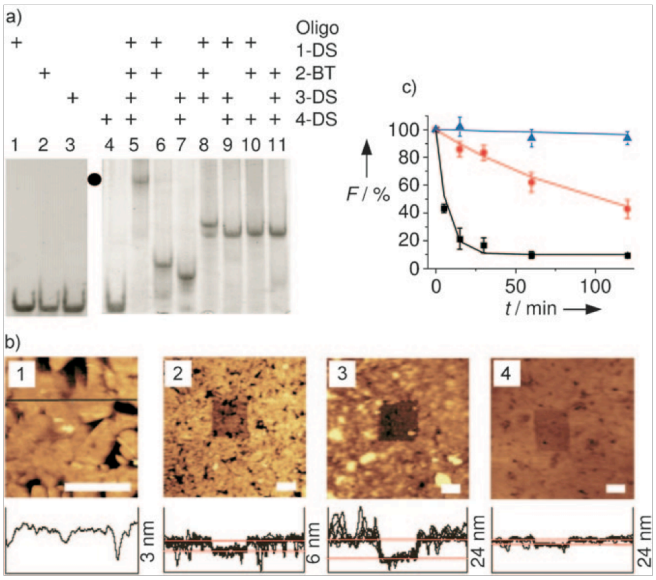


Figure 2. Generation and surface binding of DNA tetrahedra. a) Gel electrophoretic analysis on the formation of DS_3BT_1 (lane 5, dot) from four ssDNA oligonucleotides. Controls comprising monomeric ssDNA, and combinations of two and three ssDNA oligonucleotides are in lanes 1–4, 6–7, and 8–11, respectively. b) AFM topographic images and height profiles of bare TSG gold (panel 1), DS_3BT_1-coated TSG with a hole scratched into the DNA film (panel 2), a DS_3BT_1 film after incubation with streptavidin (panel 3), and a layer of BT-free tetrahedron DS_3 after incubation with streptavidin (panel 4). Scale bars 500 nm. c) Normalized fluorescence counts, *F*, of gold surfaces coated with DS_3Cy_1 (blue triangles), DS_2Cy_1 (red circles), and ssDNA DS–DNA–Cy (black squares) after incubation with DTT (10 mM) for the indicated durations, *t*. These data represent an average of three independent measurements.

increased mass and changed shape of the tetrahedron slows its movement through the gel meshwork.[10] Control experiments with different combinations of two and three ssDNA led to migration heights between ssDNA and the tetrahedron band (Figure 2a, lanes 6–11), which is consistent with the expected size of the hybridization products (see the Supporting Information). Four other DNA tetrahedra were generated: DS_3 and BT_1 (with either three DS or one BT group, respectively; Figure 1g), DS_3Cy_1 (Figure 1d), and DS_2Cy_1. The latter two tetrahedra contain three and two DS groups, respectively, and one Cy3 fluorophore.

The ability of DS_3BT_1 to bind to gold surfaces was assessed by using X-ray photoelectron spectroscopy (XPS) and atomic force microscopy (AFM). In agreement with the formation of a DNA film, XPS analysis showed the attenuation of the Au4f signal as well as the appearance of peaks for C1s, N1s, O1s, and P2p at relative ratios expected for DNA (see the Supporting Information). As would be predicted for the oriented and specific binding, AFM analysis showed that tetrahedron DS_3BT_1 yielded a monomolecular film. Template-stripped gold (TSG) was used for the topographic measurements, and exhibited flat (root mean square (rms) noise, 0.3 nm) facets 100–1000 nm in size, separated by 2–10 nm deep trenches (Figure 2b, panel 1). In comparison, gold surfaces incubated with DS_3BT_1 displayed an increased rms noise of 0.5 nm. In addition, a hole with a depth of 1.5 nm could be scratched into DS_3BT_1 surfaces (Figure 2b, panel 2). This value is lower than the DNA tetrahedron height of 4.5 nm because soft biological material can be compressed in contact mode AFM.[11] AFM analysis showed that, in agreement with the disulfide-specific tethering, gold surfaces incubated with DS-free tetrahedron BT_1 did not form films (data not shown).

AFM was also used to determine whether DS_3BT_1 was able to bind streptavidin while maintaining its adhesion to gold. A DS_3BT_1 film was formed and incubated with streptavidin. Scratching of the surface revealed a profile step of 7 nm (Figure 2b, panel 3), which is consistent with the formation of a monomolecular layer of streptavidin (protein size $4 \times 8 \times 7$ nm) on top of a DNA film. The elevated, bright features around the hole (Figure 2b, panel 3) represent streptavidin and DNA that had been removed from the film. Consistent with a specific biotin–streptavidin interaction, BT-free tetrahedron DS_3 did not capture streptavidin as indicated by the smaller step size (Figure 2b, panel 4).

After confirming the oriented binding of DS_3BT_1, we determined the affinity of the tetrahedra to the gold substrate. Because of the anticipated multivalent enhancement of individual gold–thiol interactions, the DNA structures with three DS legs should bind tighter to the surface than constructs with two legs or a single leg. The differential affinity was experimentally determined by first forming films composed of fluorescence-labeled tetrahedra DS_3Cy_1 or DS_2Cy_1, or single-stranded oligonucleotide DS–DNA–Cy, and then monitoring their time-dependent desorption by using fluorescence microscopy (see the Supporting Information). Desorption was enhanced by incubating the DNA films with dithiothreitol (DTT, 10 mM), which displaces thiolated DNA from the gold surface. The kinetics of the decrease in fluorescence are displayed in Figure 2c. During an observation window of 120 min, very little DS_3Cy_1 was removed (5% removal; Figure 2c, blue triangles) while the coverage of DS_2Cy_1 with two thiol legs fell nearly linearly to below 50% of the starting value (Figure 2c, red circles). DNA that contains a single DS group exhibited the lowest affinity, its signal decayed exponentially to 10% after 30 min (Figure 2c, black squares). The residual constant signal is due to autofluorescence of gold and background fluorescence of desorbed and solvated DNA strands. Comparative analysis of the initial rates of desorption revealed that DS_3Cy_1 (with three legs

bound) has an affinity for gold that is 5000-times higher than monothiolated DNA.

We have presented a new strategy that exploits DNA nanostructures as scaffolds to combine different chemical groups at defined geometrical distances and with tunable stoichiometry. The rationally designed structures exhibit functional properties that may be exploited for the immobilization of DNA or proteins on gold surfaces for biosensing, diagnostics, and cell biological research. Given the great variety of DNA nanostructures,[3] our approach for chemical enhancement has the potential to be extended to other nanostructures and different biochemical tags. For example, it is envisioned that the structures help prepare new templates for chemical reactions, create functional building blocks for defined multimeric enzyme complexes, or build labeling reagents that carry tunable numbers of tags.

Received: August 28, 2008
Published online: December 9, 2008

Keywords: DNA structures · nano-biotechnology · self-assembly · supramolecular chemistry · thin films

[1] C. M. Niemeyer, C. A. Mirkin, *Nanobiotechnology: Concepts, Applications and Perspectives*, Wiley, New York, **2004**.
[2] P. W. Rothemund, *Nature* **2006**, *440*, 297.
[3] F. C. Simmel, *Angew. Chem.* **2008**, *120*, 5968; *Angew. Chem. Int. Ed.* **2008**, *47*, 5884.
[4] S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, C. A. Mirkin, *Nature* **2008**, *451*, 553.
[5] D. Nykypanchuk, M. M. Maye, D. van der Lelie, O. Gang, *Nature* **2008**, *451*, 549.
[6] C. Mao, T. H. LaBean, J. H. Relf, N. C. Seeman, *Nature* **2000**, *407*, 493.
[7] C. Mao, W. Sun, Z. Shen, N. C. Seeman, *Nature* **1999**, *397*, 144.
[8] S. Rinker, Y. G. Ke, Y. Liu, R. Chhabra, H. Yan, *Nat. Nanotechnol.* **2008**, *3*, 418.
[9] Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* **2004**, *305*, 1601.
[10] R. P. Goodman, R. M. Berry, A. J. Turberfield, *Chem. Commun.* **2004**, 1372.
[11] R. P. Goodman, I. A. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, A. J. Turberfield, *Science* **2005**, *310*, 1661.

J|A|C|S COMMUNICATIONS

Published on Web 05/14/2009

Chemically Labeled Nucleotides and Oligonucleotides Encode DNA for Sensing with Nanopores

Vinciane Borsenberger, Nick Mitchell, and Stefan Howorka*

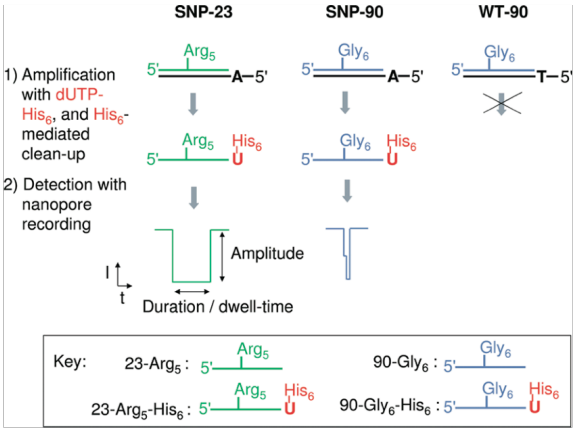
Department of Chemistry, University College London, Christopher Ingold Building, 20 Gordon Street, London WC1H OAJ, U.K.

Received March 14, 2009; E-mail: s.howorka@ucl.ac.uk

The labeling of nucleotides and oligonucleotides with reporter groups is an important tool in the sequence-specific sensing of DNA. Prominent examples include fluorescence tags that encode DNA sequences[1] or bases.[2] Reporter groups have also been used for the electronic detection of nucleic acids via cyclic voltammetry,[3] but not for the eletrical nanopore recordings. In nanopore recordings, individual DNA strands are electrophoretically driven through a nanoscale pore leading to detectable blockades of ionic current.[4] Existing nanopore approaches for the sequence-specific sensing of DNA rely on the formation of a duplex which cannot penetrate a narrow pore,[5−7] or the use of pore-linked receptors for single stranded DNA,[8,9] or individual nucleotides.[10] Here we present a new strategy which exploits tagged nucleotides[11] and oligonucleotides[12] for the sequence-specific sensing of single-point mutations. The chemical tags cause characteristic electrical signatures and thereby encode DNA sequences. By being independent of pore engineering, our approach might enhance the sensing repertoire of durable solid-state nanopores.[13]

The sensing approach was tested using the biologically relevant model system of single nucleotide polymorphisms (SNPs) in codons 23 and 90 of the HIV protease gene. The SNPs confer resistance against the HIV drug Nelfinavir.[14] The principle of the two-step sensing strategy is outlined in Scheme 1. First, the presence of the

Scheme 1. Sensing of SNPs Using Linear Amplification of Chemically Tagged Probe DNA Strands in Combination with Single-Molecule Nanopore Recordings



SNPs is encoded by chemically tagging DNA strands. Probe strands 23-Arg_5 and 90-Gly_6 bind sequence-specifically to the SNP-containing templates SNP-23 and SNP-90 (Scheme 1, step 1, templates in black) to facilitate the enzymatic addition of the hexahistidine-tagged nucleotide analogue dUTP-His_6 (red). This uridine nucleotide analogue is complementary to the resistance-conferring base A and is therefore only added for SNP-containing templates SNP-23 and SNP-90 (Scheme 1) but not wild-type WT-23 which carries the mismatched base C (Supporting Information, S6−S7) or base T for WT-90 (Scheme 1). By encoding for the presence of the SNPs, the His_6-tag facilitates the selective isolation of the SNP-derived DNA probes 23-Arg_5-His_6 and 90-Gly_6-His_6 using metal affinity purification (Scheme 1, step 1). In the second step, tagged DNAs are electrically detected in nanopore recordings (Scheme 1, step 2). Signals stemming from the two probe strands are discriminated because the covalently linked Arg_5 and Gly_6 tags give rise to two current signatures with different amplitudes and different average durations or dwell-times (Scheme 1).

The template-directed enzymatic extension was conducted with probe strands 90-Gly_6 and 23-Arg_5 (see Supporting Information, S3−S5) and dUTP-His_6 (Figure 1A). The His_6 tag of dUTP-His_6

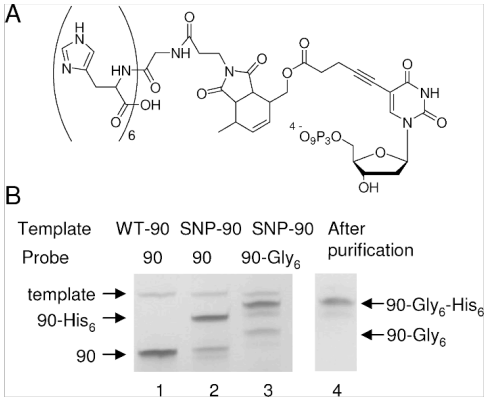


Figure 1. (A) dUTP-His_6; (B) gel electrophoretic analysis of the extension reactions. 90-Gly_6 contains a faster migrating impurity originating from the commercial DNA oligo preparation.

is positioned at C-5 of the pyrimidine base (Figure 1A) to avoid any interference with DNA polymerization.[11] The addition of the tagged nucleotide was confirmed by a visible shift in the electrophoretic migration of the bands for 90-Gly_6 and untagged control 90 (Figure 1B, lanes 3 and 2). Addition of the tagged base was specific for the presence of base A in the templates because no extension occurred for base T in codon position 90 (Figure 1B, lane 1). His_6-tag purification in spin-column format resulted in the isolation of 90-Gly_6-His_6 (Figure 1B, lane 4). Similarly, 23-Arg_5 was specifically extended for template SNP-23 and His_6-tag purified (Supporting Information, S7−S8).

23-Arg_5-His_6 and 90-Gly_6-His_6 were subjected to analysis via nanopore recordings to investigate the different current signatures. The recordings were conducted with the nonengineered version of the protein pore α-hemolysin (αHL) (Figure 2A) which is a standardized reference because of its atomically defined structure[15] and widespread use.[16,17] In the absence of DNA, the αHL pore gave rise to a unitary conductance of 1.976 ± 0.066 nS (n = 7)

10.1021/ja902004s CCC: $40.75

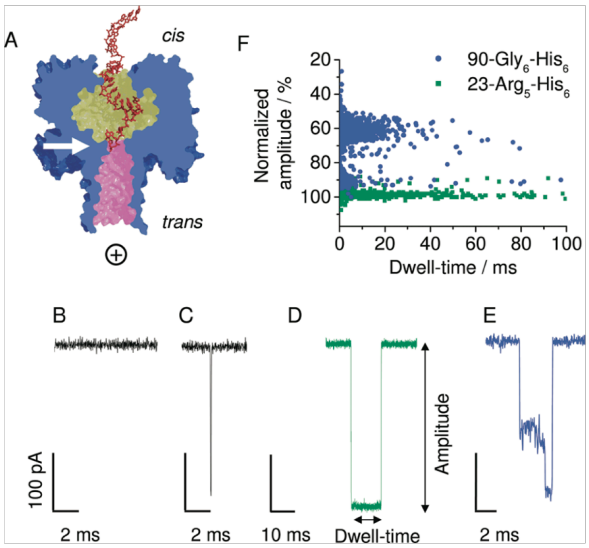


Figure 2. (A) The α-hemolysin (αHL) pore interacts with a ssDNA strand. The inner cavity and the β-barrel are highlighted in yellow and pink, respectively. The white arrow indicates the pore constriction. (B−E) Representative single-channel current traces for (B) αHL, and events of (C) untagged DNA 23, (D) tagged probes 23-Arg_5-His_6 and (E) 90-Gly_6-His_6. The traces were obtained from recordings in 2 M KCl and 20 mM TRIS HCl pH 8.0 at +150 mV, filtered, and sampled at 10 and 50 kHz, respectively. (F) Scatter plot for 23-Arg_5-His_6 and 90-Gly_6-His_6 events.

(Figure 2B). The addition of 23-Arg_5-His_6 to the cis side of the pore resulted in reversible current blockades (Figure 2D) which are most likely caused by the electrophoretically driven translocation of DNA strands to the positively polarized trans side of the pore (Figure 2A).[4,12] The blockades' characteristic average dwell-time, τ_{off}, of 9.33 ms and normalized amplitude, A, of 99% (Figure 2D, Table 1) is due to the Arg_5 tag as blockades of strands without a

Table 1. Summary of Current Blockades for 23-Arg_5-His_6 and 90-Gly_6-His_6 in Nanopore Recordings

probe strand	τ_{off} (ms)[a]	A (%)[a,b]
23-R_5-H_6	9.33	99
90-G_6-H_6	0.76	91
	1.21 (13%)/8.07 (87%)[c]	59 (80%)/72 (20%)[c]

[a] Obtained as described in the Supporting Information (S8−S11). The values are the averages from three independent recordings each for 23-Arg_5-His_6 and 90-Gly_6-His_6, and a recording in which both modified DNA probes were present. [b] Amplitude, A, was normalized to the conductance of the open unblocked channel. [c] τ_{off} and A values for midlevel events of 90-Gly_6-His_6. Two values each are provided to reflect the two components in the respective distributions. The relative proportion of the components is given in brackets.

peptide tag were much shorter (τ_{off} = 0.16 ms) and of lower amplitude (91%) (Figure 2C). The signature of 23-Arg_5-His_6 events is in line with previous observations that a positively charged peptide folds back on the negatively charged DNA thereby creating a compact bulge which slowly translocates the narrow pore constriction.[12] The terminal His_6 tag, by contrast, sequentially passes after DNA and does not cause any slow-down.[12] In comparison, 90-Gly_6-His_6 gave rise to different current blockades which were characterized by a midlevel blockade followed by a high-amplitude blockade of 91% (Figure 2E)(Table 1). The midlevel originates from DNA strands which initially reside inside the wide inner cavity of the αHL pore but eventually thread into and translocate the narrow β-barrel of the pore (Figure 2A). The high-level blockades of 90-Gly_6-His_6 were shorter and of lesser amplitude than 23-Arg_5-His_6 because the Gly_6 is smaller than Arg_5 and causes less steric blockade.[12] The events of 90-Gly_6-His_6 clustered into well-separated groups within a scatter plot where each event is represented as a data point of defined amplitude and dwell-time (Figure 2F).

In summary, our new strategy to encode sequence information exploits our previous observation that chemical tags can slow down DNA translocation through a nanopore.[12] The strategy detects SNPs and is therefore not able to sequence de-novo. The electrical nanopore approach complements existing fluorescene-based SNP-sensing strategies,[18] and may be miniaturized for point-of-care applications. The sensing method is particularly relevant for solid-state nanopores.[13] These pores are more durable and rugged than protein pores which are inserted in more fragile lipid-bilayer membranes. The abiotic pores cannot, however, be engineered to atomic precision as the biological pores, thereby limiting their potential sensing repertoire. Our approach addresses this point as it senses without pore engineering. The strategy is currently able to discriminate between two sequences. To sense multiple DNA strands, tags can be further rationally designed[12] to achieve more characteristic current signatures. We note that multianalyte detection independent of pore engineering has previously been demonstrated for proteins.[19]

Acknowledgment. This work has been supported by UCL Chemistry, UCL Business PLC, and the EPSRC. N.M. holds a PhD studentship from UCL Chemistry and a scholarship from the Postgraduate School, UCL. We thank A. B. Tabor for use of a peptide synthesizer, and Hugh Martin for preparing Figure 2A.

Supporting Information Available: Complete ref 18; experimental details on the synthesis and chemical analysis of the tagged probe strands, primer extension of probe strands, and nanopore recordings. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Piatek, A. S.; Tyagi, S.; Pol, A. C.; Telenti, A.; Miller, L. P.; Kramer, F. R.; Alland, D. *Nat. Biotechnol.* **1998**, *16*, 359–363.
(2) Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E. *Nature* **1986**, *321*, 674–679.
(3) Ferapontova, E. E.; Olsen, E. M.; Gothelf, K. V. *J. Am. Chem. Soc.* **2008**, *130*, 4256–4268.
(4) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770–13773.
(5) Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. *Nat. Biotechnol.* **2001**, *19*, 248–252.
(6) Mathe, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A. *Biophys. J.* **2004**, *87*, 3205–3212.
(7) Nakane, J.; Wiggin, M.; Marziali, A. *Biophys. J.* **2004**, *87*, 615–621.
(8) Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotechnol.* **2001**, *19*, 636–639.
(9) Iqbal, S. M.; Akin, D.; Bashir, R. *Nat. Nanotechnol.* **2007**, *2*, 243–248.
(10) Clarke, J.; Wu, H. C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. *Nat. Nanotechnol* **2009**, *4*, 265–270.
(11) Borsenberger, V.; Howorka, S. *Nucleic Acids Res.* **2009**, *37*, 1477–1485.
(12) Mitchell, N.; Howorka, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 5476–5479.
(13) Dekker, C. *Nat. Nanotechnol.* **2007**, *2*, 209–215.
(14) Shafer, R. W. *Clin. Microbiol. Rev.* **2002**, *15*, 247–277.
(15) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. *Science* **1996**, *274*, 1859–1866.
(16) Griffiths, J. *Anal. Chem.* **2008**, *80*, 23–27.
(17) Martin, C. R.; Siwy, Z. S. *Science* **2007**, *317*, 331–332.
(18) Wang, D. G.; et al. *Science* **1998**, *280*, 1077–1082.
(19) Kasianowicz, J. J.; Henrickson, S. E.; Weetall, H. H.; Robertson, B. *Anal. Chem.* **2001**, *73*, 2268–2272.

JA902004S

References

[1] Crick, F.; *Nature*; **1970**; 227; 561-563.

[2] Dahm R; *Hum. Genet.*; **2008**; 122 (6); 565–81.

[3] Levene, P. J.; *Biol Chem*; **1919**; 40 (2); 415–24.

[4] Astbury, W.; *Nucleic acid Symp. SOC. Exp. Bbl 1*; **1947**; (66).

[5] Lorenz, M. G.; Wackernagel, W.; *Microbiol. Rev.*; **1994**; 58 (3); 563–602.

[6] Hershey, A.; Chase, M.; *J Gen Physiol*.; **1952**; 36 (1); 39–56.

[7] Watson, J.; Crick, F.; *Nature*; **1953**; 171; 737 – 738.

[8] Horwitz, J. P.; Chua, J.; Noel, M.; *J. Org. Chem.*; **1964**; 29; 2076.

[9] Borsenberger, V.; Kukwikila, M.; Howorka, S.; *Org. Biomol. Chem.*; **2009**; 7; 3826 – 3835.

[10] Sanger, F.; Nicklen, N.; Coulson, A. R.; *Proc. Nati. Acad. Sci. USA*; **1977**; Vol. 74; No. 12; 5463 – 5467.

[11] Fig. taken from; Rosenblum, B. B. Lee, L. G.; Spurgeon, S. L.; Khan, S. H.; Menchen, S. M.; Heiner, C. R.; Chen, S. M.; *Nucleic Acids Research*; **1997**; Vol. 25; No. 22; 4500–4504.

[12] Fig. taken from; Griffiths, A. J. F.; Miller, J. H.; Suzuki, D. T.; Lewontin, R. C.; Gelbart, W. M.; *Genetic Analysis* 6th Ed.; W. H. Freeman and Company; 1996; 476.

[13] Russell, A. F.; Moffatt, J. G.; *Biochemistry*; **1969**; 8; 4889- 4896.

[14] Geider, K.; *Eur. J. Biochem.*; **1974**; 27; 555 – 563.

[15] McCarthy, J. R.; Robins, M. J.; Townsend, L. B.; Robins, R. K.; *J. Am. Chem. Soc.*; **1966**; 88; 1549-1553.

[16] Tener, G. M.; *J. Am. Chem. Soc.*; **1961**; 83; 159-168.

[17] Smith, L. M.; Sanders J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E.; *Nature*; **1986**; 321; (6071): 674–9.

[18] Smith, L. M.; Fung, S.; Hunkapiller, M. W.; Hunkapiller, T. J.; Hood, L. E.; *Nucleic Acids Res*.; **1985**; 13 (7); 2399–412.

[19] Fig taken from; Thum, O.; Jager, S.; Famulok, M.; *Angew. Chem. Int. Ed.*; **2001**; 40; 12; 3990.

[20] Venter, J. C. *et al*.; *Science*; **2001**; 291; 1304–1351.

[21] Staden, R.; *Nucleic Acids Res*; **1979**; 6 (7): 2601–10.

[22] Anderson, S.; *Nucleic Acids Res*.; **1981**; 9 (13): 3015–27.

[23] Stahel, R. A.; Zangemeister – Wittke, U.; *Lung Cancer*; **2003**; 41; S81 – S88.

[24] Fire, A.; Xu, S.; Montgomery, M.; Kostas, S.; Driver, S.; Mello, C.; *Nature*; **1998**; 391; 806–11.

[25] Tung, C-H.; *Bioconjugate Chemistry*; **2000**; Vol. 11; 5; 605 – 618.

[26] Ede, N. J.; Tregear, G. W.; Haralambidis, J.; *Bioconjugate Chem.*; **1994**; 5; 373-8.

[27] Basu, S.; Wickstrom, E.; *Tetrahedron Lett.;* **1995**; 36; 4943-6.

[28] Bruick, R. K.; Dawson, P. E.; Kent, S. B.; Usman, N.; and Joyce, G. F; *Chem. Biol.*; **1996**; 3, 49-56.

[29] Soukchareun, S.; Tregear, G. W.; Haralambidis, J.; *Bioconjugate Chem.*; **1995**; 6, 43-53.

[30] Arar, K.; Aubertin, A. M.; Roche, A. C.; Monsigny, M; Mayer, R.; *Bioconjugate Chem.*; **1995**; 6; 573-7.

[31] Reed, M. W.; Fraga, D.; Schwartz, D. E.; Scholler, J.; Hinrichsen, R. D.; *Bioconjugate Chem.*; **1995**; 6; 101-8.

[32] Corey, D. R.; *J. Am. Chem. Soc.*; **1995**; 117; 9373-4.

[33] Antopolsky, M.; Azhayeva, E.; Tengvall, U.; Auriola, S.; Jaaskelainen, I.; Ronkko, S.; Honkakoski, P.; Urtti, A.; Lonnberg, H.; Azhayev, A.; *Bioconjugate Chem.*; **1999**; 10; 598 – 606.

[34] Antopolsky, M.; Azhayev, A.; *Helvetica Chimica Acta*; **1999**; Vol. 82; 2130 – 2140.

[35] Juby, C. D.; Richardson, C. D.; Broussear, R.; *Tetrahedron Lett.*; **1991**; 32; 879-882.

[36] Basu, S.; Wickstrom, E.; *Tetrahedron Lett.*; **1995**; 36; 4943-6.

[37] Soukchareun, S.; Tregear, G. W.; Haralambidis, J.; *Bioconjugate Chem.*; **1995**; 6; 43-53.

[38] De La Torre, B. G.; Albericio, F.; Saison-Behmoaras, E.; Bachi, A.; Eritja, R.; *Bioconjugate Chem.*; **1999**; 10; 1005-12.

[39] Bergmann, F.; Bannwarth, W.; *Tetrahedron Lett.*; **1995**; 36; 1839-42.

[40] Tung, C. H.; Rudolph, M. J.; Stein, S.; *Bioconjugate Chem.*; **1991**; 2; 464 – 465.

[41] Kukolka, F.; Niemeyer, C. M.; *Org. Biomol. Chem.*; **2004**; 2; 2203 – 2206.

[42] Ede, N. J.; Tregear, G. W.; Haralambidis, J.; *Bioconjugate Chem.*; **1994**; 5; 373 – 378.

[43] Reed, M. W.; Fraga, D.; Schwartz, D. E.; Scholler, J.; Hinrichsen, R. D.; *Bioconjugate Chem.*; **1995**; 6; 101 – 108.

[44] Forget, D.; Boturyn, D.; Defrancq, E.; Lhomme, J.; Dumy, P.; *Curr. Eur. J.*; **2001**; 7; 3976 – 3984.

[45] Stetsenko, D. A.; Gait, M. J.; *J. Org. Chem.*; **2000**; 65; 4900 – 4908.

[46] Burley, G. A.; Gierlich, J.; Mofid, M. R.; Nir, H.; Tal, S.; Eichen, Y.; Carell, T.; *J. Am. Chem. Soc.*; **2006**; 128; 1398 – 1399.

[47] Borsenger, V.; Mitchell, N.; Howorka, S.; *J. Am. Chem. Soc.*; **2009**; 131; 7530 – 7531.

[48] Wei, Z.; Tung, C-H.; Zhu, T.; Stein, S.; *Bioconjugate Chem.*; **1994**; 5; 468 – 474.

[49] Tanaka, K.; Yamada, Y.; Shionoya, M.; *J. Am. Chem. Soc.*; **2002**; 124; 8802 – 8803.

[50] Tanaka, K.; Tengeiji, A.; Kato, Tatsuhisa, Toyama, N.; Shiro, M.; Shionoya, M.; *J. Am. Chem. Soc.*; **2002**; 124; 12494 – 12498.

[51] Clever, G. H.; Polborn, K.; Carell, T.; *Angew. Chem. Int. Ed.*; **2005**; 44; 7204 – 7208.

[52] Tanaka, K.; Clever, G. H.; Takezawa, Y.; Yamada, Y.; Kaul, C.; Shionoya, M.; Carell, T.; *Nature Nanotech*; **2006**; 1; 190 – 194.

[53] Burley, G. A.; Gierlich, J.; Mofid, M. R.; Nir, H.; Tal, S.; Eichen, Y.; Carell, T.; *J. Am. Chem. Soc.*; **2006**; 128; 1398 – 1399.

[54] Nguyen, T.; Brewer, A.; Stulz, E.; *Angew. Chem. Int. Ed.*; **2009**; 48; 1974 – 1977.

[55] Obeid, S.; Yulikov, M.; Jeschke, G.; Marx, A.; *Angew. Chem. Int. Ed.*; **2008**; 47; 6782 – 6785.

[56] Thum, O.; Jager, S.; Famulok, M.; *Angew, Chem. Int. Ed.*; **2001**; 40; 21; 3990 – 3993.

[57] Jager, S.; Famulok, M.; *Angew. Chem. Int. Ed.*; **2004**; 43; 3337 – 3340.

[58] Okamoto, A.; Taiji, T.; Taninaka, K.; Saito, I.; *Bioorg & Med. Chem. Lett*; **2002**; 12; 1895 – 1896.

[59] Harris, T. D. et al.; *Science*; **2008**; 320; 106–109.

[60] Levene, M. J.; Korlach, J.; Turner, S. W.; Foquet, M.; Craighead, H. G.; Webb, W. W.; *Science*; **2003**; 299; 682–686.

[61] Kasianowicz, J.; Brandin, E.; Branton, D.; Deamer, D. W.; *Proc. Natl. Acad. Sci. USA*; **1996**; 93; 13770 – 13773.

[62] Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E.; *Science*; **1996**; Vol. 274; 1859 – 1866.

[63] Fig. adapted from; Rhee, M.; Burns, M. A.; *Trends in Biotechnology*; **2007**; Vol. 25; No. 4; 174 – 181.

[64] Fig. adapted from; Meller, A.; Nivon, L.; Branton, D.; *Physical Review Letters*; **2001**; Vol. 86; No. 15; 3435 – 3438.

[65] Fig. adapted from; Martin, H. S. C.; Jha, S.; Howorka, S.; Coveney, P. V.; *J. Chem. Theory Comput.*; **2009**; 5; 2135 – 2148.

[66] Wu, H. C.; Bayley, H.; *J. Am. Chem. Soc.*; **2008**; 130; 6813-6819.

[67] Luchian, T.; Shin, S.; Bayley, H.; *Angew. Chem. Int. Ed.*; **2003**; 42; 1926 – 1929.

[68] Movileanu, L.; Howorka, S.; Braha, O.; Bayley, H.; *Nature Biotech*; **2000**; Vol. 18; 1091 – 1095.

[69] Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D.; *PNAS*; **2000**; Vol 97; 1079 – 1084.

[70] Howorka, S.; Cheley, S.; Bayley, H.; *Nat. Biotechnol.*; **2001**; 19; 636.

[71] Ashkenasy, N.; Sanchez-Quesada, J.; Bayley, H.; Ghadiri, M. R.; *Angew. Chem. Int. Ed.*; **2005**; 44; 1401 – 1404.

[72] Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M.; *Biotechnol.*; **2001**; 19; 248.

[73] Astier, Y.; Braha, O.; Bayley, H.; *J. Am. Chem. Soc.*; **2006**; 128; 1705 – 1710.

[74] Clarke, J.; Wu, H-C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H.; *Nature Nanotech.*; **2009**; Vol. 4; 265 – 270.

[75] Samori, B.; Zuccheri, G.; *Angew. Chemie. Int. Ed.*; **2005**; 44.; 1166 – 1181.

[76] Seeman, N.; *Nature*; **2003**; Vol. 421; 427 – 5428.

[77] Rothemund, P. W.; *Nature*; **2006**; 440; 297.

[78] Simmel, F. C.; *Angew. Chem.;* **2008**; 120; 5968.; *Angew. Chem. Int. Ed*.; **2008;** 47; 5884.

[79] Park, S. Y.; Lytton-Jean, A. R. K.; Lee, B.; Weigand, S.; Schatz, G. C.; Mirkin, C. A.; *Nature*; **2008**; 451; 553.

[80] Nykypanchuk, D.; Maye, M. M.; Van Der Lelie, V.; Gang, O.; *Nature*; **2008**; 451; 549.

[81] Mao, C.; LaBean, T. H.; Relf, J. H.; Seeman, N. C.; *Nature*; **2000**; 407; 493.

[82] Mao, C.; Sun, W.; Shen, Z.; Seeman, N. C.; *Nature*; **1999**; 397; 144.

[83] Rinker, S.; Ke, Y. G.; Liu, Y.; Chhabra, R.; Yan, H.; *Nat. Nanotech.;* **2008**; 3; 418.

[84] Gartner, J. Z.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R.; *Science*; **2004**; 305; 1601.

[85] Seeman, J.; *Theor. Biol.*; **1982**; 99; 237-247.

[86] Fu, T. J.; Seeman, N. C.; *Biochemistry*; **1993**; 32; 3211–3220.

[87] Li, X.; Yang, X.; Qi, J.; Seeman, N. C.; *J. Am. Chem. Soc.*; **1996**; 118; 6131– 6140.

[88] LaBean, T. H.; Yan, H.; Kopatsch, J.; Liu, F.; Winfree, E.; Reif, J. H.; Seeman, N. C.; *J. Am. Chem. Soc.*; **2000**; 122; 1848 –1860.

[89] Brun, Y.; Gopalkrishnan, M.; Reishus, D.; Shaw, B.; Chelyapov, N.; Adleman, L.; Building Blocks for DNA Self – Assembly; In Proceedings of the 1st Foundations of Nanoscience: Self – Assembled Architectures and Devices; (FNAN004). P. 2 – 15; Snowbird, UT, USA; April 2004.

[90] Mitchell, J. C.; Harris, J. R.; Malo, J.; Bath, J.; Turberfield, A. J.; *J. Am. Chem. Soc.*; **2004**; 126; 16342 – 16343.

[91] Liao, S.; Seeman, N. C.; *Science*; **2004**; Vol. 306; 2072 – 2074.

[92] Cohen, J. D.; Sadowski, J. P.; Dervan, P. B.; *Angew. Chem. Int. Ed.*; 46; 7956 – 7959.

[93] Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H.; *Science*; **2003**; Vol. 301; 1882 – 1884.

[94] Goodman, R.; Berry, R. M.; Turberfield, A. J.; *Chem. Comm.*; **2004**; 1372 – 1373.

[95] Goodman, R.; Schaap, I. A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J.; *Science*; **2005**; Vol. 310; 1661 – 1665.

[96] Chen, J.; Seeman, N. C.; *Nature*; **1991**; 350; 631 – 633.

[97] Shih, W. M.; Quispe, J. D.; Joyce, G. F.; *Nature*; **2004**; Vol. 427; 618 – 621.

[98] Erben, C. M.; Goodman, R. P.; Turberfield, A. J.; *Angew. Chemie. Int. Ed.*; **2006**; 45; 7414 – 7417.

[99] Bhatia, D.; Mehtab, S.; Krishnan, R.; Indi, S. S.; Basu, A.; Krishnan, Y.; *Angew. Chemie. Int. Ed.*; **2009**; 48; 4134 – 4137.

[100] Ulanovsky, L.; Bodner, M.; Trifonov, E. N.; Choder, M.; *Proc. Natl. Acad. Sci. USA*; **1986**; Vol. 83; 862 – 866.

[101] Rasched, G.; Ackermann, D.; Schmidt, T. L.; Broekmann, P.; Heckel, A.; Famulok, M.; *Angew. Chem. Int. Ed.*; **2008**; 47; 967 – 970.

[102] Schmidt, T. L.; Heckel, A.; *Angew. Chem. Int. Ed.*; **2009**; 5; 1517 – 1520.

[103] Merrifield, R. B.; *J. Am. Chem. Soc.*; **1963**; 85; 2149.

[104] Carpino, L.; Han, G. Y.; *J . Org. Chem.*; **1976**; Vol. 37; No. 62; 3404 – 3409.

[105] Sonogashira, K.; Tohda, Y.; Hagihara, N.; *Tett. Lett.*; **1975**; 16 (50); 4467–4470.

[106] Braha, O.; Walker, B.; Cheley, S.; Kasianowicz, J. J.; Song, L.; Gouaux, J. E.; Bayley, H.; *Chem. Biol.*; **1997**; 4; 497.

[107] Movileanu, L.; Howorka, S.; Braha, O.; Bayley, H.; *Nat. Biotechnol.*; **2000**; 18; 1091.

[108] Mellor, A.; Nivon, L.; Branton, D.; *Phys. Rev. Lett*.; **2001**; 86; 3435 – 3438.

[109] Mathematical model built by M. Wiggens & A. Marziali; Dept. of Astronomy and Physics; British Columbia – results to be published at a later date.

[110] Gutsmiedl, K.; Wirges, C. T.; Ehmke, V.; Carell, T.; *Org. Lett.*; **2009**; 11; 2405 – 2408.

[111] Borsenberger, V.; Kukwikila, M.; Howorka, S.; *Org. Biomol. Chem.*; **2009**; 7; 3826 – 3835.

[112] Ludwig, J.; Bruzik, K. S.; Stec, W. J.; Biophosphates and their Analogues-Synthesis, Structure, Metabolism and Activity.; **1987**; Elesevier, Amsterdam, pp. 131–133.

[113] Sowa, T; Ouchi, S.; *Bull. Chem. Soc. Jpn.*; **1975**; 48; 2084–2090.

[114] Niemeyer, C. M.; Mirkin, C. A.; Nanobiotechnology: Concepts, Applications and Perspectives, Wiley, New York, **2004**.

[115] Hegner, M.; Wagner, P.; Semenza, G.; *Surf. Sci.*; **1993**; 1993; 39.

[116] Hesse, J.; Sonnleitner, M.; Sonnleitner, A.; Freudenthaler, G.; Jacak, J.; Höglinger, O.; Schindler, H.; Schutz, J. G.; *Anal. Chem.*; **2004**; 76; 5960.

[117] Hesse, J.; Wechselberger, C.; Sonnleitner, M.; Schindler, H.; Schutz, J. G.; *J. Chromatogr. B*; **2002**; 782; 127.

[118] Kanaras, A. G.; Wang, Z.; Bates, A. D.; Cosstick, R.; Brust, M.; *Angew. Chem. Int. Ed.*; **2003**; 42; No. 2; 191 – 194.

[119] X-Ray photoelectron spectroscopy carried out by Dr. W. Chrzanowski of the Eastman Dental Institure, UCL.

[120] All AFM analysis discussed in section 3.5 & 3.6 carried out by M. Kastner et al. at the Institute of Biophysics, Johannes Keplar Uni, Linz.

[121] Franklin, R. E.; *Nature*; **1956**; 177; 929.

[122] Ganser-Pornillos, B. K.; Cheng, A.; Yeager, M.; *Cell*; **2007**; 131; 70.

[123] Tanaka, S.; Kerfeld, C. A.; Sawaya, M. R.; Cai, F.; Heinhorst, S.; Cannon, G. C.; Yeates, T. O.; *Science*; **2008**; 319; 1083.

[124] Niemeyer, C. M.; Mirkin, C. A.; Nanobiotechnology: Concepts, Applications & Perspectives; John Wiley & Sons; **2004**.

[125] Grueninger, D.; Treiber, N.; Zeigler, M. O. P.; Koetter, J. W. A.; Schulze, M-S.; Schulz, G. E.; *Science*; **2008**; 319; 206.

[126] Channon, K.; Bromley, E. H. C.; Woolfsen, D. N.; *Curr. Opin. Struc. Biol.*; **2008**; 18; 491.

[127] Seeman, N. C.; *Nature*; **2003**; 421; 427.

[128] Lata, S.; Reichel, A.; Brock, R.; Tampe, R.; Piehler, J.; *J. Am. Chem. Soc.*; **2005**; 127; 10205 – 10215.

[129] All AFM analysis discussed in section 4.8 carried out by Andreas Ebner at the Institute of Biophysics, Johannes Keplar Uni, Linz.